LETTER

doi:10.1038/nature17945

The CRISPR–associated DNA–cleaving enzyme Cpf1 also processes precursor CRISPR RNA

Ines Fonfara[1,2,3]*, Hagen Richter[2,3]*, Majda Bratovič[2,3,4], Anaïs Le Rhun[1,2,3] & Emmanuelle Charpentier[1,2,3,4]

CRISPR–Cas systems that provide defence against mobile genetic elements in bacteria and archaea have evolved a variety of mechanisms to target and cleave RNA or DNA[1]. The well-studied types I, II and III utilize a set of distinct CRISPR-associated (Cas) proteins for production of mature CRISPR RNAs (crRNAs) and interference with invading nucleic acids. In types I and III, Cas6 or Cas5d cleaves precursor crRNA (pre-crRNA)[2–5] and the mature crRNAs then guide a complex of Cas proteins (Cascade-Cas3, type I; Csm or Cmr, type III) to target and cleave invading DNA or RNA[6–12]. In type II systems, RNase III cleaves pre-crRNA base-paired with *trans*-activating crRNA (tracrRNA) in the presence of Cas9 (refs 13, 14). The mature tracrRNA–crRNA duplex then guides Cas9 to cleave target DNA[15]. Here, we demonstrate a novel mechanism in CRISPR–Cas immunity. We show that type V-A Cpf1 from *Francisella novicida* is a dual-nuclease that is specific to crRNA biogenesis and target DNA interference. Cpf1 cleaves pre-crRNA upstream of a hairpin structure formed within the CRISPR repeats and thereby generates intermediate crRNAs that are processed further, leading to mature crRNAs. After recognition of a 5′-YTN-3′ protospacer adjacent motif on the non-target DNA strand and subsequent probing for an eight-nucleotide seed sequence, Cpf1, guided by the single mature repeat-spacer crRNA, introduces double-stranded breaks in the target DNA to generate a 5′ overhang[16]. The RNase and DNase activities of Cpf1 require sequence- and structure-specific binding to the hairpin of crRNA repeats. Cpf1 uses distinct active domains for both nuclease reactions and cleaves nucleic acids in the presence of magnesium or calcium. This study uncovers a new family of enzymes with specific dual endoribonuclease and endonuclease activities, and demonstrates that type V-A constitutes the most minimalistic of the CRISPR–Cas systems so far described.

Our previous analysis of the intracellular human pathogen *Francisella novicida* U112 by small RNA (sRNA) sequencing identified sRNAs expressed from two CRISPR–Cas loci[13,16] (Extended Data Fig. 1a). As well as for the type II-B locus[13], we detected sRNAs from a CRISPR–Cas locus that resembled the minimal architecture of type II systems but lacked a *cas9* gene. Upstream of the *cas1*, *cas2* and *cas4* genes[17], FTN_1397 was identified as a *cas* gene encoding a protein distinct in sequence from known Cas proteins; this was later named *cpf1* (*cas* gene of *Pasteurella, Francisella*)[17]. This system was recently classified as a type V-A system belonging to the class 2 CRISPR–Cas systems[18,19]. The CRISPR array contains a series of nine spacer sequences separated by 36-nucleotide (nt) repeat sequences. The mature RNAs are composed of a repeat sequence in 5′ and spacer sequence in 3′, similar to the repeat-spacer composition of types I and III systems but distinct from the spacer-repeat composition of type II systems[2,14,20] (Extended Data Fig. 1b). As in type I, the repeat forms a hairpin structure at its 3′ end[20]. Neither the presence of a Cas6 homologue nor the expression of a tracrRNA-like sRNA could be detected in the vicinity of the *F. novicida* type V-A locus, indicating that Cpf1 uses a distinct mode of crRNA biogenesis compared to the mechanisms that have been described thus far[2,4,14].

We investigated whether Cpf1 acts as the single effector enzyme in pre-crRNA processing in type V-A systems. Recombinant *F. novicida* Cpf1 protein was overexpressed, purified and biochemically characterized. In contrast to the recently reported formation of Cpf1 dimers in solution[16], our data reveal a molecular weight of 187 kDa (Extended Data Fig. 2), indicating that Cpf1 is a monomer. This result is corroborated by another study showing the crystal structure of Cpf1 from *Lachnospiraceae* bacterium (LbCpf1). No oligomerization of Cpf1 was observed in the crystals, analytical ultracentrifugation experiments or electron microscopy[21]. The monomeric nature is consistent with Cpf1 forming a complex with the guide crRNA to bind and cleave target DNA because if the active protein was a dimer[16], it would probably require a tandem DNA target site, or alternatively, two different crRNAs targeting the top and bottom strand of the DNA.

In vitro cleavage assays show that Cpf1 processes a pre-crRNA consisting of a full-length repeat-spacer, yielding a 19-nt repeat fragment, and a 50-nt repeat-spacer crRNA intermediate (Fig. 1). Only RNAs with full-length repeat sequences were processed, indicating that the RNA cleavage activity is repeat-dependent (Extended Data Fig. 3a). The observed cleavage site is in good agreement with the data obtained by RNA-seq (Extended Data Fig. 1b) and a recent study[16]. The crRNAs produced *in vitro* represent intermediate forms that undergo further processing at the 5′ and 3′ ends by a nonspecific mechanism *in vivo*. Cpf1 cleaves pre-crRNA four nucleotides upstream of the stem-loop (Fig. 1). The cleavage site is reminiscent of many Cas6 enzymes and Cas5d, which recognize the hairpin of their respective repeats[2,4,5,20]. Cpf1, however, does not cleave directly at the base of the stem-loop, suggesting that the structure is not the only requirement for processing of pre-crRNA. Northern blot analysis using an inducible *Escherichia coli* heterologous system also demonstrates processing of pre-crRNA upon Cpf1 expression (Extended Data Fig. 3b), resulting in the expected RNA fragments.

To investigate the importance of the repeat and its hairpin structure in successful Cpf1 processing, we designed RNAs with mutations that yield either an altered repeat sequence keeping the stem-loop structure or an unstructured repeat. In contrast to the wild-type RNA substrate containing an intact repeat, none of the mutated RNAs was cleaved by Cpf1 (Extended Data Fig. 4a, b). We further designed repeat variants with either single nucleotide mutations between the cleavage site and the stem-loop (a region referred to as repeat recognition sequence (RRS)) or different sizes of the loop and stem regions (Extended Data Fig. 4a). Single nucleotide mutations in the RRS yielded repeat variants that were not, or only poorly, cleaved by Cpf1 (Extended Data Fig. 4c), indicating that these residues between the stem and the cleavage site have a role in processing of the substrate. This can be explained by the distinct secondary structure of crRNA in complex with Cpf1, where

[1]The Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå Centre for Microbial Research (UCMR), Department of Molecular Biology, Umeå University, Umeå 90187, Sweden. [2]Helmholtz Centre for Infection Research, Department of Regulation in Infection Biology, Braunschweig 38124, Germany. [3]Max Planck Institute for Infection Biology, Department of Regulation in Infection Biology, Berlin 10117, Germany. [4]Hannover Medical School, Hannover 30625, Germany.
*These authors contributed equally to this work.

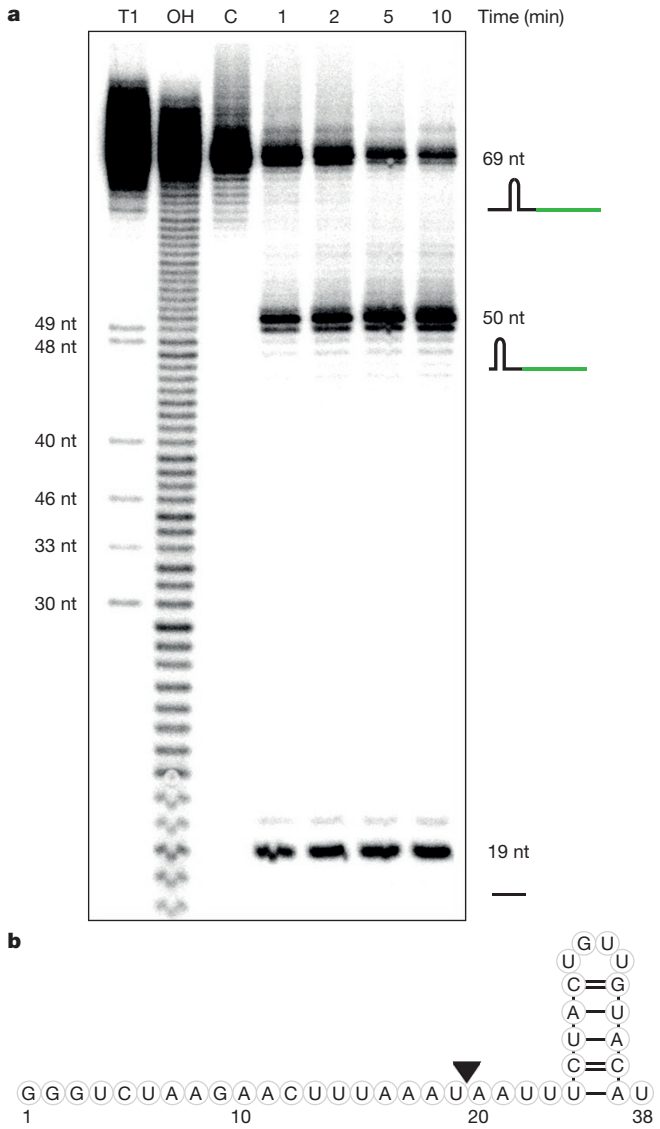


Figure 1 | Cpf1 processes pre-crRNA upstream of the repeat stem-loop structure. a, Denaturing polyacrylamide gel showing the processing of internally labelled 69-nt pre-crRNA (200 nM) by Cpf1 (1 μM) in the presence of 10 mM $MgCl_2$ over 10 min. T1, RNase T1 ladder; OH, alkaline hydrolysis ladder; C, control reaction without Cpf1. Shown is a representative of three independent experiments. **b**, Schematic representation of pre-crRNA repeat structure. The Cpf1 cleavage site is indicated by a black triangle.

the RRS folds back to make contacts with the stem-loop[21]. Changes in the loop region of the repeat structure resulted in reduced cleavage activity for a shorter loop, whereas an increased loop length did not influence cleavage (Extended Data Fig. 4d). Extensive contacts of Cpf1 to the stem-loop of the crRNA[21] explain why alterations of the stem structure yielded non-cleavable substrates. These results highlight the requirement of a stem-loop structure specific in length and sequence for recognition by Cpf1. Thus, the repeat cleavage reaction is highly sequence- and structure-dependent.

To determine the ion dependency of Cpf1 processing activity, we tested a variety of divalent metal ions in RNA cleavage assays. The activity of Cpf1 in pre-crRNA processing was highest when Mg^{2+} was added to the reaction (Extended Data Fig. 5a). Addition of Ca^{2+}, Mn^{2+} or Co^{2+} also mediated cleavage, although not to the level of specificity observed with Mg^{2+}. Equimolar addition of EDTA markedly reduced Cpf1 processing activity. The dependency on Mg^{2+} is in contrast to the ion-independent reaction of Cas6 (types I and III)[2,20] or Cas5d (type I-C)[5]. A Mg^{2+} ion is coordinated in the structure of the crRNA[21]. Whether this ion is required for catalysis or only for stabilization of the tertiary structure has not yet been determined. Thus, our study highlights a novel crRNA biogenesis mechanism in which Cpf1 is a metal-dependent endoribonuclease that cleaves pre-crRNA in a sequence- and structure-specific manner. Similarities in the pre-crRNA processing mechanisms of Cpf1 and Cas6 enzymes of type I and type III systems indicate potential evolution of these ancestral CRISPR–Cas systems through transposition events[18]. This hypothesis is supported by our finding that Cpf1 functions as the endoribonuclease of type V-A systems together with the repeat-spacer composition of mature crRNAs and the requirement for a hairpin structure in the repeat. Bioinformatic analyses indicate that type V systems may be ancestral versions of type II systems. Type V may be considered as a link between class 1 and class 2 systems, which is supported by the recent discovery of a subtype V-B that encodes tracrRNA[18,19].

It was previously shown that Cpf1 acts as the DNA endonuclease guided by crRNA to cleave double-stranded (ds)DNA site-specifically[16]. In accordance with that study, we show that only crRNA containing an intact stem-loop and a sequence complementary to the target DNA mediated Cpf1 DNA cleavage that resulted in a staggered cut producing a 5-nt 5′ overhang (Fig. 2a, b; processed crRNAs (RNA1–3), full-length pre-crRNAs (RNA4–6), mutated crRNAs (RNA7 and 8), Extended Data Figs 6 and 7). Surprisingly, a crRNA with a spacer-repeat arrangement also mediated cleavage by Cpf1, albeit with less efficiency than the wild type. Although the RNA processing activity of Cpf1 is highly dependent on the repeat sequence (sequence mutant, Extended Data Fig. 4a, b), a similar RNA resulted in residual DNA cleavage activity (RNA7, Extended Fig. 6). This might be due to the 3′ end nucleotide of the repeat, which was not mutated and was recently reported to be crucial for DNA targeting[16] and for maintaining the specific tertiary structure of crRNA[21].

Given that Cpf1 can process pre-crRNA, it is not surprising that RNAs with the full-length repeat-spacer (RNA4 and RNA6, Extended Data Fig. 6) mediate similar cleavage activities as the mature crRNA form. RNA containing the full-length repeat-spacer led to the most efficient DNA binding and nuclease activity of Cpf1 (compare RNA4 to RNA3 and RNA6, Extended Data Figs 8a and 6a, b). The processed form of crRNA (RNA3, Extended Data Fig. 6) was constructed on the basis of sRNA sequencing results (Extended Data Fig. 1) before the exact RNA processing of Cpf1 was known (Fig. 1), which resulted in a 3-nt shorter 5′ end. Binding to and processing of pre-crRNA induces conformational changes in Cpf1, causing the enzyme to change into an active endonucleolytic state[21]. Similarly, an induced-fit mechanism is used by Cas9, which undergoes large conformational rearrangements upon binding to tracrRNA–crRNA[22].

A seed sequence of 3–5 nt at the PAM-proximal side of the protospacer has been reported for Cpf1 (ref. 16). Using plasmids with single mismatches between spacer and protospacer along the target sequence, we observed that Cpf1 was sensitive to mismatches within the first eight PAM proximal nucleotides, and would not tolerate four consecutive mismatches. Furthermore, Cpf1 was sensitive to mismatches around the cleavage site (position 1–4 on the PAM-distal site), but to a lesser extent (Fig. 2e, Extended Data Table 1a). In the co-crystal structure of Cpf1 and crRNA, the targeting region of crRNA was not resolved, indicating that Cpf1 does not bind and stabilize this part of crRNA[21]. This is in contrast to Cas9, which makes extensive contact with the guide portion of the RNA, possibly explaining its longer seed region[22,23]. Together with the recent Cpf1 characterization[16], our results indicate that there may be additional factors influencing the specificity, such as the base content of the target sequence. The results highlight similarities between Cpf1 and Cas9 (refs 24, 25), which first recognizes the PAM and subsequently probes the crRNA complementary to the target DNA. Mismatches around the target site might disturb correct positioning of the catalytic residues and therefore reduce cleavage activity.

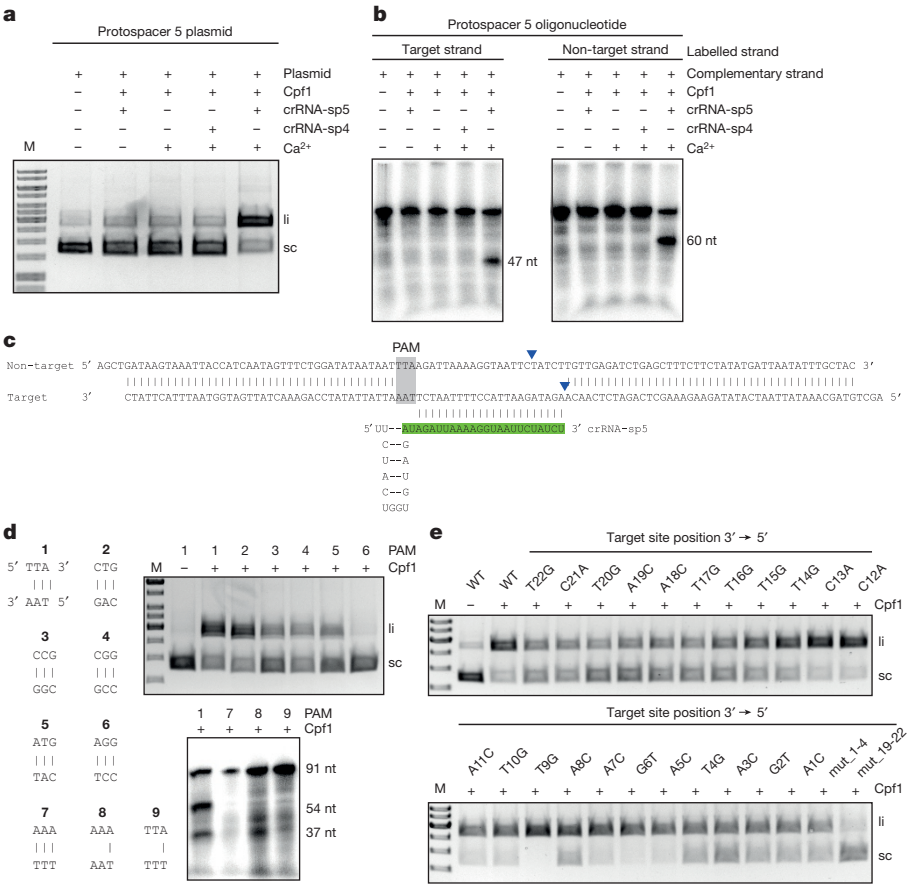


Figure 2 | Cpf1 cleaves target DNA specifically at the 5′-YTN-3′ PAM-distal end to generate 5-nt 5′ overhangs in the presence of Ca^{2+}. **a**, **b**, Cpf1-mediated target plasmid DNA cleavage (**a**) and Cpf1-mediated oligonucleotide duplex cleavage (**b**), dependent on the crRNA containing spacer 4 or 5 (crRNA-sp4 or crRNA-sp5), in the absence or presence of Ca^{2+}. **c**, Schematic representation of the protospacer 5 sequence in the DNA (top), and the structure of crRNA-sp5 used in **a**, **b**, **d** and **e** (bottom). Cleavage sites corresponding to fragments obtained in **b** and confirmed by sequencing (Extended Data Fig. 7) are indicated by blue triangles. The PAM is marked in grey. **d**, Plasmid DNA containing the PAMs 1–6, or 5′-radiolabelled double-stranded oligonucleotide containing PAMs 1, 7–9 were cleaved by Cpf1 in the presence of 10 mM $CaCl_2$ (upper and lower panel, respectively). **e**, Plasmids containing protospacer 5 and single or quadruple mismatches (mut_1-4 and mut_19-22) along the target strand were tested for cleavage by Cpf1 programmed with crRNA-sp5 in the presence of 10 mM $MgCl_2$. Quantification of three independent experiments are shown in Extended Data Table 1a. li, linear; sc, supercoiled; M, 1 kb ladder. Data in **a**, **b**, **d** and **e** are representatives of at least three independent experiments.

Aligning the two predicted protospacer sequences of the *F. novicida* U112 type V-A CRISPR–Cas revealed a conserved 5′-TTA-3′ sequence located on the non-target strand upstream of the protospacer. To verify the potential PAM, protospacer 5 was cloned without its flanking region yielding a 5′-CTG-3′ sequence. Both plasmids were cleaved equally well by Cpf1, indicating that the second position in this sequence is critical (Fig. 2d, Extended Data Fig. 7d). Mutagenesis of all three nucleotides followed by DNA cleavage analysis shows that Cpf1 recognizes a PAM, defined as 5′-YTN-3′, upstream of the crRNA-complementary DNA sequence on the non-target strand. This result expands on the already reported 5′-TTN-3′ PAM[16]. To analyse strand specificity of PAM recognition, we designed oligonucleotide substrates with either AAN or TTN on both strands. These substrates were not cleaved by Cpf1, indicating that the PAM needs to be double-stranded and is probably recognized on both strands (Fig. 2d, lower panel).

We next investigated the metal ion dependency of DNA cleavage by Cpf1. Notably, we observed that in addition to Mg^{2+} and Mn^{2+}, which were shown to mediate activity in Cas9 (ref. 15), Cpf1 also cleaves DNA in the presence of Ca^{2+} (Extended Data Fig. 5b, Extended Data Table 1b). To investigate potential differences in cleavage with Mg^{2+} or Ca^{2+}, we carried out DNA cleavage reactions in the presence of either of these ions (Fig. 2, Extended Data Fig. 7). In Cas9, two active motifs, HNH and RuvC, are responsible for cleavage of the target and non-target strand, respectively[15]. The HNH motif of Cas9 from *Neisseria meningitidis* is Ca^{2+}-dependent[26]. If there were two active sites in Cpf1, each coordinating one of the metal ions and cleaving one of the DNA strands, we would expect a difference in cleavage of target and non-target strands depending on the ion used. In contrast, we did not observe differences in the efficiency of target or non-target strand cleavage by Cpf1 in the presence of Ca^{2+} or Mg^{2+} (Fig. 2b, Extended Data Fig. 7b). This finding indicates the presence of only one catalytic motif in Cpf1 that is responsible for cleaving both DNA strands and can coordinate Mg^{2+} as well as Ca^{2+} ions.

Our experiments show for the first time that Cpf1 exhibits dual (RNA and DNA) cleavage activity. To determine the respective cleavage motifs, we performed mutagenesis of conserved residues along the Cpf1 amino acid sequence (Supplementary Fig. 2). Alanine substitution of residues H843, K852, K869 and F873 had no effect on DNA cleavage activity (Fig. 3a, upper panel), but resulted in decreased *in vitro* RNA cleavage activity (Fig. 3a, middle panel). To further confirm the involvement of these residues in RNA processing *in vivo,* a heterologous *E. coli* assay co-expressing pre-crRNA (repeat–spacer–repeat) and Cpf1, or a variant thereof, was established. Northern blot analysis was performed with total RNA extracted following induced expression of the Cpf1 variant (Fig. 3a, lower panel, Extended Data Fig. 3b). Pre-crRNA is more abundant in the presence of Cpf1, indicating possible protection from degradation by Cpf1. Expression of wild-type Cpf1 results in the production of a distinct band of around 65 nt, which corresponds to a mature crRNA formed by two cleavage events within the repeats. In the presence of Cpf1(H843A), this band is absent; however, two additional longer RNAs appear due to changed processing by this mutant, as observed *in vitro* (Fig. 3a, middle panel). Mutants K852A and K869A also resulted in the production of the 65-nt fragment, but with less intensity compared to the wild type, and two additional products with longer sizes. *In vitro*, these mutants have almost no RNA processing activity. RNA-binding experiments with Cpf1(K852A) and Cpf1(K869A) (Extended Data Fig. 8b) indicated a slightly higher affinity for RNA than wild-type Cpf1, which may explain the cleavage products observed *in vivo*. The residual activity of these Cpf1 mutants produces processed RNA, which is likely to be bound tighter to the protein and therefore better protected from degradation. Cpf1(F873A) had reduced RNA cleavage activity *in vitro,* which could not be detected *in vivo.* Mutation of the aforementioned residues did not negatively affect RNA binding (Extended Data Fig. 8b), indicating that the identified residues of Cpf1 are potentially responsible for RNA cleavage. Analysis of the co-crystal structure of *Lachnospiraceae* bacterium Cpf1

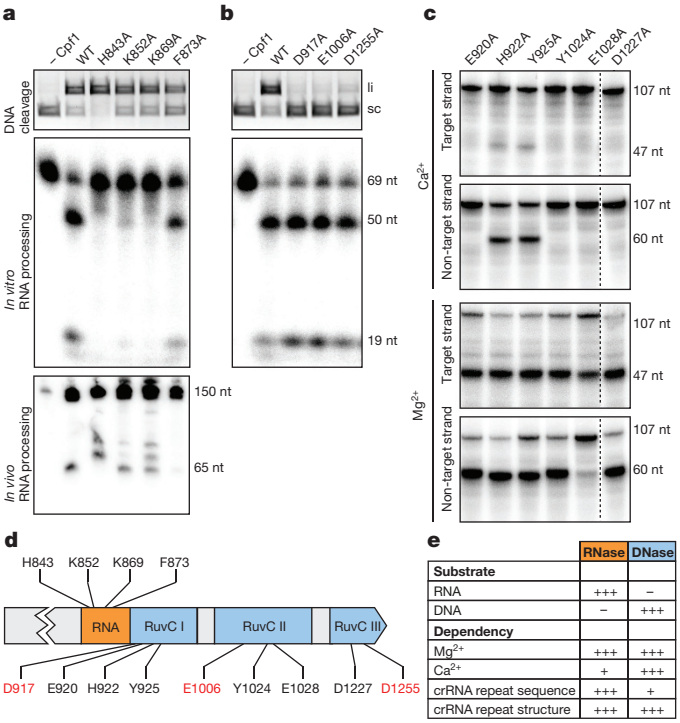


Figure 3 | Cpf1 contains active centres for RNA and DNA cleavage. **a**, RNase motif mutants were tested for DNA plasmid cleavage activity (agarose gel, upper panel), *in vitro* pre-crRNA cleavage activity (denaturing polyacrylamide gel, middle panel) and *in vivo* pre-crRNA processing activity (northern blot, lower panel). *In vitro* cleavage was performed in the presence of 10 mM $MgCl_2$. **b**, DNase motif mutants were tested for plasmid DNA cleavage activity (agarose gel, upper panel) and *in vitro* pre-crRNA cleavage activity (denaturing polyacrylamide gel, lower panel). **c**, Additional RuvC motif mutants were tested for DNA cleavage of double-stranded oligonucleotide substrates in 10 mM $CaCl_2$ (upper two panels) or $MgCl_2$ (lower two panels). Target or non-target strand was 5′ radiolabelled before annealing to the non-labelled complementary strand. **d**, Schematic representation of Cpf1 amino acid sequence (N terminus not shown for clearer visualization) with the active domains for RNA and DNA cleavage highlighted in orange and blue, respectively. Mutated amino acids are indicated with the DNase motif shown in red. **e**, Summary of recognized substrates, metal ion dependency and crRNA requirements for both RNase and DNase motifs of Cpf1. −, no activity; + residual activity; +++ full activity. Data in **a**–**c** are representatives of at least three independent experiments.

	RNase	DNase
Substrate		
RNA	+++	–
DNA	–	+++
Dependency		
Mg^{2+}	+++	+++
Ca^{2+}	+	+++
crRNA repeat sequence	+++	+
crRNA repeat structure	+++	+++

revealed that the identified residues are located in close proximity to the 5′ of the processed crRNA[21].

Mutagenesis of D917, E1006 and D1255 in the split RuvC motif resulted in loss of DNA cleavage activity[16] (Fig. 3b, upper panel), but did not influence the RNA processing activity of Cpf1 (Fig. 3b, lower panel), nor did it affect binding affinity to the DNA target (Extended Data Fig. 8c).

While screening for active site residues, we observed differences in DNA cleavage for some mutants depending on the metal ion present. Mutants E920A, Y1024A and D1227A showed no DNA cleavage in the presence of Ca^{2+}, but wild-type activity when Mg^{2+} was present (Fig. 3c). These residues are located in close proximity to the three identified catalytic residues and may be responsible for coordination of the Ca^{2+} ion. Mutating residue E1028 also led to loss of Ca^{2+}-promoted DNA cleavage and additionally reduced cleavage of the non-target strand in the presence of Mg^{2+}, indicative of its involvement in non-target strand cleavage. In contrast, mutation of residues H922 and Y925 resulted in markedly reduced cleavage of the target strand in the presence of Ca^{2+}, whereas these mutants showed wild-type levels of DNA cleavage activity in the presence of Mg^{2+}. These findings suggest that H922 and Y925 are involved in Ca^{2+} coordination and target-strand cleavage.

We show that two aspartates (D917, D1255) and one glutamate (E1006) form the catalytic site, which is in good agreement with the recent characterization of Cpf1 and other RuvC/RNaseH motifs[16]. These kind of catalytic motifs generally use a two-metal-ion mechanism for DNA cleavage, as shown for Cas9 from *Streptococcus pyogenes*[23]. Enzymes with a two-metal-ion mechanism have more specificity for metal ions, Mg^{2+} in particular[27]. In contrast, enzymes using a one-metal-ion mechanism for cleavage (for example, HNH nucleases) are more flexible in their specificity for metal ions. For example, KpnI cleaves DNA with high fidelity in the presence of Ca^{2+}, but less specifically in the presence of Mg^{2+} (ref. 28). As mentioned before, the HNH motif of Cas9 from *N. meningitidis* is active in the presence of Ca^{2+} (ref. 26). In addition to the identified RNA processing activity of Cpf1, this enzyme may also represent a new type of DNA nuclease using two-metal-ion catalysis, with the ability to utilize Mg^{2+} or Ca^{2+} ions. The physiological relevance of Cpf1 using both ions for DNA cleavage remains undetermined and requires further investigation.

In summary, Cpf1 is an enzyme with two separate catalytic moieties that cleave RNA or DNA (Fig. 3d). The RNase motif is specific for the ribose and unable to cleave DNA. This specificity can be explained by specific interactions of Cpf1 to 2′-OH groups of crRNA[21]. The DNase motif shows cleavage activity only against double-stranded and single-stranded target DNA, but no activity against single-stranded RNA, double-stranded RNA or RNA–DNA heteroduplexes (Fig. 3e, Extended Data Fig. 9). There are other nucleases reported to have certain promiscuity towards RNA and DNA cleavage activity, but one of the two activities is usually highly unspecific[29,30]. To our knowledge, Cpf1 is the first enzyme with two specificities, cleaving RNA in a sequence- and structure-dependent manner, and also performing DNA cleavage in the presence of the RNA that is produced in the first reaction. In the context of CRISPR immunity, type V-A appears to be the most minimalistic system described thus far, using only one enzyme, Cpf1, to process pre-crRNA and then using this RNA to specifically target and cut invading DNA. Evolution of one protein to perform these two specific reactions leads to a more effective mechanism, and also makes this system ideal for horizontal gene transfer. Finally, this mechanism opens new avenues for sequence-specific genome engineering, silencing and facilitates multiplexing.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 November 2015; accepted 30 March 2016.
Published online 20 April 2016.

1. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526,** 55–61 (2015).
2. Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22,** 3489–3496 (2008).
3. Ebihara, A. *et al.* Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Sci.* **15,** 1494–1499 (2006).
4. Charpentier, E., Richter, H., van der Oost, J. & White, M. F. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.* **39,** 428–441 (2015).
5. Nam, K. H. *et al.* 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 (2012).
6. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Struct. Mol. Biol.* **18,** 529–536 (2011).
7. Mulepati, S., Heroux, A. & Bailey, S. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* **345,** 1479–1484 (2014).
8. Plagens, A., Richter, H., Charpentier, E. & Randau, L. DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS Microbiol. Rev.* **39,** 442–463 (2015).
9. van der Oost, J., Westra, E. R., Jackson, R. N. & Wiedenheft, B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature Rev. Microbiol.* **12,** 479–492 (2014).

10. Zhang, J. *et al.* Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol. Cell* **45,** 303–313 (2012).
11. Staals, R. H. *et al.* RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Mol. Cell* **56,** 518–530 (2014).
12. Samai, P. *et al.* Co-transcriptional DNA and RNA cleavage during type III CRISPR-Cas immunity. *Cell* **161,** 1164–1174 (2015).
13. Chylinski, K., Le Rhun, A. & Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* **10,** 726–737 (2013).
14. Deltcheva, E. *et al.* CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).
15. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
16. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163,** 759–771 (2015).
17. Schunder, E., Rydzewski, K., Grunow, R. & Heuner, K. First indication for a functional CRISPR/Cas system in *Francisella tularensis*. *Int. J. Med. Microbiol.* **303,** 51–60 (2013).
18. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nature Rev. Microbiol.* **13,** 722–736 (2015).
19. Shmakov, S. *et al.* Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell* **60,** 385–397 (2015).
20. Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329,** 1355–1358 (2010).
21. Dong, D. *et al.* The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* http://dx.doi.org/10.1038/nature17944 (20 April 2016).
22. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A. Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348,** 1477–1481 (2015).
23. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).
24. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507,** 62–67 (2014).
25. Szczelkun, M. D. *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl Acad. Sci. USA* **111,** 9798–9803 (2014).
26. Zhang, Y., Rajan, R., Seifert, H. S., Mondragon, A. & Sontheimer, E. J. DNase H activity of *Neisseria meningitidis* Cas9. *Mol. Cell* **60,** 242–255 (2015).
27. Yang, W. An equivalent metal ion in one- and two-metal-ion catalysis. *Nature Struct. Mol. Biol.* **15,** 1228–1231 (2008).
28. Vasu, K. *et al.* Increasing cleavage specificity and activity of restriction endonuclease *Kpn*I. *Nucleic Acids Res.* **41,** 9812–9824 (2013).
29. Nam, K. H. *et al.* Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *J. Biol. Chem.* **287,** 35943–35952 (2012).
30. Punetha, A., Sivathanu, R. & Anand, B. Active site plasticity enables metal-dependent tuning of Cas5d nuclease activity in CRISPR-Cas type I–C system. *Nucleic Acids Res.* **42,** 3846–3856 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Schmidt, F. Hille and A. Escalera Maurer for technical help. This work was funded by the Alexander von Humboldt Foundation (AvH Professorship), the German Federal Ministry for Education and Research, the Helmholtz Association, the German Research Foundation, the Max Planck Society, the Göran Gustafsson Foundation (Göran Gustafsson Prize from the Royal Swedish Academy of Sciences), the Swedish Research Council and Umeå University (all to E.C.), and the Helmholtz Postdoc Programme (to H.R.).

Author Contributions I.F. and H.R. conducted the biochemical characterization of the DNase and RNase activities, M.B. performed binding studies and seed sequence characterization and A.L.R. performed and analysed RNA sequencing. I.F., H.R. and E.C. designed the research. I.F., H.R., M.B., A.L.R. and E.C. analysed and interpreted the data; I.F., H.R., and E.C. wrote the paper, which M.B and A.L.R. commented on.

Author Information RNA sequencing data have been deposited at NCBI under accession number SRP071054. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.C. (charpentier@mpiib-berlin.mpg.de).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Small RNA sequencing. Small RNA sequencing data of *Francisella novicida* U112 (Supplementary Table 1a) used in this study were obtained previously[13]. Briefly, a cDNA library of RNAs (treated with tobacco acid pyrophosphatase) of *F. novicida* U112 grown to mid-logarithmic phase was prepared using the ScriptMiner Small RNA-Seq Library Preparation Kit (Multiplex, Illumina compatible) and sequenced at the Campus Science Support Facilities GmbH (CSF) Next Generation Sequencing (NGS) Unit of the Vienna Biocentre. After adaptor removal and quality trimming, the reads were mapped to the *F. novicida* U112 genome (GenBank: NC_008601, 48205 mapped reads) using Bowtie. The read coverage was calculated using BEDTools (version 2.15.0.)[31] and a normalized wiggle file was created and visualized using the Integrative Genomics Viewer[32,33].

Production and purification of recombinant Cpf1. The *cpf1* (FTN_1397) gene was amplified from genomic DNA of *F. novicida* U112 and cloned into the expression vector pET-16b to facilitate expression of Cpf1 with an N-terminal 6× His-tag (Supplementary Table 1b, c). For the production of the protein in *E. coli* (NiCo21 (DE3)), the cells containing the overexpression plasmid were grown at 37 °C to reach an optical density $(OD)_{600}$ of 0.6–0.8. Expression was induced by addition of 0.5 mM isopropylthio-β-D-galactoside (IPTG) and the cultures were further incubated overnight at 18 °C. After collection, the cell pellet was resuspended in lysis buffer (20 mM HEPES (pH 7.5), 500 mM KCl, 25 mM imidazole, 0.1% Triton X-100) followed by 6 min of sonication (0.5 s pulses) for cell disruption. The lysate was cleared by centrifugation (47,800*g*, 30 min, 4 °C) and the supernatant was applied to Ni^{2+}-NTA-Sepharose resin in a drop column. After washing steps with 10 ml of lysis buffer followed by 10 ml wash buffer (20 mM HEPES (pH 7.5), 300 mM KCl, 25 mM imidazole), the protein was eluted with elution buffer (20 mM HEPES (pH 7.5), 150 mM KCl, 250 mM imidazole, 0.1 mM DTT, 1 mM EDTA). The eluates were analysed by SDS–PAGE followed by Coomassie blue staining. Fractions containing Cpf1 were pooled for cation-exchange chromatography (HiTrap Heparin; GE-Healthcare) using a FPLC Äkta-Purification system (GE-Healthcare) and Cpf1 was eluted with a linear gradient of 100–1000 mM KCl. Peak fractions were analysed by SDS–PAGE and Coomassie blue staining. Cpf1-containing fractions were pooled and directly applied to an equilibrated (20 mM HEPES (pH 7.5), 150 mM KCl) prepgrade Superdex 200 size-exclusion column (GE-Healthcare) and purified via fast protein liquid chromatography (FPLC), followed by analysis by SDS–PAGE and Coomassie blue staining. Molecular weight calibration of the column was performed using molecular weight markers, as described in the manufacturer's protocol (Kit for Molecular Weights, Sigma-Aldrich). The protein was dialysed against dialysis buffer (20 mM HEPES (pH 7.5), 150 mM KCl, 50% glycerol) and stored at −20 °C until use.

Site-directed mutagenesis of Cpf1. Oligonucleotides for the site-directed mutation of Cpf1 (Supplementary Table 1c) were designed using the QuickChange Primer Design tool of Agilent and produced by Sigma-Aldrich. A series of PCRs was performed to obtain the desired mutation. Briefly, the overexpression vector containing wild-type *cpf1* was amplified in two reactions with either the forward or reverse QuickChange primer. After an initial amplification, the two reactions were mixed and a second PCR was performed. Following PCR, the template plasmid was degraded with DpnI (3 h, 37 °C) and introduced by transformation into chemically competent DH5α cells. Plasmids were prepared using a plasmid Miniprep kit (Qiagen) according to the manufacturer's instructions. Successful mutagenesis was confirmed by sequencing analysis of the plasmids (SeqLab).

Generation of RNAs used in this study. The sRNAs tested in this study were generated by *in vitro* transcription using the AmpliScribe T7-Flash kit (Biozym) according to the manufacturer's protocol. In brief, oligonucleotides containing the desired sequence (Supplementary Table 1c) and a T7-promoter sequence were hybridized to an oligonucleotide containing the complementary T7-promoter sequence. The hybridization product was then used as a template for the transcription reaction according to the AmpliScribe T7-Flash kit (Biozym). To obtain internally labelled RNAs, [α-^{32}P]ATP (5000 Ci mmol^{-1}, Hartman Analytic) was added to the *in vitro* transcription reaction[34]. In order to generate end-labelled RNAs, the unlabelled transcripts were dephosphorylated with Fast-AP phosphatase (Fermentas) for 30 min at 37 °C followed by a purification using Illustra Microspin G-25 columns (GE-Healthcare). The dephosphorylated RNAs were then labelled using T4 polynucleotide kinase (Fermentas) and [γ-^{32}P]ATP (5000 Ci mmol^{-1}) according to the manufacturer's instructions, and separated using denaturing polyacrylamide gel electrophoresis (8 M urea; 1× TBE; 10% polyacrylamide). Subsequent to short exposure to an autoradiography screen (for radioactively labelled RNAs) or ethidium bromide (EtBr) staining (for unlabelled RNAs), the respective bands of the RNAs were excised. Elution of the RNAs was achieved by incubation of the gel pieces in 500 µl RNA elution buffer (250 mM NaOAc; 20 mM Tris-HCl (pH 7.5); 1 mM EDTA (pH 8.0); 0.25% SDS) and overnight incubation on ice. Following elution, RNA was precipitated with 2 vol ice-cold ethanol (100% EtOH) and 1/100 glycogen for 1 h at −20 °C. After washing with 70% EtOH, the air-dried pellets were resuspended in H_2O.

***In vitro* RNA cleavage assay.** RNA cleavage assays using indicated concentrations of Cpf1 and various RNA substrates were conducted in KGB buffer[35] (100 mM potassium glutamate, 25 mM Tris-acetate (pH 7.5), 500 µM 2-mercaptoethanol, 10 µg ml^{-1} BSA) supplemented with 10 mM $MgCl_2$ at 37 °C in a final volume of 10 µl. If not indicated otherwise, the reaction was stopped after 10 min by the addition of 2 µl proteinase K (20 mg ml^{-1}) following 10 min incubation at 37 °C to achieve protein degradation. After adding 2× loading dye (10 M urea, 1.5 mM EDTA (pH 8.0)), the samples were loaded on 12% denaturing polyacrylamide gels run in 1× TBE for 3 h at 12.5 V cm^{-1}. For the sequencing gels, the samples were precipitated before loading on 10% denaturing polyacrylamide gels. The gel electrophoresis was carried out at 40 W for 3.5 h. Visualization was achieved by phosphorimaging (Typhoon FLA 9000 Fuji). For RNA size determination, a 5′-end-labelled 69-nt long transcript consisting of a short form of pre-crRNA (repeat-spacer 5, full-length) was subjected to alkaline hydrolysis generating a single nucleotide resolution ladder and to RNase T1-specific cleavage. Each individual experiment was performed in three replicates.

***In vivo* RNA processing.** To investigate *in vivo* RNA processing by Cpf1, a heterologous system was designed in *E. coli*. A DNA fragment encoding a pre-crRNA containing a repeat-spacer-repeat structure under the control of a T7-promoter and T7-terminator was synthesized by Integrated DNA Technologies and cloned into pACYC184 using HindIII and EagI yielding pEC1690. *E. coli* BL21(DE3) was co-transformed with this plasmid and the overexpression vector of wild-type or mutant Cpf1. The empty expression vector pET-16b served as a negative control. The bacterial cells were grown in the presence or absence of 0.1 mM IPTG at 37 °C to reach early exponential phase (OD_{600} = 0.4). RNA was extracted using TRIzol (Sigma-Aldrich) according to the manufacturer's protocol followed by northern blot analysis as described previously[36–38]. In brief, RNA was separated on denaturing 10% polyacrylamide gels (8 M urea, 1× TBE) and transferred by semi-dry blotting on a nylon membrane (Hybond TM N+, GE Healthcare). Chemical crosslinking was performed for 1 h at 60 °C with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride. Oligonucleotides were radioactively labelled with [γ-^{32}P]ATP (5000 Ci mmol^{-1}) and T4 polynucleotide kinase (Fermentas) as described above and purified using Illustra Microspin G-25 columns (GE Healthcare). The hybridization of the probe against the spacer in the pre-crRNA (Supplementary Table 1c) was performed in Rapid-hyb buffer (GE-Healthcare) by incubation overnight at 42 °C. The radioactive signal was visualized using phosphorimaging. Each individual experiment was performed in three replicates.

Generation of DNA substrates. To determine the target cleavage site of Cpf1, spacer sequences of the *F. novicida* U112 type V-A CRISPR array were analysed by BLAST[39]. Potential targets for spacer 4 and spacer 5 were identified in *F. novicida* 3523, located in the intergenic region between coding sequence AEE26308.1 and AEE26307.1, and in AEE26301.1, respectively. Target protospacer containing a sequence complementary to spacer 5 including 42 bp up- and downstream sequences was synthesized as oligonucleotides containing HindIII overhangs. Following hybridization of the oligonucleotides, the fragments were cloned into pUC19 using HindIII yielding plasmid pEC1664 (protospacer 5 and flanking region). The same protospacer sequence without flanking regions was cloned into pUC19, yielding pEC1688 (protospacer 5). In order to identify the PAM, mutagenesis was performed by applying the described protocol for site-directed mutagenesis on pEC1688. Plasmid preparation was done using Miniprep kit (Qiagen) according to the manufacturer's instructions and DNA integrity was confirmed by sequencing analysis (SeqLab). Oligonucleotides containing the protospacer (Supplementary Table 1c) were ordered at Sigma and hybridized before radioactive labelling. Alternatively, a single-stranded oligonucleotide was labelled and hybridized with the complementary non-labelled oligonucleotide. 5′-end-labelling reactions were performed using [γ-^{32}P]ATP (5000 Ci mmol^{-1}) and T4 polynucleotide kinase (Fermentas) according to the manufacturer's instructions. The labelled oligonucleotides were purified using Illustra Microspin G-25 columns (GE healthcare).

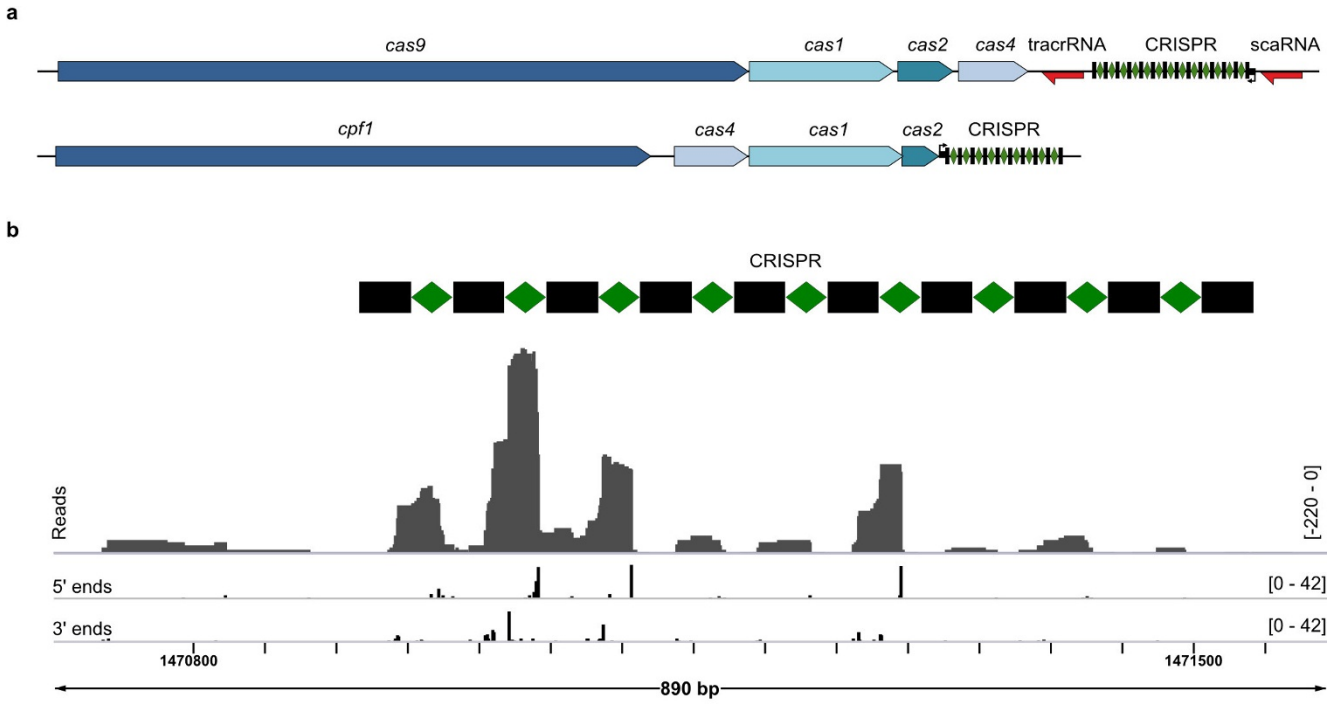
***In vitro* DNA cleavage assay.** Plasmid DNA cleavage assays were performed by pre-incubating 100 nM Cpf1 with 200 nM RNA in KGB buffer supplemented with either 10 mM $MgCl_2$ or 10 mM $CaCl_2$ for 15 min at 37 °C. Plasmid DNA (10 nM) was added to the reaction to yield a final volume of 10 µl and further incubated for 1 h at 37 °C. Reactions were stopped by the addition of 1 µl proteinase K (20 mg ml^{-1}) and 5 min incubation at 37 °C. Before separation of the reaction, 3 µl of 5× DNA loading buffer (250 mM EDTA, 1.2% SDS, 25% glycerol, 0.01% bromophenol blue) were added and the samples were loaded on 0.8% agarose gels

(1× TAE buffer). Cleavage products were visualized by EtBr staining. In cleavage assays using radioactively labelled substrates, 5 nM of 5′-labelled double-stranded oligonucleotides were added to the pre-formed complex of Cpf1 and RNA, and incubated at 37 °C for 1 h. After proteinase K treatment, 10 μl of 2× denaturing loading buffer (95% formamide, 0.025% SDS, 0.5 mM EDTA, 0.025% bromophenol blue) were added. Oligonucleotides of the size of the expected cleavage products were 5′-radiolabelled as described above and mixed with an equal volume of 2× denaturing loading buffer to serve as size markers. After 5 min incubation at 95 °C, the samples were loaded on 12% denaturing polyacrylamide gels and run in 1× TBE for 70 min at 14 V cm^{-1}. Cleavage was visualized using phosphorimaging. Each individual experiment was performed in three replicates.

Electrophoretic mobility shift assays. Substrates for electrophoretic mobility shift assays (EMSAs) were generated as described above. For DNA binding reactions, Cpf1 was pre-incubated in binding buffer (20 mM Tris-HCl (pH 7.4), 100 mM KCl, 1 mM DTT, 5% glycerol) containing two molar excess of crRNA. After 15 min at 37 °C, 1 nM labelled DNA substrate was added. The reaction was then carried out at 37 °C for 1 h before the samples were loaded on a native 5% polyacrylamide gel running at 10 V cm^{-1} for 50 min in 0.5× TBE to separate protein–DNA complexes from unbound DNA. For RNA binding reactions, the crRNA was dephosphorylated using Fast AP (Fermentas) and 5′-radiolabelled with [γ-^{32}P] ATP (5000 Ci $mmol^{-1}$) and T4 polynucleotide kinase (Fermentas) according to the manufacturer's instructions. A total of 0.5 nM radiolabelled RNA were incubated with Cpf1 in binding buffer (20 mM Tris (pH 7.5), 150 mM KCl, 10 mM $CaCl_2$, 1 mM DTT, 5% glycerol, 0.01% Triton X-100, 10 μg ml^{-1} BSA) for 1 h at 37 °C and loaded on 4% native polyacrylamide gels running at 10 V cm^{-1} for 30 min in 0.5× TBE. The gels were exposed on an autoradiography film overnight and visualized by phosphorimaging. Fractions of bound and unbound nucleic acids were determined densitometrically and the percentage of bound nucleic acid was plotted against the protein concentrations. The dissociation constant, K_d, was determined using a nonlinear regression analysis.

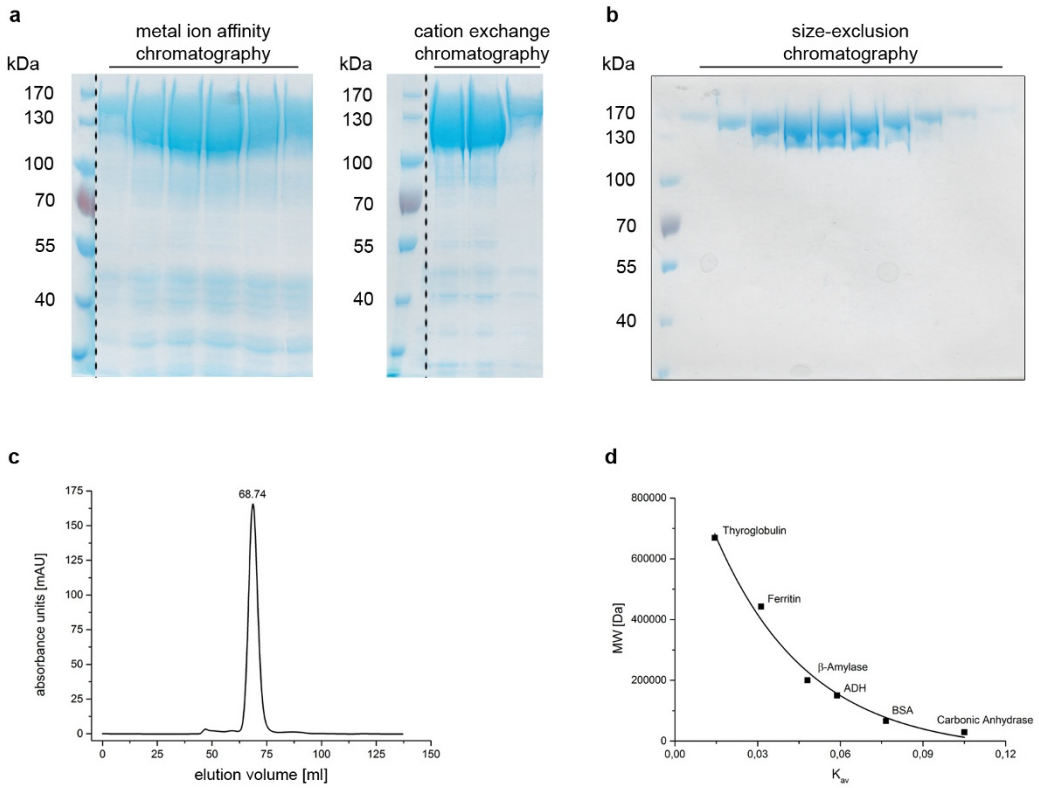
Multiple sequence alignment of Cpf1 orthologues. Cpf1 orthologous sequences were derived by BLAST[39] search of the NCBI database using Cpf1 of *F. novicida* U112 as a query. A multiple sequence alignment of 52 orthologous sequences was generated using MUSCLE[40]. The alignment of nine of the sequences was visualized with Jalview[41].

31. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
32. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnol.* **29,** 24–26 (2011).
33. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14,** 178–192 (2013).
34. Sampson, J. R. & Uhlenbeck, O. C. Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed *in vitro*. *Proc. Natl Acad. Sci. USA* **85,** 1033–1037 (1988).
35. McClelland, M., Hanish, J., Nelson, M. & Patel, Y. KGB: a single buffer for all restriction endonucleases. *Nucleic Acids Res.* **16,** 364 (1988).
36. Herbert, S., Barry, P. & Novick, R. P. Subinhibitory clindamycin differentially inhibits transcription of exoprotein genes in *Staphylococcus aureus*. *Infect. Immun.* **69,** 2996–3003 (2001).
37. Urban, J. H. & Vogel, J. Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res.* **35,** 1018–1037 (2007).
38. Pall, G. S. & Hamilton, A. J. Improved northern blot method for enhanced detection of small RNA. *Nature Protocols* **3,** 1077–1084 (2008).
39. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
40. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5,** 113 (2004).
41. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).
42. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnol.* **29,** 24–26 (2011).
43. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36,** 3429–3431 (2008).
44. Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25,** 1974–1975 (2009).

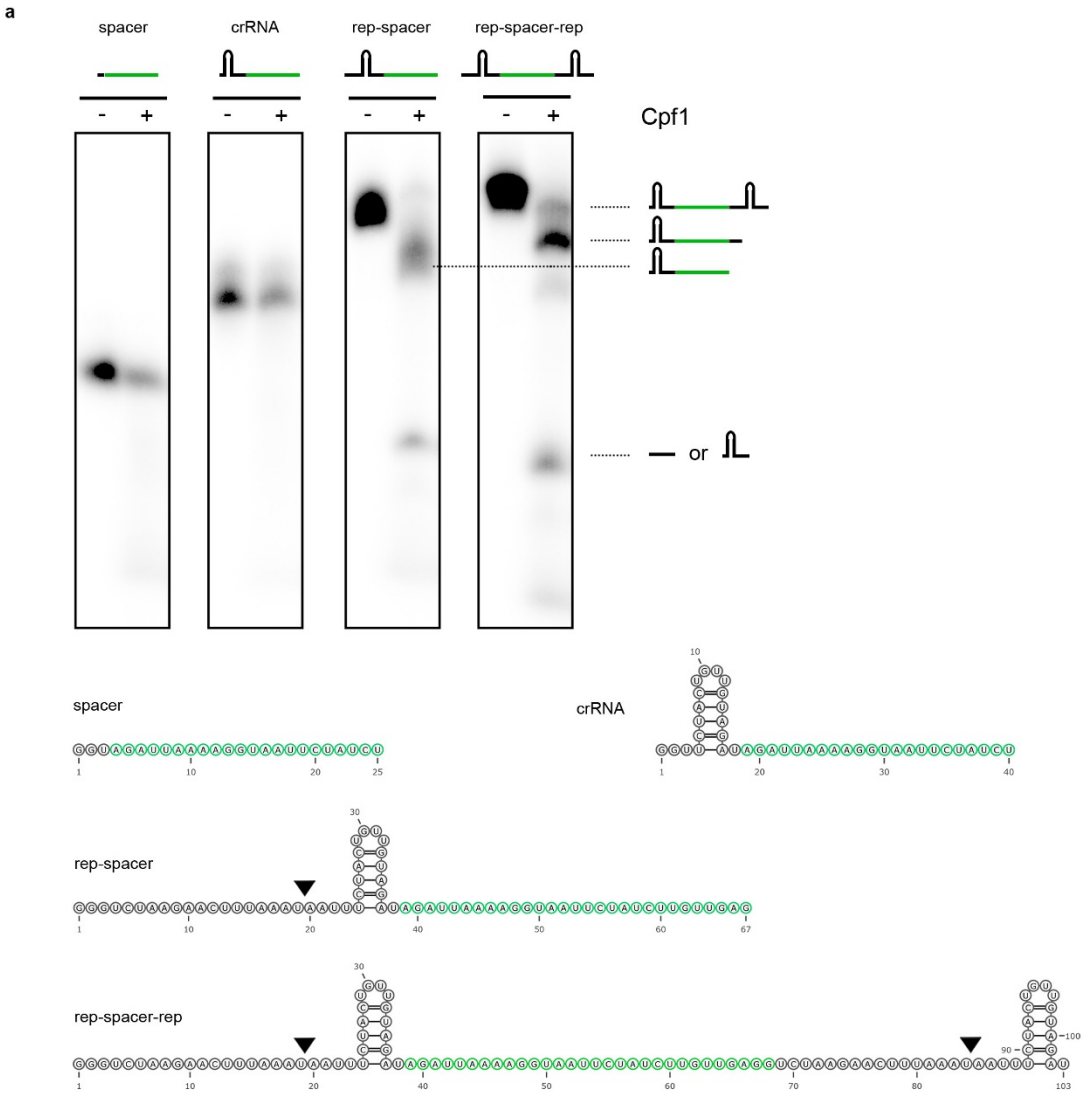


atgcgauucauagagaacaagagguuguuuuuauagacuaaaaauugcaaaccuuagucuuuauguuaaaauaacuacuaaguucuuaga
gauauuuaaaaauaugacuguuguuauauaucaaaaugcuaaaaaaaucauagauuuuaggucuuuuuuugcugauuuaggcaaaaacgg
GUCUAAGAACUUUAAAUAAUUUUCUACUGUUGUAGAUgagaagucauuuaauaaggccacuguuaaaa
GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAUgcuacuauuccugugccuucagauaauuca
GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAUgucuagagccuuuuguauuaguagccg
GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAUuagcgauuuaugaaggucauuuuuuu
GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAUagauuaaaagguaauucuaucuuguugag
GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAUuaccuaguagauacgcuuacugauaacaa
GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAUaaacuuucauuuaugauauaaaguuuuuu
GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAUucaaaaggcaagagagacggaaauaaauggac
GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAUuuguuugauugcuugcauuga

Extended Data Figure 1 | *F. novicida* U112 expresses short mature type V-A crRNAs composed of repeat-spacer. a, In-scale representation of type II-B (*cas9*) and type V-A (*cpf1*) CRISPR–Cas loci in *F. novicida* U112. Blue arrows, *cas* genes; black arrows, putative pre-crRNA promoters; thick black line, CRISPR leader sequence; black rectangles, CRISPR repeats; green diamonds, CRISPR spacers; red arrows, tracrRNA or scaRNA (small, CRISPR–Cas-associated RNA). **b**, Expression of type V-A crRNAs determined by sRNA sequencing is represented with a grey bar chart. The coverage of the reads is indicated in brackets and reads starting (5′ end) and ending (3′ end) at each position are shown (image captured from Integrative Genomics Viewer)[33,42]. The genomic coordinates and size of the CRISPR array in base pairs are indicated. Above the chart bar, a schematic representation shows repeats (black boxes) and spacers (green diamonds) of the *F. novicida* U112 CRISPR array. The sequence of the type V-A CRISPR array from the leader sequence to the last repeat is shown below. Black bold sequences, repeats; green sequences, spacers. The boxed sequences correspond to the mature crRNAs for which we could determine the 5′ and 3′ ends. The mature crRNAs are composed of part of the repeat in 5′ and part of the spacer in 3′.

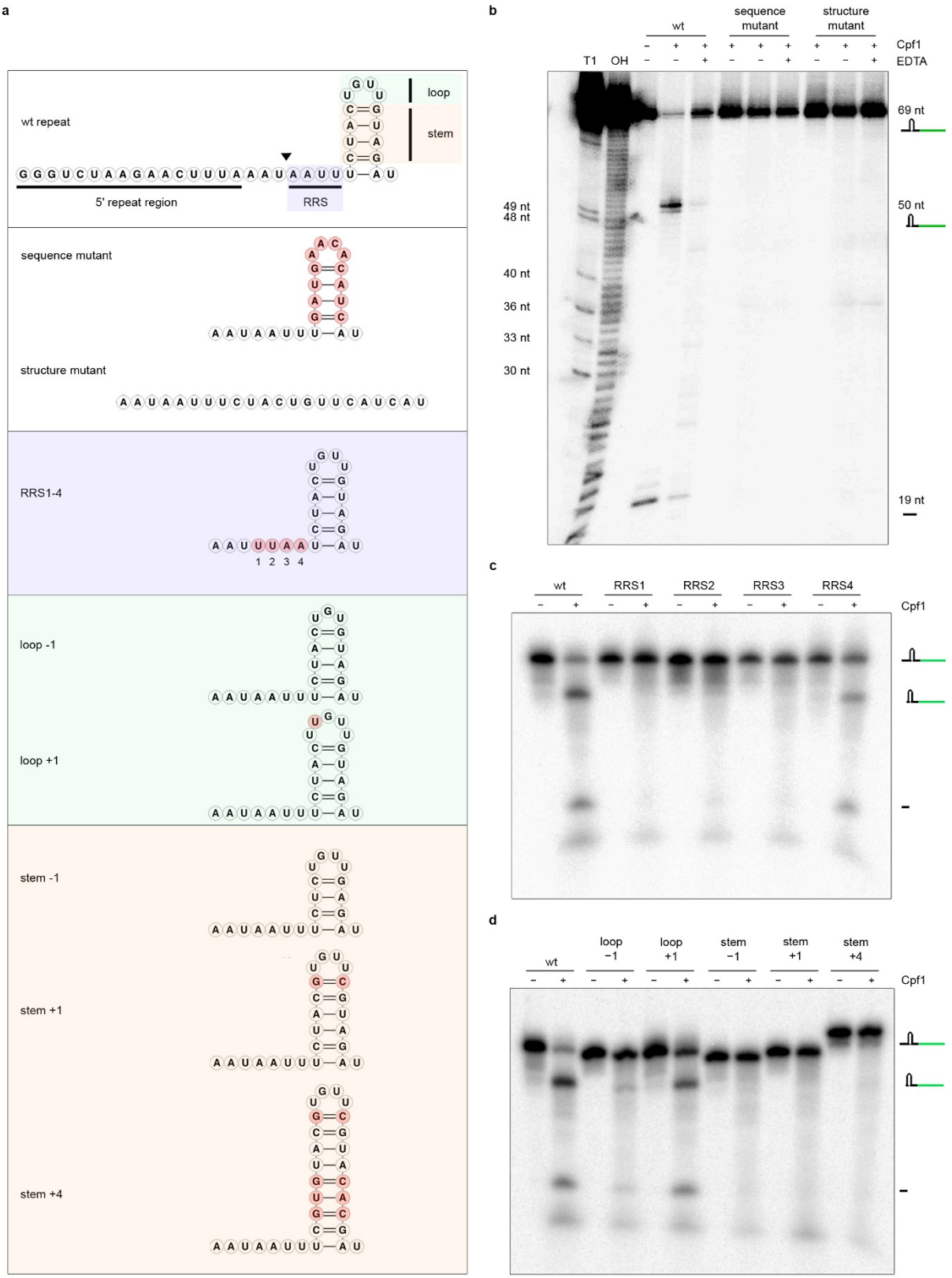


Extended Data Figure 2 | Wild-type Cpf1 purifies as a monomer in solution. Recombinant Cpf1 of *F. novicida* U112 purified via affinity and cation-exchange chromatography was applied to a Superdex 200 size-exclusion column. **a**, SDS–PAGE of protein eluates obtained by nickel-affinity purification (left panel), which were further purified by cation-exchange chromatography (right panel). **b**, Protein samples obtained by size-exclusion chromatography were separated by SDS–PAGE (8% polyacrylamide) and visualized with Coomassie staining. **c**, Elution profile of the size-exclusion chromatography of wild-type Cpf1. The partition coefficient K_{av} for Cpf1 was calculated as 0.0538 by using the equation $K_{av} = (V_e - V_0)/(V_t - V_0)$, with V_e, elution volume; V_0, void volume (elution volume of blue dextran, 45.171 ml) and V_t, geometric column volume (482.5 ml). **d**, Calibration curve of proteins with known molecular weights (thyroglobulin (669 kDa), ferritin (443 kDa), β-amylase (200 kDa), alcohol dehydrogenase (ADH; 150 kDa), bovine serum albumin (BSA; 66 kDa), carbonic anhydrase (29 kDa); Molecular Weight Marker Kit, Sigma-Aldrich). The molecular weight of these proteins was plotted against their calculated K_{av} and fitted by exponential regression analysis. On the basis of the calculation, the K_{av} of Cpf1 results in a molecular weight of 187 kDa, indicating a monomeric form of Cpf1 in solution. Assuming that the protein does not adopt a perfect globular shape, this result is in accordance with the theoretical molecular weight of Cpf1 (153 kDa).



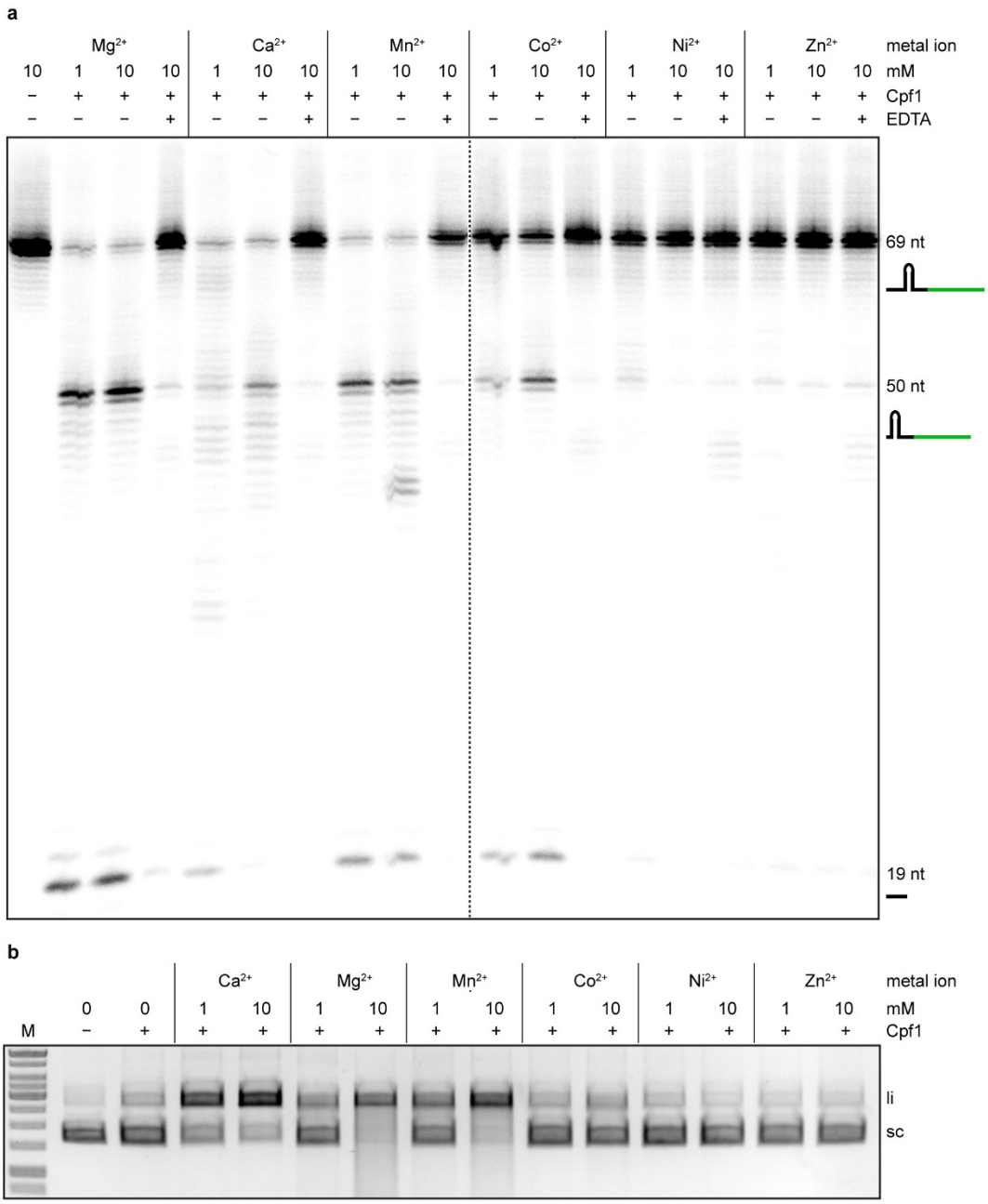
Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | The endoribonucleolytic activity of Cpf1 is dependent on the presence of an intact repeat sequence. a, Cleavage assays were performed by incubating 100 nM of internally labelled RNA constructs corresponding to different repeat and spacer sequence variants of pre-crRNA-sp5 (pre-crRNA-containing spacer 5) with 1 μM of Cpf1 for 30 min at 37 °C. The cleavage reaction was analysed by denaturing polyacrylamide gel electrophoresis and phosphorimaging. The cleavage products are represented schematically. The sequence compositions of the RNAs used as substrates are shown. RNA structures were generated with RNAfold[43] and visualized using VARNA[44] software. Cpf1 cleaved only the RNA templates containing a full-length repeat sequence. The substrate containing two repeats was cleaved twice resulting in more than two fragments, whereas cleavage of RNA with only one repeat resulted in two fragments, consistent with the determined cleavage site (see Fig. 1). **b**, Northern blot analysis of total RNA extracted from *E. coli* co-transformed with a plasmid encoding pre-crRNA and either the empty vector or overexpression vectors encoding wild-type (wt) Cpf1 and variants. Cpf1 expression was induced (+) or not induced (−) with IPTG. The northern blot was probed against the spacer sequence of the tested pre-crRNA. In the absence of Cpf1 (empty vector or not induced), the amount of transcript is reduced compared to reactions with Cpf1 present, suggesting stabilization of pre-crRNA upon binding of Cpf1. Expression of Cpf1, Cpf1(K852A) and Cpf1(K869A) results in the production of a distinct processed transcript of 65 nt, whereas expression of Cpf1(H843A), Cpf1(K852A) or Cpf1(K869A) results in the production of additional higher transcripts. Expression of Cpf1(F873A) results in almost no detectable processed transcript.

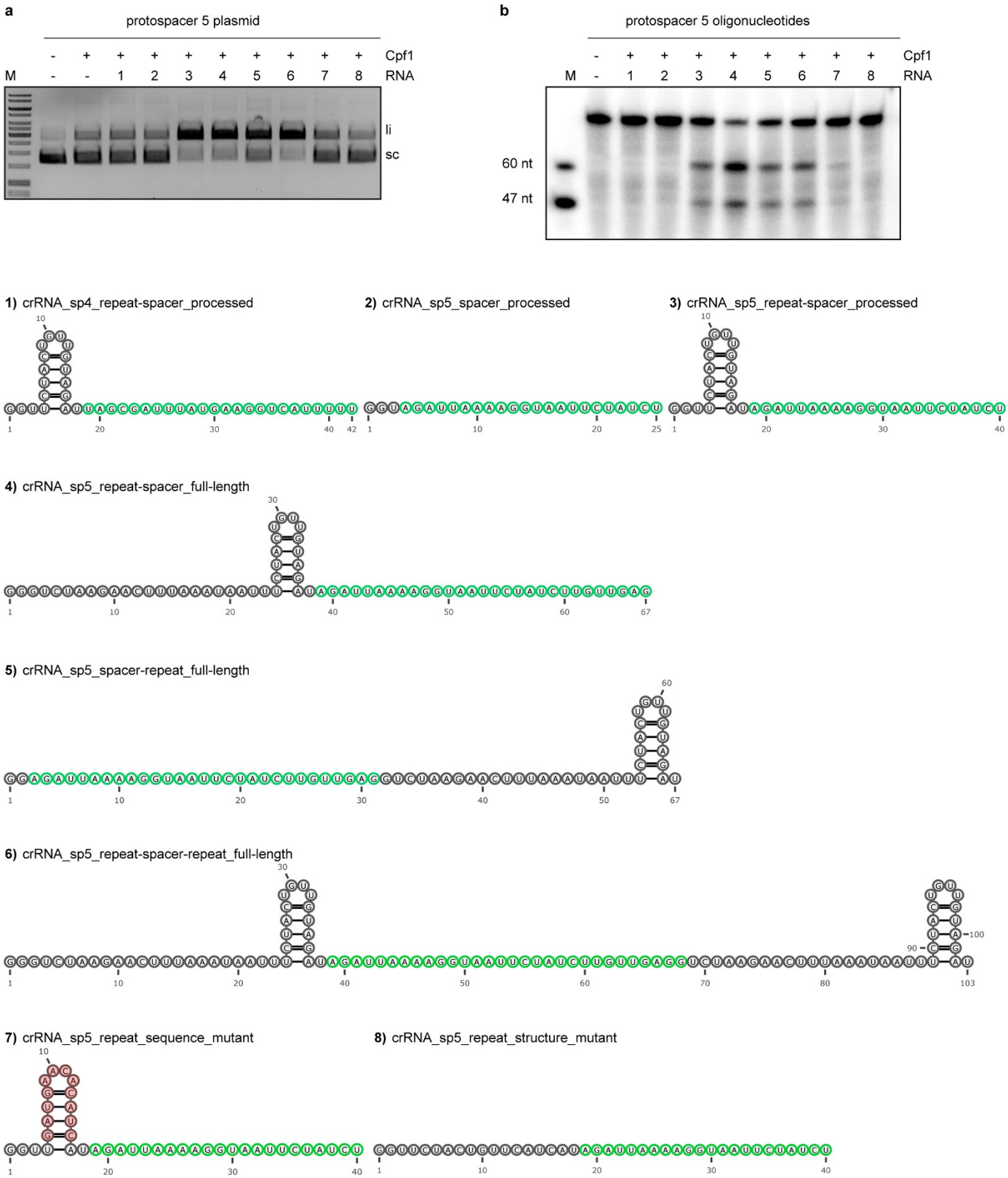


Extended Data Figure 4 | See next page for caption.

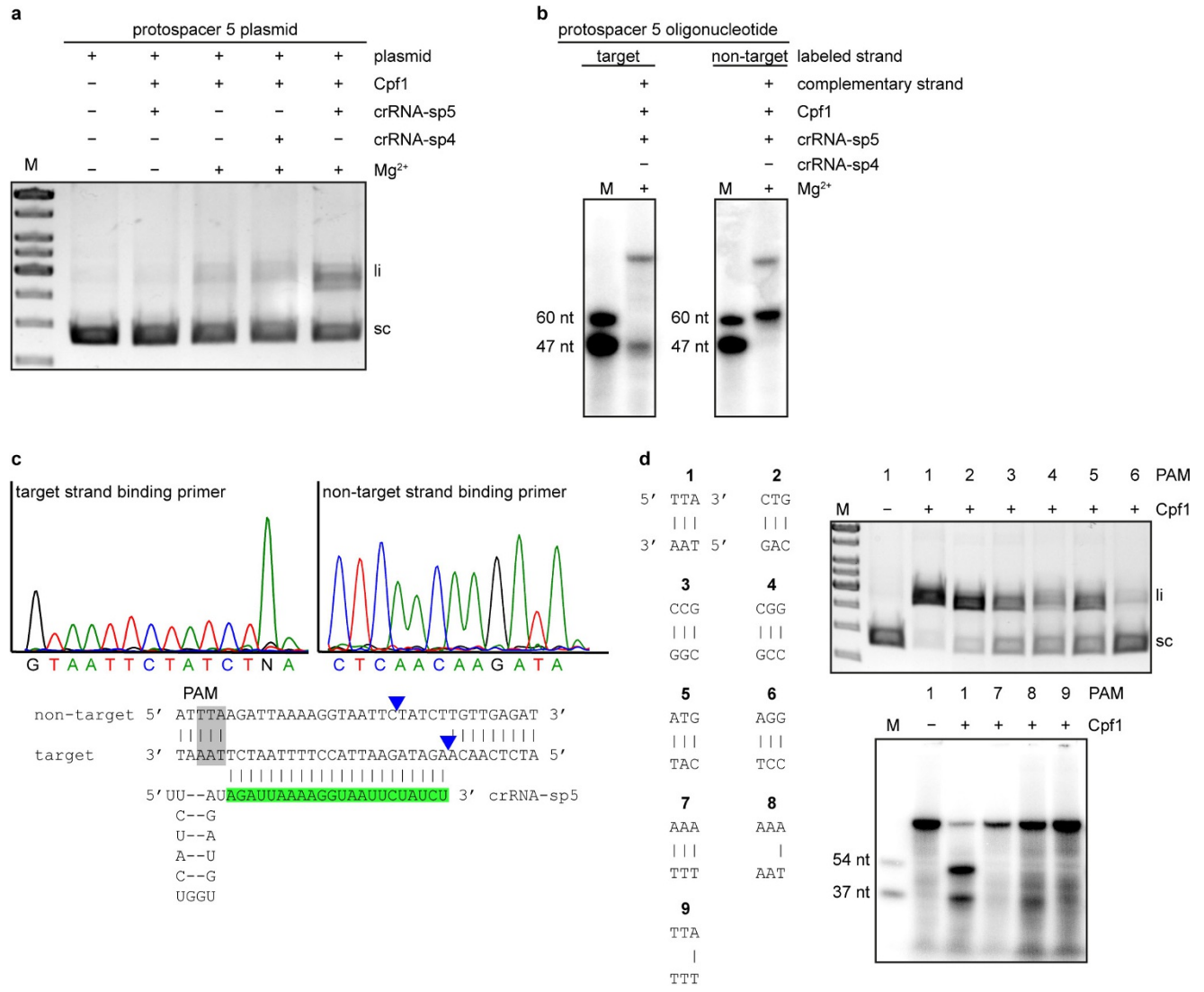
Extended Data Figure 4 | Cpf1 is a sequence- and structure-specific endoribonuclease. a, Design of various repeat variants of pre-crRNA-sp5 (pre-crRNA with spacer 5) with an altered repeat sequence, a destroyed repeat structure, single nucleotide exchanges (1–4) in the RRS (purple) and changed loop (green) and stem sizes (yellow). Note that the 5′ repeat region of the wild-type repeat is not shown in the different variants. Red circles highlight the mutated or added residues. The RNA structures were generated with RNAfold[43] and visualized using VARNA[44] software. **b**, Internally labelled pre-crRNAs containing a wild-type repeat sequence, an altered repeat sequence or a destroyed repeat structure were obtained by *in vitro* transcription. The 5′ end-labelled wild-type substrate was used to generate an alkaline hydrolysis ladder (OH) and an RNase T1 digest (T1) for size determination of the RNA fragments (Life Technologies). Cpf1 cleaved only the pre-crRNA template containing the wild-type repeat sequence yielding a small 19-nt 5′ repeat fragment and a 50-nt intermediate crRNA. **c**, Substrates with serial single mutations of the four RRS nucleotides (1–4, counting from the cleavage site) were tested for processing by Cpf1. Changes of the first three nucleotides were not tolerated for Cpf1-mediated processing, whereas changing the fourth nucleotide yielded a substrate that was processed with less efficiency compared to the wild-type substrate. **d**, The influence of loop variations in the repeat was tested with substrates containing +1 or −1 nucleotide in the loop. Both substrates were processed by Cpf1. Stems with +1 or −1 base pair, or +4 base pairs were used to determine length requirements of the stem. Cpf1 did not cleave any of the three substrates tested. The RNA cleavage reactions were performed by incubating 1 μM of Cpf1 with 200 nM of RNA variant at 37 °C for 5 min in the presence of 10 mM $MgCl_2$. The cleavage products were analysed by denaturing polyacrylamide gel electrophoresis and phosphorimaging. RNA fragments are represented schematically and fragment sizes are indicated in nucleotides.



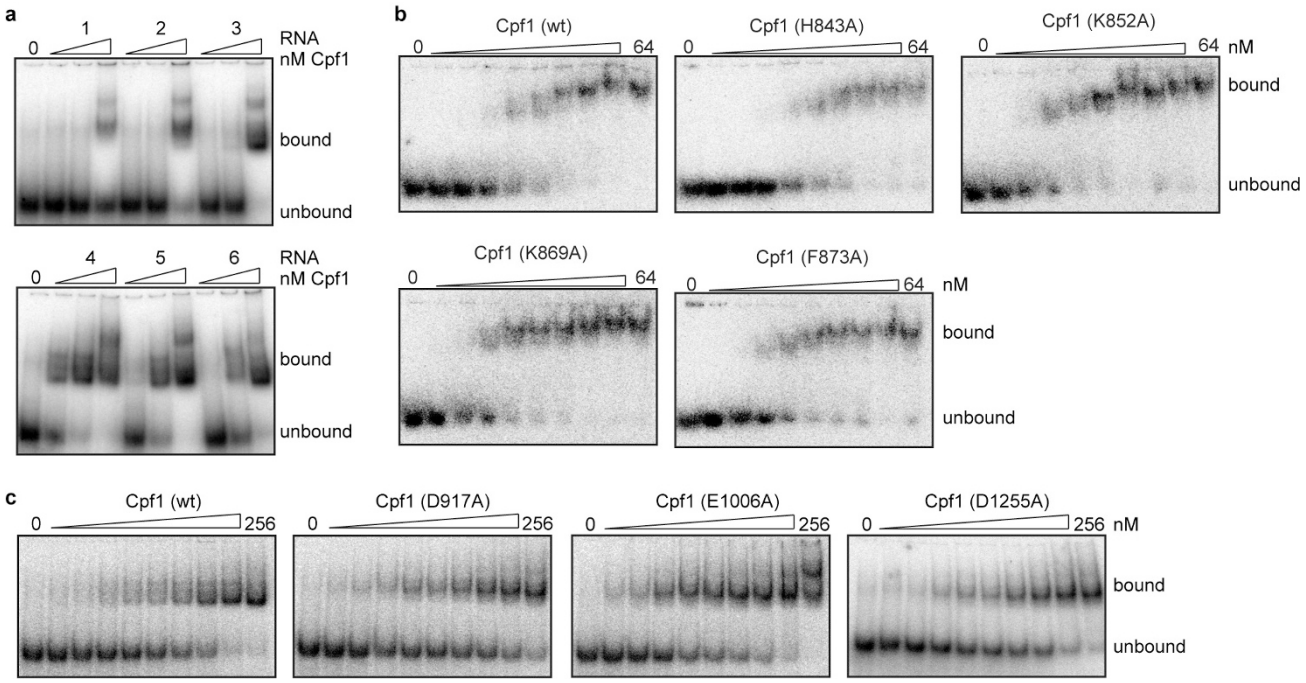
Extended Data Figure 5 | RNA and DNA cleavage activities of Cpf1 are dependent on divalent metal ions. a, Cleavage assays of pre-crRNA-sp5 (repeat-spacer 5, full-length, RNA 4, Extended Data Fig. 6) by Cpf1 in KGB buffer supplemented with different concentrations of divalent metal ion (indicated in mM) or EDTA (10 mM). Cleavage products were analysed by denaturing polyacrylamide gel electrophoresis and visualized by phosphorimaging. RNA fragments are represented schematically and fragment sizes are indicated in nucleotides. Specific RNA cleavage was observed in the presence of $MgCl_2$. Less specific cleavage was detected with $CaCl_2$, $MnCl_2$ and $CoCl_2$. No cleavage of pre-crRNA-sp5 was detected in presence of $NiCl_2$ and $ZnCl_2$. **b**, Cleavage assays of supercoiled plasmid DNA containing protospacer 5 by Cpf1 programmed with crRNA-sp5 (repeat-spacer 5, processed, RNA3, Extended Data Fig. 6) in KGB buffer supplemented with different concentrations of divalent metal ions (indicated in mM). Cleavage products were analysed by agarose gel electrophoresis and visualized by EtBr staining. DNA cleavage was observed in the presence of $MgCl_2$ and $MnCl_2$. A more specific cleavage was observed in the presence of $CaCl_2$. The addition of $CoCl_2$, $NiCl_2$ or $ZnCl_2$ to the reaction did not result in DNA cleavage. li, linear; sc, supercoiled; M, 1 kb ladder (Fermentas). Quantification of three independent experiments shown in Extended Data Table 1b.



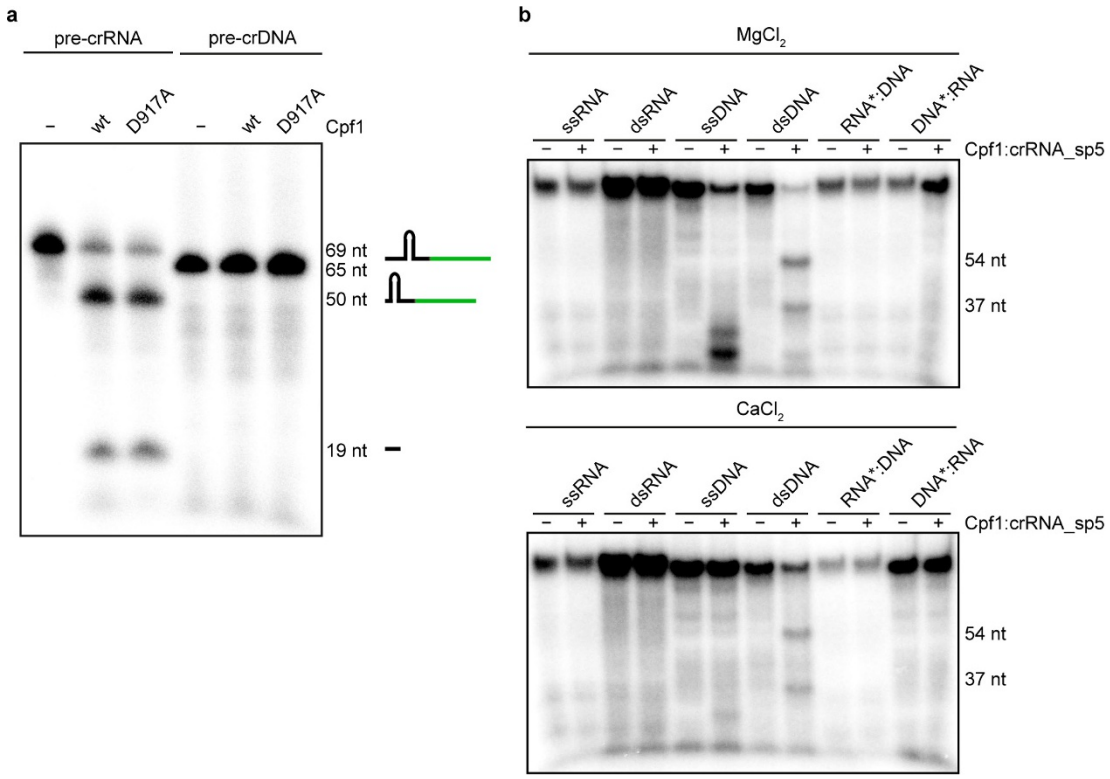
Extended Data Figure 6 | Cpf1 requires crRNA with an intact repeat structure to specifically cleave DNA. a, Cleavage assays of protospacer 5 containing supercoiled plasmid DNA by Cpf1 programmed with different RNA constructs (1, RNA construct containing spacer 4; 2–8, RNA constructs containing spacer 5) in the presence of 10 mM $CaCl_2$. Cleavage products were analysed by agarose gel electrophoresis and visualized by EtBr staining. **b**, Cleavage of 5′-radiolabelled oligonucleotide duplexes containing protospacer 5 in the presence of 10 mM $CaCl_2$. Cleavage products were analysed by denaturing polyacrylamide gel electrophoresis and visualized by phosphorimaging. Fragment sizes are indicated in nucleotides. RNA structures were generated with RNAfold[43] and visualized using VARNA[44] software. Only the RNAs containing a full-length repeat and a spacer complementary to the target mediate DNA cleavage by Cpf1.



Extended Data Figure 7 | Analysis of target DNA cleavage by crRNA-programmed Cpf1 in presence of Mg^{2+}. **a**, Cleavage assays of protospacer 5 containing supercoiled plasmid DNA by Cpf1 programmed with crRNA-sp4 or crRNA-sp5 (crRNA-sp4, repeat-spacer 4, processed, RNA 1, Extended Data Fig. 6; crRNA-sp5, repeat-spacer 5, processed, RNA 3, Extended Data Fig. 6) in absence or presence of 10 mM $MgCl_2$. Plasmid DNA cleavage was observed only with Cpf1 programmed with crRNA-sp5 in presence of Mg^{2+}. **b**, Oligonucleotide cleavage assays using Cpf1 programmed with crRNA-sp5 (repeat-spacer 5, processed, RNA 3, Extended Data Fig. 6) in presence of 10 mM $MgCl_2$. Either the target or the non-target strand was 5′-radiolabelled before annealing to the non-labelled complementary strand to form the duplex substrate. **c**, Sequencing analysis of the cleavage product obtained in **a**. The termination of the sequencing reaction indicates the cleavage site. Note that an enhanced signal for adenine is a sequencing artefact. **d**, Plasmid DNA containing protospacer 5 and the PAMs 1–6, or 5′-radiolabelled double-stranded oligonucleotide containing protospacer 5 and PAMs 1, 7–9 were subjected to cleavage by Cpf1 programmed with crRNA-sp5 (repeat-spacer, full-length, RNA 4, Extended Data Fig. 6) in the presence of 10 mM $MgCl_2$ (upper and lower panel, respectively). Oligonucleotide cleavage products are indicated in nucleotides.



Extended Data Figure 8 | Binding studies of Cpf1. a, EMSAs of 5′-radiolabelled double-stranded oligonucleotides containing protospacer 5 by Cpf1 programmed with RNA 1–6 (see Extended Data Fig. 6). The protein concentrations used were 8, 52 and 512 nM. Reactions were analysed by native PAGE and phosphorimaging. Unbound and bound DNAs are indicated. Higher DNA binding affinities are observed when Cpf1 is programmed with an RNA containing an entire repeat sequence. **b**, EMSAs of 5′-radiolabelled crRNA-sp5 (repeat-spacer 5, processed, RNA 3, Extended Data Fig. 6) by wild-type Cpf1, Cpf1(H843A), Cpf1(K852A), Cpf1(K869A) and Cpf1(F873A). The protein concentrations used were 2, 4, 8, 12, 16, 24, 32, 48 and 64 nM. Reactions were analysed by native polyacrylamide gel electrophoresis and phosphorimaging. Unbound and bound RNAs are indicated. Shown are representatives of at least three individual experiments. The bound and unbound RNA fractions were quantified, plotted against the enzyme concentration and fitted by nonlinear regression analysis. The calculated K_d values (± s.d.) were 16 ± 1 nM (wild type), 17 ± 0.5 nM (H843A), 12 ± 1 nM (K852A), 10 ± 1 nM (K869A) and 17 ± 1 nM (F873A). There are no differences between the RNA binding affinities of wild-type and mutant Cpf1. **c**, EMSAs of 5′-radiolabelled double-stranded oligonucleotides containing protospacer 5 targeted by wild-type Cpf1, Cpf1(D917A), Cpf1(E1006A) and Cpf1(D1255A) in complex with crRNA-sp5 (repeat-spacer 5, full-length, RNA 4, Extended Data Fig. 6). The protein concentrations used were 8, 16, 32, 42, 52, 64, 74, 128 and 256 nM. Reactions were analysed by native polyacrylamide gel electrophoresis and phosphorimaging. Unbound and bound DNAs are indicated. Shown are representative of at least three individual experiments. The bound and unbound DNA fractions were quantified, plotted against the enzyme concentration and fitted by nonlinear regression analysis. The calculated K_d values (± s.d.) were 50 ± 3 nM (wild type), 48 ± 8 nM (D917A), 40 ± 8 nM (E1006A) and 52 ± 6 nM (D1255A). There are no differences between the RNA-mediated DNA binding affinities of wild-type and mutant Cpf1. The reduced K_d for E1006A can be explained by the removal of the large negatively charged amino acid, which might facilitate interaction of Cpf1 with the DNA.



Extended Data Figure 9 | Processing activity of Cpf1 is specific for pre-crRNA and crRNA-mediated targeting of Cpf1 is directed only against single- and double-stranded DNA. a, Cpf1 processing activity was tested against pre-crRNA and pre-crDNA. Wild-type Cpf1 or Cpf1(D917A) (1 μM) was incubated with 200 nM internally labelled pre-crRNA-sp5 (repeat-spacer 5, full-length, RNA 4, Extended Data Fig. 6) or a 5′-labelled ssDNA (pre-crDNA-sp5) construct with the same sequence as the RNA in KGB buffer with 10 mM $MgCl_2$ for 5 min at 37 °C. Incubation of wild-type Cpf1 and DNase inactive mutant (Cpf1(D917A)) with the RNA construct, but not the DNA construct, resulted in the expected cleavage products of a 19-nt repeat fragment and a 50-nt intermediate crRNA, indicating that the processing activity of Cpf1 is specific for RNA. **b**, crRNA-mediated DNA cleavage activity of Cpf1. Cpf1 (100 nM) in complex with crRNA-sp5 (repeat-spacer 5, full-length, RNA 4, Extended Data Fig. 6) was incubated with 10 nM of 5′-radiolabelled ssRNA, dsRNA, ssDNA, dsDNA or RNA–DNA hybrids in KGB buffer with either $MgCl_2$ (10 mM; upper panel) or $CaCl_2$ (10 mM; lower panel) for 1 h at 37 °C. The oligonucleotide DNA substrates contained the sequence for protospacer 5 targeted by the tested crRNA. For DNA–RNA hybrids, the 5′-radiolabelled target strand is indicated with an asterisk. Only ssDNA and dsDNA substrates were cleaved, indicating that the crRNA-mediated cleavage activity of Cpf1 is only directed against DNA substrates. The cleavage products for ssDNA, however, vary from those expected or observed for dsDNA. Cleavage reactions were analysed by denaturing polyacrylamide gel electrophoresis and phosphorimaging. RNA cleavage products are indicated schematically. RNA and DNA fragment sizes are given in nucleotides.

Extended Data Table 1

a. Quantification of Figure 2e.

substrate	wt	T22G	C21A	T20G	A19C	A18C	T17G	T16G	T15G
% cleavage	83 ± 15	37 ± 1	41 ± 2	22 ± 3	30 ± 2	33 ± 4	28 ± 11	39 ± 18	57 ± 2
substrate	T14G	C13A	C12A	A11C	T10G	T9G	A8C	A7C	G6T
% cleavage	69 ± 9	77 ± 13	87 ± 6	68 ± 12	79 ± 5	100 ± 0	65 ± 25	79 ± 16	92 ± 14
substrate	A5C	T4G	A3C	G2T	A1C	Mut_1-4		Mut_19-22	
% cleavage	75 ± 35	55 ± 27	62 ± 19	66 ± 24	64 ± 24	47 ± 25		0	

Percent cleavage is the result of three independent experiments ± standard deviation.

b. Quantification of Extended Data Figure 5b.

ion	Ca^{2+}		Mg^{2+}		Mn^{2+}		Co^{2+}		Ni^{2+}		Zn^{2+}	
concentration	1 mM	10 mM	1 mM	10 mM	1 mM	10 mM	1 mM	10 mM	1 mM	10 mM	1 mM	10 mM
% cleavage	44 ± 17	82 ± 8	13 ± 10	84 ± 10	39 ± 17	86 ± 2	0	0	0	0	0	0

Percent cleavage is the result of three independent experiments ± standard deviation.