RNA:

A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of crRNAs in *Haloferax volcanii*

Jutta Brendel, Britta Stoll, Sita J. Lange, Kundan Sharma, Christof Lenz, Aris-Edda Stachler, Lisa-Katharina Maier, Hagen Richter, Lisa Nickel, Ruth A. Schmitz, Lennart Randau, Thorsten Allers, Henning Urlaub, Rolf Backofen and Anita Marchfelder

J. Biol. Chem. published online January 23, 2014

JBC Papers in Press. Published on January 23, 2014 as Manuscript M113.508184
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M113.508184

A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of crRNAs in *Haloferax volcanii*

Jutta Brendel[1], Britta Stoll[1], Sita J. Lange[2], Kundan Sharma[3], Christof Lenz[3,8], Aris-Edda Stachler[1], Lisa-Katharina Maier[1], Hagen Richter[4], Lisa Nickel[5], Ruth A. Schmitz[5], Lennart Randau[4], Thorsten Allers[6], Henning Urlaub[3,8], Rolf Backofen[2,7] and Anita Marchfelder[1]

[1] From the department of Biology II, Ulm University, 89069 Ulm, Germany
[2] Bioinformatics Group, Department of Computer Science, University of Freiburg, Georges-Köhler-Allee 106, 79110, Freiburg, Germany
[3] Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany
[4] Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Straße 10, 35043 Marburg, Germany
[5] Institute for General Microbiology, Christian-Albrechts-Universität Kiel, 24118 Kiel, Germany
[6] School of Life Sciences, University of Nottingham, Queen´s Medical Centre, Nottingham NG7 2UH, United Kingdom
[7] Centre for Biological Signalling Studies (BIOSS), Cluster of Excellence, University of Freiburg, Germany
[8] Bioanalytics, Institute of Clinical Chemistry, University Medical Center Göttingen, Robert-Koch-Strasse 40, 37075 Göttingen, Germany

Running title: *crRNA biogenesis and maintenance in Haloferax*

To whom correspondence should be addressed: Anita Marchfelder, Biology II, Ulm University, 89069 Ulm, Germany, Tel.: +49-731-5022658, Fax: +49-731-50-12-22658, E-mail: anita.marchfelder@uni-ulm.de

* This work was supported by the DFG in the frame of the Research Group 1680 "Unravelling the prokaryotic immune system".

Keywords: CRISPR-Cas, Cas6, *Archaea*, *Haloferax volcanii*, crRNA, type I-B

Background: The Cas6 protein is required for generating crRNAs in CRISPR-Cas I and III systems.
Results: The Cas6 protein is necessary for crRNA production, but not sufficient for crRNA maintenance in *Haloferax*.
Conclusion: A Cascade-like complex is required in the type I-B system for a stable crRNA population.
Significance: The CRISPR-Cas system I-B has a similar Cascade complex like types I-A and I-E.

ABSTRACT

The CRISPR-Cas (clustered regularly interspaced short palindromic repeats/ CRISPR associated) system is a prokaryotic defence mechanism against foreign genetic elements. A plethora of CRISPR-Cas versions exist, with more than 40 different Cas protein families and several different molecular approaches to fight the invading DNA. One of the key players in the system is the crRNA, which directs the invader - degrading Cas protein complex to the invader. The CRISPR-Cas types I and III use the Cas6 protein to generate mature crRNAs. Here, we show that the Cas6 protein is necessary for crRNA production but that additional Cas proteins that form a Cascade-like complex are needed for crRNA stability in the CRISPR-Cas type I-B system in *Haloferax volcanii in vivo*. Deletion of the *cas6* gene results in loss of mature crRNAs and interference. However, cells that have the complete *cas* gene cluster (*cas1-8*) removed and are transformed with the *cas6* gene are not able to produce and stably maintain mature crRNAs. crRNA production and stability is rescued only if *cas5, 6* and *7* are present. Mutational analysis of the *cas6* gene reveals three amino acids (His41, Gly256 and Gly258) that are essential for pre-crRNA cleavage, while the mutation of two amino acids (Ser115 and Ser224) leads to an increase of crRNA amounts. This is the first systematic *in vivo* analysis of Cas6 protein variants. In addition, we show that the *H. volcanii* I-B system contains a Cascade-like

complex with a Cas7, Cas5 and Cas6 core that protects the crRNA.

Prokaryotes developed a variety of resistance mechanisms to defend themselves against invaders. Recently, a new defence mechanism was detected: the CRISPR-Cas system (for reviews see (1-6)). Key elements of this defence system are the Cas proteins and the CRISPR RNA. The latter consists of short repeat sequences that are separated by variable sequences (spacers). Spacer sequences are derived from previous invaders; therefore, the CRISPR locus is a memory of all previous attacks by invaders.

The defence reaction is divided into three stages. In the adaptation stage, the invader DNA is cleaved and a piece of it is selected to be integrated as a new spacer into the CRISPR locus, where it is stored as an identity tag for future attacks by this invader. During the second stage -the expression stage- the CRISPR RNA (pre-crRNA) is transcribed and subsequently processed into the mature crRNAs. In the third stage -the interference stage- Cas proteins, together with crRNAs, identify and degrade the invader.

The CRISPR-Cas systems have been sorted into three major classes (I-III) that are further subdivided into 11 subtypes (I-A-F, II-A-C and III-A-B) (7,8). In CRISPR-Cas types I and III, the mature crRNA is generally generated by a member of the Cas6 protein family (9). This endonuclease has been analysed in detail in several bacteria (10-14) and a few archaea (10-17). The Cas6 proteins contain a ferredoxin fold and a characteristic glycine-rich motif, at the C-terminus (8,18,19). All Cas6 proteins analysed to date generate a crRNA with an eight-nucleotide-long 5´ handle (16,20-23). The crRNAs generated by the Cas6 proteins usually contain a 2´-3´ phosphate group (14); only processing by the type I-F Cas6 (Cas6f) results in a non-cyclic phosphate group (24). The Cas6e and Cas6f proteins stay bound to the processed crRNA (21,25,26). Apart from these similarities, the various Cas6 proteins show several differences: they share very little sequence identity and show differences in catalytic site composition and in the molecular details of the reaction. The Cas6 protein from *P. furiosus* contains a catalytic triad consisting of a tyrosine, a histidine and a lysine (14), whereas the Cas6e protein from *Thermus thermophilus* (type I-E) contains a catalytic dyad consisting of a tyrosine and a histidine (25,27). Studies with the Cas6b protein from *Methanococcus maripaludis* (type I-B) identified two histidine residues important for catalysis (16). The *Sulfolobus solfataricus* Cas6 protein does not contain a histidine close to the catalytic site but seems to require a network of basic residues for catalytic activity (28). Cas6e (20,25,27,29) and Cas6f are known to bind to stable hairpin motifs on the repeat (21,26). The Cas6 from *Staphylococcus epidermidis* binds to a smaller hairpin structure (30), whereas the Cas6 from *P. furiosus* binds to an unstructured repeat (10,31,32). Type I-C systems do not have a Cas6 protein, as the Cas5d protein is responsible for crRNA processing (33,34).

While in system III, the Cas6 protein acts alone, in type I-A, I-E and I-F systems, the respective Cas6 protein is part of the Cascade (CRISPR-associated complex for antiviral defence) complex (9). The Cascade complex is composed of different Cas proteins, depending on the subtype, and it is involved in pre-crRNA processing. The mature crRNA remains associated with the complex, that is subsequently binds to the invader DNA. The invader is then degraded by the Cas3 protein (35,36). The composition of the type I-B Cascade complex has not yet been described.

It has been suggested that the Cas6 proteins co-evolved with the CRISPR RNA repeat sequences (37,38). Repeat sequences have been shown to have highly variable structures. Some form hairpin structures while others do not, and they have been clustered into different groups (37,39). Because the Cas6 proteins interact with the crRNAs, they have to adapt to their respective repeat sequences. Taken together, the current data about the Cas6 proteins show that these proteins are highly divergent from each other. To fully understand this class of proteins, more data about different Cas6 proteins from different subtypes and organisms are required.

Here, we report new data about the Cas6 protein from *Haloferax volcanii*. *H. volcanii* is a halophilic archaeon that requires 2.1 M NaCl for optimal growth and contains similar concentrations of salt intracellularly to cope with the high salt concentration in the medium. *H. volcanii* is studied as an archaeal model organism because it is easy to cultivate and to genetically modify. *H. volcanii* encodes a CRISPR-Cas type I-B system with three different CRISPR loci: two located in close proximity to each other on the chromosomal plasmid pHV4 (locus P1 and P2) and one on the main chromosome (C) (Figure 1) (40,41). All three loci are expressed and processed (42). The *cas* gene cluster, which encodes the Cas proteins

Cas1-8b, is located between the two plasmid-encoded CRISPR loci. The repeat sequences of the *H. volcanii* CRISPR RNA have the potential to form a short-stem loop structure (23).

Here, we present the first systematic, *in vivo* mutational analysis of a Cas6 protein and its effects on the cellular crRNA amounts, as well as the interference reaction. Furthermore, we show that the *H. volcanii* I-B system contains a Cascade-like complex consisting of at least Cas5, Cas6 and Cas7 proteins as well as crRNA.

EXPERIMENTAL PROCEDURES

Strains - *H. volcanii* strains H119Δ*cas6* (Δ*pyrE2*, Δ*leuB*, Δ*trpA*, Δ*cas6*) (this study) and H26Δ*casCluster28* (Δ*pyrE2*, Δ*leuB*, Δ*trpA*, ΔpHV4:207.288-218.340)(42) were grown aerobically at 45°C in Hv-YPC medium (42). *H. volcanii* strains H119Δ*cas6* containing plasmids with mutated *cas6* genes were grown in Hv-Ca medium supplemented with 0.25 mM tryptophan. *E. coli* strains DH5α (Invitrogen) and GM121 (43) were grown aerobically at 37°C in 2YT medium (44). *Halorubrum lacusprofundi* DL12 (45) was grown aerobically in Hv-YPC medium at 37°C.

Construction of plasmids and transformation of H. volcanii - The 3XFLAG-tag sequence was amplified by PCR using the oligonucleotides FLAGN and FLAG2 (46). The PCR fragment was digested with *Nde*I and *Hin*dIII and cloned into pTA927 (43), yielding pTA927-Flag. The *cas6* gene was amplified by PCR from genomic DNA using the oligonucleotides Cas6FlagNup and Cas6FlagNdo (primer sequences are listed in Supplementary Table 1). The resulting fragment was cloned into pBluescriptII (Stratagene) linearized with *Eco*RV yielding pBlue-NFlagcas6 (Supplementary Table 2). This construct was used as template to introduce the mutations, either by the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies) or by inverse PCR using 5´-phosphorylated oligonucleotides followed by religation of the PCR fragment, using oligonucleotides carrying the desired mutation in both cases (Supplementary Table 1). The resulting pBluescriptII constructs carrying the wild type *cas6* gene and mutated *cas6* genes were digested with *Hind*III and *EcoR*I and the *cas6* gene variants were subcloned into the plasmid pTA927-Flag yielding the constructs pTA927-*cas6* and pTA927-*cas6*mutX, respectively (Supplementary Table 2). The following constructs were generated to complement the strain H26Δ*cas*Cluster28 with different sets of *cas* genes expressed from *H. volcanii* vector pTA927: pTA927-Cas6875, pTA927-Cas685, pTA927-Cas687 and pTA927-Cas675. pTA927-Cas6875 was generated by PCR-amplification of *cas6, cas8, cas7 and cas5* from genomic DNA of the wild-type strain H119 with oligonucleotides Cas6FlagCup and 3-Cas5-BamHI, followed by digestion of the PCR fragment with *Nde*I and *BamH*I and its subsequent cloning into pTA927. pTA927-Cas685 was generated by PCR amplification using oligonucleotides Cas6FlagCup and 3-Cas5-BamHI and genomic DNA from an H119Δ*cas7* strain as a template, followed by digestion of the PCR fragment with *Nde*I and *BamH*I and its subsequent cloning into pTA927. pTA928-Cas687 was generated by PCR amplification using oligonucleotides Cas6FlagCup and 3-Csh2-HindIII and genomic DNA of the wild-type strain H119, followed by digestion of the PCR fragment with *Nde*I and *Hind*III and its subsequent cloning into pTA927. pTA927-Cas675 was generated by PCR amplification using oligonucleotides Cas6FlagCup and 3-Cas5-BamHI and genomic DNA from the H119Δ*cas8* strain as a template, followed by digestion of the PCR fragment with *Nde*I and *BamH*I and its subsequent cloning into pTA927. All plasmids were passaged through *E.coli* GM121 cells to avoid methylation and then introduced into *H. volcanii* strains H119Δ*cas6* or H26Δ*cas*Cluster, respectively, using the polyethylene glycol method (47,48). Transformants were selected on Hv-Ca plates without uracil.

Northern blot hybridisation - Total RNA was isolated from exponentially growing *H. volcanii* cells as described (49) or isolated from the FLAG-Cas7 purified protein fraction as described below. After separation of 10 µg RNA (total RNA) or 600 ng (FLAG-Cas7 fraction) on 8% denaturing gels, RNA molecules were transferred to nylon membranes (Hybond-N+, GE Healthcare) and incubated with oligonucleotides against the different spacers (primer P1.1, which is complementary to the sequence of the first spacer of locus P1; primer P1.2, which is complementary to the sequence of the second spacer of locus P1; primer P2.1, which is complementary to the sequence of the first spacer of locus P2; primer C1 which is complementary to the sequence of the first spacer of locus C), as RNA loading control a hybridisation with oligonucleotide 5S which binds to the 5S rRNA was performed. All primers were radioactively labelled at the 5´ end

with γ-^{32}P-ATP. To quantify the amount of mature crRNA the membranes were exposed to imaging plates (BAS-MS, Fujifilm) and analysed using the FLA-3000 scanner (software BASreader 3.14). The intensity of the signals was measured with ImageJ. The signals of the P1.1-detection were put into relation to the 5S rRNA signal, which was used as RNA loading control. To obtain the percentage of mature crRNA in the mutants, the signals of complementation of the *cas6* deletion strain with the wild type *cas6* gene was set to 100% and the crRNA amounts of the *cas6* mutant strains were set in relation to these data. All northern analyses were performed at least three times. In the analysis of the crRNA amounts with the Δ*cas*Cluster strain the crRNA signals in the wild type strain were set to 100% and the Δ*cas*Cluster strains transformed with the different sets of *cas* genes were set in relation to the wild type crRNA amounts.

RT-PCR and 5´ RACE - RNA was isolated as described above and 1 µg was subjected to several reverse transcription reactions with primers Csh1#2, Csh2#2, Cas5#2, Cas3#5, Cas4#2, Cas1#5 and Cas2#2. The resulting cDNA was amplified with primers used for RT-PCR and primers Cas6#4, Csh1#3, Csh2#3, Cas5#3, Cas4#4, Cas1#4. PCR fragments were cloned into pBluescriptII (digested with *Sma*I) and resulting clones were sequenced. For the 5´ RACE 1 µg RNA treated with 5′-Phosphate-Dependent Exonuclease (Epicentre) to enrich primary transcripts was reverse transcribed with primer Cas6#3 using the MINT universal kit (Evrogen) according to the manufacturer´s instructions. For the PCR reaction primer M1 from the MINT kit and primer Cas6#2 were used. PCR fragments were cloned into pBluescriptII (digested with *Sma*I) and resulting clones were sequenced.

Expression of cas6 in E. coli and H. volcanii - To express the *Haloferax cas6* in *E.coli,* the gene was amplified by PCR using the oligonucleotides 5-Cas6-NcoI and 3-Cas6-NotI. The fragment was digested with *Nco*I and *Not*I and cloned into the vectors pET28a and pET32a (Novagen), yielding the constructs pET28a-*cas6* (N-terminal 6XHis tag) and pET32a-*cas6* (N-terminal 6XHis, S-, and Trx-tag). For the construct pET30a-*cas6*-FlagN, the *cas6* gene together with the N-terminal 3XFLAG-tag was removed from the construct pTA927-*cas6* by digestion with *Nde*I and *Eco*RI and cloned into the vector pET30a (digested with *Nde*I and *Eco*RI). To obtain a construct with a C-terminal tag, a PCR reaction was performed using the oligonucleotides 5-Cas6-NcoI and 3-Cas6-HindIII, followed by digestion of the resulting fragment with *Nco*I and *Hind*III and cloning into the vector pET30a, yielding pET30a-*cas6*o.stop (C-terminal 6XHis-tag). To improve the expression, the *cas6* gene was optimised to the *E.coli* codon usage by GeneArt (Life Technologies). The optimised gene was isolated from the original plasmid by digestion with *Nco*I and *Xho*I and cloned into different pET vectors, yielding the constructs pET28a-*cas6*h and pET30a-*cas6*h. *E.coli* strains BL21AI and BL21 codon plus (Novagen) were transformed with the respective constructs and the protein was expressed according to the manufacturer´s protocol. To improve the solubility of the recombinant protein several different expression conditions (growth temperature, amount of IPTG and/or arabinose for induction) were tested. To express the *cas6* gene in *Haloferax* cells, the strain H119Δ*cas6* was transformed with the construct pTA927-*cas6* (see above). To obtain a construct with 6XHis-tag, the *cas6* gene was amplified by PCR using the oligonucleotides 5-Cas6-PciI and 3-Cas6-NotI. The fragment was digested with *Pci*I and *Not*I and cloned into the vector pTA963 (43), yielding the construct pTA963-*cas6*. The *Haloferax* strain H1424 (50) was transformed with this construct. To express the tagged Cas6 protein the respective strains were grown in Hv-Ca medium containing 0.25 mM tryptophan. At exponential growth phase tryptophan was added to a final concentration of 3 mM and the culture was grown for additional 2-16 h. To prepare soluble extracts the cells were harvested and washed in enriched PBS buffer (2.5 M NaCl, 150 mM $MgCl_2$, 1XPBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM K_2HPO_4, pH 7.4)). The cells were resuspended in lysis buffer (1 M NaCl, 100 mM Tris/HCl pH 7.5, 1 mM EDTA, 10 mM $MgCl_2$, 1 mM $CaCl_2$,8 U/µl DNase RQ1 (Promega), 13 µl/ml protease inhibitor cocktail (Sigma)) and lysed by ultrasonification. Insoluble cell debris was removed by centrifugation at 48,000 x g. For affinity purification via the 3XFLAG-tag, anti-FLAG M2 affinity gel (Sigma) was equilibrated with ice-cold washing buffer (50 mM Tris/HCl pH 7.5, 1 M NaCl) and added to the extract. After incubation over night at 4°C the affinity gel was washed extensively with washing buffer and the protein was eluted by washing buffer to which 3XFLAG peptide (Sigma) was added to a final concentration of 150 ng/µl. His purification was performed according to the manufacturer´s

protocol with Protino Ni-NTA Agarose (Macherey-Nagel), using buffers with different concentrations of NaCl from 300 mM to 2 M.

Purification of a FLAG-tagged Cas7 protein

To obtain the construct pTA927-Cas68[HisFlag]75, that contains all four *cas* genes *cas5-8* with the *cas7* gene fused to cDNAs for a His- and a FLAG-tag, first the 3XFLAG sequence was amplified by PCR using the primers FlagSnaBI and Flag2. The product was digested with *SnaB*I and *Hind*III and cloned into the vector pCDF1b (Novagen) digested with *Pml*I and *Hind*III, resulting in the construct pCDF1b-Flag. This construct was digested with *Xho*I, treated with *Pfu* polymerase to generate blunt ends and then digested with *Hind*III. The genes *cas7* and *cas5* were amplified from genomic *Haloferax* DNA using primers 5-Csh2-HindIII and 3-Cas5-BamHI, the resulting fragment was digested with *Hind*III and cloned into the pCDF1b-Flag vector, yielding the construct pCDF1b-75HisFlag. The genes for Cas7 and Cas5 together with the N-terminal 6XHis- and 3XFLAG-tag were amplified from this construct using the primers HisFlagBamHI and 3-Cas5-XbaI and digested with *BamH*I and *Xba*I. The genes *cas6* and *cas8* were amplified by PCR from genomic *Haloferax* DNA with the primers Cas6FlagCup and 3-Cas8-NcoIBamHI, digested with *Nde*I and *BamH*I and cloned into the vector pTA927, digested with the same enzymes. This construct was digested with *BamH*I and *Xba*I and the insert Cas75HisFlag was ligated into it, yielding the construct pTA927-Cas68[HisFlag]75. The sequence was verified by sequencing and subsequently *Haloferax* strain H26*casCluster28* was transformed with the plasmid. Cells were grown to exponential phase in HvCa medium containing 0.25 mM tryptophan. Expression of the *cas* genes and FLAG purification was performed as described above for *cas6* gene expression in *Haloferax*. The resulting purified protein fraction was analysed using SDS PAGE and mass spectrometry.

Isolation of RNA from the FLAG-Cas7 purified fraction - The FLAG-Cas7 purified fraction, was incubated with 20 µg of proteinase K for 30 min at 37 °C in 100 µl of buffer (100 mM Tris/HCl, pH 7.5, 12.5 mM EDTA, 150 mM NaCl, 0.2% SDS), subsequently the solution was extracted with phenol-chloroform-isoamylalcohol. RNA was precipitated from the aqueous phase, and the resulting pellet was dissolved in water. The RNA fraction was analysed using a northern blot as described above.

Mass spectrometry - For mass spectrometric analysis proteins were in gel digested with trypsin according to (51). Extracted peptides were analysed by liquid chromatography tandem mass spectrometry (LC-MS/MS) on an Orbitrap XL instrument (ThermoFisherScientific) under standard conditions. Peptide fragment spectra were searched against a target decoy database for *H. volcanii* (www.halolex.mpg.de)(52) using MASCOT as a search engine. Peptides with a peptide score lower than 20 were not considered specific.

Quantification of Cascade subunits using intensity based absolute quantification (iBAQ) - 0.25 µg of the proteins co-purified with FLAG-Cas7 as described above were dried in an Eppendorf tube and dissolved in 10 µl aqueous solution of Universal Proteomics Standard 2 (UPS2, Sigma-Aldrich) in 1% Rapigest (Waters) to obtain a nominal protein concentration ratio of 0.25:1 µg total (Cas proteins:UPS2). The sample was reduced with dithiothreitol (10 µl, 50 mM in 100 mM triethylammonium bicarbonate (TEAB), 37°C, 1hr), alkylated with iodacetamide (10 µl, 100 mM in 100 mM TEAB, 37°C, 1hr), diluted with 70 µl 100 mM TEAB and digested with porcine trypsin (Promega, 1:20, 37°C, 16 hr). Following digestion, the Rapigest detergent was cleaved by acidifying the solution with trifluoroacetic acid (20 µl, 5%, 37°C, 2 hr) and pelleting the released fatty acids by centrifugation (13.000 rpm, 30 min, RT). The supernatant containing tryptic peptides was transferred to a fresh tube, taken to dryness in a Speedvac and dissolved in LC-MS loading buffer (30 µl, 5% acetonitrile, 0.1% formic acid). Samples were analysed in triplicate (3 x 5 µl) on a nanoflow liquid chromatography system (1100 series, Agilent) coupled to an LTQ-Orbitrap Velos mass spectrometer in a vented column setup at an analytical flow rate of 300 nl/min achieved through passive splitting. Samples were desalted on an RP-C18 pre-column (20 mm, 0.15 mm ID, ReproSil-Pur C18-AQ 5 µm, Dr. Maisch). Separation was achieved on an RP-C18 column (150 mm, 0.075 mm ID, ReproSil-Pur C18-AQ 3 µm, Dr. Maisch) packed into a SilicaTip emitter (FS360-75-10-N, New Objective) using a 50 min linear gradient of 3-36% acetonitrile with 0.1% formic acid as modifier). MS data were acquired using a Top8 method with CID fragmentation, using a normalized collision energy of 45%.

Results were analysed using MaxQuant software 1.2.7.4 (53). The MS data was matched against a *Haloferax volcanii*-filtered UniProt/trEMBL

protein sequence database (v 2013-11) supplemented with the sequences of the 48 proteins contained in the UPS2 standard. iBAQ values from replicate values were averaged and the standard deviation calculated to judge the precision of analysis (Suppl. Table 3). A total of 20 UPS2 standard proteins were observed in three out of three replicate analyses and used for calibration. A calibration curve was obtained by linear regression from a double logarithmic plot (log(iBAQ) vs. log(amount), Suppl. Fig. 1). The calibration function was then used to calculate the respective Cas5-7 amounts in the co-purified fraction.

Generation of a cas6 gene deletion strain - The deletion of the *cas6* gene was achieved by using the pop-in/pop-out method as described previously (43,44,46). The *cas6* gene was PCR amplified with flanking regions (approximately 500 base pairs each) from the chromosomal DNA of *H. volcanii* strain H119 using primers Cas6KOup and Cas6KOdo. The resulting 1915-nt PCR fragment was subsequently cloned into the vector pTA131 (*Eco*RV), yielding pTA131-cas6geneupdo. To remove the *cas6* gene, an inverse PCR was performed on pTA131-cas6geneupdo with primers IPCas6up and IPCas6do2; the resulting linear PCR product was ligated, yielding pTA131-cas6updo. Plasmids were passaged through *E. coli* GM121 to prevent methylation, and *H. volcanii* strain H119 was subsequently transformed with this construct to allow integration (pop-in) of the plasmid into the genome. The subsequent selection for loss of the *pyrE2* marker by plating on 5-fluoroorotic acid (5-FOA) revealed pop-out mutants. Chromosomal DNA was isolated from the wild type and potential *cas6* deletion mutants. Southern blot hybridisation was performed as described (47), with the following modifications: 10 µg of *Sal*I-digested DNA was separated on a 0.8% agarose gel and transferred to a nylon membrane (Hybond^{TM}-N, GE Healthcare). The 500-base pair *cas6* downstream region was amplified using primers IPCas6KOdo2 and Cas6KOdo, labelled using the PCR DIG Probe Synthesis Kit (Roche) and used as a hybridisation probe. Hybridisation and detection were performed according to the DIG manual (DIG Luminescent Detection Kit, Roche).

Cloning of the heterologous cas6 genes - The *cas6* gene from *M. mazei* was isolated from the original plasmid pRS714 (13) by PCR using primers Cas6Mmazfwd and Cas6Mmazrev, after digestions with *Nde*I and *Eco*RI the gene was cloned into pTA927 (digested with *Nde*I and *Eco*RI) yielding pTA927-Mmazcas6. The *cas6* gene from *M. maripaludis* was isolated from the original plasmid pHR6 (16) by digestion with *Nde*I and *Hin*dIII and was cloned into the pTA927 vector, which was previously digested with *Nde*I and *Hin*dIII, thus yielding pTA927-Mmarcas6. The *cas6* genes Hlac3333 and Hlac3572 from *Hrr. lacusprofundi* were ordered from Geneart (Life Technologies). Clones for Hlac3333 and Hlac3572 were digested with *Nco*I and *Bam*HI and *Nco*I and *Eco*RI, respectively, and resulting fragments were then cloned into vector pTA1228 yielding pTA1228-Hlac3333 and pTA1228-Hlac3572. pTA1228 is an improved version of the pTA963 over-expression plasmid (43), it features a unique *Nsp*I site in the polylinker (downstream of the 6xHis tag). This *Nsp*I site is compatible with *Sph*I ends and allows for the cloning of genes where the second codon starts with C. pTA1228 was derived from pTA963 by the removal of two existing *Nsp*I sites. The *Nsp*I site (bp 1033) located between the pBluescript backbone and the pHV2 origin was inactivated by cutting with *Nsi*I and filling-in with Klenow. The *Nsp*I site (bp 7165) located in the *hdrB* selectable marker was inactivated by replacement of a 762 bp *Bsp*EI-*Nco*I fragment (comprising the 5´ part of *hdrB* and the 3´ part of *pyrE2*) with a near-identical 762 bp *Bsp*EI-*Nco*I PCR fragment that features a synonymous G->A mutation (GAC->GAT) at the *Nsp*I site. The internal hdrB-dNsp primers with the G->A mutation (shown in lower case) were: hdrB-dNsp F and hdrB-dNsp R. The sequence and map of pTA1228 are available upon request from TA.

Modelling of cas6 3D structure - To get a prediction of the tertiary structure of the *Haloferax* Cas6 protein, the amino acid sequence was submitted to the Phyre server (protein homology/analogy recognition engine) (54). According to the Phyre database the *Pyrococcus* Cas6 protein showed the closest related structure. The 3D structure was modelled with the help of the graphical software PyMOL.

Interference tests - *H. volcanii* strain H119Δ*cas6* was transformed with the plasmids containing the *cas6* wild type gene (pTA927-*cas6*) or the *cas6* mutants (pTA927-*cas6*mutX). In a second transformation, these strains were transformed with the invader plasmids pTA352-PAM3 or pTA352-PAM9 (Supplementary Table 2) (42). Transformants were selected on Hv-Ca plates without leucine and uracil. As a positive control, the respective strains were transformed with the vector pTA352. To confirm a successful

interference reaction, *H. volcanii* cells were transformed at least three times with one of the plasmid-invader constructs. As has been observed in similar studies (42,55), it is difficult to accurately determine transformation rates; therefore, we defined only those sequences that led to at least a 100-fold reduction in transformation rates (reduced by factor 0.01) in this plasmid assay as a successful interference reaction.

RESULTS

The family of Cas6 proteins and the *Haloferax* Cas6 protein

The *Haloferax* Cas6 protein shows very little sequence similarity to other Cas6 proteins from type I systems and to most systems in general (Figure 2). As visible in the alignment only three positions are conserved in all Cas6 proteins analysed (Gly122, Gly258, Gly258). Only one amino acid, shown to be important for catalytic acitivity in other organisms, is conserved in the *Haloferax* Cas6 protein (His41). To analyse the family of Cas6 proteins and to determine Cas6 protein conservation, we first performed a BLAST search with the *H. volcanii* Cas6 (YP_003533663.1) and subsequently quantified similarity via pairwise global sequence alignments (using the NCBI implementation of the Needleman-Wunsch algorithm (56)). Close homologs of the *Haloferax* Cas6 protein were only found in closely related Haloarchaea with global percent identities between 41 - 44 % and BLAST E-values < $1x10^{-60}$, including *Haloferax mediterranei* (ATCC 33500), *Haloarcula marismortui* (ATCC 43049), *Halorhabdus utahensis* (DSM 12940), *Halorubrum lacusprofundi* (ATCC 49239), *Natrinema* (sp. J7-2) and *Halomicrobium mukohataei* (DSM 12286). Apart from the Haloarchaea, the similarity to other Cas6 proteins is very low, with pairwise similarities of only 13 - 22 % to other Cas6-like proteins (Figure 3). Even the Cas6 protein from *P. furiosus*, which of all the Cas6 proteins investigated is the closest to *H. volcanii*, has an equally low similarity of 18 %. The Cas6 proteins from other systems show comparably low levels of sequence similarities between each other.

Conservation of the *H. volcanii* repeats is analogous to the conservation of the Cas6 protein; the *Haloferax* repeat sequences are very different from other published systems, except in Haloarchaea. Repeats from haloarchaeal CRISPR-Cas systems are, in fact, well conserved and can form a minimal hairpin structure with three base pairs (25). When using the web server CRISPRmap (39), which compares all available repeats according to sequence and structure similarity, the haloarchaeal CRISPR repeats cluster together into the sequence family F19. The uniqueness of the haloarchaeal family F19 is highlighted by the fact that it is located on an isolated branch of the hierarchical CRISPRmap tree of all repeat sequences (Figure 3). Thus, in the large family of Cas6-like proteins, not only is the sequence similarity low but the individual proteins are highly specific to their associated repeat sequences and have different functional mechanisms. Obviously, the full spectrum of this family has not yet been explored. Therefore, we analysed the *Haloferax* Cas6 protein in detail to provide insights into this extremely diverse family.

Cas *genes show a low level of expression*

Earlier studies showed that the CRISPR RNAs are constitutively expressed in *Haloferax* (42). To analyse whether the *cas* genes are likewise constitutively expressed, we used northern blot analyses to determine the amount and the length of the *cas* mRNAs. No signals were detected for the *cas* mRNAs on northern blots with RNA isolated from cells grown under different conditions (standard conditions, low and high temperatures, low and high salt concentrations) (data not shown). Therefore, we used RT-PCR to determine if the genes are transcribed at all, revealing that the mRNA for all *cas* genes is present (data not shown). Employing 5´ RACE, we determined the 5´ end of the Cas6 mRNA, which is located 39 nucleotides upstream of the Cas6 protein start codon. Expression of the *cas* genes was confirmed in a parallel study in which the *Haloferax* proteome was investigated (Sharma, Stoll, Pfeiffer, Marchfelder and Urlaub, in preparation). Here, it was shown that all eight Cas proteins were present in the proteome in both the exponential and stationary phases. Taken together, we could show that the *cas* mRNA is present albeit in low amounts.

Cas6 is necessary for crRNA production

To investigate the biochemical characteristics of the Cas6 protein, we expressed recombinant Cas6 protein in *E. coli* and *H. volcanii* using different approaches (for details see Experimental Procedures). The *cas6* gene was cloned into different expression vectors, and *Haloferax* and different *E. coli* strains were transformed with these constructs. All attempts to produce an active recombinant Cas6 protein failed, therefore we next used an *in vivo* approach to investigate the Cas6 protein. To

determine the *in vivo* function of Cas6 in *Haloferax,* we generated a *cas6* deletion strain. Because the *cas6* gene overlaps with one nucleotide of the downstream *cas8b* gene (Figure 1), we deleted the complete reading frame with the exception of the last nucleotide that is shared with the downstream *cas8b* gene. The *cas6* deletion strain showed no visible phenotype in comparisons with the respective wild type strain when grown under different conditions, such as different temperatures and salt concentrations (data not shown). To investigate whether *Haloferax* is able to produce mature crRNAs without the Cas6 protein, we isolated RNA from wild type and *cas6* deletion strains and analysed the amount of crRNAs by northern blot. Hybridisation with probes against spacer1 from locus P1 (Figure 4) and spacers from the other CRISPR loci (P1.2, P2.1. and C1) (data not shown) revealed that the wild type strain contains mature crRNAs but the deletion strain does not. To confirm that the failure to produce or stabilise mature crRNAs is due to the missing Cas6 protein and not to some effect on the downstream-encoded *cas* genes, we complemented the *cas6* deletion strain with the FLAG-Cas6 over-expression construct (described in Experimental Procedures) (Figure 4). Complementation restores the wild type amount of crRNAs, verifying that Cas6 is required for efficient crRNA production. To investigate whether *cas6* genes from other organisms would be able to complement the deletion mutant, we transformed the deletion mutant with the *cas6* genes from another halophilic archaeon, *Halorubrum lacusprofundi.* Northern analysis showed that these genes are not able to complement the deletion mutant strain (data not shown). Furthermore, we complemented the *cas6* gene deletion mutant with type I-B *cas6* genes from two non-halophilic archaea, *Methanococcus maripaludis* (16) and *Methanosarcina mazei* (13). Neither of the two heterologous *cas6* genes was able to complement the *Haloferax cas6* gene deletion mutant (data not shown). The repeat sequences of the CRISPR RNAs from the four different organisms show almost no sequence similarities, which might be the reason that the heterologous Cas6 proteins cannot process the *Haloferax* repeat. Taken together we could show that without the *cas6* gene crRNAs cannot be generated or maintained *in vivo* in *Haloferax*.

Single amino acid mutations in Cas6 reveal both a loss and a gain in crRNA levels

To identify amino acids necessary for generating or stabilising crRNAs, we introduced 21 point mutations in the *cas6* gene to change the original amino acid to alanine (Figure 2). We have chosen amino acids for mutation that according to the alignment in Figure 2 either directly align with amino acids shown in other Cas6 proteins to be important for function or are in the vicinity of such amino acids. In addition, amino acids which were shown to be conserved in at least three Cas6 proteins were chosen. The *Haloferax* H119Δ*cas6* strain was transformed with the 21 *cas6* mutant constructs, and RNA was subsequently isolated for northern analyses. Hybridisation with a probe against spacer 1 from the CRISPR locus P1 showed that most of the mutants were still able to generate the crRNA (Figure 5 and Table 1). Only three mutations resulted in reduced (less than 50 % of wild type amount) crRNA amounts. Changing the amino acid His41 to Ala reduced the amount of crRNA to 37% of the wild type amount, while changing Gly256 or Gly258 reduced it to 12% and 3% of the wild type amount, respectively. Interestingly, we also observed a gain of function in two mutants (mutants Ser115 and Ser224) that resulted in higher crRNA amounts of 129% and 164%, respectively. To investigate whether the severely reduced crRNA levels in mutants Gly256 and Gly258 are due to parts of the protein being in the insoluble fraction due to improper folding, we isolated whole cell extract from the transformed cells and analysed the insoluble and soluble protein fraction for the presence of the Cas6 protein mutants. Both Cas6 protein mutants were detectable in the soluble fraction and the amount of protein did not change significantly (data not shown). Therefore the reduced activity is not due to production of an insoluble protein.

Cas6 mutations can abolish the interference reaction

Type I Cas6 proteins have been shown to be involved in the generation of crRNAs. They have also been shown to stay attached to the crRNA after processing. To determine whether an intact Cas6 is important for the interference reaction and which amino acids are important for this activity, we subjected all *cas6* mutants to an interference test, that we previously established in *Haloferax* by using a plasmid-based invader (42). The invader plasmid contains a protospacer that is recognised by one of the *Haloferax* crRNAs and PAM motifs that are recognised by the *Haloferax* CRISPR-Cas system and trigger the defence reaction. Plasmid invaders were introduced into a *Haloferax* strain, that cannot

grow without uracil as the strain lacks the *pyrE2* gene. Transformants were selected by uracil prototrophy conferred by the *pyrE2* gene on the plasmid invader. Plasmid elimination via the CRISPR-Cas defence of the host cell were detected by a drastically reduced transformation rate. Cells with an active interference system degrade the invader plasmid efficiently, reducing the transformation rate by at least a factor of 0.01. Transformation of the *cas6* deletion strain with the invader plasmid resulted in normal transformation rates, showing that deletion of the *cas6* gene abolishes the CRISPR interference reaction of the cells (Table 2). If the *cas6* deletion strain is complemented with the *cas6* wild type gene and then transformed with the invader plasmid, only a few transformants are obtained (the transformation rate is reduced by a factor of 0.01) (Table 2), showing that the interference is working again. Next, the *cas6* gene deletion strain was transformed, first with the mutated *cas6* genes and then with the invader plasmid, to investigate which Cas6 protein mutations would interfere with the defence reaction. Three mutants showed reduced or no interference reactions: His41, Gly256 and Gly258 (Table 2). Transformation of the *cas6* deletion strain with the His41 mutant and the invader plasmid resulted in reduction of the transformation rate by only a factor of 0.2; the plasmids were not degraded very efficiently. Mutants Gly256 and Gly258 showed a normal transformation rate; thus, the interference reaction was not active and was not able to degrade the plasmid. The increased crRNA amounts in the Ser mutant strains did not change the behaviour in the interference reaction. The interference reaction was working efficiently, reducing the transformation rate by a factor of at least 0.001 (Ser224) and 0.002 (Ser115) (Table 2).

Taken together, we could show that *cas6* mutations, which reduce the crRNA amounts, likewise reduce the effectiveness of the interference reaction.

A strain with only cas6 is not active in crRNA production

Next, we investigated whether Cas6 alone is sufficient for crRNA maturation. We used a *Haloferax* strain in which all *cas* genes were deleted (H26Δ*cas*Cluster28 (42)) for transformation with the wild type *cas6* gene to investigate whether the Cas6 protein alone is sufficient for pre-crRNA processing and stabilisation. Northern analysis of this strain showed no crRNA maturation, suggesting that the *cas6* gene is necessary for crRNA production (because deletion of the *cas6* gene results in loss of crRNA processing), but not sufficient for crRNA maintenance. Therefore, one or more of the other Cas proteins are also required. The Cas proteins that are potentially also involved in crRNA maintenance are Cas5, Cas7 and Cas8b, which are encoded together with Cas6 at the 5´ end of the *cas* genes locus (Figure 1). To analyse whether the presence of all four of these genes rescues crRNA generation, we transformed strain H26Δ*cas*Cluster28 with the complete 5´ part of the *cas* gene cluster (encoding the Cas5-8b proteins) (Figure 6A). RNA from this strain was analysed with northern blots for crRNA generation revealing that crRNAs were generated and stably maintained (Table 3, Figure 6B). To determine if one of the four Cas proteins can be omitted, we next transformed the H26Δ*cas*Cluster28 strain with the genes for (1) Cas5, Cas6 and Cas7; (2) Cas5, Cas6 and Cas8b; and (3) Cas6, Cas7 and Cas8b. The combination of Cas6, Cas7 and Cas8b was not able to stably maintain crRNAs, while the combination of Cas5, Cas6 and Cas8b resulted in some amounts of crRNA (Table 3, Figure 6B). Only the combination of Cas5, Cas6 and Cas7 yielded normal crRNA amounts. We also tested all dimeric combinations of the Cas proteins 5-8b, but none of these were active in crRNA protection (Figure 6B) (Table 3). Taken together, these results indicate that the Cas6 protein is necessary for generating crRNAs but not sufficient for its maintenance. *In vivo,* at least Cas5 and Cas7 are additionally required for crRNA generation or stabilisation.

A Cascade like complex in Haloferax

To confirm an association of the Cas5, Cas6 and Cas7 proteins, we expressed a FLAG-Cas7 fusion protein and purified this protein together with all potential interaction partners (Figure 7). The SDS gel of the purified protein fraction shows co-purification of two additional proteins. Mass spectrometry analyses of the proteins showed that these proteins are Cas5 and Cas6. To further investigate the stoichiometry of Cas5, Cas6 and Cas7 in the complex by mass spectrometry, we used the iBAQ quantification approach (53,57). The isolated complex was spiked into a mixture of quantified standard proteins (UPS2), which spanned a concentration range of five orders of magnitude, and the complete protein mixture was then digested in solution. The iBAQ values obtained for the Cas proteins were

calibrated using a linear function derived from the iBAQ values for the standard proteins. The derived protein concentrations indicate a Cas5:Cas6:Cas7 stoichiometry of 1.7: 1: 8.5. This low:low:high type stoichiometry is generally in agreement with stoichiometries previously observed for Cascade-type protein complexes in *E. coli* (58) and *P. aeruginosa* (24). It should be noted that bottom-up mass spectrometric methods like iBAQ show limited accuracy for determining high stoichiometries in protein complexes, as the ratio accuracy is largely limited by the accuracy of determining the higher protein copy numbers in complexes.

To investigate whether crRNAs are likewise part of the complex, we isolated RNA from the FLAG-Cas7 purified protein fraction and determined the nature of the RNA using a northern blot (Figure 7C). Hybridisation with a probe against spacer 1 from CRISPR locus P1 showed that the RNA is indeed crRNA. In addition to the crRNA of about 68 nucleotides length a shorter crRNA of about 51 nucleotides is detected, suggesting a secondary processing event. These shorter crRNAs are also visible in other northern analyses (Figure 6, (42)) but not as prominent as in the FLAG-Cas7 purified fraction.

In summary, we could show that Cas5, Cas6 and Cas7 -as well as crRNA- co-purify and that this Cascade-like complex has a similar composition as the *E. coli* Cascade complex.

DISCUSSION

The Cas6 family of proteins is very diverse; only some structural elements, such as the ferredoxin fold and the glycine-rich loop, are conserved. The binding of the pre-crRNA and the catalytic site, however, differ from protein to protein. To learn more about this protein family, more data are required regarding several additional Cas6 proteins. We report here the first *in vivo* systematic analysis of Cas6 protein mutants that investigates the effect of the mutations on the crRNA amount and the interference reaction.

Only three out of 21 tested mutations had a severe effect on the detectable crRNA amount: His41Ala, Gly256Ala and Gly258Ala. Based on the alignment with other Cas6 proteins, His41 is conserved in the archaeal Cas6 proteins from *M. maripaludis* (16) and *P. furiosus* (10), where it was shown to be important for catalysis in *in vitro* analyses. While all of the Cas6 proteins investigated had a histidine essential for catalytic activity, one of the Cas6 proteins from *S. solfataricus* (Sso1437) has been shown to lack such a catalytic histidine (15,17). Instead, several amino acids have been shown to be involved in catalysis: Lys25, Lys28, Lys51 and Arg231. Because *Haloferax* His41 aligns with the catalytic histidines from the Cas6 proteins from *M. maripaludis* (16) and *P. furiosus* (10), it is likely that His41 is also important for catalysis in *Haloferax*. According to the potential structure (Figure 5C) where the *H. volcanii* Cas6 protein is modelled according to the *P. furiosus* Cas6 structure, His41 is located at the proposed catalytic site.

The *in vitro* analyses with *M. maripaludis* and *P. furiosus* proteins illustrated the differences between these proteins because the additional amino acids identified to be important for *in vitro* catalysis were different: His38 and Tyr47 in *M. maripaludis* (16) and Tyr31 and Lys52 in *P. furiosus* (10). The additional histidines in *Haloferax* Cas6 (H21 and H45) were also mutated in this analysis, but no effect was detected. Likewise additional tyrosines and a lysine were mutated: Y15, Y19, Y51 and K22, but, again, the mutations did not change the amount of crRNA or interference. Therefore, *in vivo* either only His41 is important for catalysis or additional amino acids that were not mutated in this study are required.

Mutation of all four glycines in the glycine-rich loop of the *P. furiosus* Cas6 revealed that this mutant had no RNA cleavage activity anymore but maintained a weak RNA binding activity (32). The authors hypothesised that the path of the RNA on the Cas6 protein overlaps with the G-rich loop, explaining the effect of mutations in the glycine-rich loop. In *Haloferax*, two glycines from the glycine-rich loop are also important for generation of crRNAs or crRNA stability *in vivo*; the other two glycines are not important and, likewise, neither is Phe257, which is located in the glycine-rich loop between the two important glycines (Gly256 and Gly258). This finding might suggest that only the two important glycines (out of all amino acids from the glycine rich loop) interact with parts of the crRNA.

Two mutations, Ser115 and Ser224, resulted in a higher amount of crRNA. According to the structural model, the two serines might be located on the surfaces where the path of the crRNA is predicted to be in the *P. furiosus* Cas6 (Figure 5C). This model would suggest an interaction of the two serines with the crRNA. A mutation in an amino acid important for crRNA binding might weaken product binding and thereby enhance product turnover as observed in

the *T. thermophilus* Cas6e protein (14,25).
The experiments to reveal the impact of the mutations on the interference reaction showed that the amino acids identified as important for cleavage and crRNA stability were also important for interference. This result was expected because crRNAs are essential for the interference reaction. The His41Ala mutation resulted in reduced amounts of crRNA (37%) (Table 1) and the interference reaction was only reduced by factor 0.2. This might suggest that the amount of crRNA is important for an efficient interference reaction.
Amino acids that are not required for catalysis but might be important for the interaction of the Cas6 protein with the Cascade complex were not identified. A final conclusion on the nature of the catalytic site and binding residues of the Cas6 protein will only be reached by structure analysis of the *H. volcanii* Cas6 protein.

The Cascade I-B complex requires Cas5, Cas6 and Cas7 for maintaining a stable crRNA population

A *Haloferax* strain in which all *cas* genes, with the exception of *cas6*, have been removed is not able to stably maintain crRNAs. crRNAs can only be stably maintained if Cas5, Cas6 and Cas7 are present. This finding suggests the presence of a Cascade-like complex for crRNA processing and stabilisation in the *H. volcanii* type I-B system, similar to the type I-A (Cascade I-A) and I-E (Cascade I-E) systems. The Cascade I-B complex clearly requires the Cas proteins 5-7 for maintaining a stable crRNA population. This hypothesis is confirmed by the co-purification of Cas5, Cas6 and crRNA molecules with a FLAG-tagged Cas7.
Interestingly, the combination of Cas proteins Cas5, Cas6 and Cas8 results in some amounts of crRNAs (32 %, Table 3). Thus, the crRNA is generated but the amount of crRNA is not as high as it is in wild type strains. This result might suggest that the missing Cas7 protein is required for efficient stabilisation of the crRNA. In addition, this observation suggests that Cas8b, the large subunit of the type I-B system (59), can protect the crRNA to some extent. The large subunit of the *E. coli* type I-E system (Cse1) was shown to be located close to the crRNA 5´ end, interacting with the PAM sequence of the target DNA (60,61). Further, it was suggested that Cse1 interacts with two bases in the 5´ handle of the crRNA (62). Thus the *Haloferax* Cas8b might protect the crRNA from degradation by binding to its 5´ end. Taken together this suggests that the large subunits of the I-B and the I-E system have similar roles in binding the crRNA and PAM recognition.
A comparison of the Cascade complexes from the different type I systems (IA (22), IE (58,63,64) and IF (24,63)) shows that they all contain multiple copies of the Cas7 protein (in type I-F this is probably the Csy3 (59,63)) and that the Cas7 protein forms the backbone of the complex (14). In addition, type I-A and I-E complexes contain the Cas5 protein that interacts with Cas7. The Cas6 protein is also associated with the type I systems either as an integral part (I-E and I-F) or transiently (I-A) (14). Furthermore, these complexes can contain a small subunit (Csa5 in I-A and Cse2 in I-E) and/or a large subunit (Cas8 in type I-A , Cse1 in type I-E and Csy1 in type I-F) (14). The data presented here for the *Haloferax* type I-B Cascade complex fit very well to these general observations for the type I Cascade complexes. Cas5, Cas6 and Cas7 co-purify, showing that the *Haloferax* complex consist of these subunits. Furthermore, the three Cas proteins are present in almost the same ratio as in the *E. coli* Cascade complex. The crRNA obviously needs Cas5-7 proteins to be generated and to be stably maintained. Structural analysis will show whether the Cas7 protein likewise forms the backbone of the complex binding and protecting the crRNA and whether the Cas8b protein represents the large subunit of the complex.
In addition, these data clearly show that the Cas1-4 proteins are not involved in this reaction. This is in contrast to the results observed with the type I-B system in *Haloferax mediterranii*, in which the deletion of the *cas1*, *cas3* and *cas4* genes resulted in lower levels of crRNA (12).

Acknowledgements
We like to thank Elli Bruckbauer for expert technical help, in addition we thank all members of the Marchfelder lab and members of the Forschergruppe 1680 for helpful discussions.

REFERENCES

1. Al-Attar, S., Westra, E. R., van der Oost, J., and Brouns, S. J. (2011) Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes, *Biol Chem* **2011**, 277-289
2. 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.
3. Bhaya, D., Davison, M., and Barrangou, R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation, *Annu Rev Genet* **45**, 273-297.
4. Garneau, J. E., Dupuis, M. E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A. H., and Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA, *Nature.* **468**, 67-71.
5. Garrett, R. A., Vestergaard, G., and Shah, S. A. (2011) Archaeal CRISPR-based immune systems: exchangeable functional modules, *Trends Microbiol.* **19**, 549-556. Epub 2011 Sep 2022.
6. Marchfelder, A., Fischer, S., Brendel, J., Stoll, B., Maier, L. K., Jäger, D., Prasse, D., Schmitz, R., and Randau, L. (2012) Small RNAs for Defence and Regulation in Archaea, *Extremophiles* **16**, 685-696
7. 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. doi: 710.4161/rna.24321. Epub 22013 Apr 24325.
8. 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., van der Oost, J., and Koonin, E. V. (2011) Evolution and classification of the CRISPR-Cas systems, *Nature Reviews Microbiology* **9**, 467-477. Epub 2011 May 2019.
9. Charpentier, E., van der Oost, J., and White, M. F. (2013) crRNA biogenesis. in *CRISPR-Cas systems* (Barrangou, R., and van der Oost, J. eds.), Springer - Verlag, Berlin - Heidelberg. pp 115-144
10. Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M., and Terns, M. P. (2010) Binding and cleavage of CRISPR RNA by Cas6, *RNA.* **16**, 2181-2188. doi: 2110.1261/rna.2230110. Epub 2232010 Sep 2230130.
11. 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.
12. Li, M., Liu, H., Han, J., Liu, J., Wang, R., Zhao, D., Zhou, J., and Xiang, H. (2013) Characterization of CRISPR RNA biogenesis and Cas6 cleavage-mediated inhibition of a provirus in the haloarchaeon Haloferax mediterranei, *J Bacteriol.* **195**, 867-875. doi: 810.1128/JB.01688-01612. Epub 02012 Dec 01614.
13. Nickel, L., Weidenbach, K., Jager, D., Backofen, R., Lange, S. J., Heidrich, N., and Schmitz, R. A. (2013) Two CRISPR-Cas systems in Methanosarcina mazei strain Go1 display common processing features despite belonging to different types I and III, *RNA Biol* **10**, 5

14. Reeks, J., Naismith, J. H., and White, M. F. (2013) CRISPR interference: a structural perspective, *Biochem J.* **453**, 155-166. doi: 110.1042/BJ20130316.
15. Reeks, J., Sokolowski, R. D., Graham, S., Liu, H., Naismith, J. H., and White, M. F. (2013) Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing, *Biochem J.* **452**, 223-230. doi: 210.1042/BJ20130269.
16. Richter, H., Zoephel, 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. doi: 9810.1093/nar/gks9737. Epub 2012 Aug 9888.
17. Shao, Y., and Li, H. (2013) Recognition and cleavage of a nonstructured CRISPR RNA by its processing endoribonuclease Cas6, *Structure.* **21**, 385-393.
18. Haft, D. H., Selengut, J., Mongodin, E. F., and Nelson, K. E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes, *PLoS Comput Biol.* **1**, e60. Epub 2005 Nov 2011.
19. Voß, B., Georg, J., Schon, V., Ude, S., and Hess, W. R. (2009) Biocomputational prediction of non-coding RNAs in model cyanobacteria, *BMC Genomics.* **10**, 123.
20. Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, 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.
21. 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.
22. Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copié, V., Young, M. J., White, M. F., and Lawrence, C. M. (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
23. 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 Haloferax volcanii CRISPR/Cas system I-B, *RNA Biol* **10**, 5
24. 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. Epub 12011 May 10092.
25. Sashital, D. G., Jinek, M., and Doudna, J. A. (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3, *Nat Struct Mol Biol.* **18**, 680-687. Epub 2011 May 2015.
26. Sternberg, S. H., Haurwitz, R. E., and Doudna, J. A. (2012) Mechanism of substrate selection by a highly specific CRISPR endoribonuclease, *RNA.* **18**, 661-672. doi: 610.1261/rna.030882.030111. Epub 032012 Feb 030816.

27. Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M., and Macmillan, A. M. (2011) Recognition and maturation of effector RNAs in a CRISPR interference pathway, *Nat Struct Mol Biol.* **18**, 688-692. Epub 2011 May 2015.
28. Reeks, J., Graham, S., Anderson, L., Liu, H., White, M. F., and Naismith, J. H. (2013) Structure of the archaeal Cascade subunit Csa5: Relating the small subunits of CRISPR effector complexes, *RNA Biol.* **10**, 762-769. doi: 710.4161/rna.23854. Epub 22013 Apr 23822.
29. Juranek, S., Eban, T., Altuvia, Y., Brown, M., Morozov, P., Tuschl, T., and Margalit, H. (2012) A genome-wide view of the expression and processing patterns of Thermus thermophilus HB8 CRISPR RNAs, *RNA.* **18**, 783-794. Epub 2012 Feb 2021.
30. Hatoum-Aslan, A., Maniv, I., and Marraffini, L. A. (2011) Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site, *Proc Natl Acad Sci U S A.* **108**, 21218-21222. Epub 22011 Dec 21212.
31. Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., and Terns, M. P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex, *Cell.* **139**, 945-956.
32. Wang, R., Preamplume, G., Terns, M. P., Terns, R. M., and Li, H. (2011) Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage, *Structure.* **19**, 257-264.
33. 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. doi: 2010.1261/rna.033100.033112. Epub 032012 Sep 033124.
34. 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. doi: 1510.1016/j.str.2012.1506.1016. Epub 2012 Jul 1526.
35. Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system, *EMBO J* **2011**, 22
36. Westra, E. R., van Erp, P. B., Kunne, T., Wong, S. P., Staals, R. H., Seegers, C. L., Bollen, S., Jore, M. M., Semenova, E., Severinov, K., de Vos, W. M., Dame, R. T., de Vries, R., Brouns, S. J., and van der Oost, J. (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3, *Mol Cell.* **46**, 595-605. doi: 510.1016/j.molcel.2012.1003.1018. Epub 2012 Apr 1019.
37. Kunin, V., Sorek, R., and Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats, *Genome Biol.* **8**, R61.
38. Wang, R., Zheng, H., Preamplume, G., Shao, Y., and Li, H. (2012) The impact of CRISPR repeat sequence on structures of a Cas6 protein-RNA complex, *Protein Sci.* **21**, 405-417. doi: 410.1002/pro.2028. Epub 2012 Feb 1009.

39. Lange, S. J., Alkhnbashi, O. S., Rose, D., Will, S., and Backofen, R. (2013) CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems., *Nucleic Acids Res*, in press
40. Maier, L. K., Fischer, S., Stoll, B., Brendel, J., Pfeiffer, F., Dyall-Smith, M., and Marchfelder, A. (2012) The immune system of halophilic archaea, *Mob Genet Elements.* **2**, 228-232.
41. Maier, L. K., Stoll, B., Brendel, J., Fischer, S., Pfeiffer, F., Dyall-Smith, M., and Marchfelder, A. (2013) The ring of confidence: a haloarchaeal CRISPR/Cas system, *Biochem Soc Trans.* **41**, 374-378. doi: 310.1042/BST20120263.
42. 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**, 33351-33365
43. 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 Haloferax volcanii, *Appl Environ Microbiol* **76**, 1759-1769
44. Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, N. Y.
45. Fröls, S., Dyall-Smith, M., and Pfeifer, F. (2012) Biofilm formation by haloarchaea, *Environ Microbiol.* **14**, 3159-3174. doi: 3110.1111/j.1462-2920.2012.02895.x. Epub 02012 Oct 02812.
46. Fischer, S., Benz, J., Späth, B., Maier, L.-K., Straub, J., Granzow, M., Raabe, M., Urlaub, H., Hoffmann, J., Brutschy, B., Allers, T., Soppa, J., and Marchfelder, A. (2010) The archaeal Lsm protein binds to small RNAs, *J Biol Chem* **285**, 34429-34438
47. Allers, T., Ngo, H. P., Mevarech, M., and Lloyd, R. G. (2004) Development of additional selectable markers for the halophilic archaeon Haloferax volcanii based on the leuB and trpA genes, *Appl Environ Microbiol* **70**, 943-953
48. Cline, S. W., Schalkwyk, L. C., and Doolittle, W. F. (1989) Transformation of the archaebacterium Halobacterium volcanii with genomic DNA, *J Bacteriol.* **171**, 4987-4991.
49. Nieuwlandt, D. T., Palmer, J. R., Armbruster, D. T., Kuo, Y.-P., Oda, W., and Daniels, C. J. (1995) A rapid procedure for the isolation of RNA from Haloferax volcanii. in *Archaea: A laboratory manual* (Robb, F. T., Place, A. R., Sowers, K. R., Schreier, H. J., DasSarma, S., and Fleischmann, E. M. eds.), Cold Spring Harbour Press, New York. pp 161
50. Stroud, A., Liddell, S., and Allers, T. (2012) Genetic and Biochemical Identification of a Novel Single-Stranded DNA-Binding Complex in Haloferax volcanii, *Front Microbiol.* **3:224.**, 10.3389/fmicb.2012.00224. Epub 02012 Jun 00218.
51. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, *Anal Chem.* **68**, 850-858.
52. 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. Epub 2008 Jul 2001.
53. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein

quantification, *Nat Biotechnol.* **26**, 1367-1372. doi: 1310.1038/nbt.1511. Epub 2008 Nov 1330.

54. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web: a case study using the Phyre server, *Nat Protoc.* **4**, 363-371. doi: 310.1038/nprot.2009.1032.
55. Gudbergsdottir, S., Deng, L., Chen, Z., Jensen, J. V., Jensen, L. R., She, Q., and Garrett, R. A. (2011) Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers, *Mol Microbiol* **79**, 35-49
56. Needleman, S. B., and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins, *J Mol Biol.* **48**, 443-453.
57. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011) Global quantification of mammalian gene expression control, *Nature.* **473**, 337-342. doi: 310.1038/nature10098.
58. 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., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J., van der Oost, J., and Brouns, S. J. (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade, *Nat Struct Mol Biol* **3**, in press
59. Makarova, K. S., Aravind, L., Wolf, Y. I., and Koonin, E. V. (2011) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems, *Biol Direct.* **6:38.**, 10.1186/1745-6150-1186-1138.
60. 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. doi: 610.1016/j.molcel.2012.1003.1020. Epub 2012 Apr 1019.
61. 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. doi: 410.1038/nature10402.
62. Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., and Brouns, S. J. (2013) Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition, *PLoS Genet.* **9**, e1003742. doi: 1003710.1001371/journal.pgen.1003742. Epub 1002013 Sep 1003745.
63. van Duijn, E., Barbu, I. M., Barendregt, A., Jore, M. M., Wiedenheft, B., Lundgren, M., Westra, E. R., Brouns, S. J., Doudna, J. A., van der Oost, J., and Heck, A. J. (2012) Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced shot-palindromic-repeats (CRISPR)-associated protein complexes from Escherichia coli and Pseudomonas aeruginosa, *Mol Cell Proteomics.* **11**, 1430-1441. doi: 1410.1074/mcp.M1112.020263. Epub 022012 Aug 020223.
64. Wiedenheft, B., Sternberg, S. H., and Doudna, J. A. (2012) RNA-guided genetic silencing systems in bacteria and archaea, *Nature* **482**, 331-338
65. Randau, L. (2012) RNA processing in the minimal organism Nanoarchaeum equitans, *Genome Biol.* **13**, R63.
66. Scholz, I., Lange, S. J., Hein, S., Hess, W. R., and Backofen, R. (2013) CRISPR-Cas systems in the cyanobacterium Synechocystis sp. PCC6803

exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein, *PLoS One.* **8**, e56470. doi: 56410.51371/journal.pone.0056470. Epub 0052013 Feb 0056418.

Table and Figure Legends

Table 1. Effect of amino acid changes on the amount of crRNA.
The amounts of crRNA detectable in the wild type *H. volcanii* strain, the *cas6* gene deletion strain, the *cas6* gene deletion strain complemented with the wild type *cas6* gene as well as the *cas6* mutants were determined and assessed relative to the amounts of 5S rRNA detected (Experimental Procedures). The amount of crRNA detected in the *cas6* deletion strain complemented with the wild type *cas6* gene on a plasmid was set to 100 %.

Table 2. Cas6 mutations that result in reduced transformation rates. The *cas6* deletion strain was first transformed with the *cas6* wild-type gene or the *cas6* mutants. Next, this strain was transformed with the invader plasmid (pTA352-PAM3 and pTA352-PAM9) to trigger the defence reaction. The invader plasmid contains the protospacer sequence that is detected by a crRNA from *Haloferax* and the PAM sequence (PAM3: TTC and PAM9:ACT) (42). The factor by which the transformation rate was changed was determined.

Table 3. Cas proteins important for crRNA production and stability. The strain, for which the complete *cas* gene cluster was deleted (H26ΔcasCluster28), was transformed with the different combinations of *cas* genes. The amount of crRNA was determined and assessed relative to the amount of 5S rRNA detected (Experimental Procedures). The amount of crRNA in the wild type strain was set to 100%.

Figure 1. Location of *H. volcanii cas* genes and CRISPR loci. The *cas* gene cluster is located on the mini chromosome pHV4 and encodes the Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7 and Cas8b proteins, the latter being the signature protein for subtype I-B. The cluster is flanked by two CRISPR RNA loci (P1 and P2). CRISPR locus P1 contains 17 repeats and 16 spacers (shown in dark grey and light grey, respectively), and locus P2 contains 12 repeats and 11 spacers. The third CRISPR RNA locus is encoded on the main chromosome and contains 25 repeats and 24 spacers. The leader region (L) at the 5′ end of the CRISPR locus contains the promoter. The 5′ end of the *cas6* mRNA was mapped with 5′ RACE and is indicated in the box above the *cas* gene cluster by an arrow. The overlap of the *cas6* and *cas8b* genes is shown in the box below the *cas* gene cluster.

Figure 2. Alignment of Cas6 proteins. Cas6 proteins from *Haloferax, H. lacusprofundi, M. maripaludis, P. furiosus, E. coli* and *P. aeruginosa* are aligned to identify conserved amino acids. Amino acids from *M. maripaludis, P. furiosus, E. coli* and *P. aeruginosa* that were shown to be essential for catalysis are indicated by red boxes. The amino acids mutated in this study are indicated by asterisks. Amino acids shown to be important for the stable crRNA population in this study are boxed in red (mutations result in reduced crRNA amounts) or green (mutations result in higher crRNA amounts).

Figure 3. The haloarchaeal CRISPR-Cas systems are different from other systems. The haloarchaeal CRISPR-Cas systems are distinct from published systems where the Cas6 protein has been functionally characterised. The circular hierarchical tree represents the sequence and structure similarity of repeats from all publicly available genomes, taken from the CRISPRmap web server (39). The locations of repeats associated with previously characterised Cas6 are highlighted with red lines: *Clostridium thermocellum* (16), *P. furiosus* (10,31,32), *E. coli* (20),

Thermus thermophilus (25,27,29), *P. aeruginosa* (21,26), *Nanoarchaeum equitans* (65), *Synechocystis* (66), *Methanosarcina mazei* (13), *Staphylococcus epidermidis* (30) and *Methanococcus maripaludis* (16). The pairwise alignment percent identities in comparison to the Cas6 protein in *H. volcanii* are given in square brackets. For the CRISPRmap tree, brown branches represent CRISPRs from bacteria, the blue-green branches represent CRISPRs from archaea, the inner annotation circle represents different conserved structure motifs, the middle circle represents conserved sequence families and the outer circle represents the six superclasses.

Figure 4. The *cas6* deletion strain does not generate crRNAs. To determine the biological function of the Cas6 protein, we deleted the reading frame of this gene in *H. volcanii* and isolated RNA from the deletion (Δ*cas6*) and wild-type strains, which was subsequently separated on 8% PAGE and transferred to a membrane. Hybridisation with a probe against the spacer1 from locus P1 showed that no crRNAs were generated in the Δ*cas6* strain. Complementation of this strain with the *cas6* gene on a plasmid resulted in the rescue of crRNA production.

Figure 5. Effect of *cas6* gene mutations on crRNA amounts. Mutations were introduced into the *cas6* gene (Figure 2), and the *cas6* deletion strain was transformed with the mutant genes. **A & B.** RNA from all strains was isolated, separated on 8% PAGE and subsequently transferred to northern membranes that were then hybridised with a probe against spacer 1 from locus P1. Determination of the amount of crRNA in relation to the amount of RNA loaded (measured by the 5S rRNA hybridisation) showed that only transformation with three variants (His41, Gly256 and Gly258) resulted in reduced amounts of crRNAs. Higher amounts of crRNA were determined in mutants Ser115 and Ser224. On the left, a DNA size marker in nucleotides is shown. The hybridisation with the spacer 1 from CRISPR locus P1 is shown at the top, and hybridisation with a probe against 5S rRNA is shown at the bottom. Lane marker: the DNA size marker; lane wt: RNA from wild type cells; lane Δcas6: RNA from the *cas6* gene deletion strain; lane Δ*cas6* + *cas6*: RNA from the *cas6* gene deletion strain complemented with the wild type *cas6* gene from a plasmid; lanes + Y15A, +Y19A, +H21A, +K22A, +R26A, +W28A, +H41A, +H45A, +F49A, +Y51A, +R66A, +S115A, +T116A, +L172A, S224A, W226A, +G248A, G250A, +G256A, +F257A, G258A: RNA from the *cas6* deletion strain complemented with the mutant *cas6* gene from a plasmid. **C. Model of the *H. volcanii* Cas6 protein.** Similarity searches in the Phyre database (54) show the closest structures related to the *H. volcanii* Cas6 protein to be the *P. furiosus* Cas6 structure (32). The *H. volcanii* protein was modelled according to the published *P. furiosus* structure, and the amino acid mutations that changed the amounts of crRNA are shown. H41 is located where the catalytic site in the Pfu Cas6 protein was proposed; according to this model, the two glycines and serines might be located on the surface where the crRNA could be located.

Figure 6. Cas6 alone is not sufficient for crRNA processing and maintenance. A. Constructs used for complementation. To determine which *cas* genes are required for crRNA maturation, different *cas* gene constructs were generated for transformation of the H26Δ*cas*Cluster28 strain which has the complete *cas* gene cluster deleted. **B. Effect of different Cas proteins on crRNA amounts.** RNA isolated from the H26Δ*cas*Cluster28 strains, transformed with the different *cas* gene combinations, was transferred to membranes that were subsequently hybridised with a probe against spacer1 from CRISPR locus P1. On the left, a DNA size marker (in nucleotides) is shown. The hybridisation with the spacer 1 from CRISPR locus P1 is shown at the top, hybridisation with

a probe against 5S rRNA is shown at the bottom. Lane marker: DNA size marker; lane wt: RNA from wild type cells; lane Δ*cas*: RNA from the *cas* genes deletion strain; lane Δ*cas* + 68: RNA from the *cas* genes deletion strain complemented with the *cas6* and *cas8* genes from a plasmid; lanes + 67, +65, +87, +85 +75, +675, +685, +687, +6875: RNA from the *cas* genes deletion strain complemented with the different *cas* genes from a plasmid as indicated.

Figure 7. A. and B. Cas5 and Cas6 co-purify with a FLAG-tagged Cas7 protein. A FLAG-tagged Cas7 protein was generated in *Haloferax* cells and purified together with all potential interaction partners using the FLAG-tag. The purified fraction was investigated using a western blot (**A**) and an SDS PAGE (**B**). Three proteins are visible on the Coomassie stained SDS gel. The largest protein is, according to the western blot (that was probed with the FLAG antibody), the FLAG-Cas7 fusion protein. According to mass spectrometry analyses, the two smaller proteins are Cas5 and Cas6, respectively. Lane A: western blot of the FLAG-purified fraction, probed with an anti-FLAG antibody from a plasmid lane B: silver stained SDS PAGE of the FLAG-purified fraction. The protein size marker is given on the left in kDa, the nature of the respective proteins is indicated on the right. **C. The Cascade I-B complex contains crRNA.** RNA was isolated from the FLAG-Cas7 purified protein fraction, separated by 8% PAGE and transferred to a membrane. Hybridisation with a probe against spacer 1 from CRISPR locus P1 detected two RNAs with approx. 68 nucleotides and 51 nucleotides in length, which correspond to the *Haloferax* crRNAs (42). Shown on the left is a DNA size marker in nucleotides.

Table 1

mutant	processing activity [%]	standard deviation (+/-)	
wt	61	3	
Δ6	3	4	
Δ6+6	100	-	
Y15A	98	6	
Y19A	110	16	
H21A	69	5	
K22A	94	7	
R26A	100	7	
W28A	105	3	
H41A	**37**	**2**	↓
H45A	81	7	
F49A	84	1	
Y51A	71	2	
R66A	104	2	
S115A	**129**	**6**	↑
T116A	89	2	
L172A	68	3	
S224A	**164**	**6**	↑
W226A	108	3	
G248A	89	5	
G250A	90	5	
G256A	**12**	**2**	↓
F257A	111	3	
G258A	**0**	**0**	↓

Table 2.

first transformation of Δ*cas6* with *cas6* gene and *cas6* gene mutants	**second transformation with invader plasmid**	**reduction of transformation by factor**
---	pTA409-PAM3 and PAM9	no reduction (no interference)
pTA927-*cas6*	pTA352-PAM3 and PAM9	0.003 (interference restored)
pTA927-H41A	pTA352-PAM3 and PAM9	0.2 (reduced interference)
pTA927-S115A	pTA352-PAM3 and PAM9	0.002 (interference restored)
pTA927-S224A	pTA352-PAM3 and PAM9	0.001 (interference restored)
pTA927-G256A	pTA352-PAM3 and PAM9	no reduction (no interference)
pTA927-G258A	pTA352-PAM3 and PAM9	no reduction (no interference)

Table 3

strain	crRNA amount [%]
wt	100
Δ*cas*	2,3
Δ*cas6 + 68*	2,4
Δ*cas6 + 67*	2,6
Δ*cas6 + 65*	3,2
Δ*cas6 + 87*	0
Δ*cas6 + 85*	0,4
Δ*cas6 + 75*	3,8
Δ*cas6 + 675*	96,8
Δ*cas6 + 685*	32,2
Δ*cas6 + 687*	8,0
Δ*cas6 + 6875*	97,9

Figure 1

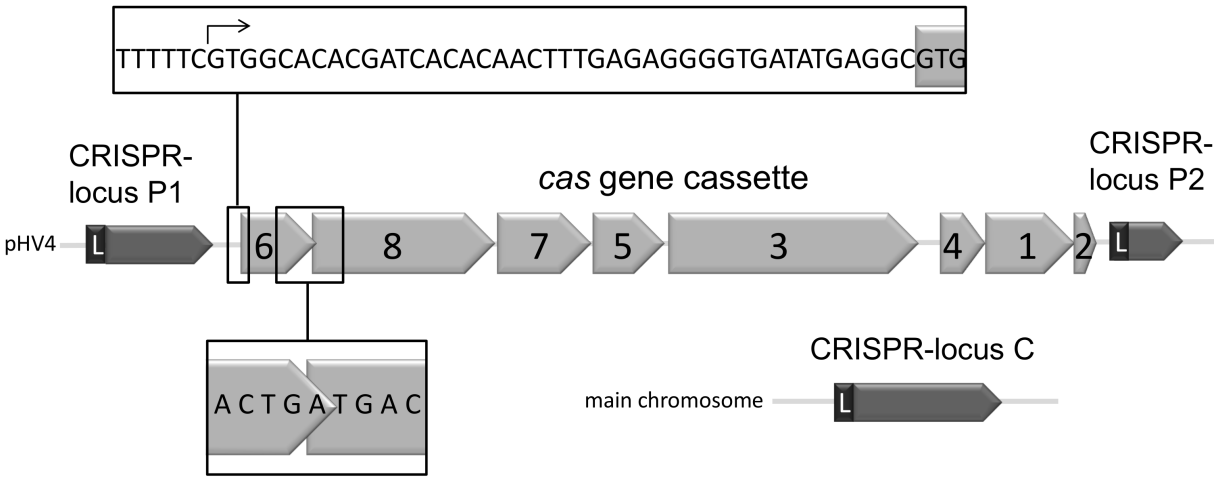

Figure 2

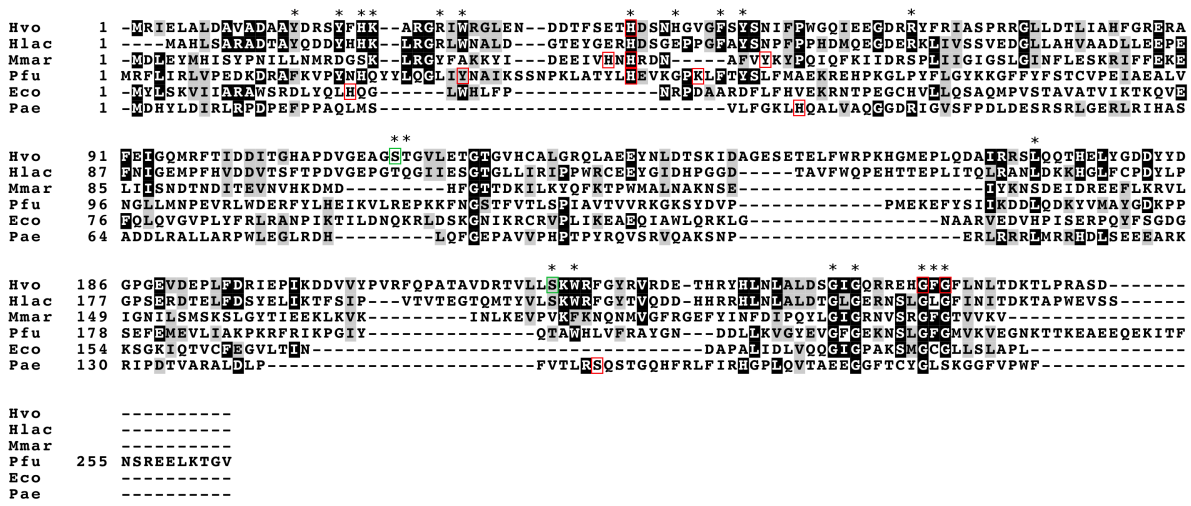

Figure 3

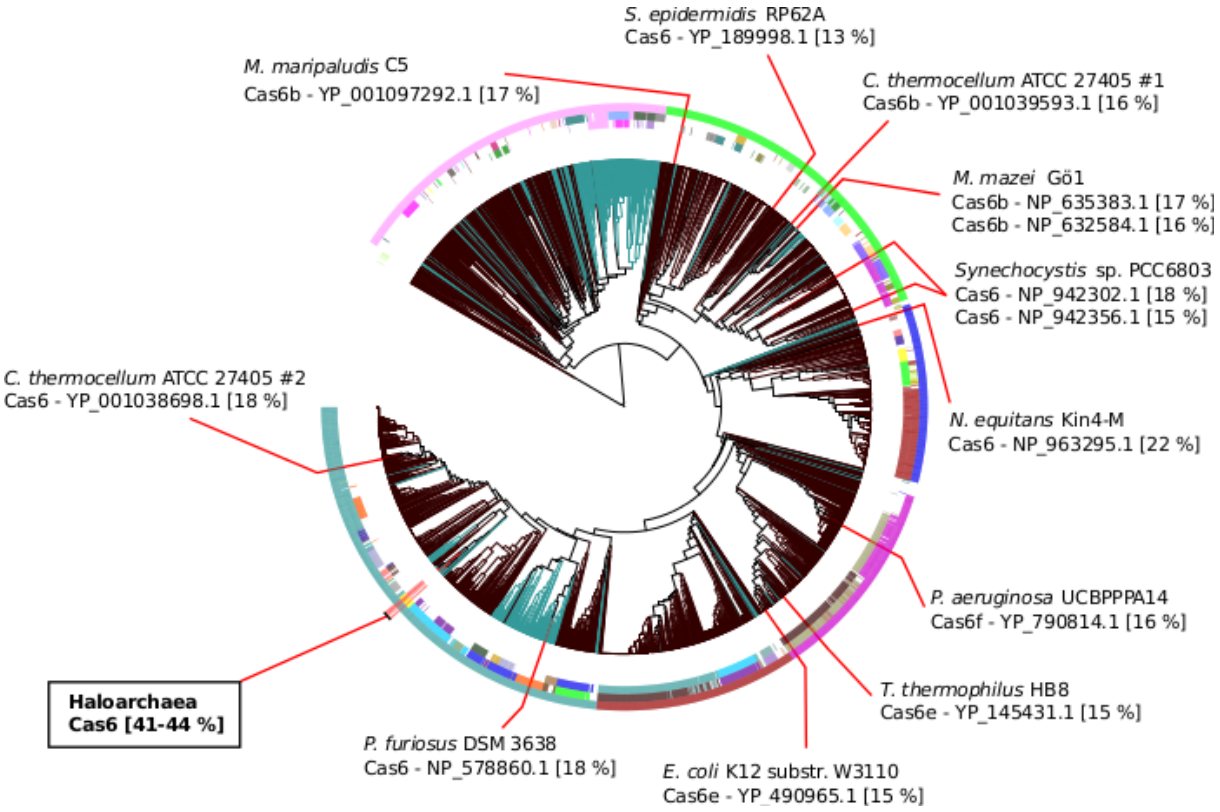

Figure 4

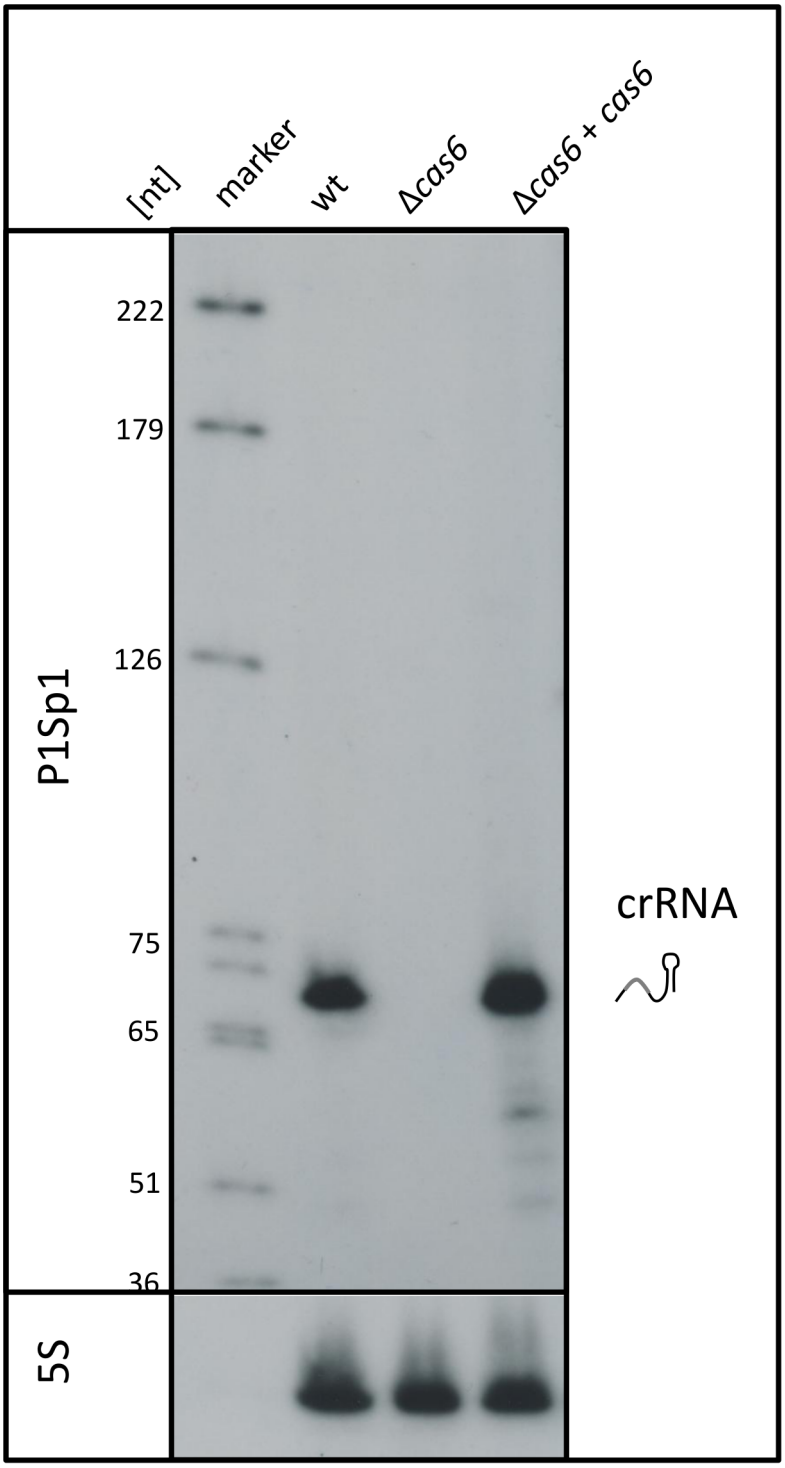

Figure 5A

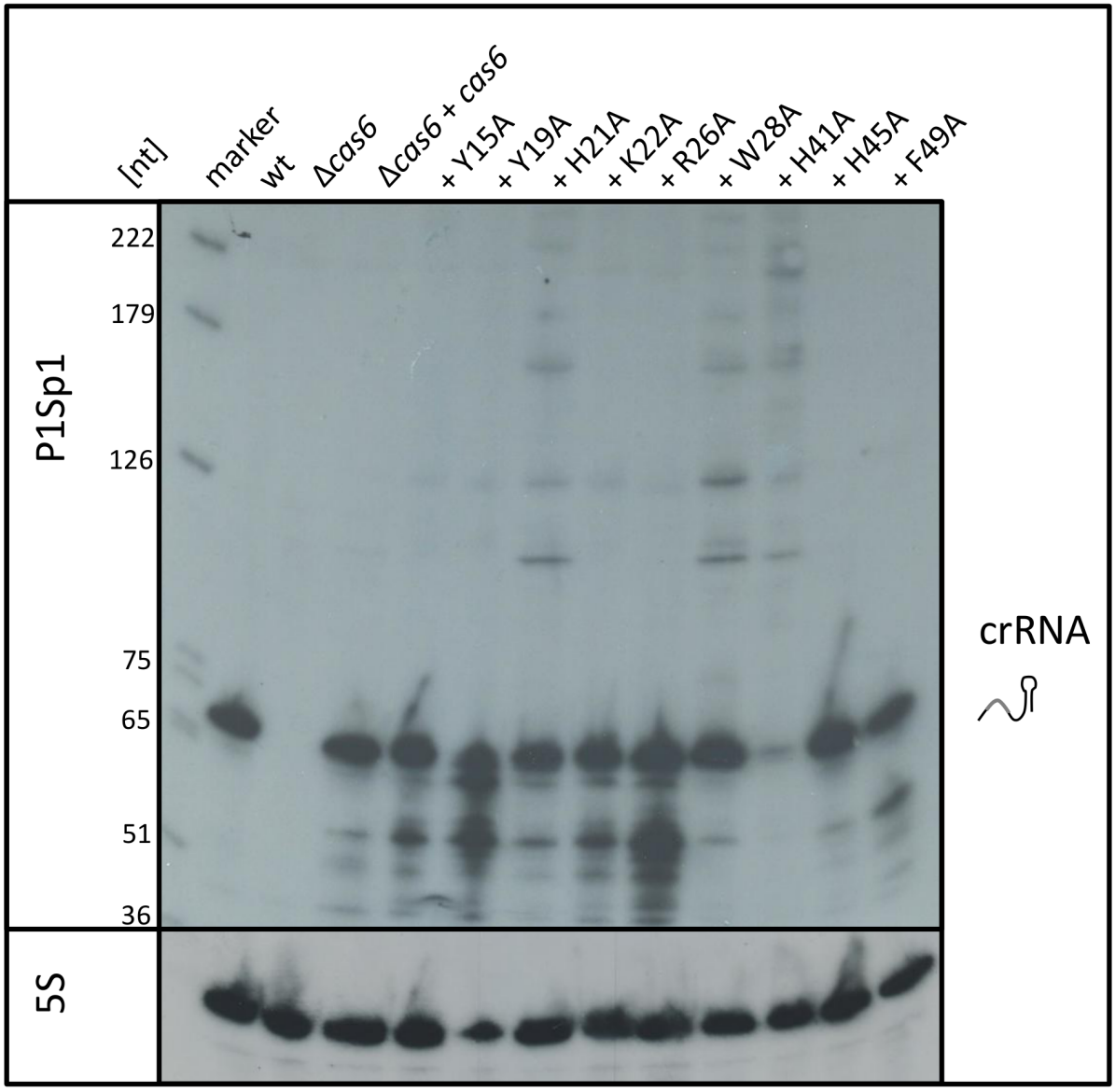

Figure 5B

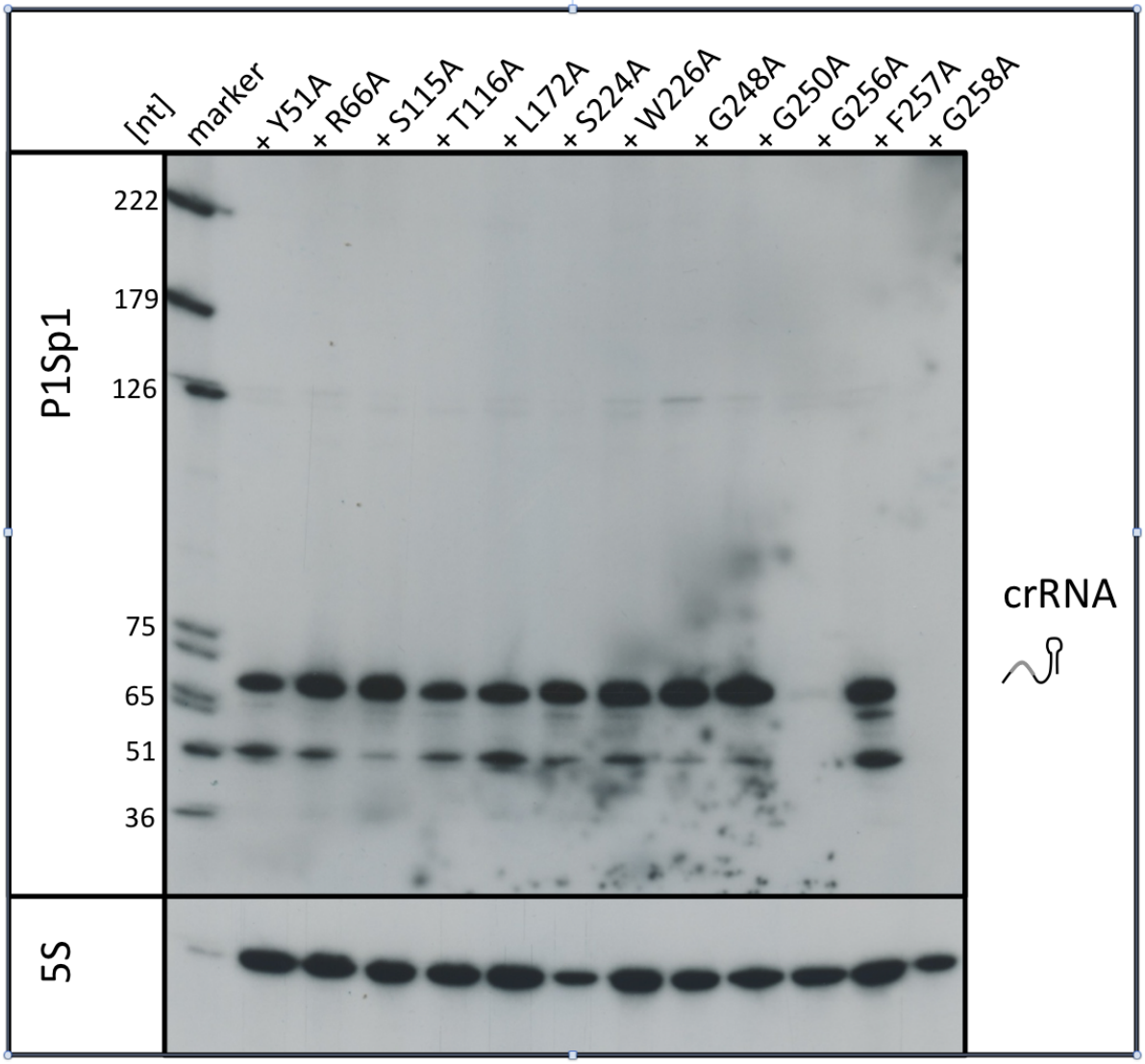

Figure 5C

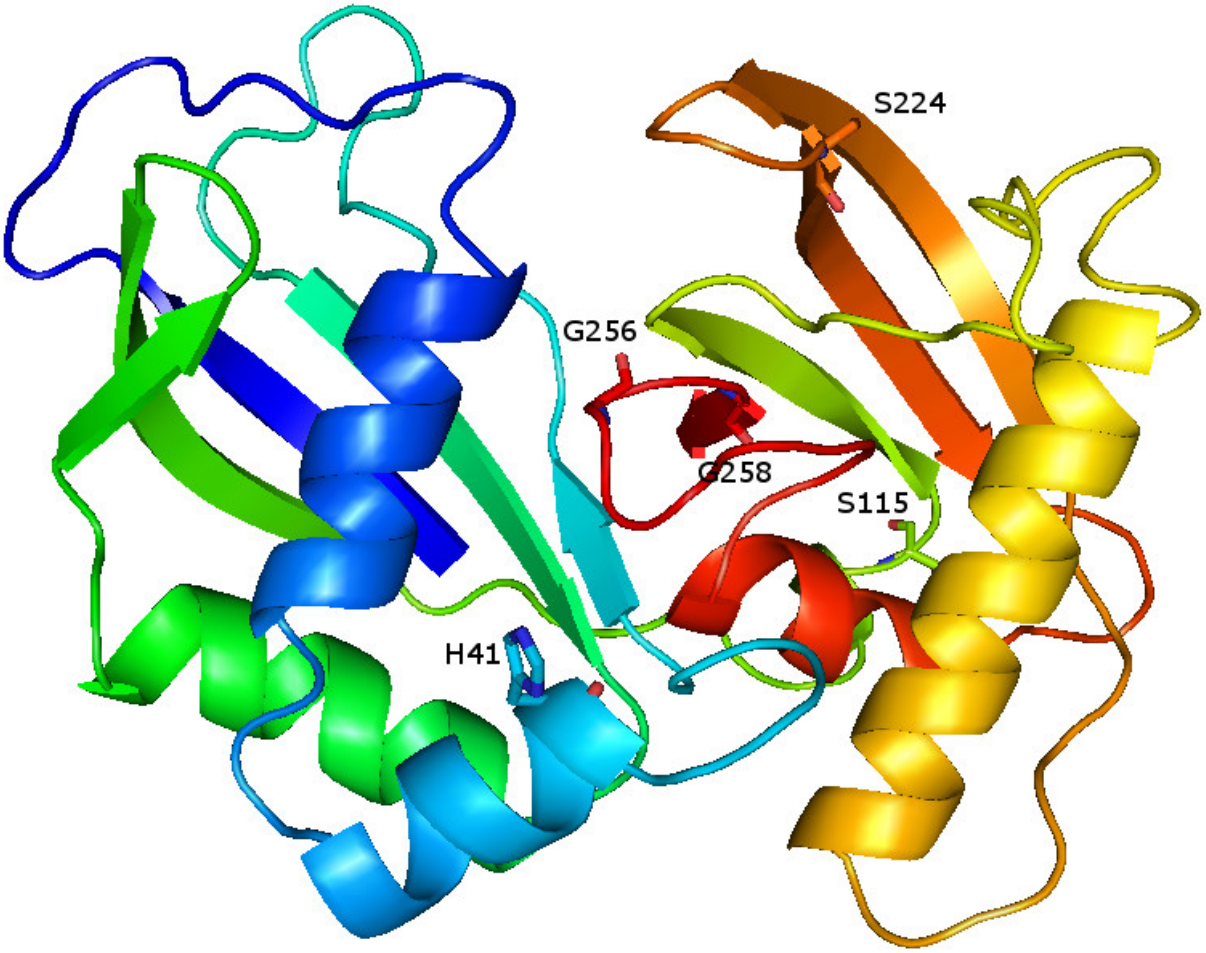

Figure 6A

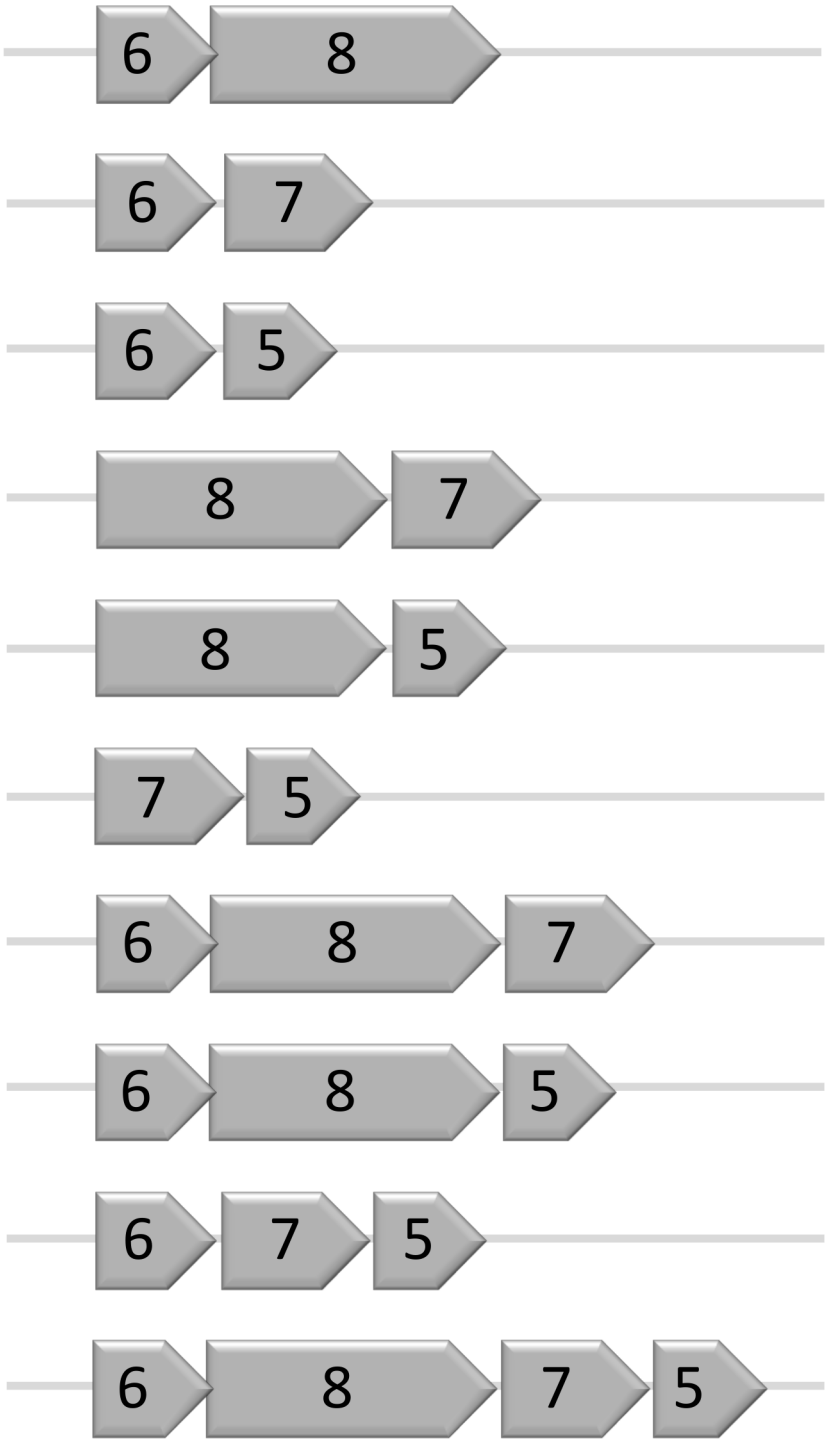

Figure 6B

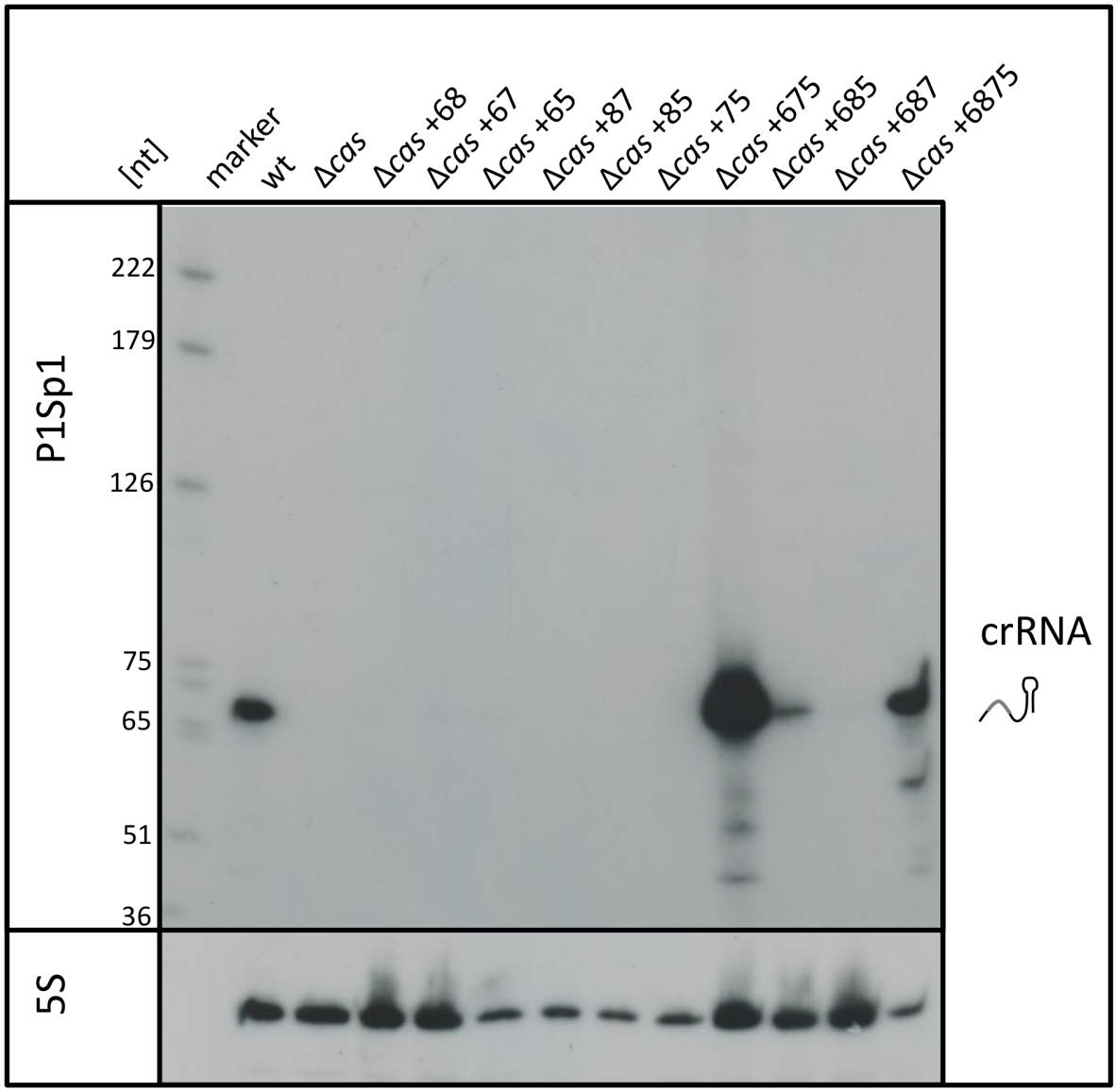

Figure 7 A & B

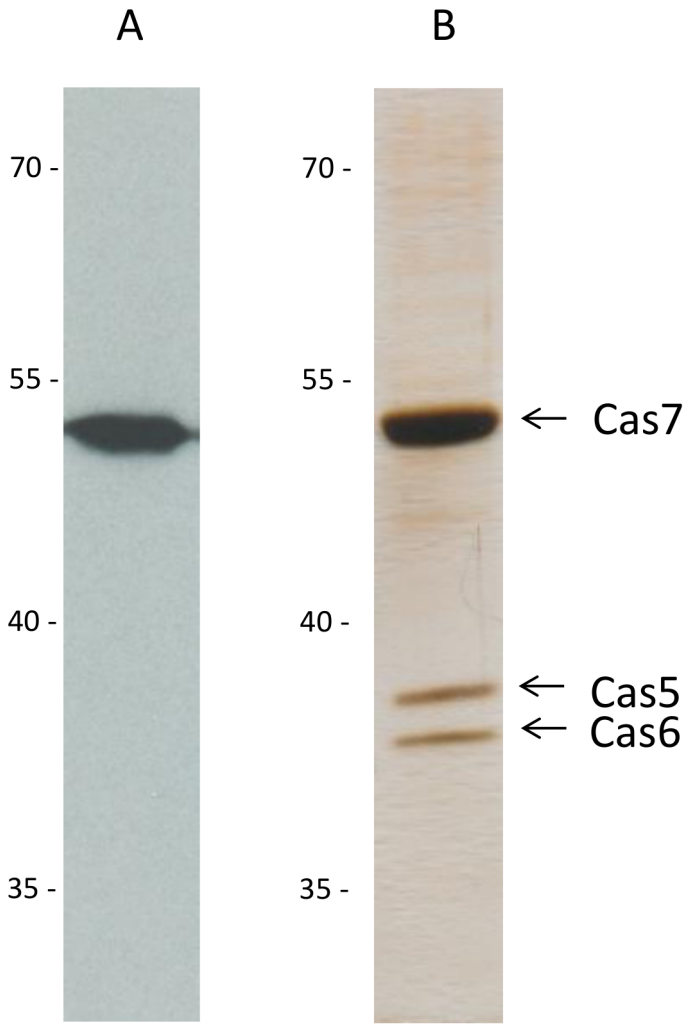

Figure 7C