nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-017-0103-5

The host-encoded RNase E endonuclease as the crRNA maturation enzyme in a CRISPR–Cas subtype III-Bv system

Juliane Behler[1], Kundan Sharma[2,3], Viktoria Reimann[1], Annegret Wilde[4,5], Henning Urlaub[2,3] and Wolfgang R. Hess[1,6*]

Specialized RNA endonucleases for the maturation of clustered regularly interspaced short palindromic repeat (CRISPR)-derived RNAs (crRNAs) are critical in CRISPR–CRISPR-associated protein (Cas) defence mechanisms. The Cas6 and Cas5d enzymes are the RNA endonucleases in many class 1 CRISPR–Cas systems. In some class 2 systems, maturation and effector functions are combined within a single enzyme or maturation proceeds through the combined actions of RNase III and *trans*-activating CRISPR RNAs (tracrRNAs). Three separate CRISPR–Cas systems exist in the cyanobacterium *Synechocystis* sp. PCC 6803. Whereas Cas6-type enzymes act in two of these systems, the third, which is classified as subtype III-B variant (III-Bv), lacks *cas6* homologues. Instead, the maturation of crRNAs proceeds through the activity of endoribonuclease E, leaving unusual 13- and 14-nucleotide-long 5′-handles. Overexpression of RNase E leads to overaccumulation and knock-down to the reduced accumulation of crRNAs in vivo, suggesting that RNase E is the limiting factor for CRISPR complex formation. Recognition by RNase E depends on a stem-loop in the CRISPR repeat, whereas base substitutions at the cleavage site trigger the appearance of secondary products, consistent with a two-step recognition and cleavage mechanism. These results suggest the adaptation of an otherwise very conserved housekeeping enzyme to accommodate new substrates and illuminate the impressive plasticity of CRISPR–Cas systems that enables them to function in particular genomic environments.

Many bacteria and archaea possess an RNA-based adaptive and inheritable immune system against viruses and other foreign genetic elements[1–5]. These systems are known as clustered regularly interspaced short palindromic repeats (CRISPRs) consisting of an array of alternating identical repeats, varying spacer sequences and CRISPR-associated (Cas) proteins[3,4,6]. Based on gene conservation and locus organization, two major classes of CRISPR–Cas systems, six major types (I–VI) and at least 19 subtypes are differentiated[5–7]. Transcription of the CRISPR repeat-spacer array leads to a long precursor transcript (pre-crRNA)[8,9] that is cleaved into short CRISPR-derived RNAs (crRNAs)[6]. The maturation of crRNAs is essential for the functionality of CRISPR–Cas systems and allows the interference complex to target and destroy invader nucleic acids upon base pairing[6,10]. In type I, III, V and VI systems, a specialized endoribonuclease mediates processing of the pre-crRNAs[9,11–13]. In several type I and III systems, these endoribonucleases belong to the Cas6 family[6,9,14], while Cas5d is found in subtype I-C systems lacking Cas6[10,15,16]. The enzyme for the ribonucleolytic step removing the 3′-end stem-loop structure in type III systems with a 6-nucleotide periodicity is not known[10,17]. Type II systems rely on *trans*-activating CRISPR RNAs (tracrRNAs) that mediate pre-crRNA maturation via the housekeeping RNase III in the presence of Cas9 (ref. [10]). In some class 2 CRISPR–Cas systems, single effector enzymes—namely Cpf1, C2c1, C2c2 and C2c3—have been discovered[18]. These enzymes have recently been renamed into Cas12a, Cas12b, Cas13a and Cas12c, respectively[19]. Cas12a is a dual-nuclease that performs crRNA biogenesis and targets DNA during interference in subtype V-A systems[20]. The effector enzymes Cas12b and Cas12c, which were recently identified in subtype V-B and V-C systems, are distantly related to Cas12a[18]. The biogenesis of mature crRNAs by Cas12b relies on a tracrRNA, and DNA cleavage depends on both the crRNA and tracrRNA[18], similar to type II systems. Cas13a is characteristic of type VI CRISPR–Cas systems and maintains two distinct RNase activities[18,21]. One is responsible for tracrRNA-independent crRNA maturation, while the other is involved in the cleavage of single-stranded RNA targets[21,22]. Nevertheless, there are some CRISPR–Cas systems that lack any known maturation ribonuclease[5]. These systems might depend on other CRISPR–Cas systems in the same organism[5] or on ribonucleases that have yet to be identified.

The cyanobacterium *Synechocystis* sp. PCC 6803 (*Synechocystis* 6803) harbours three separate CRISPR–Cas systems—CRISPR1, CRISPR2 and CRISPR3—located on the pSYSA plasmid. CRISPR1 and CRISPR2 are classified as subtypes I-D and III-D systems, respectively[23,24]. In contrast, CRISPR3 resembles a subtype III-B system, but possesses a *cmr1–cmr6* fusion and lacks an obvious *cas6* homologue (Fig. 1a), characterizing it as a subtype III-B variant (III-Bv). CRISPR1 and CRISPR2 encode one or two system-associated Cas6 endonucleases (the genes *cas6-1* (*slr7014*), *cas6-2a* (*slr7068*) and *cas6-2b* (*sll7075*))[23]. However, CRISPR3 lacks *cas6* and any other obvious *cas* genes involved in crRNA maturation[23]. Here, we show that mature CRISPR3 crRNA possesses an unusually long

[1]Genetics and Experimental Bioinformatics, Faculty of Biology, University of Freiburg, Freiburg, Germany. [2]Bioanalytics Research Group, Department of Clinical Chemistry, University Medical Centre, Göttingen, Germany. [3]Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. [4]Molecular Genetics, Faculty of Biology, University of Freiburg, Freiburg, Germany. [5]BIOSS Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany. [6]Freiburg Institute for Advanced Studies, University of Freiburg, Freiburg, Germany.
*e-mail: wolfgang.hess@biologie.uni-freiburg.de

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

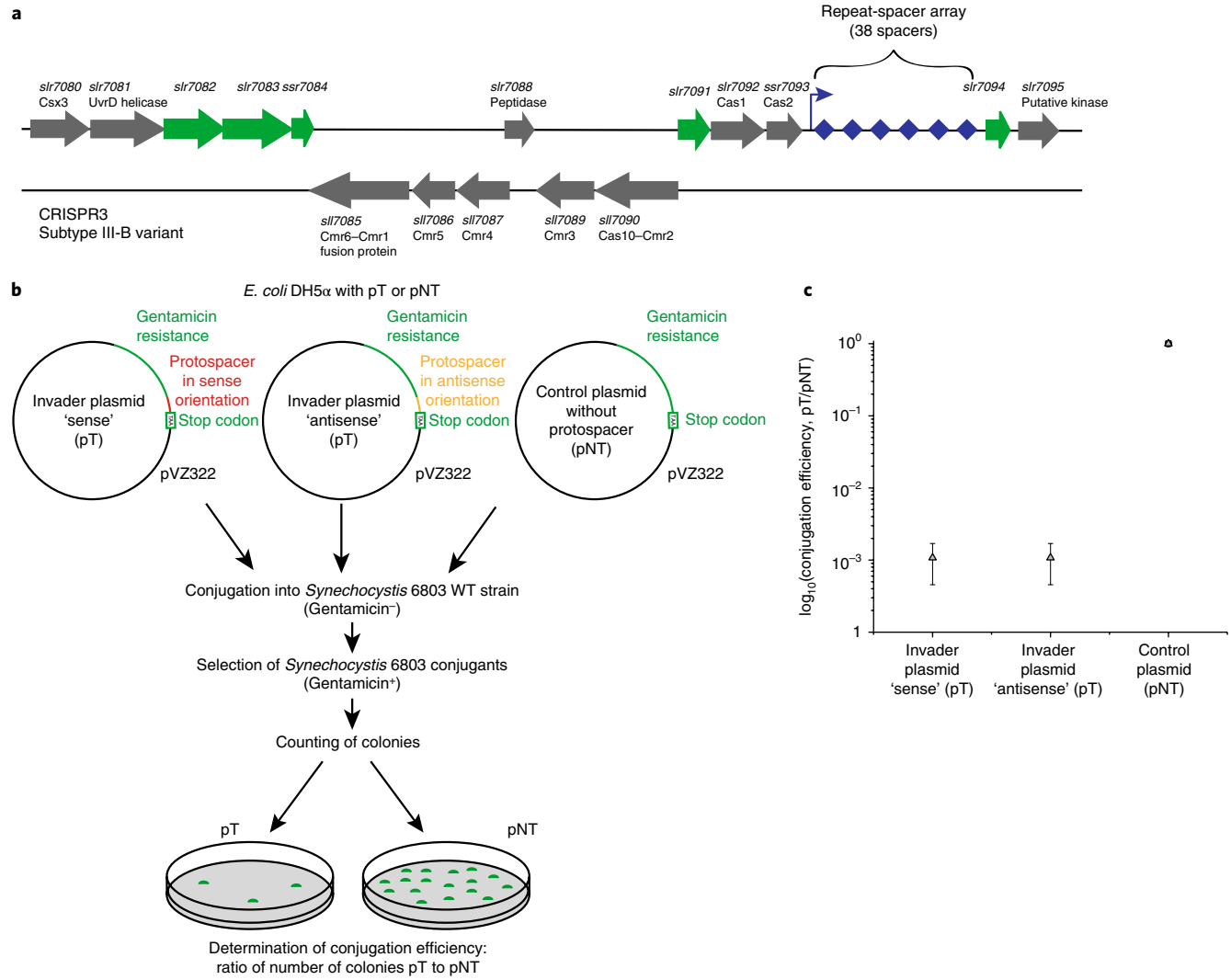


Fig. 1 | Organization of *Synechocystis* 6803 CRISPR3 *cas* genes, repeat-spacer array and interference assay. a, The subtype III-B variant CRISPR–Cas system is located on the pSYSA plasmid. Several *cas* genes are located upstream of the CRISPR array, which contains 39 repeat and 38 spacer sequences. Arrows in green illustrate genes coding for hypothetical proteins, whereas arrows in grey represent annotated *cas* genes. The promoter (blue arrow) 5′ of the CRISPR array is followed by the leader and the repeat (blue rectangles)-spacer array. **b**, Experimental setup of a conjugational interference assay for the detection of CRISPR3 interference activity in *Synechocystis* 6803. Three plasmids were used during the conjugation assay: two target plasmids (pT, plasmid target), containing a protospacer fused to the selection marker, and a control plasmid (pNT, plasmid non-target) harbouring only the selection marker without the protospacer sequence. The used protospacer corresponds to endogenous spacer 2 from CRISPR3 in either the sense (native orientation of spacer 2 in CRISPR3) or antisense (reverse complement orientation of spacer 2 in CRISPR3). If only the sense protospacer is targeted during interference, the system can successfully attack a DNA invader. In contrast, RNA is targeted when the antisense protospacer is affected by degradation. After conjugation, the number of *Synechocystis* 6803 colonies carrying the plasmid with the selection marker was counted and the conjugation efficiency was determined. Interference is observed if the ratio of the number of pT to pNT colonies is <1 in triplicate assays. **c**, Interference activity of CRISPR3 in *Synechocystis* 6803. The conjugation efficiencies are calculated by the ratio of pT to pNT. The conjugation efficiency of the control plasmid was set to 1 and the number of colonies for the pT sense and pT antisense plasmids was normalized to the control plasmid pNT. Data points are shown as dot plots and standard deviations were calculated for three independent biological replicates. Data points with 0 values are not shown because log(0) is not defined.

5′-handle and that crRNA maturation is independent of any *cas* genes. We demonstrate that the CRISPR3 repeat sequence is a suitable substrate for the housekeeping endoribonuclease E and present in vivo and in vitro evidence for crRNA biogenesis mediated by RNase E in *Synechocystis* 6803.

Results

The CRISPR3 system is functional. To test the activity and functionality of the CRISPR3 system, we performed an interference test using a conjugative vector with a protospacer motif in two orientations and the gentamicin resistance gene as a reporter. When the reporter gene is lost due to the degradation by the effector proteins, conjugants become sensitive to gentamicin. When the system is not active, transconjugants maintain the invader plasmid and show resistance. We observed a difference of several orders of magnitude in the numbers of colonies between vectors containing the protospacers and the control (Fig. 1b,c). We conclude that the *Synechocystis* 6803 CRISPR3 system is fully functional despite the Cmr1–Cmr6 fusion and the lack of an associated *cas6* gene.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

In vivo mapping of CRISPR3 crRNA processing sites reveals unusual 13- and 14-nucleotide 5′-repeat handles. In vivo and in vitro data showed 8-nucleotide 5′-handles for the mature crRNAs in the *Synechocystis* 6803 CRISPR1 and CRISPR2 systems, which resulted from processing by the two associated endoribonucleases, Cas6-1 and Cas6-2a[23,24]. CRISPR3 differs in its repeat sequence, length and secondary structure from the two other systems, positioning the hairpin at the 5′- rather than 3′-end of the repeat[23]. Therefore, we investigated the in vivo CRISPR3 crRNA processing sites by primer extension. Prominent signals indicated the presence of unusually long 13- and 14-nucleotide 5′-handles within the mature crRNAs, with the 13-nucleotide 5′-handles appearing as major maturation products (Fig. 2). Additional 5′-ends were observed within the spacer sequences and in the repeat sequences leading to 19- and 20-nucleotide 5′-extensions and a 2-nucleotide CRISPR3 leader sequence. Spacer processing sites were mainly situated at the 3′-end of the spacers. Interestingly, the majority of CRISPR3 crRNA processing sites occurred in a 6-nucleotide periodicity or multiples of it. The 13- and 14-nucleotide 5′-handles found for the shortest extension fragments correspond to the mature crRNAs with a single spacer sequence. These data were also in agreement with other primer extension experiments (Supplementary Fig. 1). The identification of 5′-handles longer than the 8 nucleotides that are typically found for Cas6-type enzymes[14,25–27] suggested the involvement of an unknown endoribonuclease.

Deletion experiments verify that non-cognate Cas6 endonucleases are not involved in CRISPR3 maturation. To test whether any genetic components associated with the CRISPR1 or CRISPR2 systems mediate CRISPR3 pre-crRNA processing, we deleted the entire loci separately from each other and together. The resulting *Synechocystis* ΔC1, ΔC2 and ΔC1ΔC2 strains were devoid of the respective full-length CRISPR–Cas loci including the *cas* genes and repeat-spacer arrays. Northern blot analysis showed no differences in the pattern of CRISPR3-derived crRNAs and intermediates in ΔC1 and ΔC2 or in ΔC1ΔC2 compared with the wild-type (WT) strain (Supplementary Fig. 2a). Thus, neither CRISPR1 nor CRISPR2 genes are involved in crRNA maturation in the CRISPR3 system. To screen for additional candidates, single knock-out mutants were generated and analysed by northern hybridization for the genes *slr7080–slr7091* (see Fig. 1a for gene functions), *slr7095* (putative nucleotide kinase), *slr0346* (RNase III), *slr1646* (RNase III) and *slr0945* (mini-RNase III). We further screened a conditional double knock-out of *slr0346* and *slr1646*, as well as quintuple mutants *slr7096–slr7100* (*slr7097* encodes α-ketoglutarate-dependent dioxygenase AlkB, while the remaining genes encode hypothetical proteins) and *slr7101–7105* (encoding hypothetical proteins, except *sll7103*, which encodes exodeoxyribonuclease V α-chain and Slr7102 containing a YddW domain). However, none of these mutants was found to be involved in the pre-crRNA processing of CRISPR3 (examples are shown in Supplementary Fig. 2b). Therefore, an alternative, host-encoded endoribonuclease must be responsible for CRISPR3 pre-crRNA maturation.

CRISPR3 repeat RNA resembles an RNase E substrate. RNase E is a widely conserved and essential endoribonuclease that exhibits certain sequence and structural preferences for efficient cleavage[28–30]. RNA-Seq data analysis suggested[23] that cleavage in the CRISPR3 repeat occurs in an adenylate–uridylate-rich region downstream of a stem-loop element that could provide the secondary structure required for efficient RNase E binding and cleavage[29]. The first cleavage in CRISPR3 maturation occurs at the 3′-end of the spacer, before repeat processing, yielding 5′-monophosphorylated crRNA intermediates[23], which are the preferred substrates for RNase E[31]. Furthermore, a consensus motif was described[32], revealing a 2-nucleotide uridine ruler-and-cut mechanism in which RNase E senses a uridine located at the +2 position downstream of the cleavage site. For comparison, we show here the previously identified RNase E cleavage sites within the *psbA2* messenger RNA leader from *Synechocystis* 6803 (ref. [29]) (Fig. 3a). Whereas *psbA2* harbours two major and six minor cleavage sites, the CRISPR3 crRNA contains two sites. Both CRISPR3 crRNA and *psbA2* are cleaved within adenylate–uridylate-rich sequences downstream of a short hairpin. Furthermore, both cleavage sites in the CRISPR3 repeat and both major sites in *psbA2* are followed by a uridine at the +2 position. Together, these results indicate that the CRISPR3 repeat is a candidate for cleavage by RNase E.

RNase E cleavage assays with native and mutated CRISPR3 repeat variants. To directly test CRISPR3 repeats as possible RNase E substrates, in vitro cleavage assays were performed. RNase E processes the CRISPR3 WT repeat yielding 22- and 23-nucleotide fragments (Fig. 3b). These fragments correspond to the longer 5′-segment of the repeat harbouring the hairpin (Fig. 3a and Supplementary Fig. 3) and are in agreement with the in vivo determined 5′-handles. Modifications located within or near the cleavage site did not abolish RNase E cleavage activity entirely, but resulted in less-efficient cleavage or cleavage at secondary sites (Fig. 3b and Supplementary Fig. 4). The substitution of the ribose sugar to deoxyribose on the uracil within the cleavage site ($C3dU_{22}$) suppressed the formation of the shorter cleavage fragment. In contrast, mutations within the stem ($C3C_{15}C_{17}$) abolished cleavage, indicating that RNase E recognition and activity relies on the stem of the CRISPR3 repeat. In contrast, cleavage site mutations ($C3A_{21}A_{22}$) and mutations interfering with the 2-nucleotide ruler-and-cut mechanism[32] ($C3G_{23}G_{24}$, $C3mA_{23}mA_{24}$, $C3G_{25}$ and $C3G_{25}G_{26}$) weakened and changed the cleavage pattern as if the substrate was recognized, but could not be cleaved at the proper site. Thus, the performed in vitro assays confirmed RNase E as a processing endonuclease for CRISPR3 repeats and identified nucleotides required for proper cleavage.

Manipulated RNase E expression affects CRISPR3 transcript accumulation. To investigate whether RNase E expression levels affect the accumulation of CRISPR3 crRNAs, the enzyme was overexpressed from a conjugative plasmid. RNase E overexpression led to elevated amounts of CRISPR3 pre-crRNA, intermediates and mature crRNAs over time (Fig. 4a). The highest crRNA levels were detected after 120 h of overexpression. This result provides in vivo evidence for a direct link between increased crRNA processing by RNase E overexpression and increased ribonucleoprotein complex formation for stabilizing the mature crRNA. RNase E is essential in *Synechocystis* 6803, and attempts to delete its gene remained futile[33,34]. Therefore, we created a conditional RNase E knock-down using a CRISPR interference-based approach and a partial RNase E deletion mutant. Northern blot analysis revealed the lowest CRISPR3 crRNA levels after 144 h of single-guide RNA (sgRNA)/dead Cas9 (dCas9) induction. Reduced crRNA accumulation was also observed in the partial RNase E deletion (Fig. 4b). These observations strengthen the existence of a direct link between crRNA processing and RNase E availability, which appears to be the limiting step for CRISPR3 crRNA maturation.

Protein–RNA cross-linking to identify CRISPR3 interacting sites in RNase E. To identify the CRISPR3 repeat-binding regions within RNase E, we used ultraviolet light-induced protein–RNA cross-linking coupled to mass spectrometry (MS) analysis[35,36]. We identified four different amino acids as cross-linked sites—namely, F64, K98, K494 and K512 (Fig. 5a and Supplementary Fig. 5). Two of these four residues—F64 and K98—are located within the amino (N)-terminal S1 RNA-binding domain of RNase E[28]. K494 and K512 are positioned within the carboxy (C)-terminal third of the enzyme in a non-conserved flexible

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

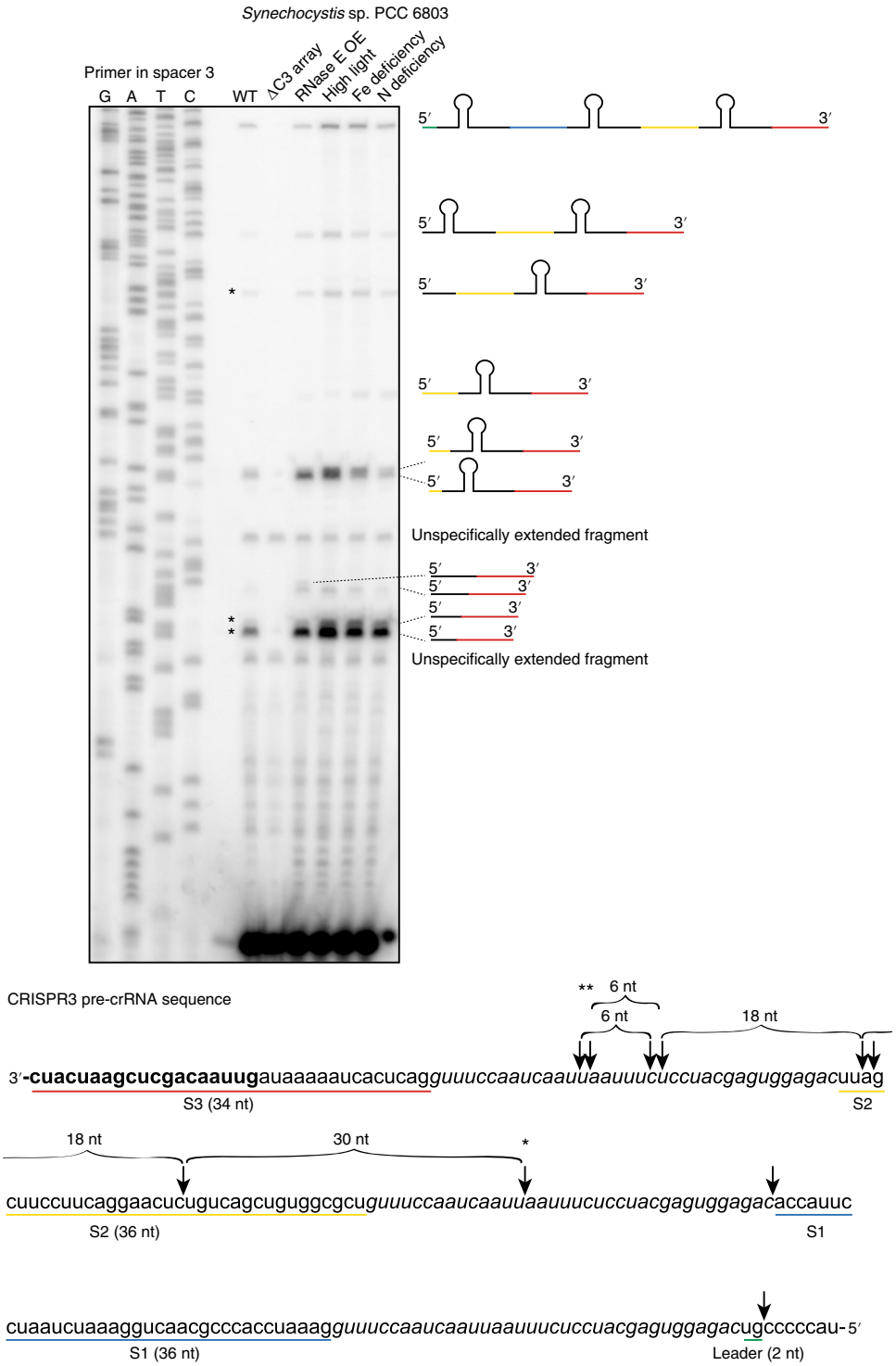


Fig. 2 | Mapping of in vivo processing sites within CRISPR3 pre-crRNA by primer extension. Total RNA was isolated from WT *Synechocystis* 6803 (WT), the CRISPR3 repeat-spacer deletion mutant (ΔC3 array, negative control) and an RNase E overexpression mutant (RNase E OE) under standard growth conditions, as well as the WT exposed to high light, iron or nitrogen deficiency and subjected to primer extension analysis using the 5′-end-labelled oligonucleotide primerExt_S3_rev (Supplementary Table 1) and 2 μg of total RNA per reaction. To test whether the length of the 5′-handle is influenced by stress conditions, RNA isolated from cells that were exposed to such conditions was investigated. The detected sites are indicated by cartoons (upper part, repeat sequences in black; spacer and leader sequences in colour) or arrows alongside the reverse complement pre-crRNA sequence (lower part, primer sequence in bold). Mapping of in vivo cleavage sites within CRISPR3 repeats revealed unusual 13- and 14-nucleotide (nt) 5′-handles (highlighted by asterisks) and a 2-nt-long leader sequence. The majority of additional processing sites followed a 6-nt periodicity or multiples of it, as marked by the curly braces. A representative of two independent experiments is shown.

region that corresponds to the previously defined variable subregion V2[37]. While the N-terminally located sites provide protein–RNA interactions, the C-terminal residues could play a role in mediating CRISPR specificity due to their location in

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

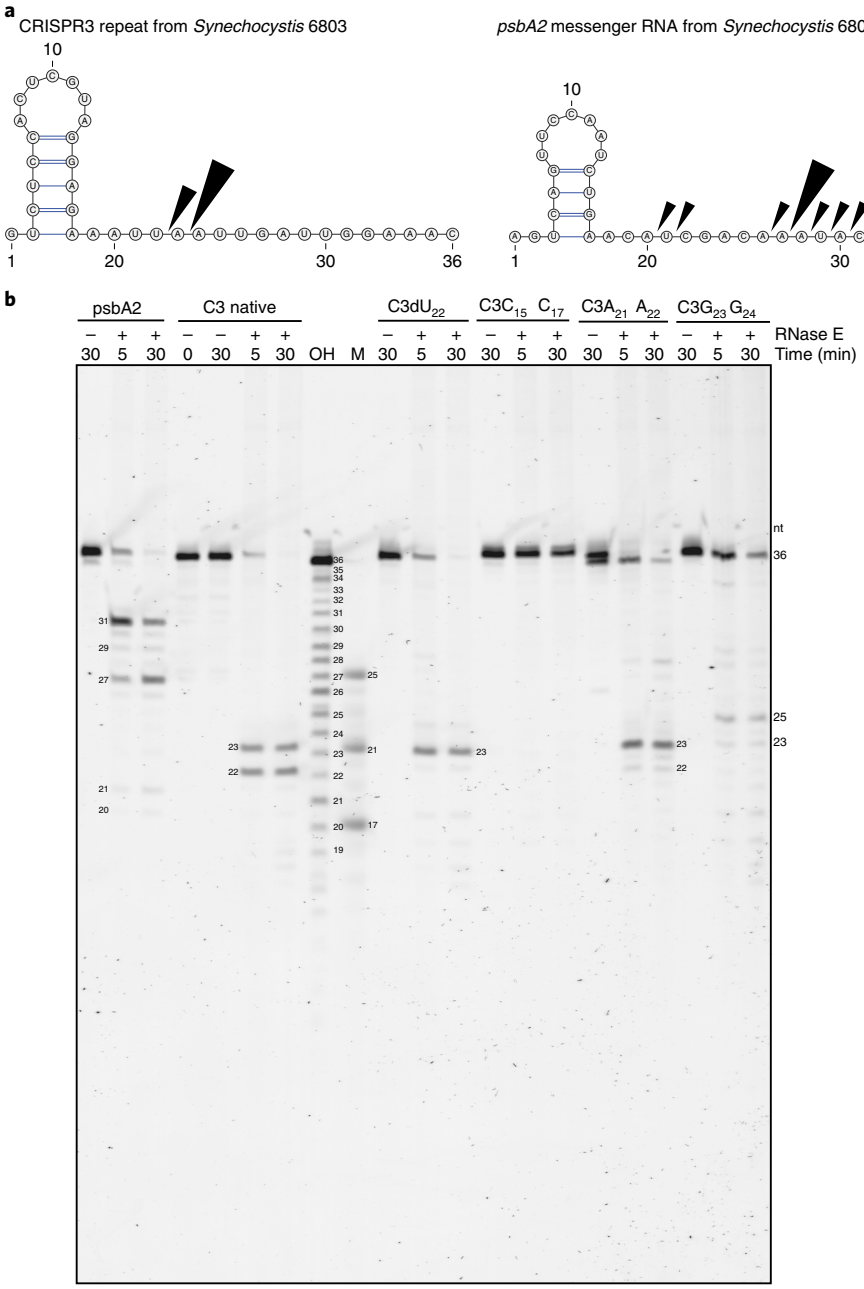


Fig. 3 | RNase E substrates and cleavage sites in *Synechocystis* 6803. a, The CRISPR3 repeat RNA forms a hairpin at its 5′-end and cleavage sites are located within an adenylate–uridylate-rich sequence downstream of the stem-loop. These features resemble the verified RNase E substrate within the *psbA2* of *Synechocystis* 6803 (ref. [29]). RNA structures were predicted by RNAfold[57] or adapted from published work[29] and visualized using VARNA[58]. Black wedges indicate RNase E cleavage sites and larger wedges mark preferred cleavage sites. **b**, In vitro RNase E cleavage assay of the native CRISPR3 repeat RNA and mutated versions. Cleavage of the native CRISPR3 repeat resulted in two fragments of 22 and 23 nucleotides (nt), which correspond to the repeat 5′-fragments including the hairpin. Substitution of the ribose sugar to deoxyribose on the uracil at the cleavage site ($C3dU_{22}$) prevented the formation of the shorter cleavage fragment, whereas mutations within the stem ($C3C_{15}C_{17}$) completely abolished cleavage. Mutations around the cleavage site and at position +2 downstream of the processing site revealed secondary targets. The *psbA2* RNA oligonucleotide was used as a positive control. OH is the alkaline hydrolysis ladder derived from the CRISPR3 repeat and M is a microRNA Marker (NEB). A representative of three independent experiments is shown.

the non-conserved region of the enzyme. The stretch containing these two residues is conserved in the RNase E protein of the closely related strain *Synechocystis* 6714 (Fig. 5b). This strain shares on plasmid pSYLA the same type of CRISPR–Cas system lacking Cas6 (ref. [38]), pointing to the possibility that the use of RNase E as the CRISPR maturation endoribonuclease is not restricted to *Synechocystis* 6803.

Biochemical analysis of mutagenized RNase E variants in vitro. To characterize the relevance of the four cross-linked sites, we constructed an F64A–K98A mutant, a K494A–K512A double mutant and an F64A–K98A–K494A–K512A quadruple mutant. In vitro cleavage assays showed that the F64A–K98A and F64A–K98A–K494A–K512A mutations led to significant loss of RNA cleavage (Fig. 5c). The RNase E variant with the K494A–K512A double mutation

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

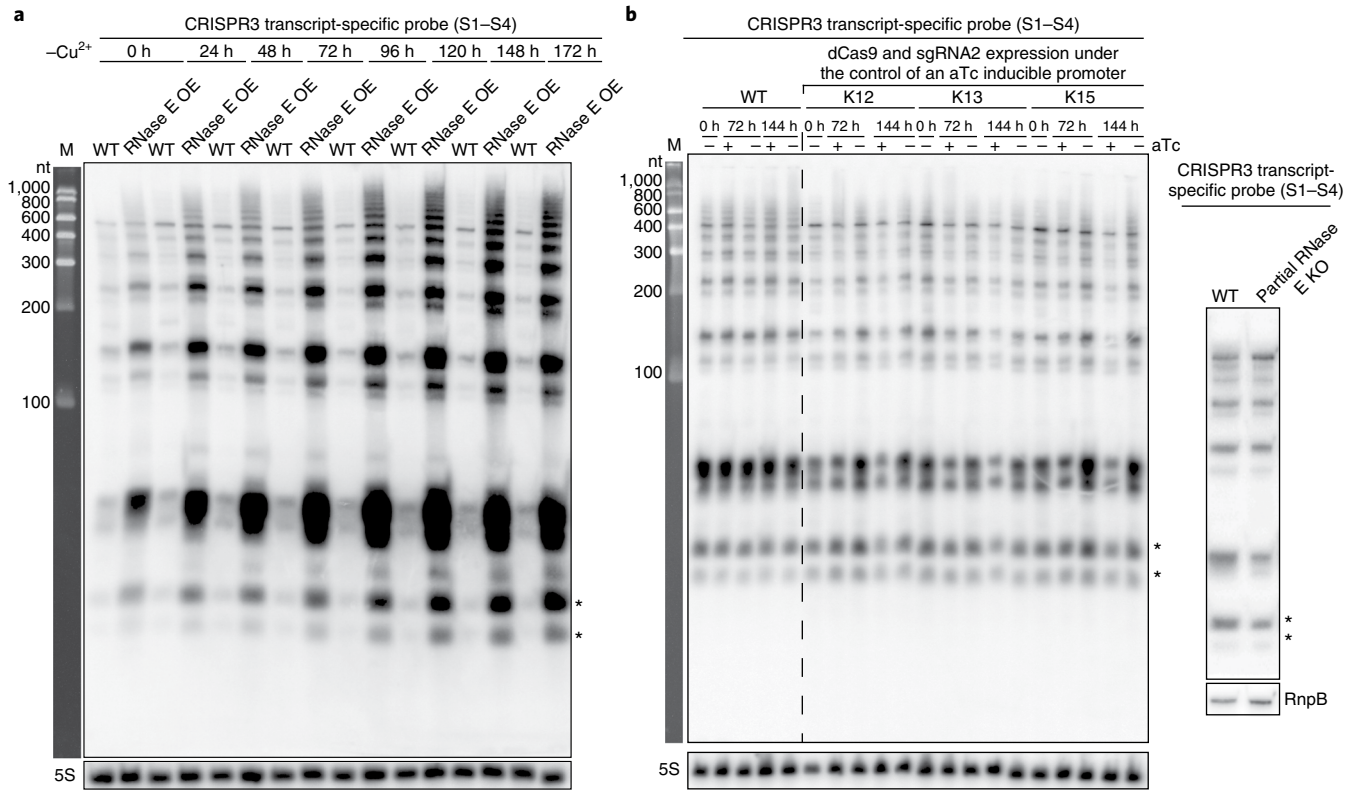


Fig. 4 | Manipulation of RNase E expression affects CRISPR3 crRNA accumulation. Northern hybridization using a radioactively labelled transcript probe spanning CRISPR3 spacers 1–4. Mature crRNAs are indicated by asterisks. **a**, Increasing transcript levels of CRISPR3-derived crRNAs were observed upon RNase E overexpression from the copper-repressible promoter P_{petJ} over time. The highest expression levels were detected approximately 120 h after copper deficiency. A representative of two independent experiments is shown. **b**, Left: conditional knock-down of RNase E transcription in three independent biological replicates. The combined expression of dCas9 and sgRNA from the anhydrotetracycline (aTc)-inducible promoter P_{L22} led to a decrease in the CRISPR3-derived crRNA accumulation over time. In contrast, WT and non-induced control samples showed unchanged crRNA levels. A representative of three independent experiments is shown. Right: analysis of a partial RNase E knock-out (KO) mutant exhibiting lowered CRISPR3 crRNA accumulation compared with the WT strain. A representative of two independent experiments is shown. Hybridization against 5S rRNA (5S) or RNase P RNA (RnpB) was performed to control for equal loading. M, RiboRuler Low Range RNA Ladder; RNase E OE, RNase E overexpression mutant.

showed RNA binding and cleavage similar to the native enzyme, suggesting that these residues are not directly involved in catalysis. However, we noticed two arginine-rich patches (seven arginine residues in a row) closely before and after K494 and K512 (five out of eight residues) (Fig. 5b and Supplementary Fig. 5). Therefore, we cannot exclude that these residues are in close proximity to an RNA-binding element but not directly involved, consistent with findings from similar protein–RNA cross-linking studies[39,40].

Discussion

Some subtype III-A and III-B CRISPR–Cas systems lack any associated Cas6 or other known enzyme for crRNA maturation[5]. It was speculated that other CRISPR–Cas systems co-existing in the same organism might provide this activity[5].

Here, we show that this is not the case for the subtype III-Bv CRISPR3 system from *Synechocystis* 6803. Possible substrate promiscuity for the endogenous Cas6-1, Cas6-2a or Cas6-2b activities was excluded by analysis of deletion mutants for both the CRISPR1 and CRISPR2 loci, demonstrating that CRISPR3 crRNA maturation depends on the activity of another endoribonuclease. We present in vivo and in vitro evidence for the housekeeping endoribonuclease RNase E to perform this pre-crRNA processing. We identified unusually long 13- and 14-nucleotide 5′-handles within the mature crRNAs pointing at substantial differences compared with other known crRNA biogenesis pathways. Furthermore, some crRNA processing sites occurred in regular 6-nucleotide intervals or multiples of it indicating a ruler mechanism involved in the maturation of pre-crRNAs. This observation was also made for other type III systems and was described as a backbone-mediated 5′-ruler mechanism[17,33]. Cas6-based maturation typically produces 8-nucleotide-long 5′-handles[6,24,41,42] while Cas5d generates 11-nucleotide-long 5′-handles[6]. Even longer 5′-handles have been observed for some class 2 enzymes. The Cas13a-dependent *Leptotrichia shahii* DSM 19757 locus exhibits mature crRNAs with 28-nucleotide 5′-handles and the Cas13a locus from *Listeria seeligeri serovar* 1/2b str. SLCC3954 produces crRNAs with a 29-nucleotide 5′-handle[18,22]. The Cas12a nuclease from *Francisella novicida* generates 19-nucleotide 5′-handles in vitro, whereas they are only 14 to 16 nucleotides long in vivo[20].

For type III systems, base-pairing potential between the crRNA 5′-handle and the 3′-flanking region of the target DNA appears to play an important role in discriminating self from non-self nucleic acids[41–43]. Typically, base pairing of two or three nucleotides in the 5′-handle is sufficient to prevent DNA degradation and autoimmunity[41,43]. In the *Synechocystis* 6803 subtype III-Bv system, the nucleotides 11–13 are essential and another one is required to efficiently prevent autoimmunity (Supplementary Fig. 6). Therefore, we conclude that at least four nucleotides of the unusually long 5′-handle are involved in the discrimination between self and non-self targets. Other nucleotides might be involved in additional CRISPR-related functions; for example, the recruitment of crRNAs to the Cmr ribonucleoprotein complex. Our results indicate that

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

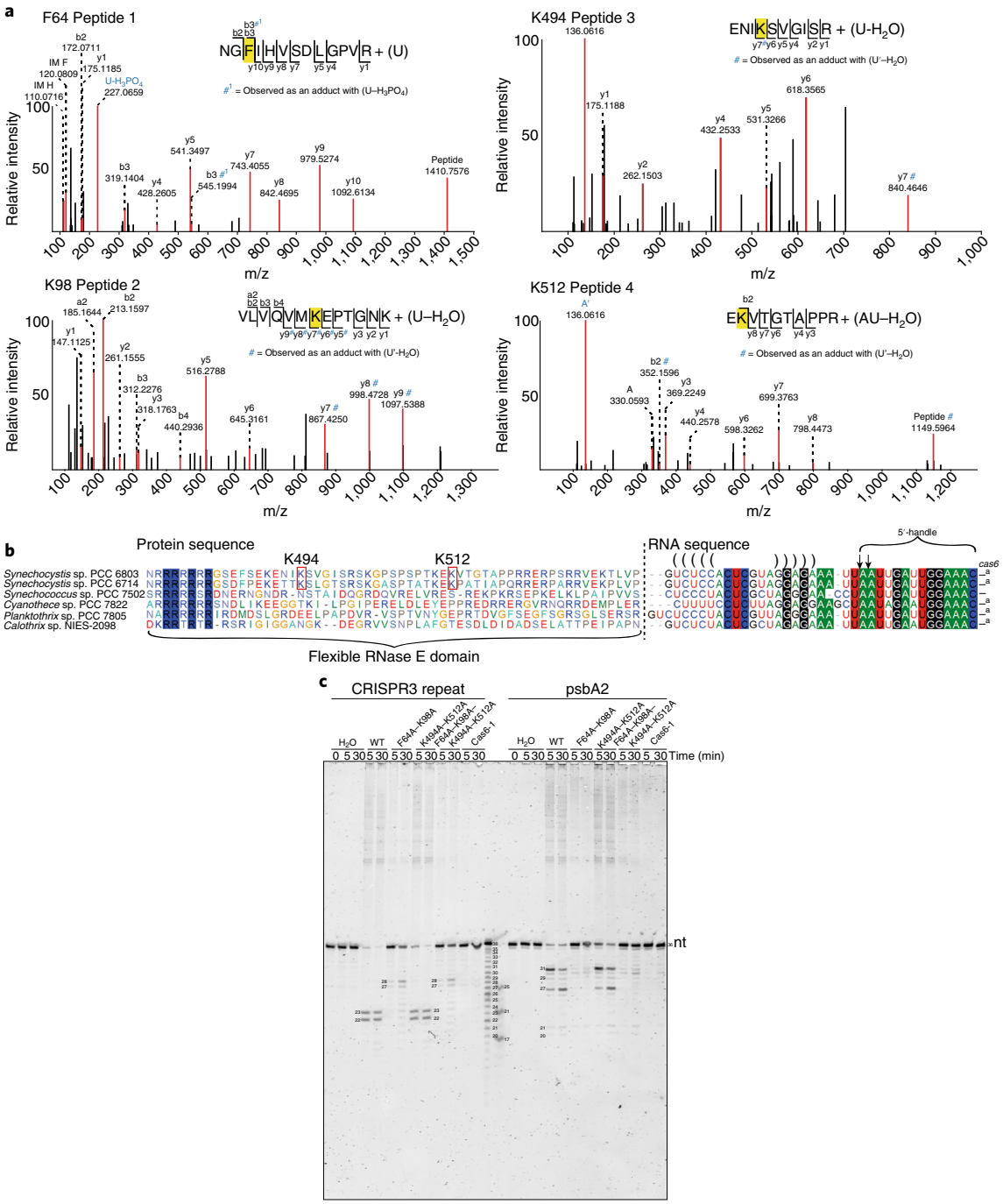


Fig. 5 | Identification and analysis of RNase E amino acids interacting with CRISPR3 repeat RNA. a, Tandem MS fragmentation spectra of protein–RNA cross-links identified within the RNase E–CRISPR3 repeat complex in vitro. The sample spectra for the four peptides that were observed to be cross-linked with an RNA moiety are shown. In each spectrum, the cross-linked peptide sequence and its corresponding y- and b-type fragment ions are indicated at the top, referring to the ions that retain the charge on the N or C terminus, respectively. The directly cross-linked amino acid residues are highlighted. Some of the b- and y- ions were observed with a mass shift labelled # and #[1] corresponding to U′-H_2O and U-H_3PO_4. The cross-linked nucleotides are given in brackets. IM, immonium ions; U′, base of U with a marker ion of 112.0273 Da. **b**, Multiple sequence alignment of the flexible RNase E domain of various cyanobacteria together with the respective CRISPR repeat sequence (hairpin indicated by brackets) and information about the absence of *cas6* genes (– indicates no *cas6* detectable; –[a] indicates no *cas6* associated with the respective system). The flexible sequence stretch containing the mapped K494 and K512 sites (boxed in red) of *Synechocystis* 6803 is conserved at this position only in *Synechocystis* 6714. However, several other positively charged residues (coloured in blue) are clustered close by, also in the other compared sequences pointing to their possible relevance as part of an RNA-binding element. **c**, In vitro cleavage assay of the CRISPR3 repeat RNA oligo and mutagenized RNase E variants. The repeat was incubated with WT and mutagenized RNase E enzyme variants. Enzyme concentrations were normalized to the WT enzyme concentration via western blotting and quantification using Quantity One software. WT RNase E and the K494A–K512A mutant exhibited comparable cleavage performance. The F64A–K98A double mutant and F64A–K98A–K494A–K512A quadruple mutant showed significant deficiencies in RNA cleavage. The *psbA2* RNA oligonucleotide was used as a positive control. Since the results with the two different substrates were consistent, no further replications were performed. Cas6-1 was expressed and purified as described for RNase E and served as a negative control. OH is the CRISPR3 repeat after alkaline hydrolysis, while M is the microRNA Marker (NEB).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

immunity was achieved by DNA destruction and, even if we cannot conclude about RNA targeting, along with the observations of 6-nucleotide periodicity in crRNA length and lack of immunity in the presence of anti-tag targets, suggest that the *Synechocystis* 6803 type III-Bv system is fully functional and operates similarly to other type III systems.

Multiple RNase E cleavage sites located in close proximity have been reported for diverse substrates and organisms, which are consistent with our data[29,30,32]. Moreover, RNase E seems to have a preference for a distinct cleavage site when multiple sites are available, resulting in the accumulation of fragments of varying abundances[29,32]. For the CRISPR3 repeat, we observed an equal ratio between 13- and 14-nucleotide-long 5′-handles in vitro, whereas the preference for the fully matured 13-nucleotide 5′-handle was only observed in vivo. We cannot rule out that this resulted from another in vivo maturation step.

We identified four RNase E amino acid residues that were cross-linked to CRISPR3 repeat nucleotides, indicating their possible importance for substrate recognition, binding and cleavage. Our results demonstrate that the mutation of F64A–K98A within the S1 RNA-binding domain prevents CRISPR3 repeat cleavage. These residues are highly conserved among RNase E and RNase G sequences from *Arabidopsis thaliana*, rice and tomato to *Synechocystis* 6803 and *Escherichia coli*, and play a crucial role in substrate binding and catalytic activity in *Arabidopsis* RNase E[28]. Here, we verified the importance of the two cross-linked sites in the S1 domain of *Synechocystis* 6803 RNase E, which correlated well with the observations made by Schein et al.[28] in terms of RNA binding.

Interestingly, mutations in RNA bases located within and in proximity to the cleavage site did not prevent RNase E cleavage entirely but led to secondary processing fragments. In contrast, mutations within the stem led to a complete loss of processing, highlighting the importance of this secondary structure. These observations point to the following two-step mechanism: (1) the repeat needs to be properly recognized; and (2) cleavage occurs 3′ to the stem-loop in the adenylate–uridylate-rich element (Fig. 6). However, when this element was mutated, secondary sites could be accepted for cleavage, albeit at lower efficiency.

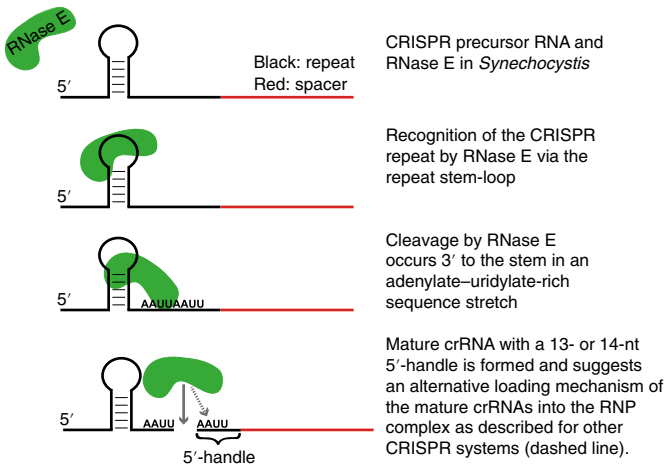


Fig. 6 | Model for RNase-E-mediated crRNA recognition and processing. First, RNase E needs to properly recognize the CRISPR repeat sequence via its stem. In the second step, RNase E cleaves the repeat sequence in the adenylate–uridylate-rich sequence located 3′ to the stem, thus generating the mature crRNA with a 13- or 14-nt 5′-handle. These findings suggest that the loading of mature crRNAs into the CRISPR complex might differ from other CRISPR systems. RNP, ribonucleoprotein.

Our data point directly at *Synechocystis* 6714—a related but not identical strain (2,838 shared and ~800 unique protein-coding genes[44]) that is probably using the same mechanism for maturation of crRNAs from its CRISPR3* system, which is homologous to the here investigated CRISPR3 of *Synechocystis* 6803[38]. To identify further candidates, we searched for CRISPR repeats resembling the CRISPR3 repeat sequence of *Synechocystis* 6803. If the corresponding CRISPR–Cas system lacked a *cas6* gene, we included the respective RNase E sequence in our comparison. This strategy also yielded *Synechocystis* 6714, but—in addition—one system each in four further cyanobacteria (Fig. 5b). Hence, this can be interpreted as evidence that the use of RNase E as the CRISPR maturation endoribonuclease is not restricted to the genus *Synechocystis*.

The diversity of crRNA processing mechanisms and implicated factors is impressive. In type II systems lacking both Cas6 and Cas5d nucleases, the general housekeeping nuclease RNase III cleaves tracrRNA and pre-crRNA in the presence of Cas9[10,45]. Therefore, host-encoded RNase III fulfills the function as an alternative to Cas6 nucleases for a distinct crRNA maturation mechanism[46]. RNase III is involved in diverse biological processes and is evolutionarily conserved[47]. This results in the interchangeability of *Streptococcus pyogenes* RNase III with the enzymes from other selected bacteria, indicating that there is no species specificity for tracrRNA:pre-crRNA cleavage by RNase III[48]. The conservation of tracrRNA:pre-crRNA co-processing by host-encoded RNase III leads to high system flexibility, which allows for functionality of dual RNA:Cas9 systems in multiple species upon horizontal gene transfer[48]. Moreover, in *Listeria monocytogenes*, the endogenous polynucleotide phosphorylase (PNPase) was described as another host-encoded, non-Cas protein responsible for crRNA maturation[49]. A function of PNPase in crRNA maturation—to mediate crRNA 3′-end trimming—was proposed for a subtype III-A system from *Staphylococcus epidermidis* RP62a[50]. In *Synechocystis* 6803, PNPase is associated and closely cooperating with RNase E[37]. Therefore, it can, at present, not be excluded that it is involved in crRNA maturation as well, possibly together with RNase E as a 3′–5′ exoribonuclease following the RNase E-mediated endoribonucleolytic step.

In summary, RNase E acts as an additional host-encoded, non-Cas protein engaged in the pre-crRNA processing aside from the previously reported RNase III[10,45,46] and PNPase[49]. The advantage of using conserved RNase E as a pre-crRNA processing factor might be similar to RNase III, maintaining flexibility of the system through functionality in various species when CRISPR–Cas systems are spread via horizontal gene transfer[48]. However, further CRISPR–Cas systems that rely on RNase E as the crRNA processing factor remain to be described. Together, RNase III, PNPase and RNase E are non-Cas nucleases involved in crRNA processing, which illustrates the complex interplay between host housekeeping functions and the CRISPR–Cas system.

Methods

Interference assay. The self-replicating conjugative vector pVZ322 and the gentamicin resistance cassette were used for the construction of invader plasmids to check interference activities in *Synechocystis* 6803. The gentamicin resistance cassette was polymerase chain reaction (PCR)-amplified (all primers are listed in Supplementary Table 1) using Q5 High-Fidelity DNA Polymerase (NEB), purified from agarose gels via the PCR clean-up and Gel extraction kit (Macherey-Nagel) and subcloned into pUC19 via Gibson assembly to facilitate the introduction of protospacer sequences by inverse PCR. Because of its high expression level compared with other CRISPR3 spacers, the spacer 2 was selected to derive an appropriate protospacer sequence. The protospacer sequence was fused in frame to the gentamicin resistance (reporter) gene just upstream of its stop codon using inverse PCR with Q5 DNA polymerase and appropriate primers. The PCR product was *Dpn*I-treated to remove the undesired template DNA, transformed into chemically competent *E. coli* DH5α cells and selected on gentamicin plates. GoTaq DNA polymerase (Promega) and 3% NuSieve agarose gels (Biozym) were used to screen for positive transformants by colony PCR. After successful fusion of the protospacer to the reporter gene, the gentamicin resistance cassette was

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

PCR amplified using Q5 DNA polymerase and primers overlapping pVZ322 and gel-purified. Overnight digestion of pVZ322 was performed using the restriction enzymes *Xmn*I and *Xba*I (NEB). The linearized backbone was again gel-purified and assembled with the gentamicin resistance cassette with and without (negative control) the protospacer sequence in both the sense and antisense orientation (target plasmids) with respect to the endogenous spacer orientation. Sanger sequencing (GATC Biotech) was used to verify sequence integrity. Mean values of interference efficiency and standard deviations were calculated by the ratio of the number of conjugants of target plasmids to non-target plasmids in biological triplicates and in parallel for the control plasmid. Interference activity was observed if the interference efficiency was <1. The interference efficiency of the non-target plasmid was set to 1, which corresponds to no interference activity.

Self- and non-self discrimination assay. To test how many and which nucleotides within the 5′-handle are important for discrimination between self and non-self targets, the construct maintaining the sense protospacer used in the interference assay was slightly modified. The full-length native CRISPR3 repeat sequence (control plasmid, pNT) and mutated repeat variants (pT) were fused in frame downstream to the protospacer sequence by inverse PCR using Q5 DNA polymerase and the pUC19-based template. After fusion of the repeat variants, the gentamicin resistance cassette including protospacer and repeat sequence was PCR amplified by Q5 polymerase for the re-assembly into *Xmn*I and *Xba*I-digested pVZ322 and sequenced for verification. The experimental procedure and statistical analysis were performed as described for the interference assay.

***Synechocystis* 6803 conjugation.** Plasmids were conjugated into *Synechocystis* 6803 by triparental mating as described[23], with the following variations. Overnight cultures of the helper strain *E. coli* J53/RP4 and the donor strain *E. coli* DH5α with the plasmid of interest were diluted to OD_{600} 0.1 with lysogeny broth medium lacking antibiotics and incubated for 2.5 h at 37 °C with shaking at 180 rpm. For each conjugation, an OD_{600} of 7.0 of the plasmid-bearing and helper cultures was harvested, resuspended in lysogeny broth and combined. The mixed culture was pelleted and resuspended in 100 µl lysogeny broth medium and incubated for 1 h at 30 °C without shaking. In parallel, an OD_{750} of 1.0 of *Synechocystis* 6803 culture was harvested, resuspended in 800 µl BG11 and combined with the mixed culture of the plasmid-bearing and helper culture. The pellet was resuspended in 30 µl BG11 and placed on a sterile filter. After overnight incubation at 30 °C with slightly covered plates, the filter was rinsed with 400 µl BG11 and 30 µl of the resulting cell suspension was plated on BG11 agar plates containing 7.5 µg ml^{-1} gentamicin. Conjugants were counted after further incubation at 30 °C for 2 weeks.

Overexpression and purification of cyanobacterial RNase E from *E. coli*. The *Synechocystis* 6803 RNase E-encoding gene *slr1129* was subjected to codon optimization at the 3′-end via PCR using the appropriate primers (Supplementary Table 1) and Q5 DNA polymerase. The PCR product was subcloned into the pJET1.2/blunt vector (Thermo Fisher Scientific), which includes a TEV protease recognition site. The codon-optimized and TEV site-fused *slr1129* gene was PCR amplified using Q5 DNA polymerase, purified from agarose gel via the PCR clean-up and Gel extraction kit (Macherey-Nagel) and subjected to Gibson assembly with the pQE70 expression vector (Qiagen). The insert and vector for Gibson assembly were joined at a 3:1 molar ratio and transformed into *E. coli* DH5α. In the resulting construct, the RNase E reading frame was prolonged by six additional histidine residues at the C-terminal end (6×His) and verified by Sanger sequencing. Overexpression and purification of *Synechocystis* 6803 recombinant RNase E from *E. coli* M15 (pREP4) was performed as described[51,52]. Briefly, 400 ml of *E. coli* culture was grown to OD_{600} 0.6–0.8 and protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside for 24 h at 22 °C. The cells were collected by centrifugation and the recombinant protein purified via a Ni^{2+}-NTA column (Qiagen).

***Synechocystis* strains, culture media and growth conditions.** *Synechocystis* 6803 cultures were grown as described[38]. For controllable overexpression of RNase E in *Synechocystis* 6803, the *slr1129* sequence was fused to the P_{petJ} promoter from vector psk9 (ref. [53]) and the bacteriophage λ *oop* terminator using primers P4 and P5 (Supplementary Table 1). The PCR fragment was subcloned and finally ligated into the *Sma*I site of vector pVZ321, followed by conjugative transfer into WT cells and selection for chloramphenicol resistance. For induction, 400 ml cultures of *Synechocystis* 6803 were grown to an OD_{750} of 0.6–0.8 in the presence of 1.5 µM Cu^{2+}, centrifuged at 3,500 *g* for 15 min at room temperature and washed twice with Cu^{2+}-free BG11 medium to induce P_{petJ}-driven *slr1129* expression. Cells were resuspended in 400 ml Cu^{2+}-free BG11 and grown for several days. Cultures were diluted daily to OD_{750} 1.0 to ensure continuous growth.

Creation of knock-out mutants and transformation of *Synechocystis* 6803. To analyse whether the CRISPR1 or CRISPR2 *cas* genes are involved in the maturation of CRISPR3-derived crRNA, the entire CRISPR1 and CRISPR2 loci were deleted separately from each other by integrating a kanamycin resistance cassette into the corresponding loci or together by inserting a streptomycin resistance cassette into the C2 locus in the ΔC1 strain. The resulting *Synechocystis* ΔC1, ΔC2 and ΔC1ΔC2 strains are devoid of the respective *cas* genes and the corresponding CRISPR repeat-spacer array(s). The up- and downstream flanking regions (1,200 and 1,800 base pairs in size) were amplified via PCR to ensure site-directed integration via homologous recombination. The resistance cassette in vector pUC19 was placed in between the upstream and downstream flanking regions by Gibson assembly. Transformation of 10 ml *Synechocystis* 6803 aliquots was performed as described[38]. Transformants were tested for full segregation of the introduced knock-out via PCR screening.

Conditional RNase E knock-down. CRISPR interference was used to create a conditional RNase E knock-down mutant. Plasmids encoding a catalytically inactive Cas9 (dCas9) and a sgRNA were a gift from E. P. Hudson[54]. An appropriate sgRNA-targeting *slr1129* was designed using CRISPy-web[55] following the protocol of Larson et al.[56] The selected sgRNA2 sequence targets the region +7 to +27 nucleotides downstream of the *slr1129* start codon and was integrated into the sgRNA template vector via inverse PCR using Q5 polymerase and appropriate primers. Both dCas9 and sgRNA2 were expressed from the anhydrotetracycline-inducible promoter P_{L22} and integrated via homologous recombination into the *psbA1* site with a spectinomycin resistance cassette and the *slr2030–slr2031* site using a kanamycin resistance cassette, respectively. Transformants were screened for segregation and verified by sequencing. Liquid 400 ml cultures were grown to an OD_{750} of 0.6–0.8, split into two parallel subcultures and either induced by feeding with 1 µg ml^{-1} anhydrotetracycline dissolved in 100% ethanol every 24 h or mock-induced by the addition of an equal volume of 100% ethanol. The expression of dCas9 was verified by western blot.

Partial RNase E knock-down. A 3.5 kb fragment bearing *slr1129* and *slr1130* was amplified by PCR using primers RnaseEf and RNaseEr and subcloned into pJET2.1 vector (Thermo Fisher Scientific). An internal *Hind*II fragment (nucleotides 126–1832 within *slr1129*) was replaced by the kanamycin resistance cassette of plasmid pUC4K and the resulting fragment used for transformation of *Synechocystis* 6803. Correct integration of the cassette by homologous recombination was analysed by sequencing, PCR and Southern blot.

RNA analysis and hybridization conditions. RNA extraction from 50 ml *Synechocystis* 6803 cultures was performed as described[38], with the following variations. Centrifugation before RNA precipitation were performed in a swing-out rotor (Beckman Coulter Allegra X-12, 3,500 rpm at room temperature). Cell pellets were resuspended in 1 ml PGTX and incubated for 15 min at 65 °C. RNA pellets were washed with 75% ethanol. Some 10 µg of total RNA per lane were separated on polyacrylamide (PAA)-urea gels (10% PAA, 0.5 g ml^{-1} urea, 1× Tris–borate–EDTA buffer) and electroblotted onto Hybond N+ membranes (GE Healthcare). RNA was cross-linked to the membrane at 120 mJ cm^{-2} for 1.5 min. The RNA sizes were estimated using the RiboRuler Low Range RNA Ladder (Thermo Fisher Scientific). Membranes were prehybridized for 10–60 min at 62 °C with hybridization buffer (50% deionized formamide, 7% SDS, 250 mM NaCl and 120 mM NaPi buffer, pH 7.2) in glass tubes under continuous rotation. Transcript probes for northern hybridization were generated by in vitro transcription from PCR-amplified templates using the Ambion MAXIscript Kit with 50-fold-reduced UTP concentration and 50 µCi ^{32}P-UTP. The reactions were incubated at 37 °C for 10 min, then supplemented with 1 µl DNase (2U µl^{-1}) and incubated for an additional 15 min at 37 °C to remove the DNA templates. Reactions were stopped by the addition of 1 µl 0.5 M EDTA (pH 8.0). Secondary structures were denatured at 80–90 °C for 5 min, kept on ice for 2 min and then added to the prehybridized membranes for hybridization at 62 °C overnight. Membranes were washed at 57 °C with washing solutions I (2× SSC and 1% SDS), II (1× SSC and 0.5% SDS) and III (0.1× SSC and 0.1% SDS) for 10 min each. The signals were detected with a storage phosphor screen (Kodak) and a GE Healthcare Typhoon FLA 9500 imaging system.

Generation of radiolabelled synthetic DNA oligonucleotides. Synthetic deoxyoligonucleotides (Supplementary Table 1) were 5′-end-labelled using 50 µCi (γ^{32}P) ATP (3,000 Ci mmol^{-1}; Hartmann Analytic) and 20 U of T4 polynucleotide kinase (Thermo Fisher Scientific) for 30 min at 37 °C in a volume of 20 µl. Reactions were stopped by the addition of 1 µl 0.5 M EDTA (pH 8.0) and incubation for 10 min at 80 °C. Unincorporated nucleotides were removed by gel filtration through Illustra MicroSpin G-25 Columns (GE Healthcare). Labelled oligonucleotides were stored at −20 °C until use.

Mapping of in vivo 5′-ends by primer extension. Total RNA (10 µg) was treated with 2 U Turbo DNase and 40 U RiboLock RNase Inhibitor (both from Thermo Fisher Scientific) at 37 °C to remove contaminating genomic DNA. After 15 min, another 2 U Turbo DNase were added and the incubation continued. DNase was removed and buffer exchanged using the RNA Clean & Concentrator-25 Kit (Zymo Research), eluting the RNA into 25 µl H_2O. Some 2 µg of DNase-treated RNA and 2 pmol of 5′-end-labelled DNA oligonucleotide were denatured in a 10 µl reaction for 5 min at 80 °C and then chilled for 1 min at room temperature and 5 min on ice. Then, 200 U SuperScript III Reverse Transcriptase (Thermo Fisher Scientific), 40 U RiboLock RNase Inhibitor, 5 mM dithiothreitol and 0.5 mM deoxynucleotides

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

(final concentrations) were added for complementary DNA synthesis in a 20 µl reaction and incubated for 1 h at 50 °C. After heat inactivation for 15 min at 70 °C, the RNA templates were digested with 5 U RNase H (NEB) for 20 min at 37 °C. The samples were mixed with loading buffer and denatured for 5 min at 95 °C and then loaded onto 6% PAA–urea gels (0.5 g ml^{-1} urea and 1 × Tris–borate–EDTA buffer) in a sequencing apparatus. The reference sequence ladder was prepared using a DNA fragment amplified by Q5 DNA polymerase using CRISPR3-specific oligonucleotides (Supplementary Table 1) and subjected to sequencing reactions using the USB Thermo Sequenase Cycle Sequencing Kit (Affymetrix). The sequencing gel was run for approximately 2 h at 65 W and subsequently incubated for 20 min in 10% acetic acid and 5 min in MilliQ H_2O before drying. Signals were detected by a storage phosphor screen (Kodak) on a GE Healthcare Typhoon FLA 9500 imager.

RNase E in vitro cleavage assays with synthetic RNA oligonucleotides. Synthetic RNA oligonucleotides were purchased polyacrylamide gel electrophoresis-purified from PURIMEX and 5′-end phosphorylated yielding the preferred substrate for RNase E cleavage. In total, 50 µM of RNA oligonucleotides were phosphorylated in the presence of 15 U T4 polynucleotide kinase (Thermo Fisher Scientific), 20 mM ATP (NEB) and 40 U RiboLock RNase Inhibitor in a 20 µl volume. The reactions were incubated for 45 min at 30 °C followed by heat inactivation for 10 min at 80 °C. T4 polynucleotide kinase removal and buffer exchange were accomplished with the RNA Clean & Concentrator-5 Kit (Zymo Research), eluting the RNA into 20 µl H_2O. Cleavage reactions were performed in a 10 µl reaction volume in RNase E cleavage buffer containing 25 mM Tris-HCl (pH 8.0), 60 mM KCl, 5 mM $MgCl_2$, 100 mM NH_4Cl and 0.1 mM dithiothreitol for 5 to 30 min at 30 °C and stopped by the addition of 2× RNA loading dye (95% formamide, 0.025% SDS, 0.025% bromophenol blue, 0.025% xylene cyanol FF and 0.5 mM EDTA, pH 8.0). Reactions were denatured for 5 min at 95 °C before loading onto 15% PAA–urea gels in a sequencing apparatus. Gels were stained with SYBR Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific) at a 1:10,000 dilution with 0.5× Tris–borate–EDTA buffer. Signals were detected with a Laser Scanner Typhoon FLA 9500 (GE Healthcare) with the following settings: excitation, 473 nm; emission filter long pass blue ≥ 510 nm; photomultiplier value, 450 or 500.

Protein–RNA cross-linking and enrichment of cross-linked peptide–RNA heteroconjugates. Upon irradiation at 254 nm, a covalent bond forms between the amino acid side chains and nucleic acid bases when both components are in close spatial proximity. Protein–RNA cross-linking was performed using ultraviolet light irradiation at 254 nm followed by the enrichment of cross-linked peptide–RNA heteroconjugates as described[35,36]. Briefly, approximately 1 nmol of the complex comprising recombinant RNase E and CRISPR3 repeat oligonucleotide was filled up to a volume of 100 µl with a buffer containing 50 mM NaH_2PO_4, 300 mM NaCl and 250 mM imidazole (pH 8.0). The complex was incubated at 26 °C for 5 min and irradiated at 254 nm for 10 min as described[35]. The samples were ethanol-precipitated and the pellet was dissolved in 4 M urea and 50 mM Tris-HCl pH 7.9. The final urea concentration was then adjusted to 1 M with 50 mM Tris-HCl pH 7.9, and the RNA was hydrolysed using 1 µg RNase A and T1 (Ambion/Applied Biosystems) for 2 h at 52 °C followed by digestion with benzonase at 37 °C for 1 h. The proteolysis was performed with trypsin (Promega) at 37 °C overnight. The sample was desalted to remove the non-cross-linked RNA fragments using an in-house prepared C18 (Dr. Maisch) column, and the cross-linked peptides were enriched on an in-house prepared TiO_2 (GL Sciences) column using existing protocols[35]. Modifications of the TiO_2 enrichment protocol included the use of 10 µm TiO_2 beads and 5% glycerol (v/v) in Buffer B (80% v/v acetonitrile and 5% v/v trifluoroacetic acid) as Buffer A. Following the enrichment of the peptide–RNA cross-linked heteroconjugates, the samples were dried and resuspended in 5% v/v acetonitrile and 1% v/v formic acid for MS analysis.

Liquid chromatography and MS analysis. The sample was injected into a nano-liquid chromatography system (Dionex UltiMate 3000; Thermo Fisher Scientific) coupled with a Q-Exactive HF instrument (Thermo Fisher Scientific)[35]. Online electrospray ionization MS was performed in data-dependent mode using the Top 20 precursor ions from the High-energy Collisional Dissociation (HCD) method. All precursor ions, as well as fragment ions, were scanned in the Orbitrap and the resulting spectra were measured with high accuracy (<5 ppm) in both the MS and tandem MS level. Data analysis was performed using a dedicated database search tool[36].

Site-directed mutagenesis of RNase E. Deoxyoligonucleotides for the site-directed mutagenesis of RNase E were designed using the NEBaseChanger tool and the Q5 Site-Directed Mutagenesis Kit (both NEB). Briefly, PCRs were performed to obtain plasmids containing the desired mutations and treated with a mix of kinases, ligases and *Dpn*I according to manufacturer's instructions. Some 5 µl of the treated PCR mix were used for transformation of chemically competent TOP10 *E. coli* cells and selected on appropriate antibiotics. Sequences were verified by Sanger sequencing.

Western blot analysis. Proteins were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis on 10% PAA gels and transferred to nitrocellulose membranes (Hybond ECL, GE Healthcare) by electroblotting (PEQLAB) for 1 h at 1.5 mA per 1 cm^2. Membranes were blocked for 1 h at room temperature or overnight at 4 °C. His-tagged recombinant RNase E was detected with the Penta·His HRP Conjugate Kit (Qiagen). For the detection of dCas9, the membrane was blocked with Tris-buffered saline (TBS) supplemented with Tween 20 (TBS-T, 10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Tween 20) containing 5% milk powder. A monoclonal anti-Cas9 antibody (Epigentek/BioCat) diluted in Tris-buffered saline supplemented with Tween 20 (TBS-T) and 0.5% milk powder was applied overnight at 4 °C, followed by a secondary anti-mouse horseradish peroxidase antibody. Membranes were sprayed with the WesternBright ECL Spray (Advansta) and signals were detected using FUSION-CAP (Vilber Lourmat) and Quantity One software (Bio-Rad).

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. The data that support the findings of this paper are contained within the Supplementary Information.

Received: 28 April 2017; Accepted: 20 December 2017;
Published online: 5 February 2018

References

1. Bhaya, D., Davison, M. & Barrangou, R. CRISPR–Cas systems in Bacteria and Archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* **45**, 273–297 (2011).
2. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinforma.* **8**, 172 (2007).
3. Jansen, R., Embden, J. D. A., van Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
4. Lange, S. J., Alkhnbashi, O. S., Rose, D., Will, S. & Backofen, R. CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucleic Acids Res.* **41**, 8034–8044 (2013).
5. Makarova, K. S. et al. An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
6. Hochstrasser, M. L. & Doudna, J. A. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends Biochem. Sci.* **40**, 58–66 (2015).
7. Westra, E. R., Dowling, A. J., Broniewski, J. M. & van Houte, S. Evolution and ecology of CRISPR. *Annu. Rev. Ecol. Evol. Syst.* **47**, 307–331 (2016).
8. Brouns, S. J. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
9. Hale, C., Kleppe, K., Terns, R. M. & Terns, M. P. Prokaryotic silencing (psi) RNAs in *Pyrococcus furiosus*. *RNA* **14**, 2572–2579 (2008).
10. 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).
11. Hale, C. R. et al. RNA-guided RNA cleavage by a CRISPR RNA–Cas protein complex. *Cell* **139**, 945–956 (2009).
12. Przybilski, R. et al. Csy4 is responsible for CRISPR RNA processing in *Pectobacterium atrosepticum*. *RNA Biol.* **8**, 517–528 (2011).
13. Karginov, F. V. & Hannon, G. J. The CRISPR system: small RNA-guided defense in Bacteria and Archaea. *Mol. Cell.* **37**, 7–19 (2010).
14. 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).
15. 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).
16. 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).
17. Staals, R. H. J. et al. Structure and activity of the RNA-targeting type III-B CRISPR–Cas complex of *Thermus thermophilus*. *Mol. Cell.* **52**, 135–145 (2013).
18. Shmakov, S. et al. Discovery and functional characterization of diverse class 2 CRISPR–Cas systems. *Mol. Cell.* **60**, 385–397 (2015).
19 Koonin, E. V. et al. Diversity, classification and evolution of CRISPR–Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
20. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521 (2016).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

21. East-Seletsky, A. et al. Two distinct RNase activities of CRISPR–C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270–273 (2016).
22. Abudayyeh, O. O. et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).
23. Scholz, I., Lange, S. J., Hein, S., Hess, W. R. & Backofen, R. 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 (2013).
24. Reimann, V. et al. Structural constraints and enzymatic promiscuity in the Cas6-dependent generation of crRNAs. *Nucleic Acids Res.* **45**, 915–925 (2017).
25. Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M. & Terns, M. P. Binding and cleavage of CRISPR RNA by Cas6. *RNA* **16**, 2181–2188 (2010).
26. Wang, R., Preamplume, G., Terns, M. P., Terns, R. M. & Li, H. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure* **19**, 257–264 (2011).
27. Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M. & MacMillan, A. M. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat. Struct. Mol. Biol.* **18**, 688–692 (2011).
28. Schein, A., Sheffy-Levin, S., Glaser, F. & Schuster, G. The RNase E/G-type endoribonuclease of higher plants is located in the chloroplast and cleaves RNA similarly to the *E. coli* enzyme. *RNA* **14**, 1057–1068 (2008).
29. Horie, Y. et al. Dark-induced mRNA instability involves RNase E/G-type endoribonuclease cleavage at the AU-box and SD sequences in cyanobacteria. *Mol. Genet. Genom.* **278**, 331–346 (2007).
30. Mohanty, B. K., Petree, J. R. & Kushner, S. R. Endonucleolytic cleavages by RNase E generate the mature 3′ termini of the three proline tRNAs in *Escherichia coli*. *Nucleic Acids Res.* **44**, 6350–6362 (2016).
31. Jiang, X., Diwa, A. & Belasco, J. G. Regions of RNase E important for 5′-end-dependent RNA cleavage and autoregulated synthesis. *J. Bacteriol.* **182**, 2468–2475 (2000).
32. Chao, Y. et al. In vivo cleavage map illuminates the central role of RNase E in coding and non-coding RNA pathways. *Mol. Cell.* **65**, 39–51 (2017).
33. Zhang, J., Graham, S., Tello, A., Liu, H. & White, M. F. Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. *Nucleic Acids Res.* **44**, 1789–1799 (2016).
34. Rott, R., Zipor, G., Portnoy, V., Liveanu, V. & Schuster, G. RNA polyadenylation and degradation in cyanobacteria are similar to the chloroplast but different from *Escherichia coli*. *J. Biol. Chem.* **278**, 15771–15777 (2003).
35. Sharma, K. et al. Analysis of protein–RNA interactions in CRISPR proteins and effector complexes by UV-induced cross-linking and mass spectrometry. *Methods* **89**, 138–148 (2015).
36. Kramer, K. et al. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* **11**, 1064–1070 (2014).
37. Zhang, J.-Y. et al. RNase E forms a complex with polynucleotide phosphorylase in cyanobacteria via a cyanobacterial-specific nonapeptide in the noncatalytic region. *RNA* **20**, 568–579 (2014).
38. Hein, S., Scholz, I., Voß, B. & Hess, W. R. Adaptation and modification of three CRISPR loci in two closely related cyanobacteria. *RNA Biol.* **10**, 852–864 (2013).
39. Loedige, I. et al. The NHL domain of BRAT is an RNA-binding domain that directly contacts the hunchback mRNA for regulation. *Genes. Dev.* **28**, 749–764 (2014).
40. Schultz, A., Nottrott, S., Watkins, N. J. & Lührmann, R. Protein–protein and protein–RNA contacts both contribute to the 15.5K-mediated assembly of the U4/U6 snRNP and the box C/D snoRNPs. *Mol. Cell. Biol.* **26**, 5146–5154 (2006).
41. Tamulaitis, G., Venclovas, Č. & Siksnys, V. Type III CRISPR–Cas immunity: major differences brushed aside. *Trends Microbiol.* **25**, 49–61 (2017).
42. Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č. & Siksnys, V. Spatiotemporal control of type III-A CRISPR–Cas immunity: coupling DNA degradation with the target RNA recognition. *Mol. Cell.* **62**, 295–306 (2016).
43. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in Bacteria and Archaea. *Nat. Rev. Genet.* **11**, 181–190 (2010).
44. Kopf, M. et al. Comparative genome analysis of the closely related *Synechocystis* strains PCC 6714 and PCC 6803. *DNA Res.* **21**, 255–266 (2014).
45. Deltcheva, E. et al. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
46. Karvelis, T. et al. crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol.* **10**, 841–851 (2013).
47. Li, H. Structural principles of CRISPR RNA processing. *Structure* **23**, 13–20 (2015).
48. Fonfara, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR–Cas systems. *Nucleic Acids Res.* **42**, 2577–2590 (2014).
49. Sesto, N. et al. A PNPase dependent CRISPR system in *Listeria*. *PLoS. Genet.* **10**, e1004065 (2014).
50. Walker, F. C., Chou-Zheng, L., Dunkle, J. A. & Hatoum-Aslan, A. Molecular determinants for CRISPR RNA maturation in the Cas10–Csm complex and roles for non-Cas nucleases. *Nucleic Acids Res.* **45**, 2112–2123 (2016).
51. Stazic, D., Lindell, D. & Steglich, C. Antisense RNA protects mRNA from RNase E degradation by RNA–RNA duplex formation during phage infection. *Nucleic Acids Res.* **39**, 4890–4899 (2011).
52. Sakurai, I. et al. Positive regulation of *psbA* gene expression by *cis*-encoded antisense RNAs in *Synechocystis* sp. PCC 6803. *Plant Physiol.* **160**, 1000–1010 (2012).
53. Kuchmina, E., Wallner, T., Kryazhov, S., Zinchenko, V. V. & Wilde, A. An expression system for regulated protein production in *Synechocystis* sp. PCC 6803 and its application for construction of a conditional knockout of the ferrochelatase enzyme. *J. Biotechnol.* **162**, 75–80 (2012).
54. Yao, L., Cengic, I., Anfelt, J. & Hudson, E. P. Multiple gene repression in Cyanobacteria using CRISPRi. *ACS Synth. Biol.* **5**, 207–212 (2016).
55. Blin, K., Pedersen, L. E., Weber, T. & Lee, S. Y. CRISPy-web: an online resource to design sgRNAs for CRISPR applications. *Synth. Syst. Biotechnol.* **1**, 118–121 (2016).
56. Larson, M. H. et al. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).
57. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).
58. Darty, K., Denise, A. & Ponty, Y. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).

Acknowledgements

We thank O. Alkhnbashi and C. Steglich for helpful comments, U. Ruppert for technical assistance, M. Asayama and E. P. Hudson for the provided plasmids. Financial support from the German Research Foundation programme FOR1680 'Unravelling the Prokaryotic Immune System' (grants HE 2544/8-2 and UR225/1-2) to W.R.H. and H.U. is greatly acknowledged.

Author contributions

J.B. and W.R.H. designed the work. The protein–RNA cross-linking experiments and identification of cross-linked peptide–RNA bonds were performed by K.S. and H.U. The RNase E overexpression and partial deletion strains were constructed by A.W. The ΔC1 and ΔC1ΔC2 deletion strains were constructed by V.R. All other experiments were carried out by J.B. J.B., K.S. and W.R.H. analysed the data. J.B. and W.R.H. wrote the paper with contributions from all authors.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-017-0103-5.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to W.R.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

nature research

Corresponding author(s): Wolfgang R. Hess

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Sample size determination was not applicable. Mostly, three biological replicates were used (e.g. interference assay and deletion mutants) or technical triplicates (e.g. in vitro cleavage assays). Please see individual figure captions for details.

2. Data exclusions

Describe any data exclusions.

No data have been excluded from the analyses.

3. Replication

Describe whether the experimental findings were reliably reproduced.

All attempts at replication were successful.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Randomization was not applicable because no animals and/or human research participants were involved. Biochemical and genetic experiments were performed in replicates to verify successful replication.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Blinding was not applicable because of absence of animals and/or human research participants. All experiments were performed without bias and/or expectations.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☒	☐	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☒	☐	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☒	☐	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Quantity One software (BIO-RAD, 4.6.6. Basic), OriginPro 2017 (64-bit, b9.4.0.220), BlastP (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins), BlastN (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch), VARNA (3.9), RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi), Chemiluminescence Imaging - Fusion-Capt Advance (v17.01), OpenMS, OMSSA, Typhoon FLA 9500 (Version 1.1, Build 1.1.0.187)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	All unique materials are available from the authors.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Western Blot analysis: Penta His HRP Conjugate Kit (Qiagen) Cat No./ID: 34460; Statement on manufacturer's website: Applications include Western, dot, and colony blotting procedures, ELISA and other immunoassay procedures, Monoclonal anti-His antibodies for sensitive detection of 6xHis-tagged proteins. QIAGEN News 1997 No. 3, 1. Monoclonal anti-Cas9 antibody (Epigentek/BioCat) Cat No./ID: A-9000-050-EP, Validation in Chew et al. (2016) A multifunctional AAV-CRISPR-Cas9 and its host response. Nat Methods, Fuller et al. (2015) Development of the CRISPR/Cas9 system for targeted gene disruption in Aspergillus fumigatus. Eukaryot Cell, Wang et al. (2015) Adenovirus-mediated somatic genome editing of Pten by CRISPR/Cas9 in mouse liver in spite of Cas9-specific immune responses. Hum Gene Ther, Yao et al. (2016) Multiple gene repression in cyanobacteria using CRISPRi. ACS Synth. Biol.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used within this study.
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used within this study.
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used within this study.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No eukaryotic cell lines were used within this study.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used in this study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	No human participants were involved in this study.