
Article:
Tamulaitis, G., Kazlauskiene, M., Manakova, E. et al. (5 more authors) (2014) Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. Molecular Cell , 56 (4). pp. 506-517. ISSN 1097-2765

https://doi.org/10.1016/j.molcel.2014.09.027

Molecular Cell

Programmable RNA shredding by the Type III-A CRIPSR-Cas system of Streptococcus thermophilus

--Manuscript Draft--

Manuscript Number:	MOLECULAR-CELL-D-14-00842R1
Full Title:	Programmable RNA shredding by the Type III-A CRIPSR-Cas system of Streptococcus thermophilus
Article Type:	Research Article
Keywords:	protein-nucleic acid interactions; RNA interference; CRISPR-Cas; Csm complex; ribonuclease
Corresponding Author:	Virginijus Siksnys Institute of Biotechnology Vilnius, LITHUANIA
First Author:	Gintautas Tamulaitis
Order of Authors:	Gintautas Tamulaitis
	Migle Kazlauskiene
	Elena Manakova
	Česlovas Venclovas
	Alison O. Nwokeoji
	Mark J. Dickman
	Philippe Horvath
	Virginijus Siksnys
Abstract:	Immunity against viruses and plasmids provided by CRISPR-Cas systems relies on a ribonucleoprotein effector complex that triggers the degradation of invasive nucleic acids (NA). Effector complexes of Type I (Cascade) and II (Cas9-dual RNA) target foreign DNA. Intriguingly, the genetic evidence suggests that Type III-A Csm complex targets DNA whereas biochemical data show that III-B Cmr complex cleaves RNA. Here we aimed to investigate NA specificity and mechanism of CRISPR-interference for the Streptococcus thermophilus Csm (III-A) complex (StCsm). When expressed in Escherichia coli, two complexes of different stoichiometry co-purified with 40- and 72-nt crRNA species, respectively. Both complexes targeted RNA and generated multiple cuts at 6-nt-intervals. The Csm3 protein, present in multiple copies in both Csm complexes, acts as endoribonuclease. In the heterologous E.coli host StCsm restricted MS2 RNA phage dependently on Csm3 nuclease. Thus, our results demonstrate that the Type III-A StCsm complex targets RNA and not DNA.
Suggested Reviewers:	Malcolm F. White, prof. University of St Andrews mfw2@st-andrews.ac.uk Expert in the structure and function of CRISPR systems including Type III
	Nadja Heidrich, dr. University of Wűrzburg nadja.heidrich@uni-wuerzburg.de Expert in crRNA transcription and processing (Vogel's lab)
	Scott Bailey, dr. Johns Hopkins University scbailey@jhsph.edu Expert in biochemistry of CRISPR-Cas systems
Opposed Reviewers:	Luciano A. Maraffinni, dr. The Rockefeller University, USA marraffini@rockefeller.edu

	competing interests
	John van der Oost, prof. Wageningen University, The Netherlands john.vanderoost@wur.nl competing interests
	Michael Terns, prof. University of Georgia mterns@bmb.uga.edu competing interests

Cover Letter

Response to reviewers

Programmable RNA shredding by the Type III-A CRIPSR-Cas system of Streptococcus thermophilus

Gintautas Tamulaitis[1], Migle Kazlauskiene[1], Elena Manakova[1], Česlovas Venclovas[2], Alison O. Nwokeoji[3], Mark J. Dickman[3], Philippe Horvath[4] and Virginijus Siksnys[1*]

[1]Department of Protein–DNA Interactions and [2]Department of Bioinformatics, Institute of Biotechnology, Vilnius University, Vilnius, Lithuania,

[3]Department of Chemical and Biological Engineering, ChELSI Institute, University of Sheffield, Sheffield, UK,

[4]DuPont Nutrition and Health, Dangé-Saint-Romain, France

[*]Corresponding author

Contact:

Tel: +370-5-2602108; Fax: +370-5-2602116; e-mail: siksnys@ibt.lt;

Keywords:

protein-nucleic acid interactions; RNA interference; CRISPR-Cas; Csm complex; ribonuclease

Running title:

Csm complex of Type III-A CRISPR-Cas system

Characters (including spaces): 54958

Figures: 7

Summary

Immunity against viruses and plasmids provided by CRISPR-Cas systems relies on a ribonucleoprotein effector complex that triggers the degradation of invasive nucleic acids (NA). Effector complexes of Type I (Cascade) and II (Cas9-dual RNA) target foreign DNA. Intriguingly, the genetic evidence suggests that Type III-A Csm complex targets DNA whereas biochemical data show that III-B Cmr complex cleaves RNA. Here we aimed to investigate NA specificity and mechanism of CRISPR-interference for the Streptococcus thermophilus Csm (III-A) complex (StCsm). When expressed in Escherichia coli, two complexes of different stoichiometry co-purified with 40- and 72-nt crRNA species, respectively. Both complexes targeted RNA and generated multiple cuts at 6-nt-intervals. The Csm3 protein, present in multiple copies in both Csm complexes, acts as endoribonuclease. In the heterologous E.coli host StCsm restricted MS2 RNA phage dependently on Csm3 nuclease. Thus, our results demonstrate that the Type III-A StCsm complex targets RNA and not DNA.

Highlights

- ✓ Streptococcus thermophilus Type III-A Csm (StCsm) complex targets RNA
- ✓ Multiple cuts are introduced in the target RNA at 6-nt-intervals
- ✓ Csm3 protein subunits are responsible for an endoribonuclease activity of the complex
- ✓ StCsm complex offers a novel programmable tool for RNA-degradation or modification

Introduction

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) together with Cas (CRISPR-associated) proteins provide RNA-mediated adaptive immunity against viruses and plasmids in bacteria and archaea (Terns and Terns, 2014). Immunity is acquired through the integration of invader-derived nucleic acid (NA) sequences as 'spacers' into the CRISPR locus of the host. CRISPR arrays are further transcribed and processed into small interfering CRISPR RNAs (crRNAs) that together with Cas proteins assemble into a ribonucleoprotein (RNP) complex which, using crRNA as a guide, locates and degrades the target NA. CRISPR-Cas systems have been categorized into three major Types (I-III) that differ by the structural organization of RNPs and NA specificity (Makarova et al., 2011b).

Type I and II systems provide immunity against invading DNA. In Type I systems crRNAs are incorporated into a multisubunit RNP complex called Cascade (CRISPR-associated complex for antiviral defense) that binds to the matching invasive DNA and triggers degradation by the Cas3 nuclease (Brouns et al., 2008; Sinkunas et al., 2013; Westra et al., 2012). In Type II systems, CRISPR-mediated immunity solely relies on the Cas9 protein. It binds a dual RNA into the RNP effector complex, which then specifically cuts the matching target DNA, introducing a double strand break (Gasiunas et al., 2012; Jinek et al., 2012). In Type I and II CRISPR-Cas systems, the target site binding and cleavage requires a short nucleotide sequence (protospacer-adjacent motif, or PAM) in the vicinity of the target (Mojica et al., 2009). Target DNA strand separation, necessary for the crRNA binding, is initiated at PAM and propagates in a directional manner through the protospacer sequence to yield the R-loop intermediate, one strand of which is engaged into the heteroduplex with crRNA, while the other strand is displaced into solution (Sternberg et al., 2014; Szczelkun et al., 2014). Thus, despite differences in their architecture, Type I and II RNP complexes share three major features: i) they act on the invasive double-stranded DNA (dsDNA), e.g., viral DNA or plasmids, ii) they require the presence

of a PAM sequence in the vicinity of the target site, and iii) they generate an R-loop as a reaction intermediate.

Intriguingly, Type III CRISPR-Cas systems are believed to target either DNA (Type III-A) or RNA (Type III-B) (Makarova et al., 2011b). In the III-B systems Cas RAMP proteins (Cmr) and crRNA assemble into a multisubunit RNP complex. Using crRNA as a guide, this complex in vitro binds single-stranded RNA (ssRNA) in a PAM-independent manner and triggers the degradation of target RNA (Hale et al., 2009; Staals et al., 2013; Zhang et al., 2012). The Cmr-effector complex is comprised of six Cmr proteins (Cmr1, Cas10, Cmr3-6) that are important for efficient target RNA cleavage, however roles of the individual Cmr proteins and the ribonuclease (RNase) component have yet to be identified. Cmr1, Cmr3, Cmr4 and Cmr6 are predicted RNA-binding proteins that share a ferredoxin-like fold and resemble RNA-recognition motifs (RRMs) identified in RNA-binding proteins (Terns and Terns, 2014).

The Cas genes encoding the RNA-targeting III-B (Cmr) and DNA-targeting Type III-A (Csm) effector complexes share a partial synteny (Makarova et al., 2011a). In Staphylococcus epidermidis the DNA-targeting Csm complex (SeCsm) is comprised of Cas10, Csm2, Csm3, Csm4, and Csm5 proteins, however the function of individual Csm proteins is unknown. The evidence that Type III-A systems target DNA remains indirect and relies on the experimental observation that Type III-A RNP complex from Staphylococcus epidermidis (SeCsm) limits plasmid conjugation and transformation in vivo, but the DNA degradation has not been demonstrated directly (Marraffini and Sontheimer, 2008, 2010). The Csm complex from the archaeon Sulfolobus solfataricus (SsCsm) binds dsDNA, however, it shows no crRNA-dependent nuclease activity in vitro (Rouillon et al., 2013). Thus, while the RNA cleavage activity of the Cmr complex has been characterised in vitro, the DNA degradation activity of the Type III-A Csm complex has yet to be demonstrated. The Csm complex so far remains the only CRISPR-Cas effector complex, for which the function is not yet reconstituted in vitro. Here we aimed to establish the

composition and mechanism of the Csm complex for Type III-A system Streptococcus thermophilus (St).

Results

Type III-A CRISPR-Cas loci in S. thermophilus

S. thermophilus strain DGCC8004 that was selected for further experimental characterization carries 13 spacers in its Type III-A CRISPR2 array (Figure 1A and S1). This strain also contains a Type II CRISPR1 system that is present in other S. thermophilus strains (unpublished data). In the CRISPR2 locus of DGCC8004 the 36-nt repeat sequences, that are partially palindromic, are conserved with the exception of the two terminal repeats (Figure 1A). An A+T rich 100-bp leader sequence is located upstream of the CRISPR2 array.

DGCC8004 CRISPR2 (Type III-A) spacers range in size between 34 and 43 nt, but 36-nt spacers are the most abundant. In total, 38 unique spacers were identified among CRISPR2-positive S. thermophilus strains and a majority (20 out of 38) of these spacer sequences have matches (protospacers) in S. thermophilus DNA phage sequences, although phage interference for the S. thermophilus CRISPR2 locus has not yet been demonstrated. Analysis of the sequences located immediately upstream and downstream of these protospacers failed to identify any consensus sequence (data not shown) as a putative PAM, either due to the relatively small number of protospacers or RNA targeting that is often PAM-independent (Hale et al., 2009). In DGCC8004, although no CRISPR2 spacer gives perfect identity with currently known sequences, 6 spacers out of 13 (S3, S4, S6, S8, S12 and S13) show strong sequence similarity with S. thermophilus DNA phages (at least 94% identity over at least 80% of spacer length). Interestingly, all phage matching protospacers were selected from the template strand. For instance, the 36-nt spacer S3 matches 34 nt of a protospacer in the S. thermophilus phage O1205 genome (Figure 1B). A corresponding crRNA would match the template DNA strand of

the protospacer S3, and will pair with the target sequence on the coding strand of phage DNA or the respective mRNA sequence. If crRNA processing in the S. thermophilus Type III-A locus is similar to that in S. epidermidis (Hatoum-Aslan et al., 2011; Hatoum-Aslan et al., 2014; Hatoum-Aslan et al., 2013), the resulting crRNA 5'-handle in the mature crRNA will be non-complementary to the protospacer S3 3'-flank in the phage DNA coding strand or mRNA (Figure 1B). In the S. epidermidis Type III-A system, which limits the spreading of plasmid DNA, the crRNA/target DNA non-complementarity outside of the spacer sequence plays a key role in invading DNA silencing and self vs non-self DNA discrimination (Marraffini and Sontheimer, 2010). Taking these elements into consideration, crRNA encoded by the spacer S3 was selected as the guide, and a complementary protospacer sequence as the NA target (DNA or RNA) (Figure 1B).

Cloning, expression and isolation of the S. thermophilus DGCC8004 Type III-A effector complex

In order to isolate the Type III-A RNP effector complex (StCsm) of the DGCC8004 we split CRISPR2 locus into the three fragments and cloned them into three compatible vectors (Figure 1C). Plasmid pCas/Csm contained a cassette including all the cas/csm genes (except cas1 and cas2), while plasmid pCRISPR_S3 carried 4 identical tandem copies of the repeat-spacer S3 unit flanked by the leader sequence and the terminal repeat. Plasmids pCsm2-Tag or pCsm3-Tag carried a StrepII-tagged variant of csm2 or csm3 genes, respectively. Next, all three plasmids were co-expressed in E. coli BL21 (DE3) and tagged Csm2 or Csm3 proteins were isolated by subsequent Strep-chelating affinity and size exclusion chromatography.

Strep-tagged Csm2 or Csm3 proteins pull-downed from E. coli lysates co-purified with other Csm/Cas proteins suggesting the presence of a Csm complex (Figure 1D). Csm complexes isolated via N-terminus Strep-tagged Csm2 (Csm2_StrepN) and the N-terminus Strep-tagged Csm3 proteins (Csm3_StrepN) were subjected to further characterization. SDS-PAGE of these complexes revealed six bands that matched the individual Cas proteins Cas6, Cas10, Csm2, Csm3, Csm4 and Csm5

(Figure 1D). The identity of proteins in these Csm complexes was confirmed by mass spectrometry (MS) analysis (Tables S1 and S2).

We next examined the Csm complexes for the presence of NA using basic phenol-chloroform extraction followed by RNase I or DNase I digestion. Denaturing PAGE analysis revealed that ~70-nt and ~40-nt RNA molecules co-purified with the Csm3_StrepN and Csm2_StrepN pulled-down Csm complexes, respectively (Figure 1E). The complex isolated via Csm2_StrepN subunit also contained ~10% of the ~70-nt RNA. When subjected to RNase I protection assay the RNA in the complexes showed no visible degradation, indicating that the RNA is tightly bound and protected along its entire length (data not shown).

Characterization of the crRNA

We used denaturing RNA chromatography in conjunction with electrospray ionization mass spectrometry (ESI-MS) to analyse the crRNA sequence and determine the chemical nature of the 5'- and 3'-termini of crRNAs co-purified with both Csm complexes. Denaturing ion pair reverse phase chromatography was used to rapidly purify the crRNA directly from the Csm complexes. The crRNA isolated from the Csm3_StrepN pull-down complex revealed a single crRNA with a retention time consistent with an approximate length of 70 nt (Figure 1F). The crRNA isolated from Csm2_StrepN pull-down complex revealed the presence of an additional crRNA, with a retention time consistent with an approximate length of 40 nt (Figure 1G). Purified crRNAs were further analyzed using ESI-MS to obtain the accurate intact masses. A molecular weight of 22 998.5 Da was obtained for RNA isolated from Csm3 and 12 602.2 Da for RNA isolated from Csm2 pull-downs, respectively. Csm2 pull-down also contained a minor component, with a molecular weight of 12 907.3 Da (data not shown). In addition, ESI MS/MS was also used to analyse the oligoribonucleotide fragments generated from RNase A/T1 digestion of the crRNAs (Figure S2). In conjunction with the intact mass analysis, these results revealed a 72-nt crRNA in the complex isolated via Csm3 (further termed Csm-72 according to

the length of crRNA) and a 40-nt crRNA in the complex isolated via Csm2 (further termed Csm-40 complex). The MS analysis of the 72-nt crRNA is consistent with the pre-CRISPR cleavage at the base of the CRISPR RNA hairpin to yield a 8-nt 5'-handle, a 36-nt spacer and a 28-nt 3'-handle with 5'-OH and 3'-P, and could represent unmaturated crRNA intermediate (Figure 1F) similar to that of Type III-A and III-B CRISPR-Cas systems (Hale et al., 2009; Hatoum-Aslan et al., 2013). Further verification of the 3'-P termini was obtained upon acid treatment of the 72-nt crRNA where no change in mass was observed using ESI-MS. Likewise, the MS analysis of the 40-nt crRNA in the Csm-40 complex revealed an 8-nt 5'-handle and a 32-nt spacer with 5'-OH and 3'-OH that would correspond to the matured crRNA (Figure 1G). The difference in the chemical nature of the 3′-end between intermediate and mature crRNAs suggests that primary processing and final maturation are achieved by distinct catalytic mechanisms as proposed by Hatoum-Aslan for the S. epidermidis model system (Hatoum-Aslan et al., 2011).

Composition and shape of the Csm complex

Evaluation of the complex composition by densitometric analysis of the SDS gels suggests the $Cas10_1$:$Csm2_6$:$Csm3_{10}$:$Csm4_1$:$Csm5_{0.14}$ stoichiometry for Csm-72, and the $Cas6_{0.10}$:$Cas10_1$:$Csm2_3$:$Csm3_5$:$Csm4_1$:$Csm5_1$ stoichiometry for Csm-40. Fraction numbers for Cas6 and Csm5 proteins are presumably due to the weak transient interactions of these proteins in the respective complexes. Protein subunits that are involved in pre-crRNA processing, e.g. Cas6, would not necessarily occur in stoichiometric amounts in the purified effector complex.

We also performed the small angle X-ray scattering (SAXS) measurements in order to characterize the molecular mass/shape of both Csm-40 and Csm-72 effector complexes in solution. M_w values obtained using SAXS are in agreement both with DLS and gel-filtration data (Table S3). Taken together these data are consistent with the stoichiometry $Cas10_1$:$Csm2_6$:$Csm3_{10}$:$Csm4_1$:$crRNA_1$ (calculated Mw 486.2

kDa including the 72-nt crRNA) for Csm-72 and $Cas10_1$:$Csm2_3$:$Csm3_5$:$Csm4_1$:$Csm5_1$:$crRNA_1$ (calculated Mw 344.8 kDa including 40-nt crRNA) for Csm-40.

SAXS measurements revealed that the Csm-40 complex in solution has elongated and slightly twisted shape. The maximal interatomic distance (D_{max}) of the complex estimated from SAXS data is 215 Å, whereas its diameter is 75-80 Å (Table S4). The shape of this effector complex (Figure 1H) is very similar to the electron microscope structure of Cmr complexes from Thermus thermophilus (Staals et al., 2013), Pyrococcus furiosus (Spilman et al., 2013) and Cascade from E. coli (Wiedenheft et al., 2011) (Figure S3E). The Csm-72 complex with D_{max} of 280 Å (Table S4) is significantly more elongated than the Csm-40 complex (Figure 1H). The lowest normalized spatial discrepancy was obtained for the end-to-end superimposition of the Csm-40 and Csm-72 models (Figure 1H).

Nucleic acid specificity of the Type III-A StCsm complex

In the CRISPR2 locus of DGCC8004, 34 out of 36 nt of the spacer S3 match a sequence present in the genome of S. thermophilus phage O1205. Therefore, to probe the functional activity of the Csm-40 complex we first designed DNA and RNA substrates that are fully complementary to the 32-nt crRNA encoded by spacer S3 and that carry phage O1205-flanking sequence. These flanking sequences lack complementarity to the 8-nt 5'-handle of the crRNA identified in the Csm-40/Csm-72 complexes (Figure 2A and Table S5). For binding analysis DNA or RNA substrates were 5'-end radioactively labeled and the Csm-40 complex binding was evaluated by an electrophoretic mobility shift assay (EMSA) in the absence of any divalent metal (Me^{2+}) ions. Csm-40 showed weak affinity for oligoduplex S3/1 DNA/DNA and DNA/RNA substrates since binding was observed only at high (100-300 nM) complex concentrations. Single-stranded S3/1 DNA (ssDNA) was bound to Csm-40 with an intermediate affinity ($K_d \approx 30$ nM), whereas a single-stranded S3/1 RNA (ssRNA) showed high affinity binding ($K_d \approx 0.3$ nM) (Figure 2B). Binding competition experiments with various nucleic acids

supported the single-stranded RNA specificity for the Csm-40 complex (Figure 2C). Cleavage data correlated with the binding affinity: S3/1 DNA/DNA, DNA/RNA and ssDNA are refractory to cleavage, whereas S3/1 ssRNA complementary to the crRNA is cut by Csm-40 in the presence of Mg^{2+} ions (Figure 2D). RNase activity of Csm-40 complex requires Mg^{2+} or other divalent metal ions (Mn^{2+}, Ca^{2+}, Zn^{2+}, Ni^{2+} or Cu^{2+}) and is inhibited by EDTA (Figure S4E).

Surprisingly, Csm-40 cuts the S3/1 RNA target at 5 sites regularly spaced by 6-nt intervals to produce 48-, 42-, 36-, 30- and 24-nt products, respectively (Figures 2D, 2E). The sequence complementarity between the crRNA in the complex and the RNA target is a key pre-requisite for the cleavage: a non-specific RNA (Figure 2E, bottom) was resistant to Csm-40. The Csm-40 cleavage pattern of the 3'-labeled S3/1 RNA substrate differs from that of the 5'-labeled variant. While the 5'-labeled substrate cleavage produces 48-, 42-, 36-, 30- and 24-nt products, short degradation products of 21, 27, and 33 nt (1 nt shift is due to an additional nucleotide added during the 3'-labeling) are visible on the gel (Figures 2D, 2E). Taken together, cleavage data for the 5'- and 3'-end labeled RNA substrates suggest that Csm-40 cuts the RNA molecule initially at its 3'-end and endonucleolytic degradation is further extended towards the 5'-end with 6-nt increments.

The Csm-72 complex carrying a 72-nt crRNA (8-nt 5'-handle plus 36 nt of the spacer S3 and 28 nt of the 3'-handle, Figure S4A) showed ~ 30-fold weaker binding affinity ($K_d \approx 10$ nM) to S3/1 RNA in comparison to the Csm-40 (Figure S4B). Nevertheless, similarly to the Csm-40 complex, in the presence of Mg^{2+} ions Csm-72 cleaved S3/1 RNA, albeit at a decreased rate which may correlate with its weaker binding affinity (Figure S4C). The 5'- and 3'-labeled S3/1 RNA cleavage pattern is identical to that of Csm-40 (Figures S4C, S4D and data not shown). Like the Csm-40 complex, Csm-72 showed no cleavage of S3/1 ssDNA, DNA/DNA or DNA/RNA substrates (data not shown). The heterogeneous Csm complex isolated from the E.coli host carrying the wt CRISPR array containing 13 spacers produces RNA cleavage products identical to those of the homogenous StCsm (Figure S4F). Taken

together, these data unambiguously demonstrate that Csm-40 and Csm-72 complexes in vitro target RNA but not DNA, and cut RNA at multiple sites regularly spaced by 6-nt intervals.

Reprogramming of the StCsm complex

To demonstrate that the Type III-A StCsm complex can be reprogrammed to cut a desired RNA sequence in vitro we designed and isolated Csm complexes loaded with crRNA(+Tc) and crRNA(-Tc) targeting, respectively, the 68 nt sense(+) and anti-sense(-) mRNA fragments obtained by in vitro transcription of the tetracycline (Tc) resistance protein gene in the pBR322 plasmid (nt 851-886) (Figure S5A and Table S5). Both Csm-40 and Csm-72 complexes guided by the crRNA(+Tc) sliced the complementary sense RNA fragment but not the antisense RNA sequence (Figure S5B). On the other hand, Csm-40 and Csm-72 complexes guided by the crRNA(-Tc) cleaved antisense RNA but not a sense Tc mRNA fragment (Figure S5B). In both cases target RNA was cleaved at multiple sites regularly spaced by 6-nt intervals (Figure S5C).

Target RNA determinants for cleavage by crRNA-guided Csm complex

We examined further whether the nucleotide context downstream or upstream of the protospacer sequence modulates RNA cleavage by the Csm complexes. To this end, we designed the S3/2 RNA substrate in which the flanking regions originating from O1205 phage DNA in the S3/1 substrate are replaced by different nucleotide stretches that are non-complementary to the 5'-handle of crRNA in the Csm-40 and Csm-72 complexes, and to the 3'-handle in the Csm-72 complex. RNA binding and cleavage data showed that despite differences in the nucleotide context of flanking sequences in the S3/1 and S3/2 substrates, cleavage patterns for the Csm-40 and Csm-72 complexes are nearly identical, except for an extra 18-nt product for the Csm-72 (Figure S4C).

Next, we addressed the question whether the base-pairing between the flanking sequences of the RNA target and 5'- and 3- handles of crRNA in the Csm-40 and Csm-72 complexes affect either the cleavage

efficiency or pattern. We designed S3/3, S3/4, and S3/5 RNA substrates that contain flanking sequences complementary to the 5'-handle (40- or 72-nt crRNA), 3'-handle (72-nt crRNA) or both 5'- and 3'-handles in 72 nt-crRNA, respectively (Figure S5A and Table S5). The cleavage analysis revealed that base-pairing between the 8-nt 5'-handle of crRNA and the 3'-flanking sequence had no effect on the cleavage pattern of the Csm-40 and Csm-72 complexes. Indeed, the S3/3 substrate is cleaved with the same 6-nt step by Csm-40, suggesting that the non-complementarity of the flanking sequences is not a necessary pre-requisite for cleavage by the Csm complex (Figure S5). Surprisingly, for the Csm-72 complex extension of the base-pairing between the 3'-handle of the 72-nt crRNA and the protospacer 5'-flanking sequence in S3/5 RNA substrate results in target RNA cleavage outside the protospacer yielding 12- and 6-nt cleavage products (Figure 3). Moreover, the S3/6 substrate, which has extended complementarity between crRNA 3'-handle and 5'-flanking sequence was cleaved at multiple positions along the full length of RNA duplex, except for the region complementary to the crRNA 5'-handle (Figure 3). The cleavage at 18 and 12 nt outside the protospacer was also detected for the Csm-40 complex on S3/4 and S3/5 RNA substrates (Figure S6). The 40-nt crRNA present in the Csm-40 complex lacks the 3'-handle and therefore cannot form RNA duplex with the 5'-flanking sequence in the S3/5 and S3/6 RNA substrates. However, the Csm-40 complex preparation still contains ~ 10% of unmatured 72-nt crRNA, and this heterogeneity results in the extra cleavage outside the protospacer (Figure S6C).

To interrogate the importance of base-pairing within the protospacer region for target RNA cleavage we designed a set of RNA substrates harboring two adjacent nucleotide mutations in the spacer region (substrates S3/7, S3/8 and S3/9, see Figure S6A and Table S5). Two nucleotide mismatches in these substrates did not compromise RNA cleavage by the Csm-40 (Figure S6) and Csm-72 complexes (data not shown), suggesting that the StCsm complex tolerates at least 2 contiguous mismatches in the protospacer region homologous to the crRNA.

To explore whether 3'- or 5'-ends of the target RNA are important for cleavage by the Csm-40 complex, we designed a set of truncated RNA substrates. In S3/10, S3/14 and S3/12 RNA substrates unpaired flaps at the 3'-, 5'- or both ends of the target RNA were truncated, while in S3/11 and S11/13 substrates the truncations extend into the region complementary to crRNA (Figure 4A). Binding affinity for the most of truncated substrates was not compromised (Figure 4B) and target RNA cleavage occurred at multiple sites spaced by 6-nt intervals at conserved protospacer positions (Figure 4C). Truncations extending into the protospacer region (S3/11 and S3/13) showed decreased binding and reduced cleavage rates. This could be a result of the decreased duplex stability, however the role of the "seed" sequence cannot be excluded. Importantly, for all RNA substrates the cleavage sites were located at a fixed distance with respect to the conserved 5'-handle of crRNA (Figure 4C).

Identification of the ribonuclease subunit in the StCsm complex

Regularly spaced cleavage pattern of the RNA target (Figures 2-4, S4, S6) implies the presence of multiple cleavage modules in the Csm complex. According to the densitometric analysis, 3 Csm2 and 5 Csm3 subunits are identified in the Csm-40 complex, while 6 Csm2 and 10 Csm3 subunits are present in the Csm-72 complex. Multiple copies of the Csm2 and Csm3 proteins in the Csm complexes make them prime candidates for catalytic subunits. StCsm2 is a small (121 aa) α-helical protein of unknown structure. StCsm3 (220 aa) contain a conserved (RRM core and is fairly closely related (~35% sequence identity) to Methanopyrus kandleri Csm3, whose crystal structure has been solved recently (Hrle et al., 2013). We reasoned that, since the catalytic activity of the StCsm complex requires the presence of Me^{2+} ions, the active site is likely to contain one or more acidic residues. We inspected multiple sequence alignments of both Csm2 and Csm3 protein families for conserved aspartic or glutamic residues. We did not find any promising candidates in StCsm2, but identified several, including D33, D100, E119, E123 and E139 in StCsm3 (Figure 5A). To probe the role of these conserved negatively charged Csm3 residues we constructed mutants having one residue at a time

replaced with alanine. We also constructed the H19A mutant, since it was shown that the corresponding mutation (R21A) in M. kandleri Csm3 abolished binding of single-stranded RNA (Hrle et al., 2013). We expressed each mutant in the context of other StCsm/Cas proteins and analysed the cleavage activity of the StCsm-40 complex containing mutant Csm3 subunits. StCsm3 H19A, D100A, E119A, E123A and E139A mutants did not compromise RNA binding, RNP complex composition or cleavage activity of Csm-40 complex (Figure 5B-5F). However, the D33A mutant impaired Csm-40 RNA cleavage (Figure 5D, 5F) without affecting RNA binding (Figure 5D). Taken together these data demonstrate that Csm3 is a RNase, which cuts RNA producing multiple cleavage patterns spaced by regular 6-nt intervals, and that the D33 residue is part of the catalytic/metal-chelating site. StCsm3 structural model based on the homologous structure of M. kandleri Csm3 is in good agreement with the identified role for this residue (Figures S7A). D33 belongs to the highly conserved surface patch that extends from the RRM core into the "lid" subdomain (Figure S7B). Part of this surface patch is positively charged, supporting the idea that it represents an RNA-binding site (Figure S7C).

In vivo RNA targeting by the StCsm complex

To demonstrate that the StCsm complex could target RNA in vivo, we employed the MS2 phage restriction assay. MS2 is a lytic single-stranded RNA coliphage which infects E. coli host via fertility (F) pilus. MS2 phage is a favorable model to investigate RNA targeting by the CRISPR-Cas system in vivo as no DNA intermediates are formed during the life cycle of this phage (Olsthoorn and van Duin, 2011). To engineer the MS2 phage targeting heterologous host, E.coli NovaBlue (DE3, F^+) strain has been transformed by two compatible plasmids: i) pCRISPR_MS2 plasmid bearing the synthetic CRISPR array of five repeats interspaced by four 36-nt spacers targeting the mat, lys, cp, and rep MS2 RNA sequences, and ii) pCsm/Cas plasmid for expression of Cas/Csm proteins (Figure 6). The phage targeting and control E. coli strains were plated and infected with series of dilutions of MS2 using drop plaque assay. The assay revealed that E. coli strain expressing wt Csm and MS2 targeting crRNAs

restricts the phage at least 1000-fold in respect to the control cells. Importantly, no MS2 phage resistance has been observed in the strain expressing the non-targeting crRNA or the cleavage deficient Csm3 variant. Taken together these data demonstrate that the StCsm complex in vivo convey resistance to RNA phage in the heterologous E. coli host.

Discussion

Here we aimed to establish the NA specificity and mechanism for the Type III-A CRISPR-Cas system of Streptococcus thermophilus. In sharp contrast to other CRISPR-Cas subtypes, the functional activity of Type III-A system so far has not been reconstituted in vitro. Cas/Csm proteins in the Type III-A CRISPR locus of the S. thermophilus DGCC8004 are homologous to those of in S. thermophilus DGCC7710 and LMD-9, and show more distant but significant similarities to Cas/Csm proteins in L. lactis, E. italicus and S. epidermidis (Marraffini and Sontheimer, 2008; Millen et al., 2012) (Figure S1).

Csm complexes of S. thermophilus

We have expressed the Type III-A CRISPR-Cas locus of the DGCC8004 in E. coli and isolated two RNP complexes termed Csm-40 and Csm-72. Both complexes share a conserved set of Cas10, Csm2, Csm3 and Csm4 proteins. In addition to this core, the Csm-40 also contains the Csm5 protein. Two distinct crRNAs of 72- and 40-nt co-purify with Csm-40 and Csm-72 complexes isolated from the heterologous E. coli host. The 72-nt crRNA comprised of an 8-nt 5'-handle, a 36-nt spacer and a 28-nt 3'-handle would result from the pre-crRNA cleavage between 28 and 29 nt within the conserved repeat region presumably by the Cas6 nuclease similarly to the III-B CRISPR-Cas system (Carte et al., 2008). The shorter 40-nt crRNA co-purified with the Csm-40 complex of S. thermophilus contains the conserved 8-nt 5'-handle and 32-nt spacer indicating that the 72-nt crRNA intermediate underwent further 3'-end processing to produce a mature 40-nt crRNA that lacks the 3'-handle and 4 nt within the

spacer region (Figure 7). The RNAse involved in the 72 nt crRNA intermediate maturation still has to be identified, however the Csm5 protein which is absent in Csm-72 but is present in Csm-40 could be a possible candidate. Indeed, csm5 gene deletion in DGCC8004 produces only unmatured Csm-72 complexes (G.T., M.K., Irmantas Mogila, unpublished data).

The crRNA processing and maturation pathway in the S. thermophilus Type III-A system (Figure 7) shows striking similarity to that in S. epidermidis. First, the SeCsm complex includes the same set of Cas10, Csm2, Csm3, Csm4 and Csm5 proteins as the StCsm-40. Furthermore, in S. epidermidis, the primary processing by Cas6 produces a 71-nt crRNA intermediate, that is subjected to further endonucleolytic processing at the 3′ end (Hatoum-Aslan et al., 2011; Hatoum-Aslan et al., 2014).

StCsm complex cuts RNA producing a regular cleavage pattern

The Csm complexes of S. epidermidis and S. solfataricus have been reconstituted and isolated, however NA cleavage activity has not been reported so far. In vivo studies in S. epidermidis suggested that the Type III-A SeCsm RNP complex targets DNA (Marraffini and Sontheimer, 2008) in a PAM-independent manner and prevents autoimmunity by checking the complementarity between the crRNA 5′-handle and the 3′-flanking sequence in the vicinity of the protospacer (Marraffini and Sontheimer, 2010). In contrast to these data we found that the StCsm-40 and StCsm-72 complexes bind ssRNA with high affinity and cut a ssRNA target in a PAM-independent manner in the presence of Me^{2+} ions, producing a regular 6-nt cleavage pattern in the protospacer region (Figures 2D and S4C-S4E). In this respect the Type III-A StCsm complex resembles the RNA-targeting Type III-B Cmr-complexes PfCmr, SsCmr and TtCmr (Hale et al., 2009; Staals et al., 2013; Zhang et al., 2012) (Figure 7) rather than DNA targeting Type I and II complexes. By targeting RNA rather than DNA StCsm complex avoids autoimmunity. We further show that the nucleotide context and non-complementarity outside the protospacer have no effect on the target RNA cleavage, demonstrating that PAM or unpaired flanking sequences of the protospacer are not required for cleavage by the StCsm (Figure S6). The complementarity of the protospacer is the only pre-requisite for the StCsm cleavage: non-matching

RNA is not cleaved, however in the complimentary protospacer S3 two contiguous mismatches and end truncations are tolerated by StCsm (Figure S6). The differences in the cleavage patterns of the 5'- and 3'-labeled RNAs (Figure 2D) imply that cleavage first occurred at 3'-end of the target RNA. However, it remains to be established whether the observed cleavage directionality is dictated by the "seed" sequence/directionality of the crRNA:target RNA duplex formation or nucleotide context-dependent cleavage rate differences.

Strikingly, we found that for the Csm-72 complex the target RNA is being cleaved at regular 6-nt intervals outside the protospacer if it retains base complementarity to the crRNA 3'-handle. Such regularly spaced cleavage pattern of the RNA target (Figures 2-4, S4 and S6) implies the presence of multiple cleavage modules in the Csm complex. The major difference between the Csm-40 and Csm-72 complexes is the number of Csm2 and Csm3 subunits. The Csm-40 contains 3 Csm2 and 5 Csm3 subunits while Csm-72 contains 6 Csm2 and 10 Csm3 subunits (Figure 1D). The size of the complexes determined by SAXS correlates with the different stoichiometry of Csm-40 and Csm-72. Indeed, both complexes show a slightly twisted elongated shape but the Csm-72 is significantly more elongated than Csm-40 complex (Figure 1H). Taken together these data suggest that the longer unmatured 72-nt crRNA intermediate in the Csm-72 complex binds additional copies of Csm2 and Csm3 subunits into a RNP filament (Figure 7).

Csm3 is a RNAse subunit in the StCsm complex

Computational analysis revealed that StCsm3 has a conserved RRM core and is fairly closely related (~35% sequence identity) to M. kandleri Csm3 (Hrle et al., 2013). StCsm3 displays close structural similarity to MkCsm3, in particular the RRM-core and insertions into RRM-core that form the "lid" subdomain (Figure S7A). In contrast, StCsm3 lacks both the N-terminal zinc binding domain and the C-terminal helical domain, making its structure more compact compared to that of MkCsm3. Thus, StCsm3 may be considered as a trimmed-down version of MkCsm3. Guided by the multiple sequence alignment and homology model of StCsm3, we selected candidate active site/metal chelating residues

of Csm3 and subjected them to alanine mutagenesis. We showed that the highly conserved D33 residue of the StCsm3 is critical for the RNA cleavage activity of the Csm complex, demonstrating that Csm3 is a RNase in the StCsm and could be universal in all Type III-A CRISPR/Cas systems (Figure 5).

Implications for other RNA-targeting CRISPR systems

Taken together, our data indicate that the StCsm complex is specific for the RNA and cuts it in a PAM-independent manner producing a regular 6-nt cleavage pattern. Furthermore, we demonstrate that the Csm3 protein, which is present in Csm-40 and Csm-72 complexes in multiple copies, acts as a RNase responsible for the target RNA cleavage. In this respect the Type III-A Csm complex of S. thermophilus closely resembles the RNA targeting Type III-B Cmr complex of T. thermophilus (TtCmr complex) that also produces a regular 6-nt cleavage pattern (Staals et al., 2013). The RNA degrading subunit in the Type III-B Cmr-module remains to be identified. Although there is currently no experimental evidence, Staals et al. suggested that Cmr4 could fulfill this role (Staals et al., 2013). Indeed, clustering of Csm3 and Cmr4 homologs by sequence similarity revealed that they form two related but separate groups (Figure S7D). On the other hand, neither Csm3 nor Cmr4 families are homogenous. They are comprised of sequence clusters of various sizes. StCsm3 is a member of a large representative group of Csm3 homologs that includes those from S. epidermidis, L. lactis and M. kandleri. Another large, but more loosely connected group does not have proteins from experimentally characterized systems, except for the Csm complex from S. solfataricus. Sso1425 and Sso1426, two of its Csm3-like proteins (Makarova et al., 2011a), are members of this group albeit they are non-typical. The Cmr4 family appears even more fragmented than Csm3. Cmr4 proteins of experimentally characterized III-B systems from T. thermophilus and P. furiosus represent one of the larger clusters, while Cmr4 from S. solfataricus is a non-typical outlier. Indeed, biochemical characterisation revealed that PfCmr and TtCmr RNA cleavage mechanism are similar and follow a 3'- or 5-' ruler mechanism, respectively (Hale et al., 2009; Staals et al., 2013). Meanwhile, SsCmr endonucleolytically cleaves both target RNA and crRNA at UA dinucleotides (Zhang et al., 2012).

Therefore, it would not be surprising if members of other, so far experimentally uncharacterized groups were part of Cmr complexes with somewhat different properties.

Naturally, we were curious if Csm3 and Cmr4 proteins may have similarly organized active site. The aligned sequences of Csm3 and Cmr4 subunits from characterized systems revealed that sequences of both families have Asp in the corresponding positions, suggesting similar active sites (Figure 5A). The exception is Sso1426. This is quite surprising, considering the composition of the S. solfataricus Csm complex. Four copies of Sso1426 were found to be present within the complex suggesting that this subunit might play a role of the Csm3 (Rouillon et al., 2013). In contrast, another Csm3-like protein Sso1425 does have the D33 counterpart suggesting it can cleave ssRNA. However, only a single copy of Sso1425 was found in the S. solfataricus complex. Taken together, these data suggest that Csm-modules in S. thermophilus and S. solfataricus have different architectures and RNA cleavage mechanisms.

Concluding remarks

We show here for the first time that the Csm effector complex of the S. thermophilus Type III-A system targets RNA and establish the mechanism of RNA cleavage. We demonstrate that in the Type III-A effector complex Cas/Csm proteins assemble into an RNP filament (Figure 7) that contains multiple copies of Csm2 and Csm3 proteins. Furthermore, we provide evidence that the Csm3 subunit acts as a RNase that cleaves target RNA at multiple sites spaced by regular 6-nt intervals (Figure 7). The number of cleavage sites number correlates with the number of Csm3 subunits in the Csm effector complex. Easy programmability of the Type III-A StCsm complex by custom crRNAs (Figure S5), paves the way for the development of novel molecular tools for RNA interference.

RNA cleavage specificity established here for the StCsm complex in vitro, is supported by in vivo experiments of the MS2 RNA phage interference in the heterologous E.coli host (Figure 6). It remains to be established whether RNA silencing by the StCsm complex can contribute to the DNA phage interference in the S. thermophilus host. Transcription-dependent DNA targeting mechanism has been

proposed recently for the Type III-B CRISPR-Cmr system (Deng et al., 2013), however it yet has to be demonstrated for S. thermophilus and other Type III-A systems.

Acknowledgements

The authors thank Giedrius Gasiunas for suggestions, Tomas Sinkunas, and Mindaugas Zaremba for plasmid vectors. G.T. acknowledges support from Lithuanian Science Council (grant MIP-40/2013). MJD acknowledges support from the Engineering and Physical Sciences Research Council (UK) and the Biotechnology and Biological Sciences Research Council (UK).

Experimental procedures

Expression and isolation of Csm complexes

Sequenced CRISPR2 locus of S. thermophilus DGCC8004 was deposited in GenBank (accession number KM222358). Heterologous E. coli BL21(DE3) cells producing the Strep-tagged Csm complexes were engineered and cultivated as described in the Supplemental Experimental Procedures. Csm-40 and Csm-72 complexes were isolated by subsequent Strep-chelating affinity and size exclusion chromatography steps (Supplemental Experimental Procedures).

Bioinformatic analysis and mutagenesis of Csm3

Putative active site residues of Csm3 were identified from multiple alignment of Csm3/Cmr4 (see Supplemental Experimental Procedures for details). Csm3 mutants were constructed using quick change mutagenesis and purified as described in Supplemental Experimental Procedures.

Extraction, HPLC purification and ESI-MS analysis of crRNA

NAs co-purified with Csm-40 and Csm-72 were isolated using phenol:chloroform:isoamylalcohol (25:24:1, v/v/v) extraction and precipitated with isopropanol. Purified NAs were incubated with 0.8 U

DNase I or 8 U RNase I (Thermo Scientific) for 30 min at 37°C. NAs were separated on a denaturing 15% polyacrylamide gel (PAAG) and visualised by SybrGold (Invitrogen) staining.

Ion-pair reversed-phased-HPLC purified crRNA architecture was determined using denaturing RNA chromatography in conjunction with electrospray ionization mass spectrometry (ESI-MS) as described in (Sinkunas et al., 2013) and Supplemental Experimental Procedures.

Small angle X-ray scattering (SAXS) experiments

SAXS data for Csm-40 and Csm-72 were collected at P12 EMBL beam-line at PETRAIII storage ring of DESY synchrotron in Hamburg (Germany). Csm-40 and Csm-72 complexes were measured in 3 different concentrations in buffer containing 20 mM Tris-HCl (pH 8.5 at 25°C), 0.5 M NaCl, 1mM EDTA and 7 mM 2-mercaptoethanol. Data collection, processing and ab initio shape modeling details are presented in Table S4 and Figure S3.

DNA and RNA substrates

Synthetic oligodeoxynucleotides were purchased from Metabion. All RNA substrates were obtained by in vitro transcription using TranscriptAid T7 High Yield Transcription Kit (Thermo Scientific). A full description of all the DNA and RNA substrates is provided in the Table S5. DNA and RNA substrates were either 5'-labeled with [γ^{32}P] ATP and PNK or 3'-labeled with [α^{32}P] cordycepin-5'-triphospate (PerkinElmer) and poly(A) polymerase (Life Technologies) followed by denaturing gel purification.

Electrophoretic mobility shift assay

Binding assays were performed by incubating different amounts of Csm complexes with 0.5 nM of ^{32}P-5′-labeled NA in the Binding buffer (40 mM Tris, 20 mM acetic acid (pH 8.4 at 25°C), 1 mM EDTA, 0.1 mg/ml BSA, 10% (v/v) glycerol). All reactions were incubated for 15 min at room temperature prior to electrophoresis on native 8% (w/v) PAAG. Electrophoresis was carried out at room temperature for 3 h at 6 V/cm using 40 mM Tris, 20 mM acetic acid (pH 8.4 at 25°C), 0.1 mM EDTA as the running buffer. Gels were dried and visualized using a FLA-5100 phosphorimager (Fujifilm).

The K_d for NA binding by Csm-72 and Csm-40 was evaluated assuming the complex concentration at which half of the substrate is bound as a rough estimate of K_d value. For binding competition assay 0.5 nM ^{32}P-labelled S3/1 RNA was mixed with 0.5-5000 nM of unlabelled competitor NA and 0.3 nM StCsm-40,and analysed by EMSA.

Cleavage assay

The Csm-40 reactions were performed at 25°C and contained 20 nM of 5'- or 3'-radiolabeled NA (Table S5) and 62.5 nM (unless stated otherwise) complex in the Reaction buffer (33 mM Tris-acetate (pH 7.9 at 25°C), 66 mM K-acetate, 0.1 mg/ml BSA and 10 mM Mg-acetate). Csm-72 reactions were performed in the same Reaction buffer at 37°C and contained 20 nM of radiolabeled NA and 125 nM of complex unless stated otherwise. Reactions were initiated by addition of the Csm complex. The samples were collected at timed intervals and quenched by mixing 10 µl of reaction mixture with 2X RNA loading buffer (Thermo Scientific) followed by incubation for 10 min at 85°C. The reaction products were separated on a denaturing 20% PAAG and visualized by autoradiography. ^{32}P-5'-labeled RNA Decade marker (Ambion) was used as size marker. To map the cleavage products oligoribonucleotide markers were generated by RNase A (Thermo Scientific, final concentration 10 ng/ml) treatment of RNA substrates for 8 min at 22°C or by alkaline hydrolysis in 50 mM $NaHCO_3$ (pH 9.5) at 95°C for 5 min.

Phage drop plaque assay

Phage drop plaque assay was conducted using LGC Standards recomendations (see Supplemental Experimental Procedures).

References

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960-964.

Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes & development 22, 3489-3496.

Deng, L., Garrett, R.A., Shah, S.A., Peng, X., and She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. Molecular microbiology 87, 1088-1099.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A 109, E2579-2586.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139, 945-956.

Hatoum-Aslan, A., Maniv, I., and Marraffini, L.A. (2011). Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. Proc Natl Acad Sci U S A 108, 21218-21222.

Hatoum-Aslan, A., Maniv, I., Samai, P., and Marraffini, L.A. (2014). Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. J Bacteriol 196, 310-317.

Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W., and Marraffini, L.A. (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. The Journal of biological chemistry 288, 27888-27897.

Hrle, A., Su, A.A., Ebert, J., Benda, C., Randau, L., and Conti, E. (2013). Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3. RNA biology 10, 1670-1678.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821.

Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E.V. (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biology direct 6, 38.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., et al. (2011b). Evolution and classification of the CRISPR-Cas systems. Nature reviews. Microbiology 9, 467-477.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322, 1843-1845.

Marraffini, L.A., and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nature reviews. Genetics 11, 181-190.

Millen, A.M., Horvath, P., Boyaval, P., and Romero, D.A. (2012). Mobile CRISPR/Cas-mediated bacteriophage resistance in Lactococcus lactis. PLoS One 7, e51663.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733-740.

Olsthoorn, R., and van Duin, J. (2011). Bacteriophages with ssRNA. In eLS (John Wiley & Sons Ltd, Chichester).

Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C.V., Spagnolo, L., and White, M.F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. Mol Cell 52, 124-134.

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. EMBO J 32, 385-394.

Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., and Stagg, S. (2013). Structure of an RNA silencing complex of the CRISPR-Cas immune system. Mol Cell 52, 146-152.

Staals, R.H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K., et al. (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. Mol Cell 52, 135-145.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62-67.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc Natl Acad Sci U S A 111, 9798-9803.

Terns, R.M., and Terns, M.P. (2014). CRISPR-based technologies: prokaryotic defense weapons repurposed. Trends in genetics : TIG 30, 111-118.

Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol Cell 46, 595-605.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature 477, 486-489.

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.V., et al. (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Mol Cell 45, 303-313.

Figure 1. Cloning, isolation and characterization of the Type III-A Csm complex of S. thermophilus DGCC8004**.** (A) Schematic organization of the Type III-A CRISPR2 locus (see also Figure S1). Repeats and spacers are indicated by diamonds and rectangles, respectively, T is for the terminal repeat, L is for the leader sequence. Arrow indicate promoter. (B) Protospacer PS3 and the 5'-flanking sequence in the S. thermophilus phage O1205 genome. (C) Strategy for expression and isolation of the StCsm complex. Four copies of the spacer S3 have been engineered into the pCRISPR_S3 plasmid to increase the yield of the Csm-crRNA complex. (D) Coomassie blue-stained SDS-PAAG of Strep-tagged Csm2 and Csm3 pull-downs. 3N – Csm3_StrepN protein, M – protein mass marker. (E) Denaturing PAGE analysis of NA co-purifying with the Csm2_StrepN and Csm3_StrepN complexes. M – synthetic DNA marker. (F) and (G) Characterization of crRNA in the isolated StCsm complexes. Cartoon models illustrate crRNA which co-purify with StCsm-72 and StCsm-40 complexes. Composition of the crRNA was determined using LC ESI MS analysis (see also Figure S2). IP RP HPLC analysis and ESI MS spectra of IP RP HPLC purified crRNA from StCsm-40 and StCsm-72 are presented. (H) Superimposed averaged dummy atom models obtained from SAXS data of StCsm-40 (yellow beads) and StCsm-72 (blue beads) (see also Figure S3).

Figure 2. Nucleic acid binding and cleavage by the Type III-A Csm complex of S. thermophilus. (A) Schematic representation of DNA and RNA substrates used for in vitro binding and cleavage assays. NA were 5′- or 3'-end labeled with ^{32}P (indicated as *). (B) EMSA analysis of DNA or RNA binding by StCsm-40. NS stands for a non-specific RNA. (C) Bing competition assay. 0.5 nM of ^{32}P-labelled S3/1 RNA was mixed with increasing amounts of unlabelled competitor NAs and 0.3 nM StCsm-40, and analysed by EMSA. (D) StCsm-40 cleavage assay. Gel-purified DNA or RNA were used as substrates in the NA cleavage assay. Triangles with corresponding numbers indicate cleavage product length. M – RNA Decade marker, R – RNase A digest marker, H – alkaline hydrolysis marker. (E) RNA cleavage products mapped on the S3/1 RNA substrate sequence. Triangles and dashed lines

indicate cleavage positions. Short vertical lines above the sequence indicate nucleotides complementary to crRNA. crRNA (StCsm-40) sequence is depicted above the matching substrate fragments.

Figure 3. The effect of the sequence complementarity outside the spacer region on the StCsm-72 cleavage pattern. (A) Schematic representation of the StCsm-72 complex and RNA substrates used in the cleavage assay. RNA substrates were 5′-end labeled with ^{32}P (indicated as *) and gel-purified. (B) StCsm-72 cleavage assay. M – RNA Decade marker. (C) RNA cleavage products mapped on the S3/2, S3/5 and S3/6 RNA substrates sequences.

Figure 4. The effect of protospacer truncations on the StCsm-40 cleavage pattern. (A) Schematic representation of the StCsm-40 complex and RNA substrates used in the cleavage assay. RNA substrates were 5′-end labeled with ^{32}P (indicated as *) and gel-purified. (B) StCsm-40 cleavage assay. M – RNA Decade marker. (C) RNA cleavage products mapped on the RNA substrates sequences.

Figure 5. Computational and mutational analysis of Csm3. (A) Alignment of Csm3 and Cmr4 sequence representatives from experimentally characterized Type III effector complexes. Identical and similar residues in more than half of sequences are shaded in blue and green correspondingly. StCsm3 positions subjected to site-directed mutations are indicated by triangles above the sequence. (B) Coomassie blue-stained SDS-PAAG of StCsm complexes containing Csm3 mutants. M – protein marker. (C) Denaturing PAGE analysis of NA co-purifying with mutant StCsm complexes. M – synthetic DNA marker. (D) EMSA analysis of S3/2 RNA binding by mutant StCsm complexes. (E) S3/2 RNA cleavage by the mutant StCsm complexes. (F) The cleavage rate constant k_{obs} values for Csm3 mutant variants.

Figure 6. Restriction of ssRNA phage MS2 in E. coli cells expressing StCsm complex. (A) Cartoon representation of assay. Arrow indicate promoter. (B) Phage plaque analysis. Serial 10-fold dilutions of MS2 were transferred onto lawns of E. coli NovaBlue (DE3, F^+) strain expressing StCsm-crRNA complex targeting the MS2 genome or control cells.

Figure 7. Structural and cleavage models of StCsm complexes. Transcript of CRISPR region is first processed into 72 nt crRNA intermediate which undergoes further maturation into 40 nt crRNA. Both crRNAs are incorporated into StCsm complexes that target RNA but differ by the number of Csm3 and Csm2 subunits. The number of RNA cleavage products correlate with the number of Csm3 nuclease subunits. Schematic models of StCsm complexes were generated based on similarity to TtCmr and PfCmr. Csm analogs of Cmr proteins according to (Makarova et al., 2011a) are colored identically.

Figure1

Click here to download high resolution image

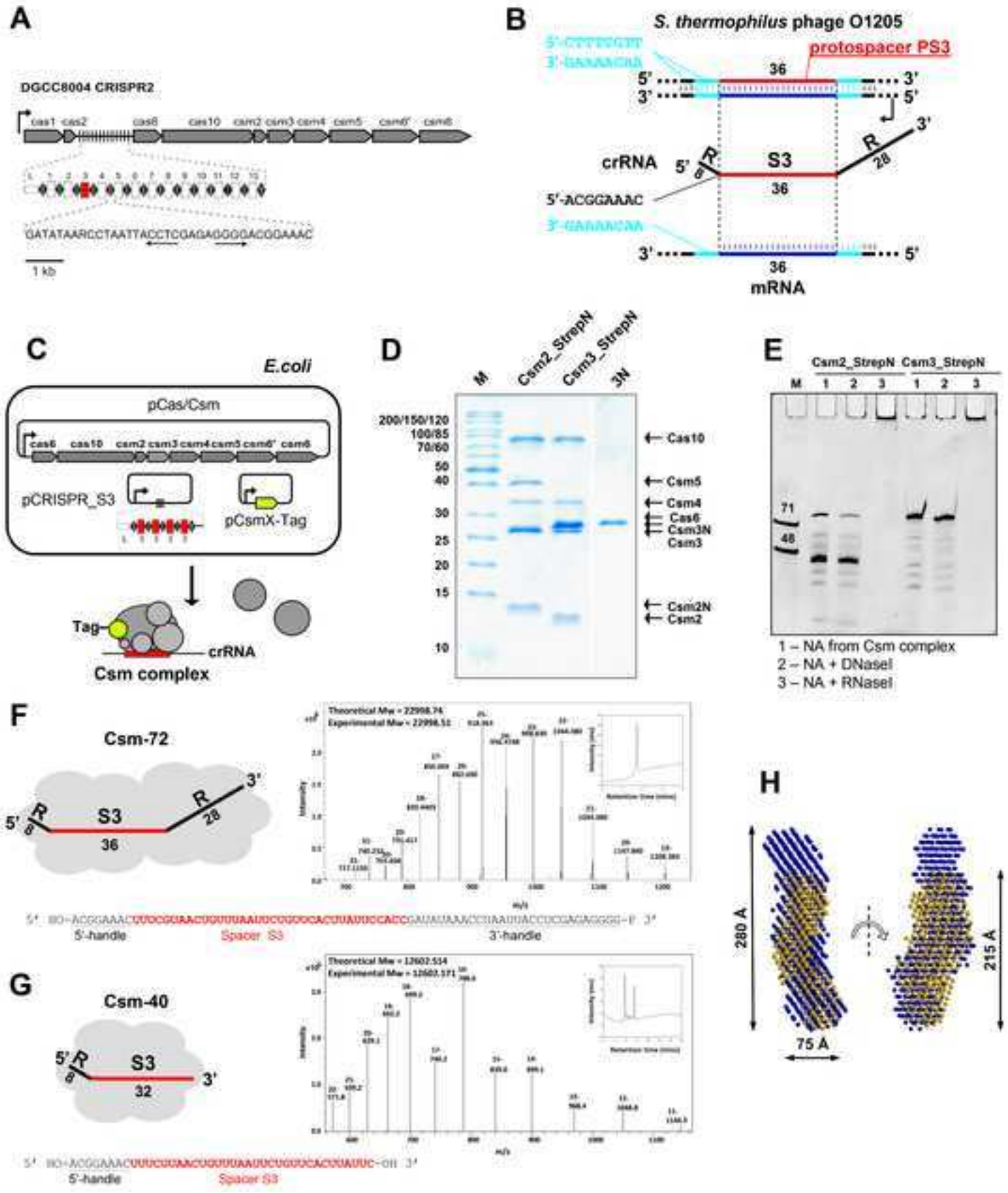


Figure2

Click here to download high resolution image

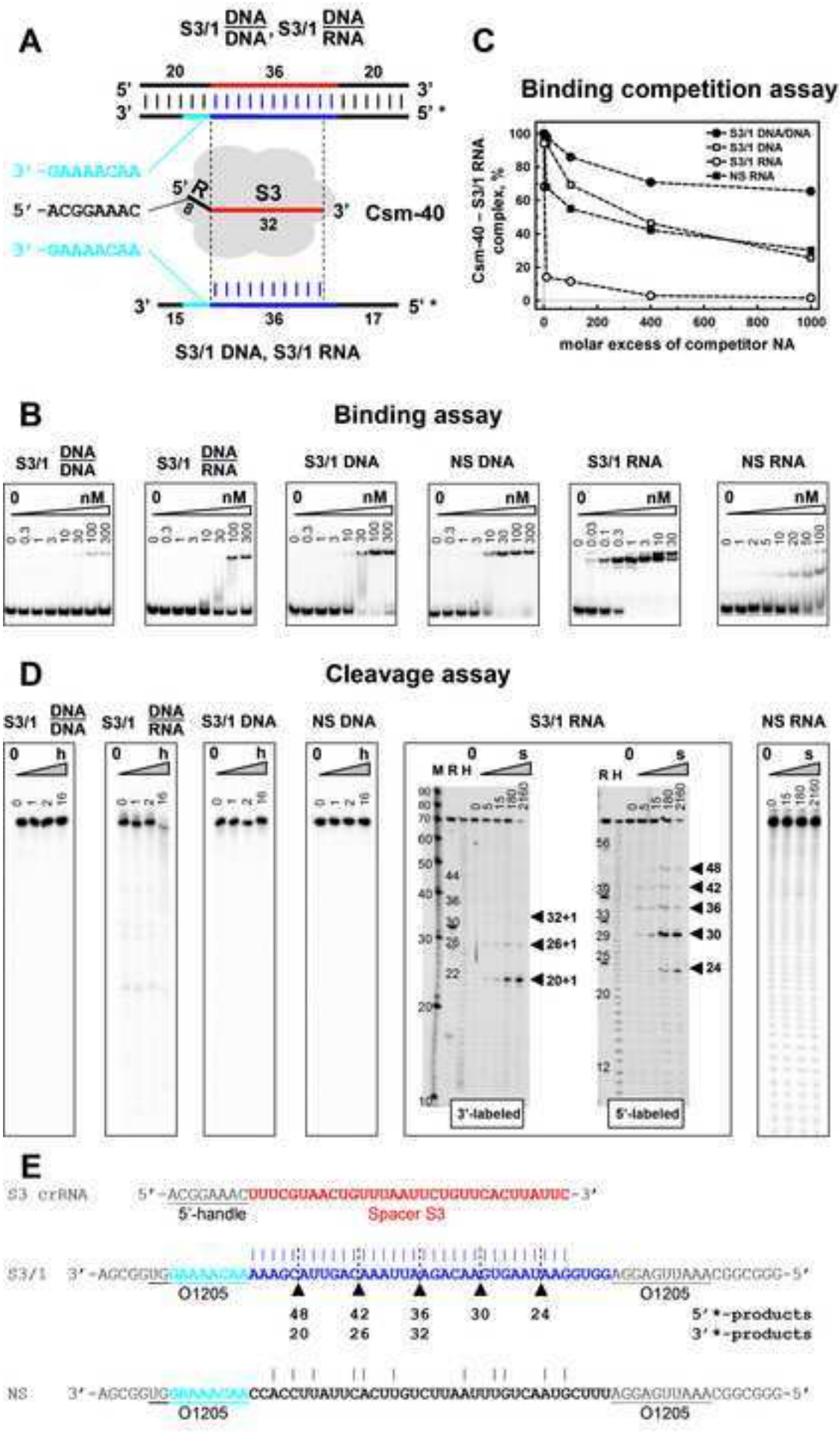


Figure3
Click here to download high resolution image

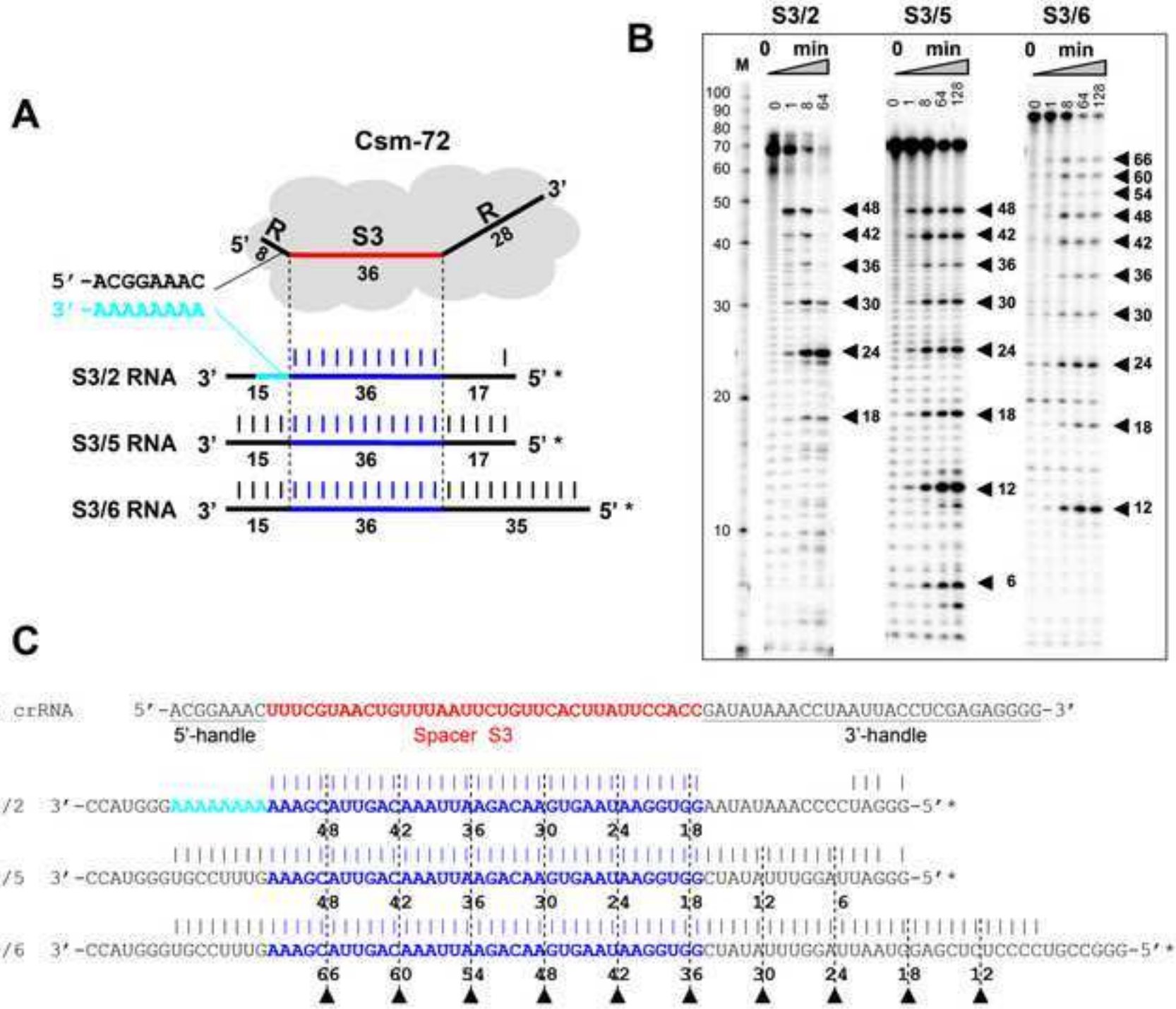


Figure4

Click here to download high resolution image

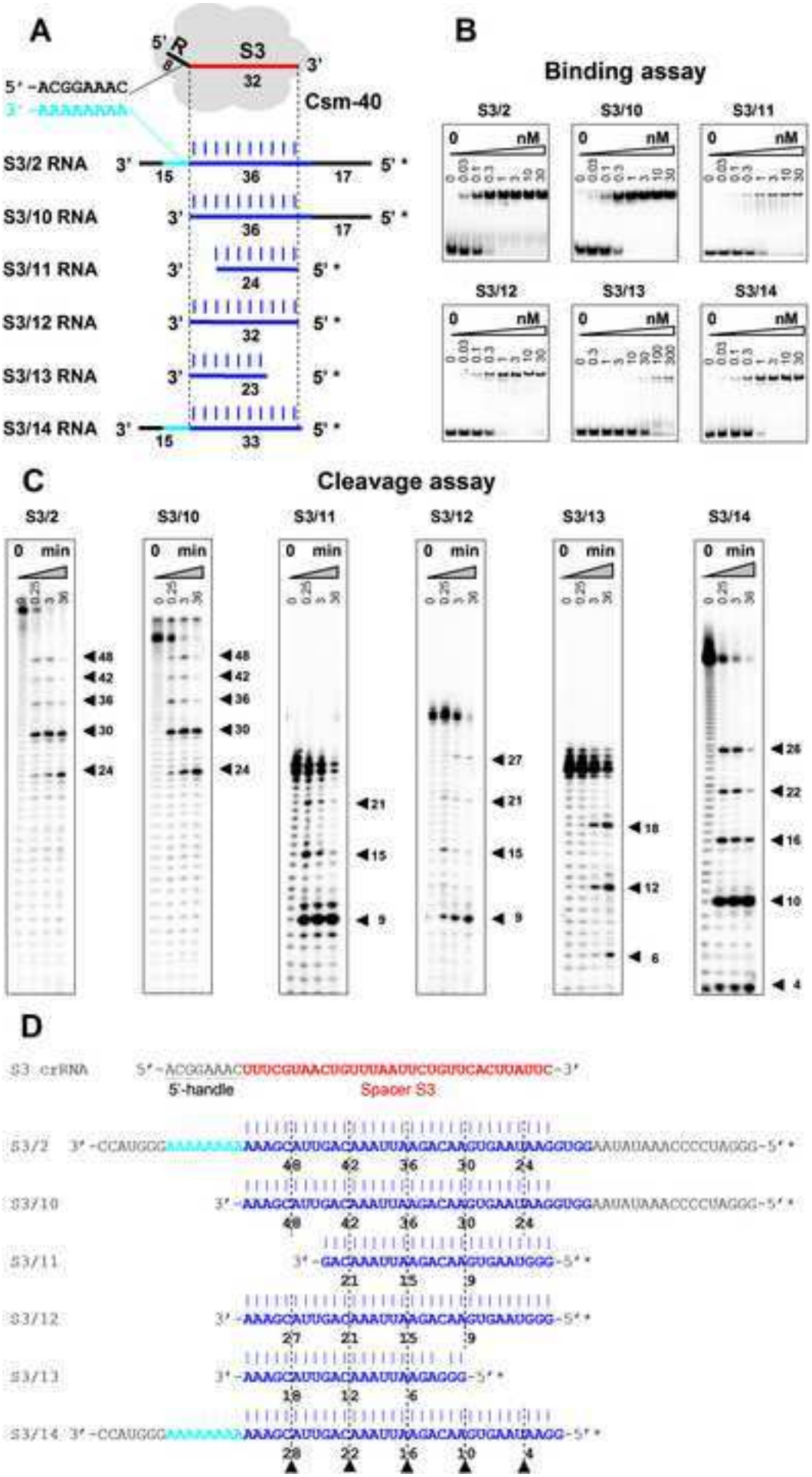


Figure5

Click here to download high resolution image

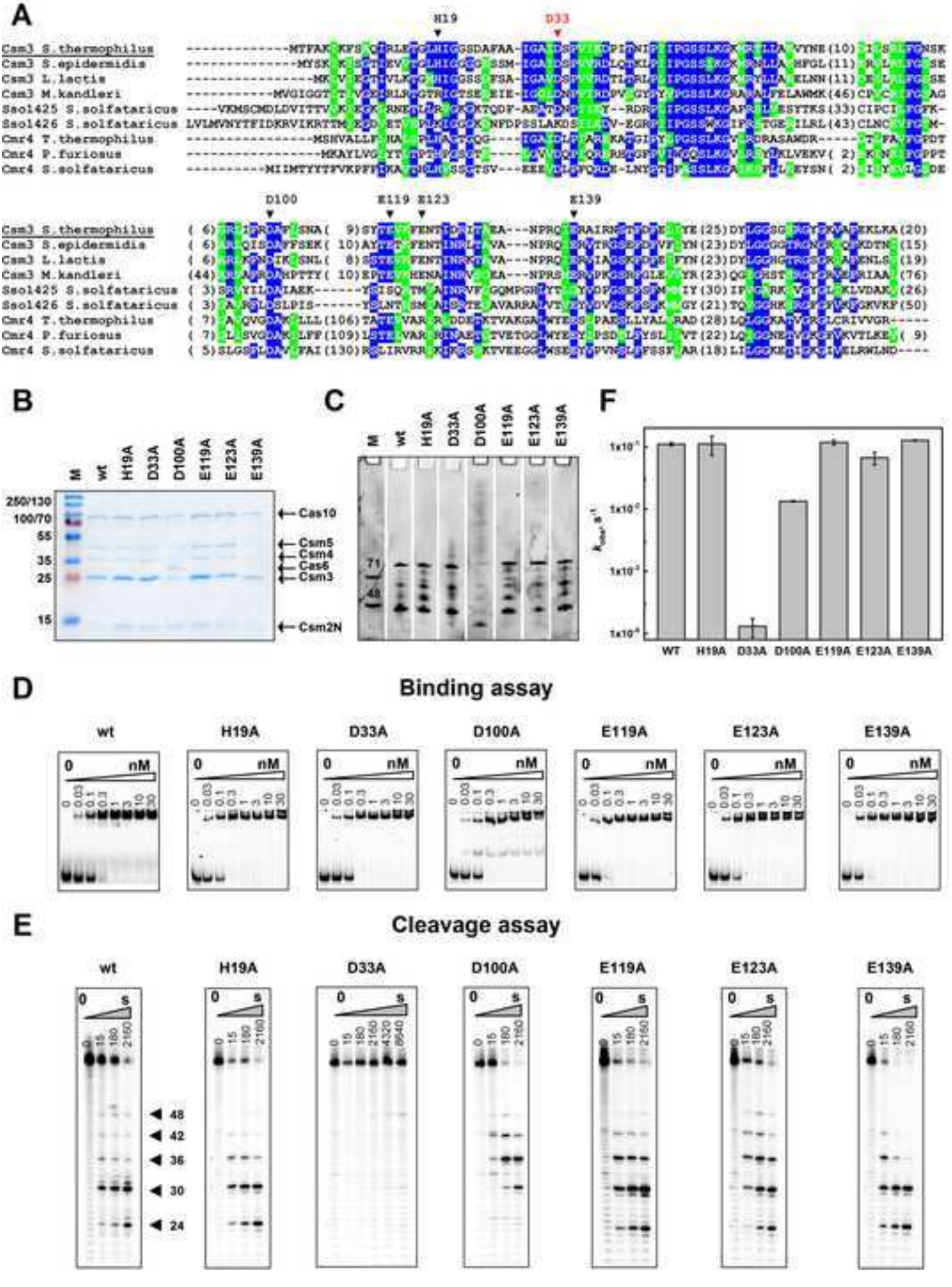


Figure6
Click here to download high resolution image

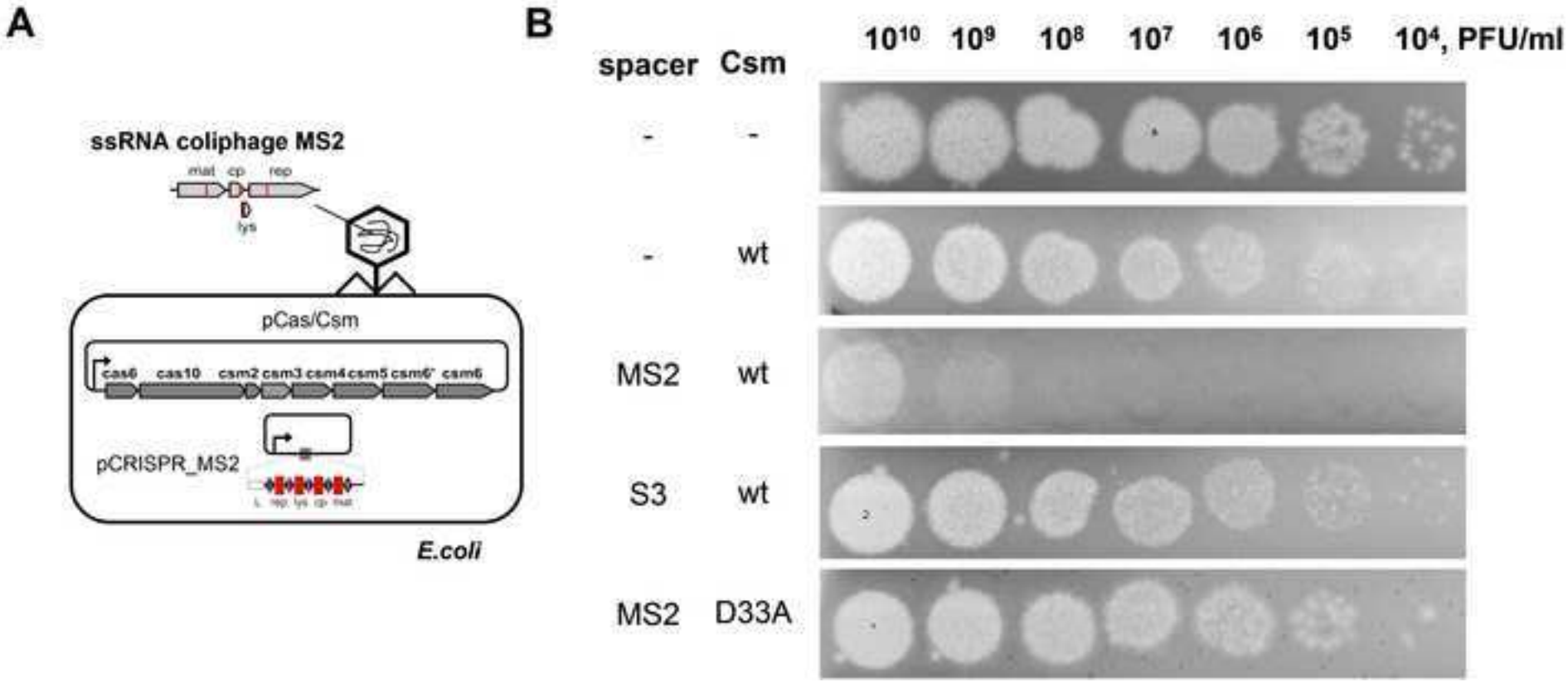
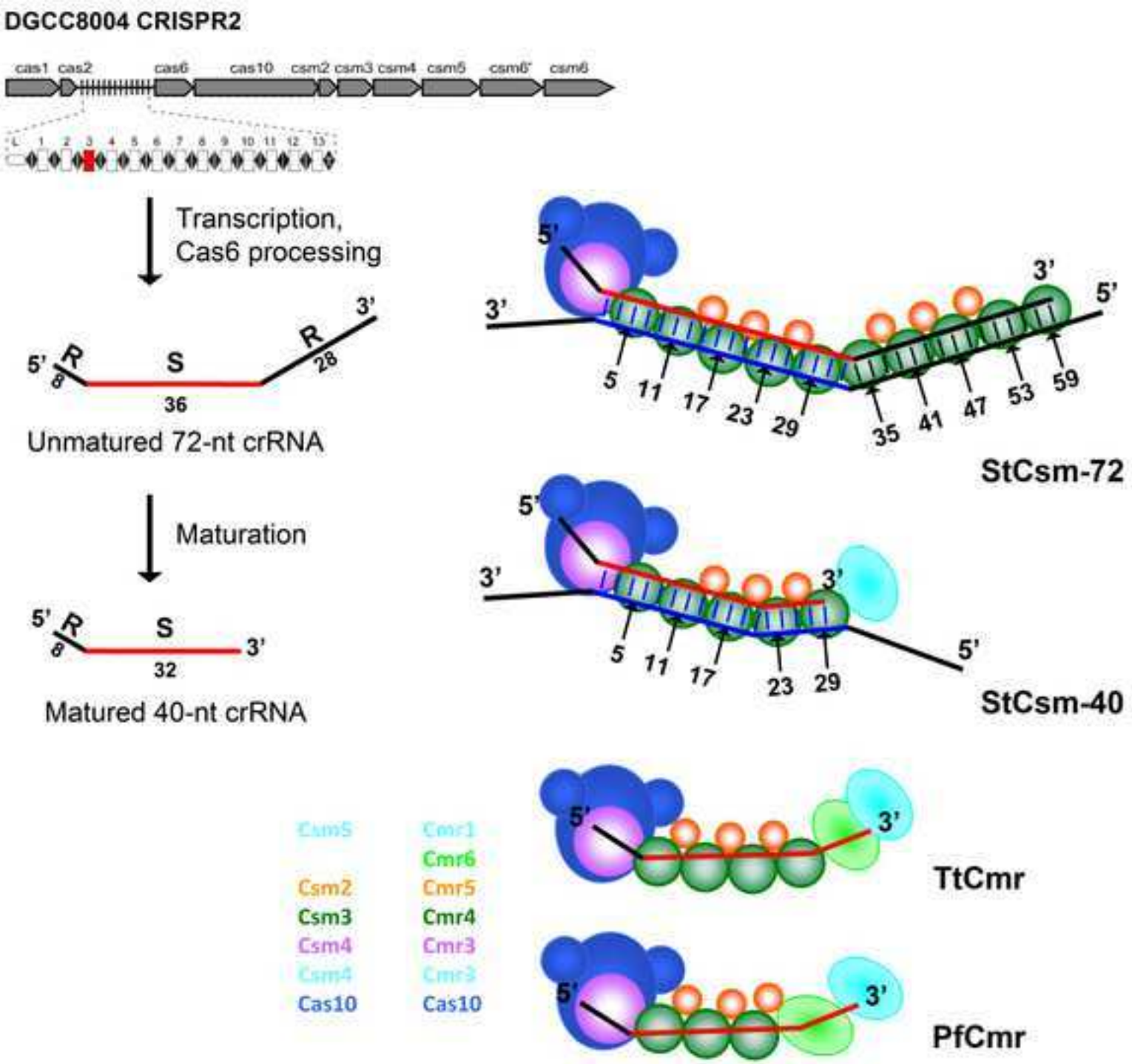


Figure7

Click here to download high resolution image



Supplemental Data

Supplemental Figures

Figure S1 (related to Figure 1):

Schematic organization of the Type III-A CRISPR-Cas systems of *Streptococcus thermophilus* DGCC8004, DGCC7710, LMD-9, *Staphylococcus epidermidis* RP62a, *Enterococcus italicus* DSM15952, *Lactococcus lactis* DGCC7167 and *Sulfolobus solfataricus* P2.

Figure S2 (related to Figure 1):

ESI MS/MS oligoribonucleotide mapping of crRNA isolated from StCsm-72 and StCsm-40.

Figure S3 (related to Figure 1):

SAXS data for StCsm complexes.

Figure S4 (related to Figure 2):

Target RNA binding and cleavage by StCsm-72 and StCsm-40 complexes.

Figure S5 (related to Figure 2):

Reprogramming of the StCsm complex to cut a desired RNA target.

Figure S6 (related to Figure 3):

The role crRNA: target RNA complementarity on the StCsm-40 cleavage pattern.

Figure S7 (related to Figure 5):

Computational analysis of Csm3.

Supplemental Tables

Table S1 (related to Figure 1):

Proteins identified following mass spectrometry analysis of StCsm-72.

Table S2 (related to Figure 1):

Proteins identified following mass spectrometry analysis of StCsm-40.

Table S3 (related to Figures 1):

M_W estimations for StCsm-40 and StCsm-72 by different methods.

Table S4 (related to Figure 1):

SAXS data collection details and structural parameters of StCsm-40 and StCsm-72 complexes.

Table S5 (related to Figures 2, 3, 4, 5 and 6):

Nucleic acid substrates used in this study.

Supplemental Experimental Procedures

Cloning, expression and purification of StCsm complexes

HPLC purification of crRNA

ESI-MS analysis of crRNA

SAXS experiments

Computational sequence and structure analysis

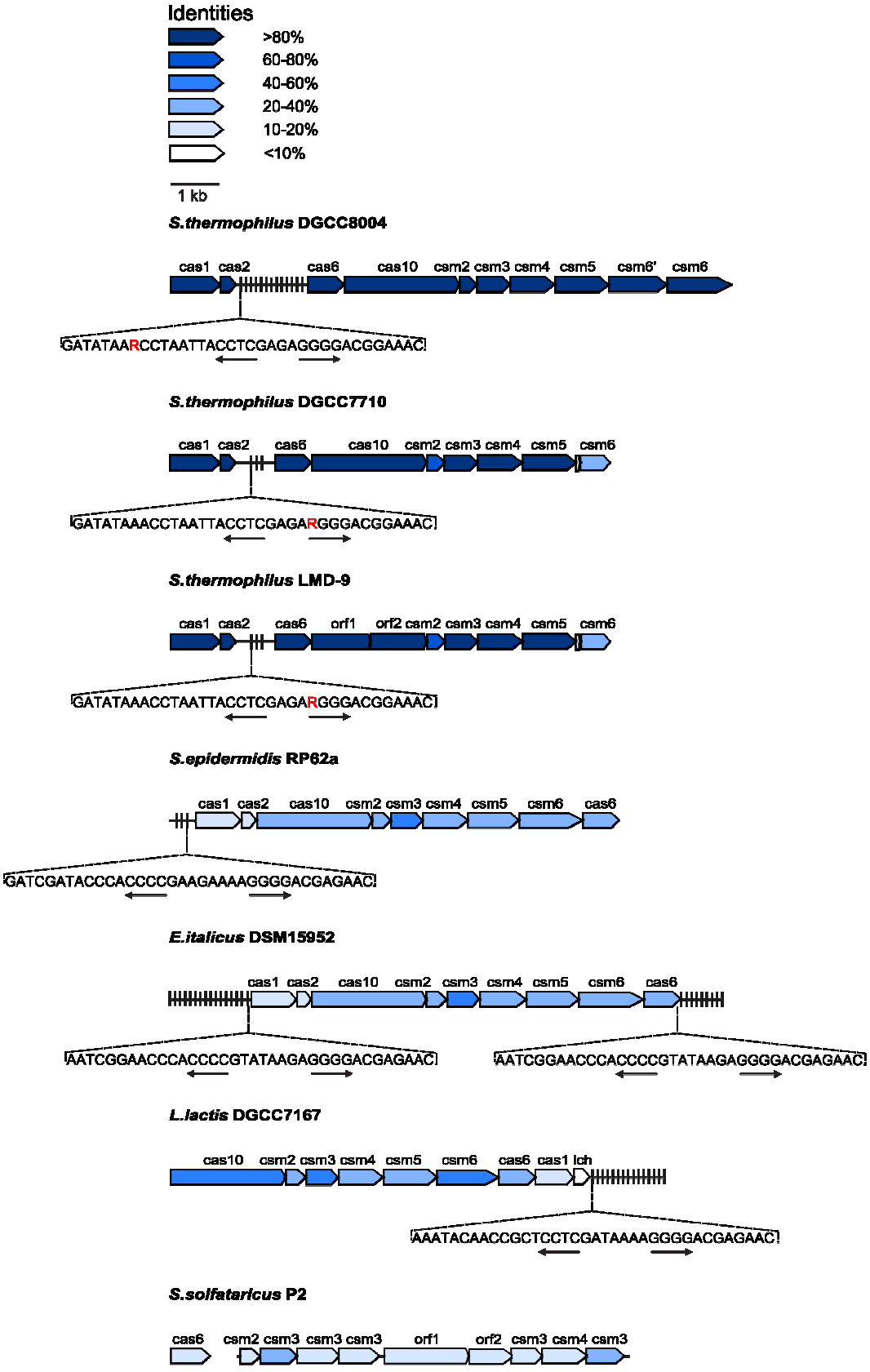
Mutagenesis of Csm3

DNA and RNA substrates

Phage drop plaque assay

Supplemental References

Figure S1 (related to Figure 1).



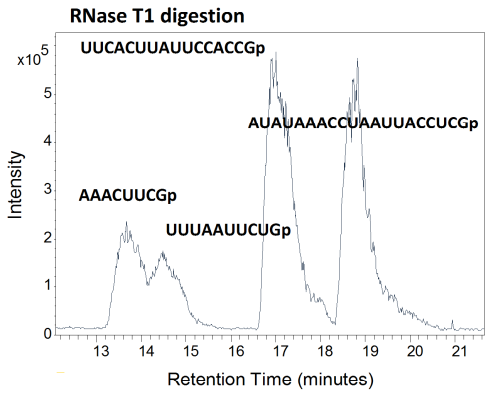
Schematic organization of the Type III-A CRISPR-Cas systems of *Streptococcus thermophilus* DGCC8004, DGCC7710 (GenBank AWVZ01000003), LMD-9 (GenBank NC008532), *Staphylococcus epidermidis* RP62a (GenBank NC002976), *Enterococcus italicus* DSM15952 (GenBank AEPV01000074), *Lactococcus lactis* DGCC7167 (GenBank JX524189) and

***Sulfolobus solfataricus* P2 (GenBank AE006641)*.** Arrows are colored according to the percentage of identical residues (Vector NTI AlignX tool) in Csm/Cas proteins in respect to the *S. thermophilus* DGCC8004. Conserved repeat sequences are shown in the inserts. Partially palindromic repeat sequences are indicated by arrows. In *L. lactis* DGCC7167 CRISPR2 system *lch* gene which shows a partial homology to the relE/parE toxin gene is present instead of *cas2* (Millen et al., 2012). In CRISPR-Cas loci of *S. thermophilus* LMD-9 and *S. solfataricus* P2 *cas10* is split in two open reading frames ORF1 and ORF2.

* The Type III-A system of DGCC8004 contains 10 *cas* genes flanking the CRISPR2 array and includes *cas1, cas2, cas6, cas10, csm2, csm3, csm4, csm5, csm6* and *csm6'* genes. The DGCC8004 CRISPR2 locus share similar gene arrangement to that of DGCC7710 (GenBank AWVZ00000000, (Horvath and Barrangou, 2010)) and LMD-9 (GenBank NC_008532, (Makarova et al., 2006)). The major difference is an additional *csm6'* gene in DGCC8004. The Csm6' protein in DGCC8004 is comprised of 386 aa and shows ~34% amino acid identity to the 428 aa Csm6 protein, suggesting a possible ancient gene duplication event followed by sequence divergence. In contrast, DGCC7710 contains only a short 117-nt ORF in front of *csm6*. The Cas/Csm proteins associated to CRISPR2 in DGCC8004 are homologous to the corresponding proteins in DGCC7710 and LMD-9 (more than 90% aa identity, except for the Csm2 protein, which shares ~70% identity). Other experimentally characterized Type III-A systems including *S. epidermidis* RP62a (GenBank NC002976, (Marraffini and Sontheimer, 2008)), *Enterococcus italicus* DSM15952 (GenBank AEPV01000074, (Millen et al., 2012)) and *Lactococcus lactis* DGCC7167 (GenBank JX524189, (Millen et al., 2012)) share with DGCC8004 a conserved arrangement of the *cas10-csm2-csm3-csm4-csm5-csm6* gene cluster, while the position of *cas6* and *cas1*/*cas2* genes differ in some strains. The Type III-A signature protein Cas10 of DGCC8004 shows ~34-40% identity (~50-55% similarity) to Cas10 of *S. epidermidis*, *E. italicus* and *L. lactis*. In LMD-9, the *cas10* gene is split into two ORFs which match to the N- and C-terminal fragments of Cas10 in DGCC8004 (> 92% identical aa). Type III-A CRISPR-Cas locus in *S. solfataricus* P2 (GenBank AE006641) has different gene organization and shows low protein sequence similarity to Cas/Csm orthologues in DGCC8004. Noteworthy, the Csm3 protein is most conserved among the Cas/Csm proteins across different strains and 5 copies of the Csm3 paralogues are present in *S. solfataricus*. Repeat sequences in *S. epidermidis*, *Enterococcus italicus* and *Lactococcus lactis* are of the same length (36 nt), however the nucleotide conservation is limited to the palindromic parts and 3'-terminal end of the repeats. The 8-nt 3'-terminal sequence of the repeat, which may contribute to the crRNA 5'-handle, shows an ACGRRAAC consensus between *S. thermophilus*, *S. epidermidis*, *E. italicus* and *L. lactis* but differs from that of *S. solfataricus* (AUUGAAG (Rouillon et al., 2013)).

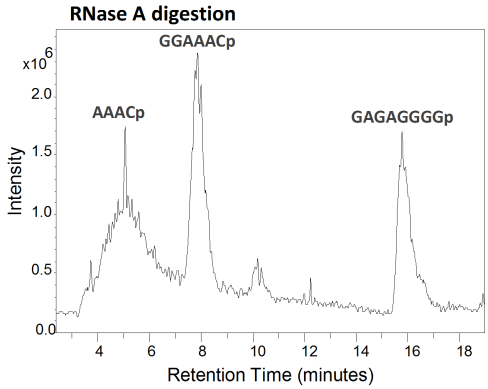
Figure S2 (related to Figure 1).

A



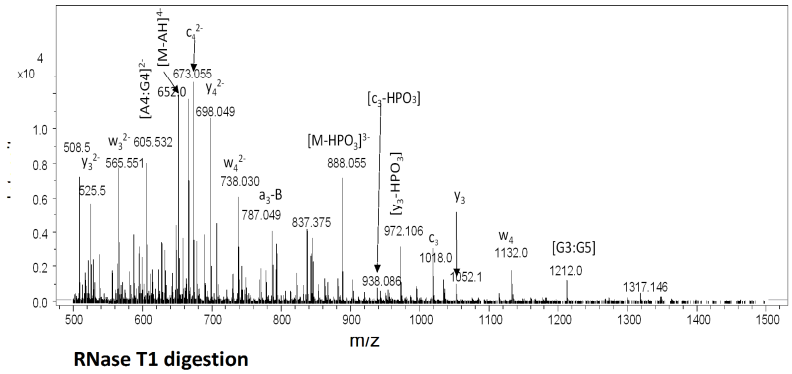
Position	Theoretical mass (Da)	Experimental Mass (Da)	Sequence
A5:G13	2879.725	2879.529	AAACUUUCGp
U20:G29	3163.839	3163.663	UUUAAUUCUGp
U30:G45	5018.969	5018.918	UUCACUUAUUCCACCGp
A46:G65	6359.830	6359.347	AUAUAAACCUAAUUACCUCGp

B

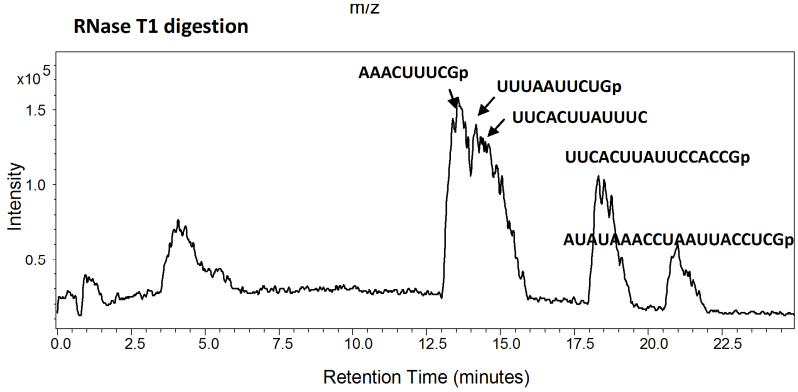


Position	Theoretical mass (Da)	Experimental Mass (Da)	Sequence
G3:C8	2001.244	2001.3	GGAAACp
A50:C53	1310.279	1310.3	AAACp
G65:G72	2747.306	2747.4	GAGAGGGGp

C



D



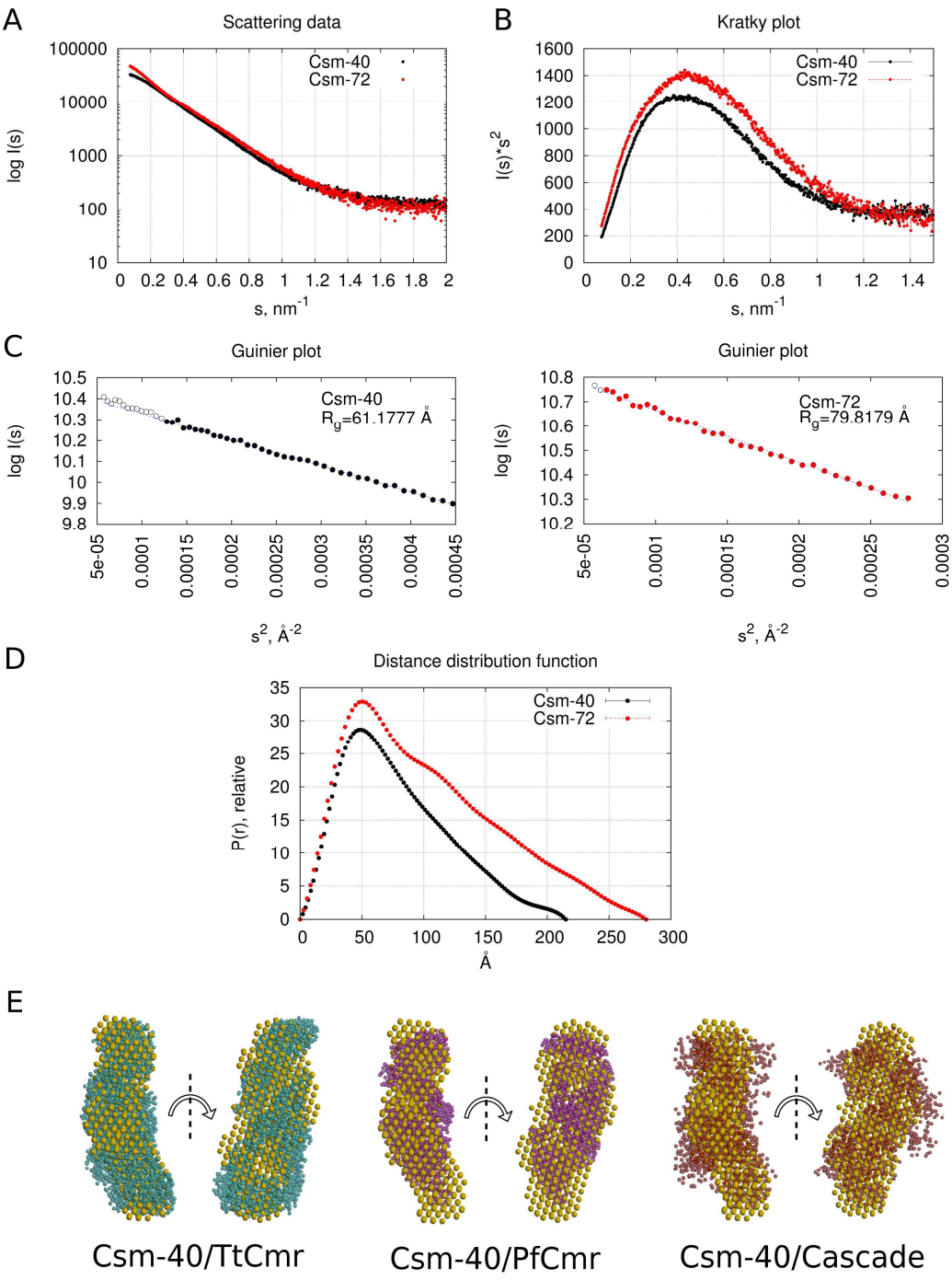
Position	Theoretical mass (Da)	Experimental Mass (Da)	Sequence
A5:G13	2879.725	2879.526	AAACUUUCGp
U20:G29	3163.839	3163.381	UUUAAUUCUGp
U30:C40	3349.019	3348.906	UUCACUUAUUC
U30:G45	5018.969	5018.478	UUCACUUAUUCCACCGp
A46:G65	6359.830	6359.333	AUAUAAACCUAAUUACCUCGp
U30:G45	5000.954	5001.082	UUCACUUAUUCCACCG>p
A46:G65	6341.815	6342.009	AUAUAAACCUAAUUACCUCG>p

ESI MS/MS oligoribonucleotide mapping of crRNA isolated from StCsm-72 and StCsm-40.

(A-C) ESI MS/MS oligoribonucleotide mapping of crRNA isolated from StCsm-72. (A) Base peak chromatogram of RNase T1 digest. RNase T1 cleaves single-stranded RNA 3’ of G residues. Predominant oligoribonucleotide peaks of the crRNA are highlighted. Masses of each oligoribonucleotide are presented in the table. The theoretical and experimental masses are shown for the oligoribonucleotides identified. (B) Base peak chromatogram of RNase A digest. RNase A cleaves single-stranded RNA 3’ of C or U residues. (C) ESI MS/MS analysis of the oligoribonucleotide GAGAGGGGp. Tandem MS was used to verify the oligoribonucleotide. The predominant fragment ions are highlighted.

(D) ESI MS/MS oligoribonucleotide mapping of crRNA isolated from StCsm-40. Base peak chromatogram of RNase T1 digest. The oligoribonucleotide UUCACUUAUUC was unique to the 40-nt crRNA. p = 3’-phosphate, >p = 2’3’-cyclic phosphate.

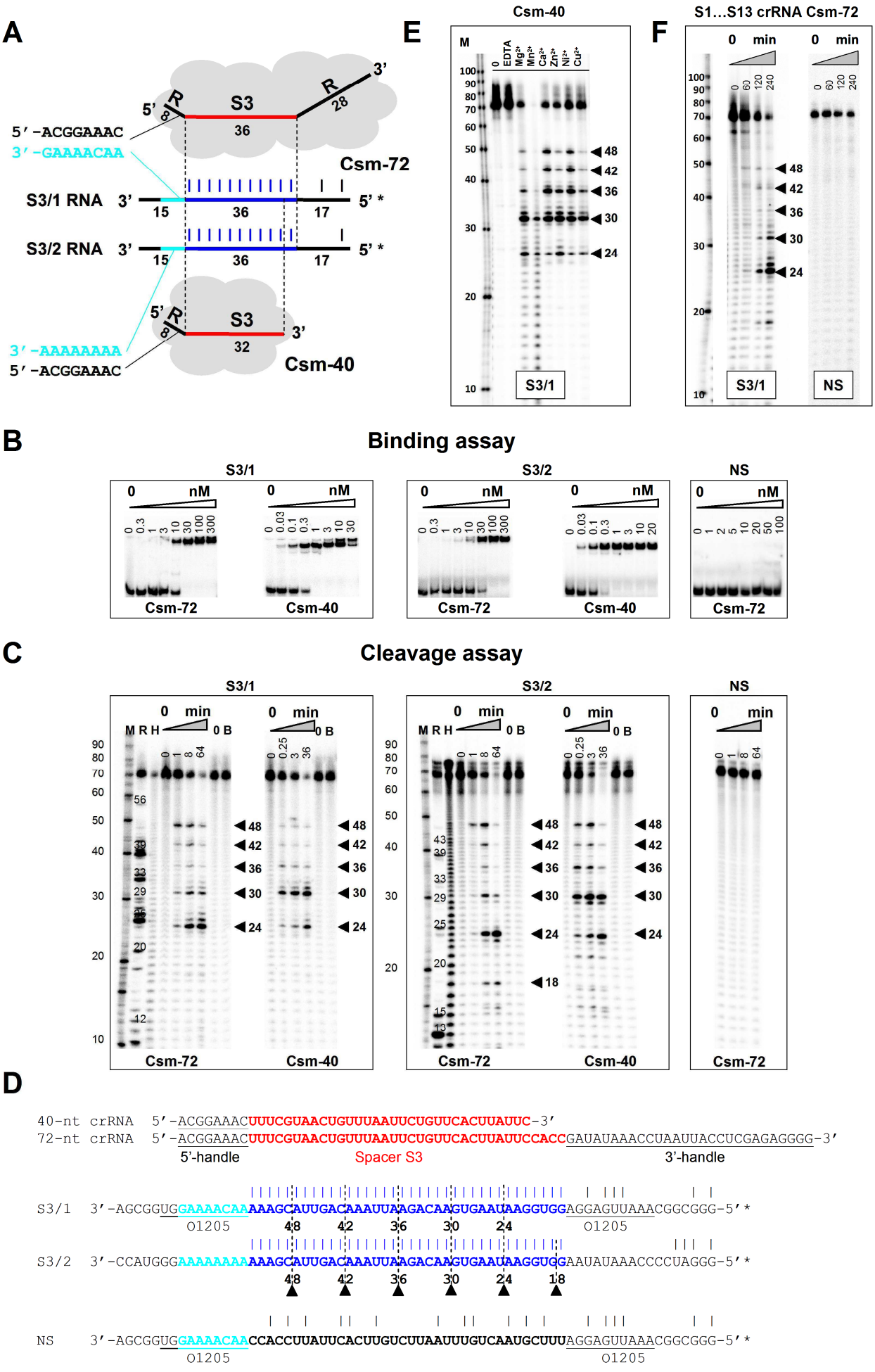
Figure S3 (related to Figure 1).



SAXS data for StCsm complexes.

SAXS data for StCsm-40 (black dots) and StCsm-72 (red dots) are shown. (A) scattering profiles shown as a logarithmic plot of scattering intensity I(s) *vs* momentun transfer $s = 4\pi \sin(\theta)/\lambda$, where 2θ is a scattering angle and λ is X-ray wavelength. (B) Kratky plot of SAXS data, $I(s)*s^2$ *vs* s. (C) Guinier plots of SAXS data, ln I(s) vs s^2 and its linear fit. The truncated first points are shown as open circles. (D) Distance distribution functions of StCsm-40 and StCsm-72 complexes calculated using GNOM (Svergun, 1992). (E) The electron density of the TtCmr (green beads), PfCmr (violet beads), and *E. coli* Cascade complexes (red beads) aligned with the StCsm-40 (yellow beads) model.

Figure S4 (related to Figure 2).



Target RNA binding and cleavage by StCsm-72 and StCsm-40 complexes.

(A) Schematic representation of S3/1 and S3/2 RNA substrates used in binding and cleavage assays. Nucleic acids were 5′-end labeled with ^{32}P (indicated as *).

(B) Electrophoretic mobility shift binding assay. The binding reactions contained ^{32}P-labeled RNA (0.5 nM) and the Csm-72 or Csm-40 at concentrations indicated by each lane. Samples were analyzed by PAGE under non-denaturing conditions. NS shows the non-specific RNA control.

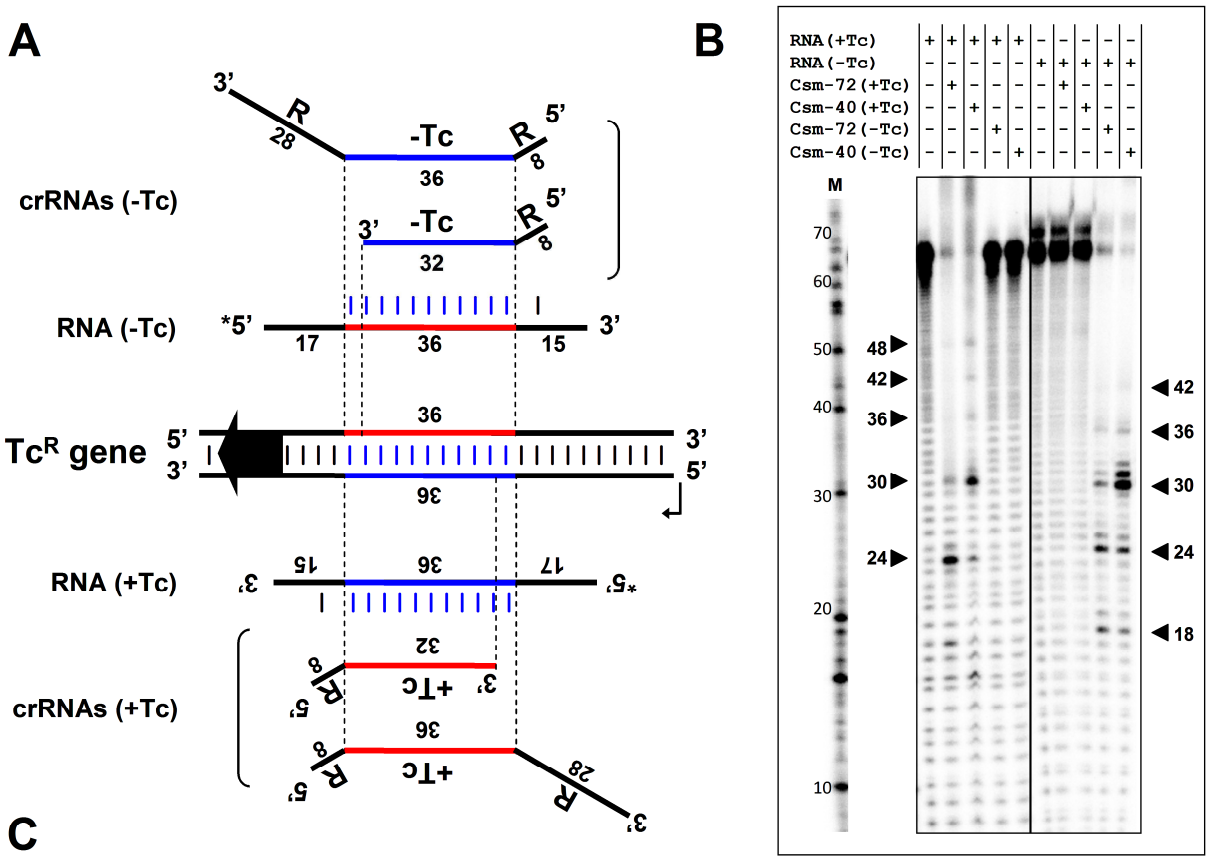
(C) Cleavage assay. Cleavage reactions were performed at 37°C for Csm-72 and 25°C for Csm-40 for indicated time intervals in the Reaction buffer supplemented with 10 mM Mg-acetate, 20 nM RNA substrate and 125 nM Csm-72 or 62.5 nM Csm-40. Samples were analyzed by denaturing PAGE, followed by phosphorimaging. In control experiments RNA substrate was incubated for 64 min at 37°C or 36 min at 25°C in the Reaction buffer alone ("lane 0") or the storage buffer was added instead of the Csm complex ("lane B"). Triangles denote the reaction products (the sizes of cleavage fragments are given near triangles). M – RNA Decade marker, R – oligoribonucleotide fragments generated from RNase A digestion of RNA, H – alkaline hydrolysis of RNA.

(D) RNA cleavage products mapped on the RNA substrates sequence. Triangles and dashed lines indicate cleavage positions. Short vertical lines above the sequence indicate nucleotides complementary to crRNA. 40-nt and 72-nt crRNAs containing spacer S3 sequences are depicted above the matching substrate fragments. SP for specific RNA, NS for non-specific RNA.

(E) Metal ion (Me^{2+}) dependency of the RNA cleavage by the Csm complex. S3/1 RNA substrate was pre-incubated with Csm-40 and reaction products were analyzed in denaturating polyacrylamide gels. Cleavage reactions were conducted at 25°C for 3 min in Reaction buffer containing 20 nM ^{32}P-5'-labelled gel purified S3/1 RNA substrate, 62.5 nM Csm-40 and 1 mM EDTA, 10 mM Mg-acetate, 10 mM $MnCl_2$, 1 mM Ca-acetate, 0.1 mM $ZnSO_4$, 0.1 mM $NiCl_2$, 1 mM or $CuSO_4$. Triangles and numbers denote the reaction products and their sizes, respectively. M – RNA Decade marker.

(F) S3/1 RNA cleavage pattern of the heterogeneous Csm-complex. To express and to purify heterogeneous Csm complex the wt CRISPR2 region containing 13 spacers of *S. thermophilus* DGCC8004 was cloned into the pACYC-Duet-1 vector. The heterogeneous Csm-72 complex was expressed and purified following the same procedure described for the homogenous Csm complex targeting the S3 protospacer. Specific S3/1 RNA substrate and non-specific NS RNA were pre-incubated with heterogeneous Csm-72. Cleavage reactions were performed at 37°C for indicated time intervals in the Reaction buffer supplemented with 10 mM Mg-acetate, 20 nM RNA substrate and 350 nM heterogeneous Csm-72. Samples were analyzed by denaturing PAGE, followed by phosphorimaging. Triangles and numbers denote the reaction products and their sizes, respectively. M – RNA Decade marker.

Figure S5 (related to Figure 2).



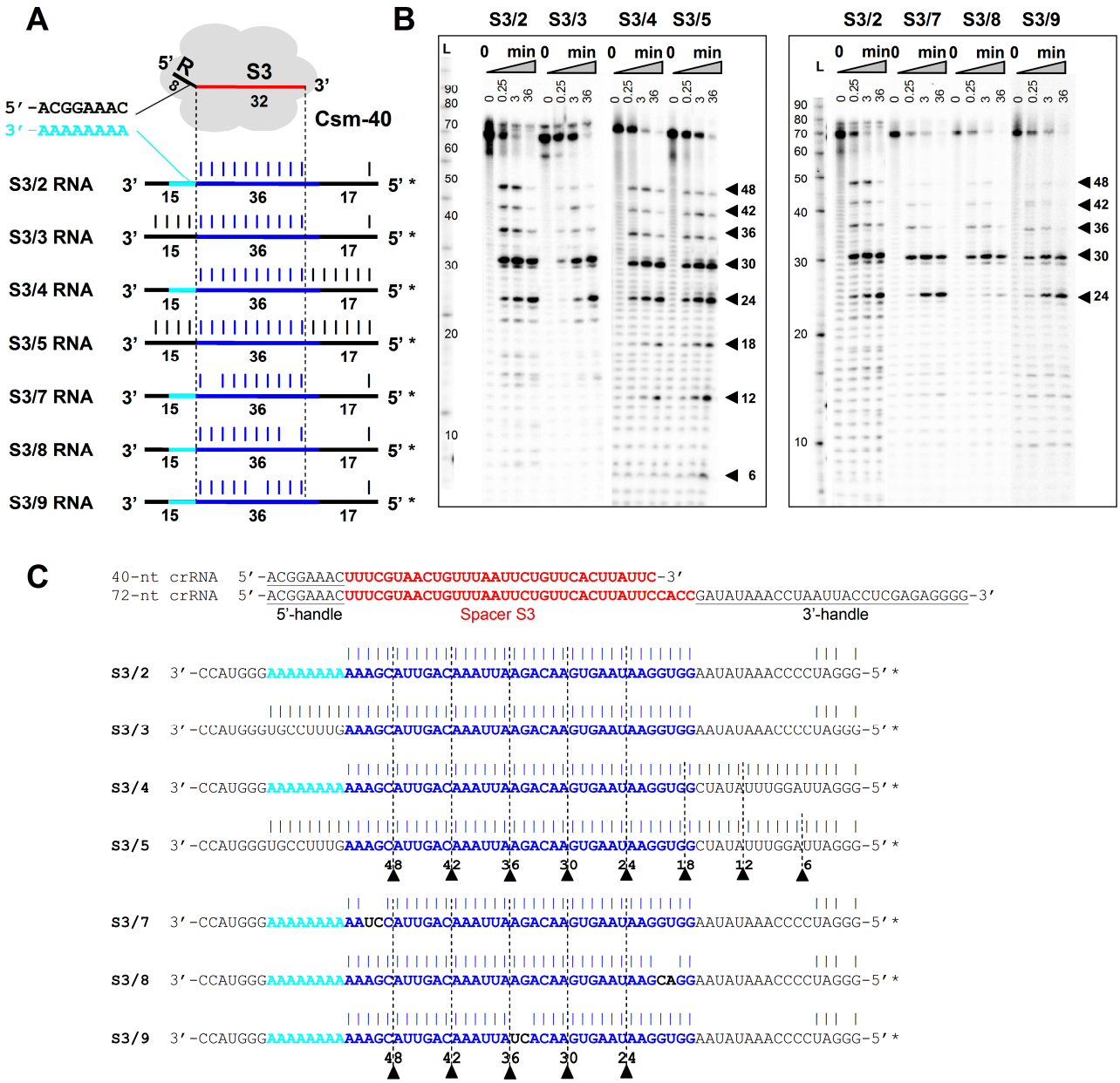
Reprogramming of the StCsm complex to cut a desired RNA target.

(A) Schematic representation (+Tc) and (-Tc) RNA substrates used in the cleavage assay. Arrow indicate Tc^R gene promoter. RNA substrates were 5′-end labeled with ^{32}P (indicated as *) and gel purified. The Tc (tetracycline resistance protein) gene transcript or RNA corresponding to the non-coding strand of Tc in the pBR322 plasmid (nt 851-886) were used as RNA targets. To reprogram a Csm complex a synthetic CRISPR locus containing five 36-nt length repeats interspaced by four identical 36-nt spacers complementary to the sense or antisense DNA strands of the Tc gene were engineered into the pACYC-Duet-1 plasmids which were independently co-expressed in *E. coli* together with plasmids pCsm/Cas and pCsmX-Tag. Csm-40 and Csm-72 complexes reprogrammed for the sence (+Tc) RNA or anti-sense (-Tc) RNA fragments were isolated similar to StCsm bearing spacer S3.

(B) Cleavage reactions were performed at 37°C for 120 min in the Reaction buffer supplemented with 10 mM Mg-acetate, 20 nM gel purified RNA substrate and 40-120 nM of Csm complex. Samples were analyzed by denaturing PAGE, followed by phosphorimaging. Triangles with corresponding numbers indicate cleavage product length. M – RNA Decade marker.

(C) RNA cleavage products mapped on the (+Tc) and (-Tc) RNA substrates sequences. The sequences of reprogrammed 40-nt and 72-nt length crRNAs are depicted above the substrates. Short vertical lines above the sequence indicate nucleotides complementary to crRNA. Triangles and dashed lines indicate cleavage positions. Translated fragment which corresponds to the tetracycline resistance protein gene RNA transcript is indicated under (+Tc) RNA substrate.

Figure S6 (related to Figure 3).



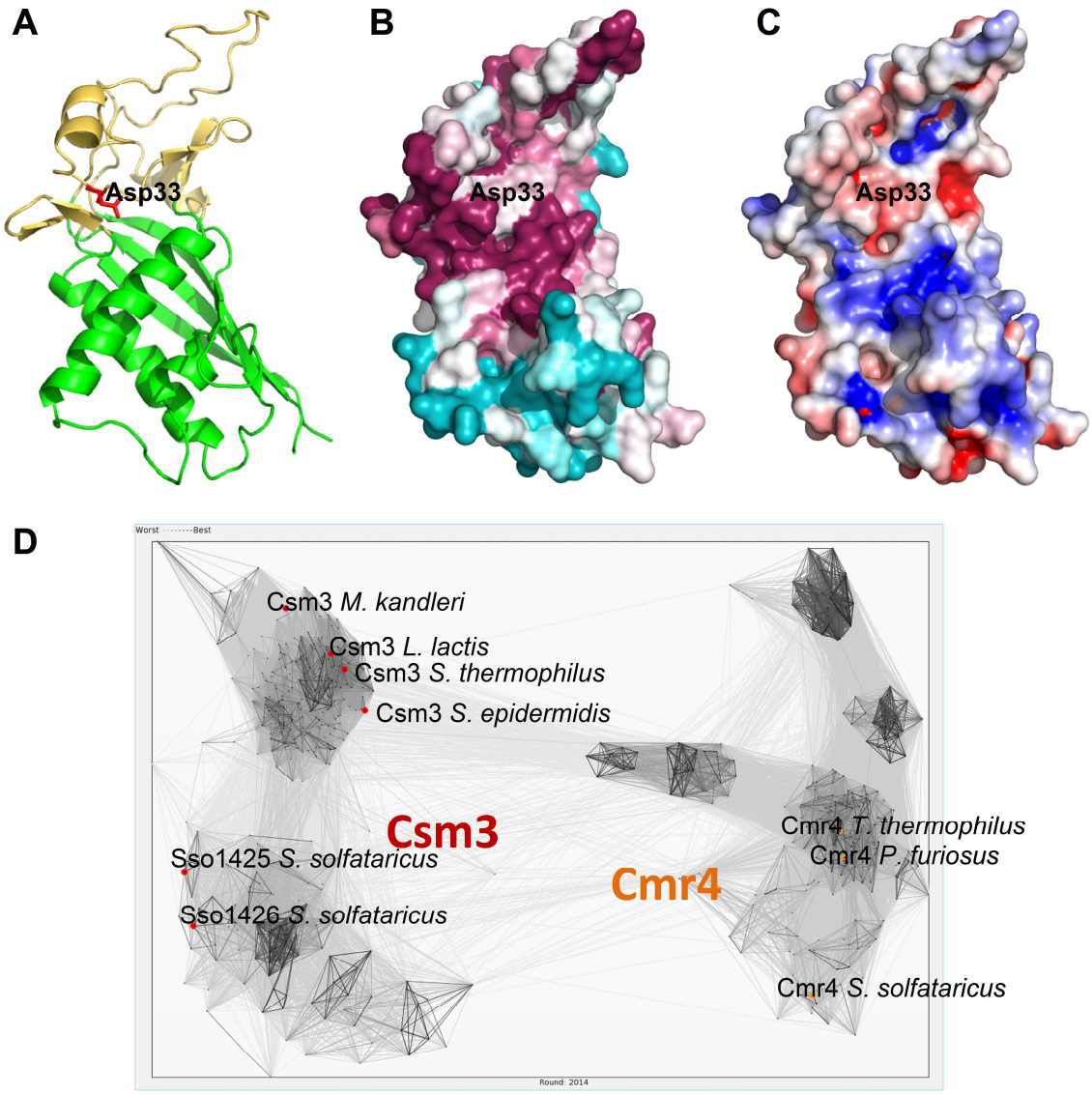
The role crRNA: target RNA complementarity on the StCsm-40 cleavage pattern.

(A) Schematic representation of the StCsm-40 complex and RNA substrates used in the cleavage assay. RNA substrates were 5′ -end labeled with ^{32}P (indicated as *) and gel purified.

(B) Cleavage reactions were performed at 25°C for indicated time intervals in the Reaction buffer supplemented with 10 mM Mg-acetate, 20 nM RNA substrate and 62.5 nM Csm-40. Samples were analyzed by denaturing PAGE, followed by phosphorimaging. Triangles with corresponding numbers indicate cleavage product length. M – RNA Decade marker.

(C) RNA cleavage products mapped on RNA substrates sequences. Short vertical lines above the sequence indicate nucleotides complementary to crRNA. Triangles and dashed lines indicate cleavage positions. The sequences of both 40-nt and 72-nt length crRNAs containing spacer S3 present in Csm-40 preparation are depicted above the substrates.

Figure S7 (related to Figure 5).



Computational analysis of Csm3.

(A-C) A structural model of StCsm3 in different representations. (A) Cartoon representation with the core RRM region shown in green and the „lid“ domain shown in yellow. Active site residue D33 is shown in red. (B) Molecular surface of the Csm3 model colored according to sequence conservation (maroon – conserved, cyan – variable). (C) Molecular surface of the Csm3 model colored according to electrostatic potential (blue – positive, red – negative).

(D) Clustering of 604 Csm3 and Cmr4 sequence homologs with CLANS. Representatives of Csm3 and Cmr4 families from experimentally characterized (Hale et al., 2009; Hatoum-Aslan et al., 2013; Hrle et al., 2013; Millen et al., 2012; Rouillon et al., 2013; Staals et al., 2013; Zhang et al., 2012) Type III CRISPR-Cas systems are labeled. Each dot represents a sequence, connecting lines represent the similarity between sequences. Thicker lines and shorted distances indicate higher sequence similarity. Only connections corresponding to P-values of 1e-12 or better are shown.

Table S1 (related to Figure 1). Proteins identified following mass spectrometry analysis of StCsm-72.

Protein	Mass (Da)	Score	Coverage (%)	Peptides
Cas10	86891	1076	36	LAYYLTR GDYAAIATR VYINQFASDK TVETLVQFEK YFKPTVLNLK YHMANYQSDK HNYKEDLFTK LYVAFGWGSFAAK DSISLFSSDYTFK DIMSELNSPESYR IDLFYGALLHDIGK DFNQFLLANFQTR FITNVYDDKLEQIR EKIDLFYGALLHDIGK GNEKDSISLFSSDYTFK IWDTYTNQADIFNVFGAQTDK SKPNFASATYEPFSKGDYAAIATR IWDTYTNQADIFNVFGAQTDKR HALVGADWFDEIADNQVISDQIR
Csm3	24541	768	46	ITAEANPR FENTIDR TLNELLTAEV ATTVFGNYDVK LLELDYLGGSGSR LKATTVFGNYDVK VAEKPSDDSDILSR DPITNLPIIPGSSLK SYTEVKFENTIDR DAFLSNADELDSLGVR FENTIDRITAEANPR
Csm4	33727	584	33	KQDLYK IFSALVLESLK DGNLYQVATTR HDQIDQSVDVK SSGFAFSHATNENYR FELDIQNIPLELSDR NQPHKDGNLYQVATTR LYIMTFQNAHFGSGTLDSSK
Cas6	28240	197	16	LVFTFK LIFQSLMQK RIDHPAQDLAVK SQGSYVIFPSMR
Csm2	14817	186	21	AQILEALK VQFVYQAGR YMEALVAYFK
Csm5	41013	138	12	LISFLNDNR NHESFYEMGK DAFGNPYIPGSSLK

Table S2 (related to Figure 1). Proteins identified following mass spectrometry analysis of StCsm-40.

Protein	Mass (Da)	Score	Coverage (%)	Peptides
Cas10	86891	1149	30	LAYYLTR GDYAAIATR VYINQFASDK YFKPTVLNLK YFFNHQDER YHMANYQSDK HNYKEDLFTK LYVAFGWGSFAAK DSISLFSSDYTFK DIMSELNSPESYR IDLFYGALLHDIGK DFNQFLLANFQTR FITNVYDDKLEQIR EKIDLFYGALLHDIGK GNEKDSISLFSSDYTFK IWDTYTNQADIFNVFGAQTDK SKPNFASATYEPFSKGDYAAIATR IWDTYTNQADIFNVFGAQTDKR HALVGADWFDEIADNQVISDQIR
Csm3	24541	801	57	ITAEANPR FENTIDR TLNELLTAEV ATTVFGNYDVK LLELDYLGGSGSR LKATTVFGNYDVK VAEKPSDDSDILSR DPITNLPIIPGSSLK SYTEVKFENTIDR DAFLSNADELDSLGVR FENTIDRITAEANPR NSTFDFELIYEITDENENQVEEDFK
Csm4	33727	554	33	KQDLYK IFSALVLESLK DGNLYQVATTR HDQIDQSVDVK SSGFAFSHATNENYR FELDIQNIPLELSDR FELDIQNIPLELSDRLTK NQPHKDGNLYQVATTR SSGFGEFELDIQNIPLELSDR
Cas6	28240	171	16	LVFTFK LIFQSLMQK RIDHPAQDLAVK SQGSYVIFPSMR
Csm2	14817	110		AQILEALK VQFVYQAGR
Csm5	41013	965	50	WDYSAK QADGILQR EFIYENK FYFPDMGK TILMNTTPK KFYFPDMGK VSDSKPFDNK LISFLNDNR NHESFYEMGK EYDDLFNAIR WNNENAVNDFGR GKEYDDLFNAIR KGKEYDDLFNAIR IEFEITTTTDEAGR LSLLTLAPIHIGNGEK DAFGNPYIPGSSLK LAEKFEAFLIQTRPNAR

Table S3 (related to Figures 1). M_W estimations for StCsm-40 and StCsm-72 by different methods.

	SDS-PAGE, kDa *	DLS, kDa **	MoW server, kDa ***	Porod volume, kDa ***	DAMMIN models, kDa ****
Csm-40	344.8	305±75	302±9	282±15	347.5
Csm-72	486.2	523±128	425±15	350±9	465.6

* Molecular mass calculated from evaluation of the complex composition by densitometric analysis of the SDS-PAGE gels.

** Molecular mass calculated from dynamic light scattering (DLS) analysis.

*** Molecular mass calculated from the SAXS data by the method described in (Fischer et al., 2010) using the SAXS MoW program run on the server http://www.if.sc.usp.br/~saxs/saxsmow.html.

**** Molecular mass was estimated using the Porod volumes calculated from SAXS data and excluded volumes of DAMMIN models as described in (Petoukhov et al., 2012).

Table S4 (related to Figure 1). SAXS data collection details and structural parameters of StCsm-40 and StCsm-72 complexes*.

Data collection parameters						
Beam line	P12					
Wavelength, nm	0.124					
Sample to detector distance, m	3.1					
Detector	Pilatus 2M					
s range, nm^{-1}	0.075786 - 4.665330					
exposure time of each frame, s	0.05					
Frames collected	20					
Sample storage temperature, °C	10					
Cell temperature, °C	20					
Structural parameters						
	Csm-40			**Csm-72**		
Sample concentrations, mg/ml	0.13	0.52	1.34	0.20	0.65	2.00
Guinier range (first-last point) as calculated by AUTORG	14-53	26-55	19-52	8-35	21-39	11-34
P(r) calculation range, $Å^{-1}$	0.0114-0.2006	0.0114-0.2006	0.0117-0.1739	0.0089-0.1076	0.0108-0.1076	0.0084-0.1049
Real space Rg, calculated by GNOM, Å	63.59 ±0.414	62.80 ±0.329	63.20 ±0.163	83.82 ±0.545	81.40 ±0.333	83.14 ±0.287
Real space Rg calculated by DATGNOM, Å	64.02	62.35	63.26	84.15	81.69	84.51
Reciprocal space Rg calculated by DATGNOM, Å	68.08	58.04	61.34	81.51	79.71	83.79
D_{max} as parameter for GNOM, Å	210	208	215	275	265	280
D_{max} calculated by DATGNOM, Å	233.2	203.1	214.7	279.2	267.0	293.3
Porod volume estimated by DATPOROD, $Å^3$	452186	501468	485803	611618	589997	581121
Excluded volume of DAMMIN models, $Å^3$ (10 models averaged)			590770± 5209			791440± 11366

Table S5 (related to Figures 2, 3, 4, 5, and 6). Nucleic acid substrates used in this study*.

<table>
<tr><td>S3 crRNA in Csm-72</td><td>5'-handle Spacer S3 3'-handle
5′-ACGGAAACUUUCGUAACUGUUUAAUUCUGUUCACUUAUUCCACCGAUAUAAACCUAAUUACCUCGAGAGGGG-3′</td></tr>
<tr><td>S3 crRNA in Csm-40</td><td>5'-handle Spacer S3
5′-ACGGAAACUUUCGUAACUGUUUAAUUCUGUUCACUUAUUC-3′</td></tr>
</table>

<table>
<tr><th>Substrates</th><th>Length, nt</th><th>Sequence</th></tr>
<tr><td>S3/1 DNA/DNA</td><td>76/76</td><td>5′-GGAGCTCGCCACCTTTTGTTTTTCGTAACTGTTTAATTCTGTTCACTTATTCCACCTCCTCAATTTGCCGGCTTGG-3′
3′-CCTCGAGCGG<u>TGGAAAACAA</u>AAAGCATTGACAAATTAAGACAAGTGAATAAGGTGG<u>AGGAGTTAAA</u>CGGCCGAACC-5′
O1205 O1205</td></tr>
<tr><td>NS DNA/DNA</td><td>73/73</td><td>5′-GACCACCCTTTTTGATATAATATACCTATATCAATGGCCTCCCACGCATAAGCGCAGATACGTTCTGAGGGAA-3′
3′-CTGGTGGGAAAAACTATATTATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′</td></tr>
<tr><td>S3/2 DNA/DNA</td><td>52/52</td><td>5′-TTTTTTTTTTTCGTAACTGTTTAATTCTGTTCACTTATTCCACCTTATATTT-3′
3′-AAAAAAAAAAAGCATTGACAAATTAAGACAAGTGAATAAGGTGGAATATAAA-5′</td></tr>
<tr><td>S3/1 DNA/RNA</td><td>76/68</td><td>5′-GGAGCTCGCCACCTTTTGTTTTTCGTAACTGTTTAATTCTGTTCACTTATTCCACCTCCTCAATTTGCCGGCTTGG-3′
3′-AGCGG<u>UGGAAAACAA</u>AAAGCAUUGACAAAUUAAGACAAGUGAAUAAGGUGG<u>AGGAGUUAAA</u>CGGCGGG-5′
O1205 O1205</td></tr>
<tr><td>S3/2 DNA/RNA</td><td>52/68</td><td>5′-TTTTTTTTTTTCGTAACTGTTTAATTCTGTTCACTTATTCCACCTTATATTT-3′
3′-CCAUGGGAAAAAAAAAAAGCAUUGACAAAUUAAGACAAGUGAAUAAGGUGGAAUAUAAACCCCUAGGG-5′</td></tr>
<tr><td>S3/1 DNA</td><td>76</td><td>|||||||||||||||||||||||||||||||||||| | |||
3′-CCTCGAGCGG<u>TGGAAAACAA</u>AAAGCATTGACAAATTAAGACAAGTGAATAAGGTGGAGGAGTTAAACGGCCGAACC-5′
O1205 O1205</td></tr>
<tr><td>NS DNA</td><td>73</td><td>3′-CTGGTGGGAAAAACTATATTATATGGATATAGTTACCGGAGGGTGCGTATTCGCGTCTATGCAAGACTCCCTT-5′</td></tr>
<tr><td>S3/2 DNA</td><td>52</td><td>||||||||||||||||||||||||||||||||||||
3′-AAAAAAAAAAAGCATTGACAAATTAAGACAAGTGAATAAGGTGGAATATAAA-5′</td></tr>
<tr><td>S3/1 RNA</td><td>68</td><td>|||||||||||||||||||||||||||||||||||| | ||| | |
3′-AGCGG<u>UGGAAAACAA</u>AAAGCAUUGACAAAUUAAGACAAGUGAAUAAGGUGG<u>AGGAGUUAAA</u>CGGCGGG-5′
O1205 O1205</td></tr>
<tr><td>NS RNA</td><td>68</td><td>| | | || | | || | | | | ||| | |
3′-AGCGG<u>UGGAAAACAA</u>CCACCUUAUUCACUUGUCUUAAUUUGUCAAUGCUUU<u>AGGAGUUAAA</u>CGGCGGG-5′
O1205 O1205</td></tr>
<tr><td>S3/2 RNA</td><td>68</td><td>|||||||||||||||||||||||||||||||||||| ||| |
3′-CCAUGGGAAAAAAAAAAAGCAUUGACAAAUUAAGACAAGUGAAUAAGGUGGAAUAUAAACCCCUAGGG-5′</td></tr>
<tr><td>S3/3 RNA</td><td>68</td><td>|||||||||||||||||||||||||||||||||||||||||||| ||| |
3′-CCAUGGGUGCCUUUGAAAGCAUUGACAAAUUAAGACAAGUGAAUAAGGUGGAAUAUAAACCCCUAGGG-5′</td></tr>
<tr><td>S3/4 RNA</td><td>68</td><td>|||||||||||||||||||||||||||||||||||||||||||||||||| |
3′-CCAUGGGAAAAAAAAAAAGCAUUGACAAAUUAAGACAAGUGAAUAAGGUGGCUAUAUUUGGAUUAGGG-5′</td></tr>
<tr><td>S3/5 RNA</td><td>68</td><td>|||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
3′-CCAUGGGUGCCUUUGAAAGCAUUGACAAAUUAAGACAAGUGAAUAAGGUGGCUAUAUUUGGAUUAGGG-5′</td></tr>
</table>

Substrates	Length, nt	Sequence
S3/6 RNA	86	3′-CCAUGGGUGCCUUUG**AAAGCAUUGACAAAUUAAGACAAGUGAAUAAGGUGG**CUAUAUUUGGAUUAAUGGAGCUCUCCCCUGCCGGG-5′
S3/7 RNA	68	3′-CCAUGGGAAAAAAAA**AAUCCAUUGACAAAUUAAGACAAGUGAAUAAGGUGG**AAUAUAAACCCCUAGGG-5′
S3/8 RNA	68	3′-CCAUGGGAAAAAAAA**AAAGCAUUGACAAAUUAAGACAAGUGAAUAAGCAGG**AAUAUAAACCCCUAGGG-5′
S3/9 RNA	68	3′-CCAUGGGAAAAAAAA**AAAGCAUUGACAAAUUAUCACAAGUGAAUAAGGUGG**AAUAUAAACCCCUAGGG-5′
S3/11 RNA	24	3′-**GACAAAUUAAGACAAGUGAAUGGG**-5′
S3/12 RNA	32	3′-**AAAGCAUUGACAAAUUAAGACAAGUGAAUGGG**-5′
S3/13 RNA	23	3′-**AAAGCAUUGACAAAUUAAGAGGG**-5′
S3/14 RNA	48	3′-CCAUGGGAAAAAAAA**AAAGCAUUGACAAAUUAAGACAAGUGAAUAAGG**-5′

crRNA (+Tc) in Csm-72

5'-handle Spacer(+Tc) 3'-handle
5′-ACGGAAAC**ACGCCAGCAAGACGUAGCCCAGCGCGUCGGCCGCCA**GAUAUAAACCUAAUUACCUCGAGAGGGG-3′

crRNA (+Tc) in Csm-40

5'-handle Spacer(+Tc)
5′-ACGGAAAC**ACGCCAGCAAGACGUAGCCCAGCGCGUCGGCC**-3′

Substrate	Length, nt	Sequence
RNA (+Tc)	68	3′-UCGGAGCGCAGCGCU**UGCGGUCGUUCUGCAUCGGGUCGCGCAGCCGGCGGU**ACGGCCGCUAUUACGGG-5′

crRNA (-Tc) in Csm-72

5'-handle Spacer(-Tc) 3'-handle
5′-ACGGAAAC**UGGCGGCCGACGCGCUGGGCUACGUCUUGCUGGCGU**GAUAUAAACCUAAUUACCUCGAGAGGGG-3′

crRNA (-Tc) in Csm-40

5'-handle Spacer(-Tc)
5′-ACGGAAAC**UGGCGGCCGACGCGCUGGGCUACGUCUUGCUG**-3′

Substrate	Length, nt	Sequence
RNA (-Tc)	68	3′-GGUAAUAGCGGCCGU**ACCGCCGGCUGCGCGACCCGAUGCAGAACGACCGCA**AGCGCUGCGCUCCGGGG-5′

Rep crRNA in Csm-72

```
5'-handle           Spacer Rep                              3'-handle
5'-ACGGAAACCAACGAGCCUAAAUUCAUAUGACUCGUUAUAGCGGAGAUAUAAACCUAAUUACCUCGAGAGGGG-3'
```

Rep crRNA in Csm-40

```
5'-handle           Spacer Rep
5'-ACGGAAACCAACGAGCCUAAAUUCAUAUGACUCGUUAUAG-3'
```

Substrate	Length, nt	Sequence
Rep RNA	72	3'-CAAGGGAUGUUGCUCGGAUUUAAGUAUACUGAGCAAUAUCGCCUGGCGCACAGACUAGGUGCCGCGUGUAAC-5'

Lys crRNA in Csm-72

```
5'-handle           Spacer Lys                              3'-handle
5'-ACGGAAACUGUCUUCGACAUGGGUAAUCCUCAUGUUUGAAUGGCGAUAUAAACCUAAUUACCUCGAGAGGGG-3'
```

Lys crRNA in Csm-40

```
5'-handle           Spacer Lys
5'-ACGGAAACUGUCUUCGACAUGGGUAAUCCUCAUGUUUGAA-3'
```

Substrate	Length, nt	Sequence
Lys RNA	72	3'-AAGAAACAACAGAAGCUGUACCCAUUAGGAGUACAAACUUACCGGCCGCAGAUAAUCAUCUACGGCCUCAAA-5'

Cp crRNA in Csm-72

```
5'-handle           Spacer Cp                             3'-handle
5'-ACGGAAACUCUUUUAGGAGACCUUGCAUUGCCUUAACAAUAAGCGAUAUAAACCUAAUUACCUCGAGAGGGG-3'
```

Cp crRNA in Csm-40

```
5'-handle           Spacer Cp
5'-ACGGAAACUCUUUUAGGAGACCUUGCAUUGCCUUAACAAU-3'
```

Substrate	Length, nt	Sequence
Cp RNA	72	3'-CCAAAGGUAGAAAAUCCUCUGGAACGUAACGGAAUUGUUAUUCGAGCGUCAGCCUUAAGCAUCGCUUUUAAC-5'

Mat crRNA in Csm-72

```
5'-handle           Spacer Mat                              3'-handle
5'-ACGGAAACAGUUUGCAGCUGGAUACGACAGACGGCCAUCUAACUGAUAUAAACCUAAUUACCUCGAGAGGGG-3'
```

Mat crRNA in Csm-40

```
5'-handle           Spacer Mat
5'-ACGGAAACAGUUUGCAGCUGGAUACGACAGACGGCCAU-3'
```

Substrate	Length, nt	Sequence
Mat RNA	72	3'-ACAGACCUUCAAACGUCGACCUAUGCUGUCUGCCGGUAGAUUGAACUACAAUCAUGGCUGGACUGCAUGCCG-5'

* Above each Table crRNAs in Csm-72 and Csm-40 are depicted for clarity. Bold lettering in crRNAs represents the spacer (guide) sequence. Non-bold regions in crRNAs is for repeat sequences. Designed 72 and 40 nt crRNAs (+Tc) are complementary to tetracycline resistance gene (Tc) transcript and are guided to cleave RNA(+Tc) substrate (sence RNA or Tc transcript). Similarly, designed 72 and 40 nt crRNAs (-Tc) are guided to cleave RNA (-Tc) (antisence RNA corresponding the non-coding strand of Tc gene) substrate. Designed Rep, Lys, Cp and Mat 72 and 40 nt crRNAs are guided to cleave ss RNA coliphage MS2 rep, lys, cp and mat transcripts, respectively.

DNA and RNA substrates used in this study are presented in the Tables. Bold lettering in substrates represents the sequence complementary to spacer (guide) of crRNA. For single stranded DNA and RNA substrates nucleotides complementary to corresponding nucleotide in crRNA are depicted by dashes. Nucleotides marked in yellow were incorporated into RNA during *in vitro* transcription. Rep, Lys, Cp and Mat RNA are RNA sequences in MS2 genome.

Supplemental Experimental Procedures

Cloning, expression and purification of StCsm complexes

Streptococcus thermophilus DGCC8004 was cultivated at 42°C in M17 broth (Oxoid) supplemented with 0.5% (w/v) lactose. Chromosomal DNA was extracted and purified using GeneJET Genomic DNA Purification Kit (Thermo Scientific). CRISPR2-Cas region was amplified by polymerase chain reaction (PCR) and sequenced using primers designed by genomic comparison with *S. thermophilus* DGCC7710 (GenBank accession number AWVZ01000003). Annotation of the predicted ORFs was performed using BLASTP at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). CRISPR region was identified through repeat sequence similarity to that of *S. thermophilus* DGCC7710. Multiple sequence alignments of *cas/csm* genes, spacers and repeats sequences were carried out with ClustalW2 (http://www.ebi.ac.uk).

Genomic DNA isolated from *S. thermophilus* DGCC8004 strain was used as the template for PCR amplification of the *cas/csm* genes. DNA fragment covering the 8.5 kb *cas6-cas10-csm2-csm3-csm4-csm5-csm6-csm6'* gene cassette was cloned into pCDFDuet-1 expression vector via NcoI and AvrII restriction sites in two separate subcloning steps to generate plasmid pCas/Csm. Individual *cas/csm* genes were cloned into pETDuet-1_N-StrepII and pETDuet-1_C-StrepII expression vectors, except of *cas10* (which was cloned into pBAD24_C-His-StrepII-His) and *csm6* or *csm6'* (that were cloned into pBAD24_N-His-StrepII-His) to generate pCsmX-Tag and pCasY-Tag plasmids, where X=2,3,4,5,6,6' and Y=6,10. A synthetic 445-nt CRISPR locus containing five 36-nt length repeats interspaced by four identical 36-nt spacers S3 of the *S. thermophilus* DGCC8004 CRISPR2 system was obtained from Invitrogen and cloned into the pACYC-Duet-1 vector to generate a plasmid pCRISPR_S3. Four copies of the spacer S3 have been engineered into the pCRISPR_S3 plasmid to increase the yield of the Csm-crRNA complex. Full sequencing of cloned DNA fragments confirmed their identity to the original sequences.

All three plasmids were co-expressed in *Escherichia coli* BL21 (DE3) grown at 37°C in LB medium supplemented with streptomycin (25 µg/µl), ampicilin (50 µg/µl), and chloramphenicol (30 µg/µl). The fresh LB medium was inoculated with an overnight culture (1/20 (v/v)), and bacteria were grown to the mid-log phase (OD_{600nm} 0.5 to 0.7), then 1 mM IPTG (and 0.2% (w/v) L-(+)-arabinose in case of Cas10, Csm6 and Csm6') was added and cell suspension was further cultured for another 4 h. Harvested cells were resuspended in a Chromatography buffer (20 mM Tris-HCl (pH 8.5), 0.5 M NaCl, 7 mM 2-mercaptoethanol, 1 mM EDTA) supplemented with 0.1 mM phenylmethylsulfonyl fluoride (PMSF), and disrupted by sonication. Cell debris was removed by centrifugation. Csm complexes were captured on the StrepTrap affinity column (GE Healthcare) and further subjected to the Superdex 200 size exclusion chromatography (prep grade XK 16/60; GE Healthcare). SDS-PAGE of individual Strep-tagged Csm2, Csm3, Csm4, Csm5, Cas6 and

Cas10 proteins isolated by affinity chromatography from *E. coli* lysates revealed co-purification of other Csm/Cas proteins suggesting the presence of a Csm complex. The abundance of the Csm complex co-purified via the Csm4-, Csm5-, Cas6- and Cas10-Strep tagged subunits was very low, and no complex was pull-downed via Csm6 or Csm6' subunits (data not shown). Therefore, Csm complexes isolated via N-terminus Strep-tagged Csm2 (Csm2_StrepN) and the N-terminus Strep-tagged Csm3 proteins (Csm3_StrepN) were subjected to further characterization. Individual Csm3-N-Strep protein was purified using StrepTrap affinity column. Csm3-N-Strep and Csm complexes eluted from the columns were dialysed against 10 mM Tris-HCl (pH 8.5) buffer containing 300 mM NaCl, 1 mM DTT, 0.1 mM EDTA, and 50% (v/v) glycerol, and stored at -20°C.

The composition of the isolated Csm-40 and Csm-72 complexes was analysed by SDS–PAGE and the sequence of Csm proteins was further confirmed by the mass spectrometry of tryptic digests. In order to estimate the stoichiometry of Csm complexes, protein bands in SDS-PAGE were quantified by densitometric analysis taking a count the different staining of Cas/Csm proteins. The molecular weights of the Csm complexes were estimated by dynamic light scattering (DLS) using Zetasizer μV (Malvern) and respective software. For DLS analysis Csm-40 and Csm-72 samples were analysed in a Chromatography buffer at 0.36 mg/ml and 0.6 mg/ml concentrations, respectively. Csm complex concentrations were estimated by Pierce 660nm Protein Assay (Thermo Scientific) using bovine serum albumin (BSA) as a reference protein. Conversion to molar concentration was performed assuming that the Csm-72 stoichiometry is $Cas10_1$:$Csm2_6$:$Csm3_{10}$:$Csm4_1$:$crRNA72_1$ and the Csm-40 stoichiometry is $Cas10_1$:$Csm2_3$:$Csm3_5$:$Csm4_1$:$Csm5_1$:$crRNA40_1$.

HPLC purification of crRNA

All samples were analyzed by ion-pair reversed-phased-HPLC (Dickman and Hornby, 2006; Waghmare et al., 2009) on an Agilent 1100 HPLC with UV260nm detector (Agilent) using a DNAsep column 50 mm x 4.6 mm I. D. (Transgenomic). The chromatographic analysis was performed using the following buffer conditions: A) 0.1 M triethylammonium acetate (TEAA) (pH 7.0) (Fluka); B) buffer A with 25% LC MS grade acetonitrile (v/v) (Fisher). The crRNA was obtained by injecting purified intact Csm-40 or Csm-72 at 75°C using a linear gradient starting at 15% buffer B and extending to 60% B in 12.5 min, followed by a linear extension to 100% B over 2 min at a flow rate of 1.0 ml/min. Analysis of the 3' terminus was performed by incubating the HPLC-purified crRNA in a final concentration of 0.1 M HCl at 4°C for 1 hour. The samples were concentrated to 10-20 μl on a vacuum concentrator (Eppendorf) prior to ESI-MS analysis.

ESI-MS analysis of crRNA

Electrospray Ionization Mass spectrometry (ESI-MS) was performed in negative mode using an Amazon Ion Trap mass spectrometer (Bruker Daltonics), coupled to an online capillary liquid

chromatography system (Ultimate 3000, Dionex, UK). RNA separations were performed using a monolithic (PS-DVB) capillary column (50 mm × 0.2 mm I.D., Dionex, UK). The chromatography was performed using the following buffer conditions: C) 0.4 M 1,1,1,3,3,3,-Hexafluoro-2-propanol (HFIP, Sigma- Aldrich) adjusted with triethylamine (TEA) to pH 7.0 and 0.1 mM TEAA, and D) buffer C with 50% methanol (v/v) (Fisher). RNA analysis was performed at 50°C with 20% buffer D, extending to 40% D in 5 min followed by a linear extension to 60% D over 8 min at a flow rate of 2 µl/min, 250 ng crRNA was digested with 1U RNase A/T1 (Applied Biosystems). The reaction was incubated at 37^{o}C for 4 h. The oligoribonucleotide mixture was separated on a PepMap C-18 RP capillary column (150 mm × 0.3 µm I.D., Dionex, UK) at 50°C using gradient conditions starting at 20% buffer C and extending to 35% D in 3 mins, followed by a linear extension to 60% D over 40 mins at a flow rate of 2 µl /min. The mass spectrometer was operated in negative mode, a capillary voltage was set at -2500 V to maintain capillary current between 30−50 nA , temperature of nitrogen 120°C at a flow rate of 4.0 L/h and N_2 nebuliser gas pressure at 0.4 bar. A mass range of 500-2500 m/z was set. Oligoribonucleotides with -2 to -4 charge states were selected for tandem mass spectrometry using collision induced dissociation.

SAXS experiments

Ab initio shape modeling of both complexes was performed with the samples having highest concentration (1.3 mg/ml for Csm-40 and 2.0 mg/ml for Csm-72). Unprocessed scattering data with subtracted buffer scattering, Guinier plots of the low s region of the scattering curves used for the shape determination and P(r) functions of the highest concentration samples of Csm-40 and Csm-72 are presented in Figure S3. Two-dimensional scattering curves were transformed and distance distribution functions P(r) were calculated using GNOM (Svergun, 1992). At this stage data were truncated to s values 0.15-0.1 Å^{-1} and calculated distance distribution function was used for following *ab initio* modeling. 10 independent bead models for both complexes were generated using DAMMIN (Svergun, 1999). These models were aligned, filtered and averaged based on occupancy using DAMAVER (Volkov and Svergun, 2003). The averaged NSD of superposition of DAMMIN models of Csm-40 complex was 0.563±0.028 (for Csm-72 models averaged NSD is 0.575±0.019), no model was rejected in both cases.

The inertia tensor was calculated for averaged models of both complexes and models were aligned along the largest principal axis so as the end points of both models coincided. After that the protruding part of the longer Csm-72 complex was truncated. Csm-40 model was aligned with truncated Csm-72 models by automatic procedure SUPCOMB (Kozin and Svergun, 2001) producing an NSD value. Then Csm-40 model was shifted along the principal axis of Csm-72 model by the fixed step (5 or 10 Å) and again Csm-40 model was aligned by SUPCOMB with the Csm-72 model after truncation of protruding parts. Thus the Csm-40 model was sequentially shifted

along the principal axis of Csm-72 model and the best superposition showed the lower NSD value (S. Grazulis, personal communication). MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Bacon, 1997) programs were used for SAXS models presented in Figure 1 and S3 preparation.

Computational sequence and structure analysis

Sequence searches were performed with PSI-BLAST (Altschul et al., 1997) against the nr80 sequence database (the NCBI 'nr' database filtered to 80% identity) using E-value=1e-03 or a more stringent inclusion threshold. Clustering of homologous sequences according to their mutual similarity was done using CLANS (Frickey and Lupas, 2004). Multiple sequence alignments were constructed with MAFFT (Katoh et al., 2002) using the accuracy-oriented mode (L-INS-i). Homology model for StCsm3 was constructed with HHpred (Söding et al., 2005) using the related structure of *M. kandleri* Csm3 (PDB code 4N0L) as a template. The analysis of surface residue conservation was performed using the ConSurf server (Ashkenazy et al., 2010). Electrostatic map of the structure surface was calculated with the APBS (Baker et al., 2001) plugin in PyMol (Schrodinger, 2010). Pictures were prepared with PyMol (Schrodinger, 2010).

Mutagenesis of Csm3

The Csm3 mutants H19A, D33A, D100A, E119A, E123A and E139A were obtained by the Quick Change Mutagenesis (QCM) Protocol (Zheng et al., 2004). First, a 3.0 kb DNA fragment containing *csm2* and *csm3* genes was subcloned from pCas/Csm plasmid into the pUC18 vector pre-cleaved with SphI and KpnI. The resulting plasmid pUC18_Csm2_Csm3 was used for Csm3 QCM mutagenesis. After QCM, the same fragment containing mutated versions of the Csm3 gene was transferred back into the pCas/Csm vector using NdeI and SpeI sites, reconstituting the gene cassette. Sequencing of the entire cloned DNA fragment for each mutant confirmed that only the designed mutation had been introduced. Csm-40 complexes containing Csm3 mutants were isolated following the procedures described for the wt StCsm-40 (see above). D100A mutant StCsm-40 was purified only by the affinity chromatography.

Preparation of DNA and RNA substrates

To assemble DNA oligoduplexes, complementary oligodeoxynucleotides were mixed at 1:1 molar ratio in the Reaction buffer (33 mM Tris–acetate (pH 7.9 at 25°C), 66 mM potassium acetate), heated to 90°C and slowly let to cool to room temperature.

For generation of S3/1-10, S3/14 RNA substrates, first pUC18 plasmids pUC18_S3/1 and pUC18_S3/2, bearing S3/1 or S3/2 sequences were constructed. For this purpose, annealed synthetic DNA oligoduplexes S3/1 or S3/2 were ligated into pUC18 plasmid pre-cleaved with SmaI. Engineered plasmids pUC18_S3/1 and pUC18_S3/2 were sequenced to persuade that only

copy of DNA duplex was ligated into the vector. Further these plasmids were used as a template to produce different DNA fragments by PCR using appropriate primers containing a T7 promoter in front of the desired RNA sequence. Purified PCR products were used in the *in vitro* transcription reaction to obtain RNA substrates. S3/11-13 RNAs were prepared by hybridizing two complementary DNA oligonucleotides, containing a T7 promoter in front of the desired RNA sequence followed by *in vitro* transcription.

DNA/RNA hybrids were assembled in similar manner annealing complementary oligodeoxynucleotide to RNA obtained by *in vitro* transcription.

pBR322 plasmid bearing the Tc gene, encoding tetracycline (Tc) resistance protein, was used to produce Tc RNA and ncTc RNA substrates using the same *in vitro* transcription reaction as described above for S3/1-10, S3/14. Prior to ^{32}P labeling RNA substrates were dephosphorylated using FastAP thermosensitive alkaline phosphatase (Thermo Scientific).

Phage drop plaque assay

Phage drop plaque assay was conducted using LGC Standards recomendations. Briefly, *E. coli* NovaBlue (DE3) [(endA1 hsdR17(r $_{K12}$− m $_{K12}$+) supE44 thi-1 recA1 gyrA96 relA1 lac (DE3) F'[proA$^+$B$^+$ lacI q ZΔM15::Tn10] (Tet^R)] was trasformed with wt pCas/Csm (Str^R) or D33A Csm3 pCas/Csm (Str^R) and pCRISPR_MS2 (Cm^R), pCRISPR_S3 (Cm^R), or pACYC-Duet-1 (Cm^R). *E. coli* cells bearing different sets of plasmids were grown in LB medium with appropriate antibiotics at 37°C to an OD 600 of 0.9 and a 0.4 ml aliquot of bacterial culture was mix ed with melted 0.5% soft nutrient agar (45°C). This mixture was poured onto 1.5% solid agar to make double layer agar plates. Both layers of agar contained appropriate antibiotics, 0,1mM IPTG, 0,1% glucose, 2mM $CaCl_2$ and 0,01mg/ml thiamine. When the top agar hardened, phage stock (5 μl) from a dilution series was spotted on each plate with the bacteria. The plates were examined for cell lysis after overnight incubations at 37°C. NovaBlue (DE3) was used as the indicator for determination of the phage titer. pCRISPR_MS2 plasmid bearing the synthetic CRISPR array of five repeats interspaced by four 36-nt spacers targeting the mat, lys, cp, and rep MS2 RNA sequences (GenBank accession number NC001417) was constructed similar to pCRISPR_S3 (see above).

Supplemental References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res *38*, W529-533.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A *98*, 10037-10041.

Dickman, M.J., and Hornby, D.P. (2006). Enrichment and analysis of RNA centered on ion pair reverse phase methodology. RNA *12*, 691-696.

Fischer, H., de Oliveira Neto, M., Napolitano, H., Polikarpov, I., and Craievich, A. (2010). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. J Appl Crystallogr *43*, 101-109.

Frickey, T., and Lupas, A. (2004). CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics *20*, 3702-3704.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945-956.

Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W., and Marraffini, L.A. (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. The Journal of biological chemistry *288*, 27888-27897.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science *327*, 167-170.

Hrle, A., Su, A.A., Ebert, J., Benda, C., Randau, L., and Conti, E. (2013). Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3. RNA biology *10*, 1670-1678.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res *30*, 3059-3066.

Kozin, M.B., and Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. J Appl Crystallogr *33*, 33-41.

Kraulis, P.J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr *24*, 945-950.

Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., *et al.* (2006). Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci U S A *103*, 15611-15616.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843-1845.

Merritt, E.A., and Bacon, D.J. (1997). Raster3D: Photorealistic molecular graphics. Methn Enzymol *277*, 505-524.

Millen, A.M., Horvath, P., Boyaval, P., and Romero, D.A. (2012). Mobile CRISPR/Cas-mediated bacteriophage resistance in Lactococcus lactis. PLoS One *7*, e51663.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr *45*, 342-350.

Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C.V., Spagnolo, L., and White, M.F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. Mol Cell *52*, 124-134.

Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.3r1.

Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res *33*, W244-248.

Staals, R.H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K., *et al.* (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. Mol Cell *52*, 135-145.
Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr *25*, 495-503
Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophysical journal *2879-2886*.
Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of it ab initio shape determination in small-angle scattering. J Appl Crystallogr *36*, 860-864.
Waghmare, S.P., Pousinis, P., Hornby, D.P., and Dickman, M.J. (2009). Studying the mechanism of RNA separations using RNA chromatography and its application in the analysis of ribosomal RNA and RNA:RNA interactions. Journal of chromatography. A *1216*, 1377-1382.
Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.V., *et al.* (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Mol Cell *45*, 303-313.
Zheng, L., Baumann, U., and Reymond, J.L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res *32*, e115.