CellPress

Molecular Cell
Article

RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of *Thermus thermophilus*

Raymond H.J. Staals,[1,13,*] Yifan Zhu,[1,13] David W. Taylor,[2,13] Jack E. Kornfeld,[2] Kundan Sharma,[3] Arjan Barendregt,[4] Jasper J. Koehorst,[5] Marnix Vlot,[1] Nirajan Neupane,[1] Koen Varossieau,[1] Keiko Sakamoto,[6] Takehiro Suzuki,[7] Naoshi Dohmae,[7] Shigeyuki Yokoyama,[8] Peter J. Schaap,[5] Henning Urlaub,[3,9] Albert J.R. Heck,[4] Eva Nogales,[2,10,11] Jennifer A. Doudna,[2,10,12] Akeo Shinkai,[6,8] and John van der Oost[1,*]

[1]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, 6703 HB Wageningen, the Netherlands
[2]Howard Hughes Medical Institute and California Institute for Quantitative Biosciences, University of California, Berkeley, Berkeley, CA 94720-3200, USA
[3]Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany
[4]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, 3584 CH Utrecht, the Netherlands
[5]Laboratory of Systems and Synthetic Biology, Wageningen University, 6703 HB Wageningen, the Netherlands
[6]RIKEN SPring-8 Center, Hyogo 679-5148, Japan
[7]Global Research Cluster, RIKEN, Saitama 351-0198, Japan
[8]Structural Biology Laboratory, RIKEN, Kanagawa 230-0045, Japan
[9]Bioanalytics Research Group, Department of Clinical Chemistry, University Medical Center, 37075 Göttingen, Germany
[10]Department of Molecular and Cell Biology, Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720-3200, USA
[11]Life Sciences Division, Lawrence Berkeley National Lab, Berkeley, CA 94720-3200, USA
[12]Physical Biosciences Division, Lawrence Berkeley National Lab, Berkeley, CA 94720-3200, USA
[13]Co-first author
*Correspondence: raymond.staals@otago.ac.nz (R.H.J.S.), john.vanderoost@wur.nl (J.v.d.O.)
http://dx.doi.org/10.1016/j.molcel.2014.10.005

SUMMARY

CRISPR-Cas is a prokaryotic adaptive immune system that provides sequence-specific defense against foreign nucleic acids. Here we report the structure and function of the effector complex of the Type III-A CRISPR-Cas system of *Thermus thermophilus*: the Csm complex (TtCsm). TtCsm is composed of five different protein subunits (Csm1–Csm5) with an uneven stoichiometry and a single crRNA of variable size (35–53 nt). The TtCsm crRNA content is similar to the Type III-B Cmr complex, indicating that crRNAs are shared among different subtypes. A negative stain EM structure of the TtCsm complex exhibits the characteristic architecture of Type I and Type III CRISPR-associated ribonucleoprotein complexes. crRNA-protein crosslinking studies show extensive contacts between the Csm3 backbone and the bound crRNA. We show that, like TtCmr, TtCsm cleaves complementary target RNAs at multiple sites. Unlike Type I complexes, interference by TtCsm does not proceed via initial base pairing by a seed sequence.

INTRODUCTION

The arsenal of prokaryotic defense mechanisms against mobile genetics elements (MGE), such as bacteriophages and (conjugative) plasmids, includes adaptive immunity that serves as a sequence-specific memory of prior infections (Barrangou and Marraffini, 2014; Gasiunas et al., 2014; Reeks et al., 2013; Terns and Terns, 2014; van der Oost et al., 2014). These systems are made up of arrays of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (*cas*) genes that are present in approximately half of sequenced bacteria and most archaea (Grissa et al., 2007; Haft et al., 2005; Makarova et al., 2011). CRISPR-Cas systems are categorized into three major types (Types I, II, and III) on the basis of their specific Cas proteins (Koonin and Makarova, 2013; Makarova et al., 2011).

CRISPR arrays are short repeated sequences (24–50 bp) interspaced by similar-sized sequences with homology to MGE (spacers). The array is preceded by a leader sequence, which contains the promoter for transcription of the array. The spacers are acquired from the MGE and inserted in the chromosomal CRISPR array of the host by a process called ''acquisition'' that requires Cas1 and Cas2 proteins (Arslan et al., 2014; Barrangou et al., 2007; Yosef et al., 2012), although the requirement for these proteins in Type III systems has not been demonstrated so far. Transcription of the CRISPR array generally results in a long pre-crRNA. In Type I and Type III systems, this pre-crRNA is subsequently processed by a Cas6-type endoribonuclease into separate crRNAs, containing part(s) of the repeat and a single spacer sequence (Brouns et al., 2008; Carte et al., 2008). Type II systems use Cas9, the host factor RNase III, and a transactivating crRNA (tracrRNA) with complementarity to the repeat for crRNA maturation (Carte et al., 2014; Deltcheva

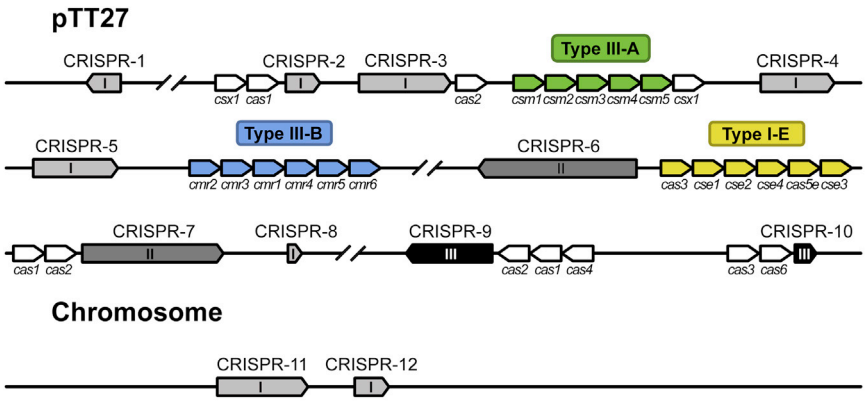


Figure 1. Schematic Representation of CRISPR Arrays and *cas* Genes on the Chromosome and Plasmid pTT27 of *T. thermophilus* HB8

CRISPR arrays (1–12) are indicated in different grayscales, depending on the repeat type (I, II, or III). *Cas*(-related) genes belonging to a particular CRISPR-Cas subtype are colored in green (subtype III-A), blue (subtype III-B), or yellow (subtype I-E). Additional *cas* genes are indicated in white. For each of these CRISPR arrays, the bottom panel summarized the genomic location, the consensus repeat sequence, repeat type, and the number of spacers. The 5′ tag sequences, as found by our RNA-seq analysis, are underlined. The figure and legend are adapted from Staals et al. (2013).

Locus	Genomic location	Consensus repeat sequence (5' to 3')	Repeat type	# Spacers
CRISPR-1	18303-18044	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	3
CRISPR-2	133766-133954	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	2
CRISPR-3	135156-136099	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	12
CRISPR-4	144129-144842	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	9
CRISPR-5	146042-146983	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	12
CRISPR-6	190947-189507	GTAGTCCCCACGCACGTGGGGATGGACCG	II	23
CRISPR-7	200811-202078	GTAGTCCCCACGCGTGTGGGGATGGACCG	II	20
CRISPR-8	210807-210842	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGATAC	I	0
CRISPR-9	228237-227324	GTTGCAAACCCCGTCAGCCTCGTAGAGGATTGAAAC	III	12
CRISPR-10	239108-239214	GTTGCAAACCTCGTTAGCCTCGTAGAGGATTGAAAC	III	1
CRISPR-11	872101-873199	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	14
CRISPR-12	874397-874734	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	4

et al., 2011). In some CRISPR-Cas systems, unknown nucleases trim the 5′ or 3′ ends of the crRNA (Deltcheva et al., 2011; Hatoum-Aslan et al., 2011; Richter et al., 2012; Scholz et al., 2013). In the interference stage, Cas protein(s) and the mature crRNA associate to form a ribonucleoprotein (RNP) complex that targets nucleic acid sequences complementary to the crRNA (the protospacer) for degradation by a *trans*-acting nuclease (Cas3) in Type I systems (Westra et al., 2012) or by intrinsic nuclease activity in Type II and Type III-B crRNP complexes (Gasiunas et al., 2012; Hale et al., 2009; Jinek et al., 2012). While Type I and Type II complexes target DNA, the Type III-B complex is the only CRISPR-Cas system characterized to date that targets RNA.

Despite these differences, recent studies have highlighted key similarities in the architecture of Type I (Cascade-like) and Type III complexes (Reeks et al., 2013; Rouillon et al., 2013; Spilman et al., 2013; Staals et al., 2013; Zhang et al., 2012), suggesting that these complexes have evolved from a common ancestor. These complexes share a "backbone" consisting of 4–6 copies of Cas7(-like) proteins and contain a smaller Cas5-like protein, which is thought to be involved in binding the 5′ end of the crRNA. An important distinction between Type I and Type III complexes concerns the large subunit positioned at the base of the backbone, which is Cas8 in most Type I systems and Cas10 in Type III systems (Reeks et al., 2013; van der Oost et al., 2014).

Thermus thermophilus HB8 is a convenient model organism to study CRISPR-Cas systems, since it has 11 CRISPR arrays (CRISPR8 is not considered a genuine CRISPR array, since it solely consists of a single repeat sequence and no spacers) and 4 different CRISPR-Cas systems: Type I-E, Type III-A, Type III-B, and an unclassified Type I system (Agari et al., 2010) (Figure 1). We previously characterized the RNA-targeting Cmr complex of the Type III-B system of *T. thermophilus* (Staals et al., 2013). As opposed to the well-characterized Type I-E system, the Type III-A system has not been extensively studied biochemically.

The Csm operon of the Type III-A system encodes five Csm proteins (Csm1–Csm5) that form an RNP complex with the mature crRNA and sometimes an additional protein (called Csm6 or Csx1) (Hatoum-Aslan et al., 2013; Rouillon et al., 2013). The Cas7-like Csm3 forms the backbone of the complex and binds RNA in a sequence-independent fashion (Hrle et al., 2013; Koonin and Makarova, 2013; Rouillon et al., 2013). After primary cleavage of the pre-crRNA, guide maturation in the Type III-A system involves secondary trimming of the 3′ end by a ruler-like mechanism (Hatoum-Aslan et al., 2011), in which each Csm3 subunit binds and extends 6 nt segments of the mature crRNA and exposes unbound 3′ ends for cleavage by an unknown nuclease (Hatoum-Aslan et al., 2013).

The Type III-A system prevents autoimmunity (e.g., targeting the CRISPR array) by a self-discrimination versus non-self-discrimination mechanism based on complementarity with the 5′ repeat-derived fragment of the crRNA (Carte et al., 2008; Marraffini and Sontheimer, 2010), i.e., by "self-inactivation" (van der Oost et al., 2014). In contrast, Type I systems use a protospacer adjacent motif (PAM)-dependent targeting mechanism, which does not rely on base pairing (Westra et al., 2013), i.e., by "non-self-activation."

Despite these initial genetic and structural insights, no in vitro activity of the Type III-A Csm complex has yet been reported. In this study, we investigated the structural and biochemical properties of the native Type III-A Csm complex isolated from *T. thermophilus*. Unexpectedly, and in contrast to previous findings, we show that Csm exhibits endoribonuclease activity using

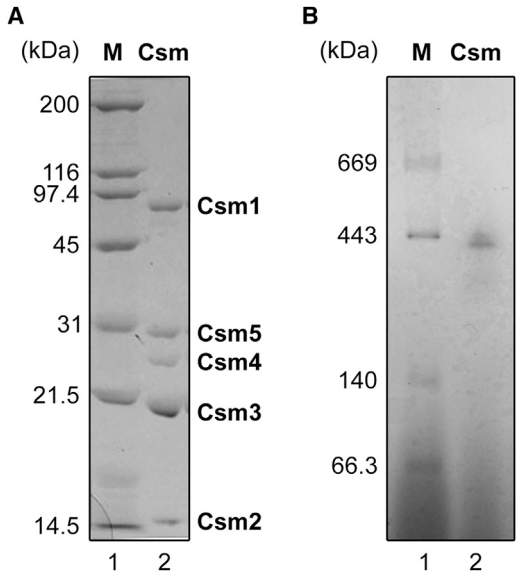


Figure 2. Purification of the Native *T. thermophilus* Type III-A Csm Complex

(A) SDS-PAGE analysis of the *T. thermophilus* Csm complex. A representative sample of the purified protein (2 μg) was analyzed on a 15% polyacrylamide gel (lane 2), followed by staining with Coomassie Brilliant Blue R-250. Each subunit is indicated. The Csm5 has a $(His)_6$-tag at its C terminus. Lane 1, molecular-mass markers.

(B) BN-PAGE analysis of the *T. thermophilus* Csm complex. Two μg of the representative sample was analyzed on a 4%–16% linear polyacrylamide gradient gel in the presence of 0.02% Coomassie Brilliant Blue G-250 (lane 2). Lane 1, molecular-mass markers. Protein concentration used was determined by the Bradford method (Bradford, 1976), using bovine serum albumin as a standard.

a cleavage mechanism similar to the Type III-B CRISPR-Cas complex Cmr.

RESULTS

Purification and Protein Composition of the TtCsm Complex

The *csm* genes of *T. thermophilus* HB8 are located in the vicinity of the CRISPR4 region on megaplasmid pTT27, comprising an operon composed of *csm1*, *csm2*, *csm3*, *csm4*, *csm5*, and *csx1* (Figure 1). We constructed a recombinant *T. thermophilus* strain that produces the Csm5 protein fused to a $(His)_6$ tag at its C terminus. The protein complex was purified to homogeneity using five subsequent column chromatography steps as described in the Supplemental Experimental Procedures available online. The purified protein complex was composed of five proteins (Figure 2A). We confirmed by mass-spectrometry-based analyses (not shown) that the proteins corresponded to TTHB147 (Csm1/Cas10), TTHB148 (Csm2), TTHB149 (Csm3), TTHB150 (Csm4), and TTHB151 (Csm5). Blue native polyacrylamide gel electrophoresis (BN-PAGE) analysis of the Csm complex revealed two major bands of 430–450 kDa, suggesting minor heterogeneity of the purified Csm complex (Figure 2B). We also constructed a recombinant *T. thermophilus* strain that produces the Csx1 protein fused with a $(His)_6$ tag at its C terminus. Under the conditions we used, Csx1 did not copurify with the Csm complex.

crRNA Content of the TtCsm Complex

Denaturing gel analysis of the copurifying nucleic acids revealed that TtCsm binds crRNAs of variable lengths (Figure 3A). RNA sequencing (RNA-seq) was performed to determine the size distribution and nature of these sequences. RNA-seq confirmed that the most abundant crRNAs varied in length from 35 to 53 nt and were enriched for the 45 and 53 nt species. Although not as clear as in other studies of Type III systems (Hale et al., 2012; Hatoum-Aslan et al., 2013; Staals et al., 2013), there is a trend of 5 or 6 nt steps in the size distribution of crRNAs isolated from TtCsm (35-40-45 nt and 35-41-47-53 nt) (Figure 3B). Similar to our previous observation with crRNAs from the *T. thermophilus* Cmr complex, the majority of the Csm-bound crRNAs (71%) contained an 8 nt repeat-derived 5′ handle: 5′-AUUGCGAC (Staals et al., 2013). Csm-bound crRNAs retained either the complete or a truncated spacer region (39–42 nt), with the larger species containing a few nucleotides (3–6 nt) of the downstream repeat sequence.

The vast majority of the reads of the RNA-seq data set could be mapped to the genome of *T. thermophilus* HB8 (94.18%) and revealed that most crRNAs were derived from CRISPRs 1, 4, and 11 (84.73%), while CRISPRs 6, 7, 9, and 10 were highly underrepresented (0.13%) (Figure 3C; Figure S1A). This bias strongly correlates with the different classes of repeat sequences (Figure 1). In addition, major variation of Csm-bound crRNAs occurs among sequences derived from the same CRISPR array: e.g., crRNA 4.5 (spacer 5 from CRISPR array 4, 15.5%) was one of the most abundant guides in the complex, while levels of crRNA 4.4 (0.1%) were extremely low (Figure 3D). Strikingly, the observed bias in the Csm-bound crRNA population closely resembles that of the recently established Cmr-bound crRNAs (Staals et al., 2013) (Figure S1B), indicating that similar crRNAs can be shared among the effector complexes of different CRISPR-Cas systems within one host. Despite these similarities, TtCsm-bound crRNAs are somewhat longer than those found in the TtCmr complex (Juranek et al., 2012; Staals et al., 2013). This finding most likely suggests that crRNAs initially assemble with premature protein complexes that differ in number of backbone subunits (Csm3 in TtCsm, Cmr4 in TtCmr) and that trimming of their 3′ overhang results in the aforementioned 5–6 nt size differences. These data showed that the Csm complex binds crRNA species of multiple lengths with a conserved 8 nt 5′ handle from a subset of CRISPR arrays and spacers.

crRNA-Protein Interactions

To study the protein-crRNA interactions within the Csm complex, we used UV-induced protein-RNA crosslinking (Kramer et al., 2011; Schmidt et al., 2012). UV crosslinking was followed by enzymatic digestion of the protein and RNA moiety, enrichment of crosslinked peptide-RNA oligonucleotides and liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. Peptide-RNA oligonucleotides were identified with their crosslinked amino acid and nucleotide by dedicated database

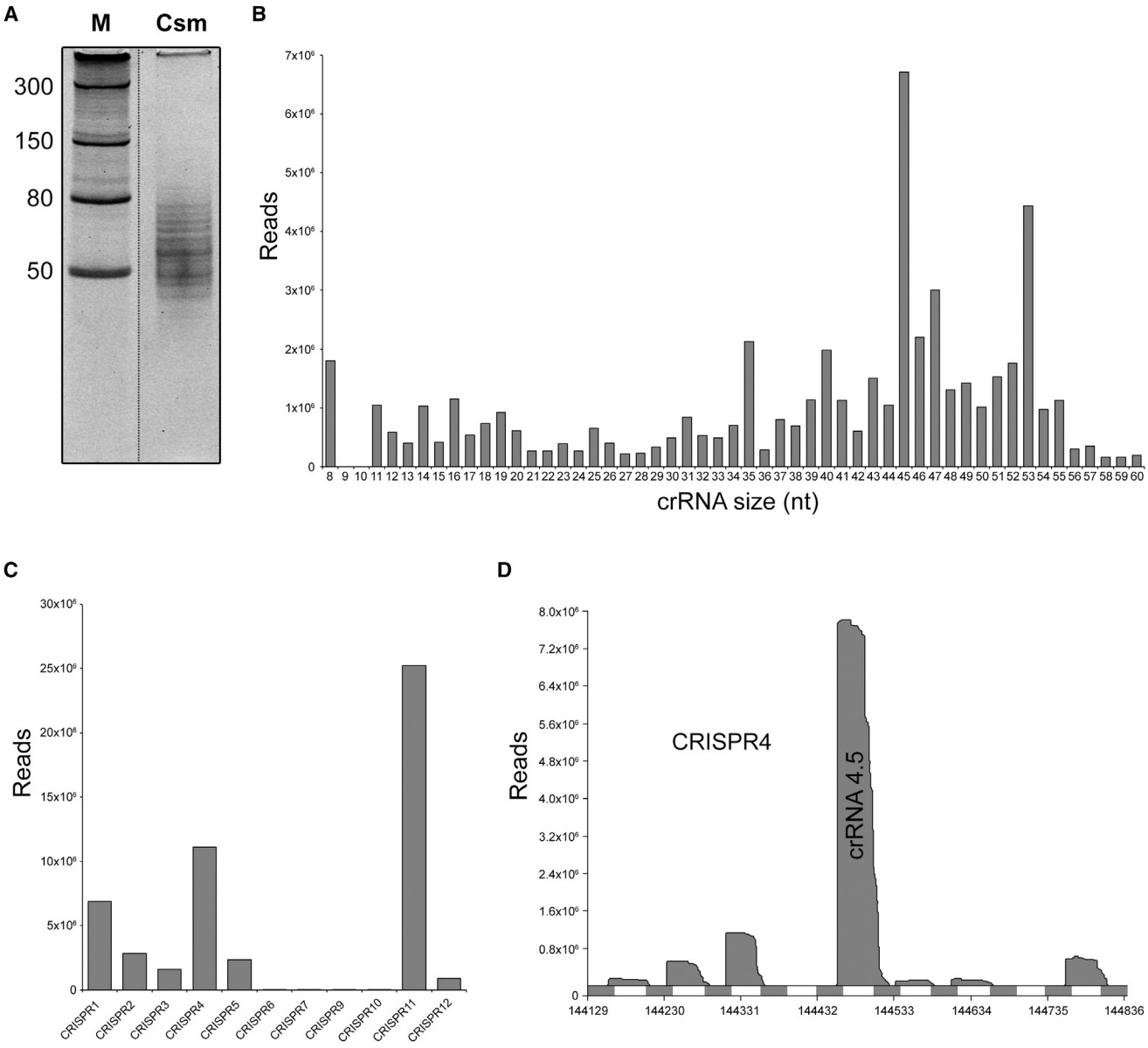


Figure 3. RNA-Seq Analysis of TtCsm-Bound crRNAs

(A) crRNAs were isolated by phenol-chloroform-isoamyl alcohol extraction and analyzed by denaturing PAGE (20% AA, 7M urea). Discontinuous gel lanes are indicated by dashed lines.

(B) Histogram of the size-distributions of the Csm-bound crRNAs.

(C) Histogram of the distribution of the Csm-bound crRNAs over the 11 CRISPR arrays.

(D) Mapping of the Csm-bound crRNAs on CRISPR4.

Overview of all mapped crRNAs and comparison of the crRNA content of the TtCsm and TtCmr complexes are provided in Figure S1.

searches (Kramer et al., 2011). Using this approach, we found 12 peptides crosslinked to different mono-, di-, and trinucleotides from the crRNAs in the TtCsm complex (Table 1). For each of the five protein subunits (Csm1 to Csm5), at least one crosslinked peptide was identified. Remarkably, six different crosslinked peptides were identified in the Csm3 subunit. By inspection of the MS/MS fragment ion spectra, we identified different amino acids as crosslinking sites (Table 1; Figure S2). The crosslinking site on the RNA was always uracil, because this is the most UV-reactive nucleotide for this technique (Kramer et al., 2014). In most cases, the same peptide sequence was found to be crosslinked to di- and trinucleotide RNAs of various compositions. However, because the sample contained a mixture of natural guides isolated from *T. thermophilus*, unambiguous identification of the exact crosslinking site on the crRNAs was not possible.

Table 1. List of RNA-Protein Crosslinks Identified in the TtCsm Complex

Protein	Peptide Sequence	RNA Moieties Observed Crosslinked to the Peptide
Csm1	[371] RLHEALAR [378]	*UUA*
Csm2	[35] LKSSQFR [41]	U, *U-H_2O*
Csm3	[21] IGMSRDQMAIGDLDNPVVR [39]	U, *UU*, UG
Csm3	[40] NPLTDEPYIPGSSLK [54]	U, U-H_2O, *UG*, UA
Csm3	[91] IFGLAPENDER [101]	*U*, UU, UC, UG
Csm3	[136] GGLYTEIKQEVFIPR [150]	U, UU, *UC*, UG, UCG, UUC, UUG
Csm3	[151] LGGNANPR [158]	*UA*, UC, UG UGG, UCA, UUA
Csm3	[159] TTERVPAGAR [168]	*U*, UG, UGG, UUG
Csm4	[69] LPPVQVEETTLRK [81]	U, *UG*, UUA
Csm4	[126] TRVGVDR [132]	UC, *UU*
Csm5	[132] SPLGAYLPGSSVK [144]	*U*, UA, UG, UUA
Csm5	[255] MVLLAETFR [263]	*U*, U-H_2O, UG

Overview of peptide-RNA oligonucleotide crosslinks identified in TtCsm complex. The positions and sequence of the crosslinked peptides as identified by MS is shown and the crosslinked amino acids are underlined. For crosslinked peptides of Csm1 (positions 371–378) and Csm3 (positions 21–39), the crosslinked amino acid could not be unambiguously identified because of the lack of corresponding fragment ions in the MS/MS carrying a nucleotide moiety (for details, see Figure S2). For peptides crosslinked to mono-, di-, or trinucleotides shown in italics, the corresponding MS/MS spectra are given in Figure S2.

Stoichiometry of the TtCsm Complex

To investigate the architecture of the TtCsm complex, we determined the composition of the Csm protein complex using native mass spectrometry as performed previously (Jore et al., 2011; Staals et al., 2013; van Duijn et al., 2012). Denaturing and tandem MS analyses provided accurate mass measurements for each protein subunit of TtCsm. The measured masses of the individual subunits were consistent with the theoretical values on the basis of their amino acid sequence (Table S1). Analysis of the intact assembly by native MS revealed the presence of two major species. From their well-resolved charge state distributions, we accurately determined their masses as 426,998 ± 217 Da and 381,896 ± 261 Da (Figure 4A; Table S1), in agreement with the estimate from native gel electrophoresis (430–450 kDa; Figure 2B). Although we could measure the masses quite accurately, the stoichiometry of the Csm subunits could not be resolved unambiguously.

The two most abundant TtCsm complexes observed (427 and 382 kDa) most likely represent the intact Csm and a Csm subcomplex lacking Csm5, respectively. The measured mass difference between the two assemblies is 46.1 kDa (the mass of a Csm5 monomer is 44,286 Da). Previously, we used collision-induced dissociation on mass-selected ions of intact Cascade complexes (van Duijn et al., 2012). However, selection and activation did not result in substantial fragmentation of TtCsm, because of the exceptional intrinsic stability of the complex. As the TtCsm assembly could not be disrupted by tandem MS, we sought to further explore the TtCsm structure using a variety of in solution dissociation experiments on intact TtCsm. As previously described for the *Sulfolobus solfataricus* Csm complex (Rouillon et al., 2013), we lowered the pH of the solvent used for electrospray and also added organic modifiers to the spray solution.

From the results of all mass spectrometric experiments, we obtained a plethora of masses for subcomplexes of TtCsm formed by elimination of individual subunits (Table S2). Using these data, the stoichiometry of the complex was narrowed down to two possible solutions, $Csm1_1Csm2_3Csm3_6Csm4_2Csm5_1crRNA_1$ (model 1; Figure 4B) or $Csm1_1Csm2_3Csm3_2Csm4_4Csm5_2crRNA_1$ (model 2). Model 1 contains multiple copies of Csm3 and has an expected mass of 427,040 Da, whereas model 2 shows a more diverse stoichiometry and has an expected mass of 427,462 Da (Table S2). Both of these models have masses that are in reasonable agreement with the measured mass of 426,998.1 Da. Although, it is not possible to distinguish between these two models on the basis of the MS data alone, we favor model 1 for two reasons. First, model 1 suggests that Csm3 (rather than Csm4) is present in multiple copies, in better agreement with the abundances observed by SDS/PAGE (Figure 2A). Second, the structural similarities with other Type I and Type III CRISPR-Cas complexes (Reeks et al., 2013; van der Oost et al., 2014) suggest that the Cas7-like Csm3 protein forms the backbone of the complex, as previously proposed (Hrle et al., 2013; Rouillon et al., 2013).

Enzymatic Activity of the TtCsm Complex

The Csm complex of *Staphylococcus epidermidis* has previously been shown in vivo to provide resistance against conjugative plasmids (Hatoum-Aslan et al., 2014; Marraffini and Sontheimer, 2008), and it has been proposed to rely on a DNA-targeting mechanism. However, its cleavage activity has not yet been demonstrated in vitro. In the present study, we performed in vitro Csm activity assays with radiolabeled single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) oligos, as well as plasmid targets complementary to abundant crRNAs in the complex. Despite numerous attempts in the presence of potential cofactors, such as Mg^{2+}, Mn^{2+}, Fe^{2+}, Cu^{2+}, Co^{2+}, Ni^{2+}, and Zn^{2+} (data not shown) and different topologies of the DNA target (oligos and plasmids), no specific activity could be detected in these assays (Figures S3A and S3B).

This prompted us to investigate the ribonuclease capabilities of the complex. A 50 nt, 5′ radiolabeled ssRNA substrate complementary to crRNA 4.5 (one of the most abundant Csm-bound crRNAs; Figure 3D; Figure S1A) was incubated with the Csm complex in a buffer containing different cofactors (Mg^{2+}, Mn^{2+}, Zn^{2+}, and Cu^{2+}) followed by denaturing gel analysis. Surprisingly, we observed Mg^{2+}- and Mn^{2+}-stimulated endoribonuclease activity, while Zn^{2+} and Cu^{2+} did not stimulate specific activity (Figure S3C). For this reason, we performed all subsequent in vitro activity assays in the presence of Mg^{2+}. Multiple degradation products of different sizes accumulated, with 21, 15, and 9 nt being the most predominant sizes observed.

Experiments with different concentrations of TtCsm (Figure 5A) or a low Mg^{2+} concentration (Figure S3D) showed that the larger degradation products were short-lived, which suggests that the target RNA is cleaved in a stepwise fashion

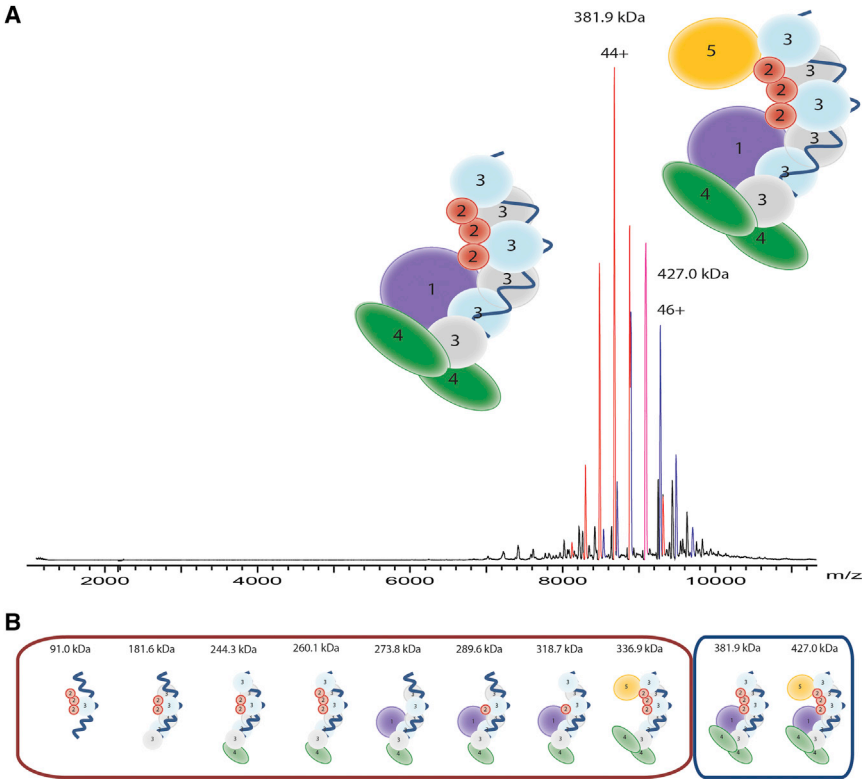


Figure 4. Subunit Composition of TtCsm

(A) Native nanoelectrospray ionization mass spectrum of the native TtCsm complex. Two main well-resolved charge state distributions are present at high *m/z* values, corresponding to complexes of 427 kDa (blue) and 382 kDa (red).

(B) TtCsm (sub)complexes analyzed by native mass spectrometry. The subcomplexes were formed after in-solution dissociation with 30% DMSO v/v or 175 mM ammonium acetate acidified with acetic acid (pH 3.6–4). More in-depth calculations of the different (sub)complexes can be found in Table S1 and Table S2.

starting at its 3′ end. To establish that this activity was Csm specific, we tested three RNA substrates (Figure S3D). The results showed that only RNA targets complementary to crRNAs loaded in the Csm complex (crRNAs 4.5 and 11.3) were degraded, while an unrelated target RNA with no complementary to Csm-bound crRNAs remained unaffected. These results demonstrated that the endoribonuclease activity in our assays was specific for Csm and not due to any copurifying contaminant nuclease. The degradation products of the 4.5 and 11.3 target RNAs had similar sizes, indicating a sequence-independent cleavage mechanism.

The pattern of cleavage products and their 6 nt periodicity had a striking resemblance to those observed with the Type III-B TtCmr complex, i.e., 33, 27, 21, and 15 nt (Figure S3D; Figure 5A) (Staals et al., 2013). These results suggested that TtCsm, like TtCmr, cleaves complementary RNAs with a 5′ ruler-like mechanism, cleaving its target RNA at 6 nt intervals measured from the 5′ end and progressing from the 3′ end. To confirm this, we performed a similar activity assay with either TtCsm or TtCmr and followed the cleavage activity in time (Figure 5B). Although TtCsm appeared to favor the formation of the smaller degradation products more quickly than TtCmr, both complexes indeed had similar cleavage patterns. In further support, activity assays with a 3′ end-labeled 50 nt ssRNA target (complementary to crRNA 4.5) with either TtCsm or TtCmr resulted in the accumulation of a degradation product of predominantly 12 nt (Figure S3E). This indicates that the cleavage of the target RNA by both complexes is initiated at the 3′ end, followed by 6 nt interspaced, periodic cleavage events progressing toward its 5′ end (the 3′ labeling reaction adds one additional nucleotide at the 3′ end). These results indicate that TtCsm cleaves complementary target RNAs with a 5′ ruler-like mechanism analogous to TtCmr.

Since complementarity to the Csm-bound crRNAs is required for activity (Figure S3D), we analyzed the determinants of target recognition in more detail, by testing activity on target RNAs with single base-pair-disrupting mismatches at positions 1 to 7 (Figure 5C, also see Figure 6A). Only the mismatch at position 5 hampered cleavage at the site directly downstream of it, as is reflected by the less abundant 39 nt band. Apart from that, it was found that single nucleotide mismatches did not affect degradation of these target RNAs, showing that perfect complementarity is not required at these positions. These results are in contrast with the stringent requirement for base-pairing interactions at positions 1–5 and 7–8 in the Type I-E Cascade complex, which make up the "seed sequence" in this system (Fineran et al., 2014; Semenova et al., 2011; Wiedenheft et al., 2011b). Experiments with mismatching nucleotides in the remaining regions of the target RNAs showed that base pairing in these regions was also dispensable for overall target degradation, suggesting that RNA targeting by TtCsm is rather flexible (Figure 5D). Interestingly, similar to the mutation at position 5, we observed that cleavage directly downstream of the mismatched region was affected, whereas cleavages at other sites proceeded normally. The results of these experiments are schematically summarized in Figure 6B.

Structural Analysis of the TtCsm Complex

We used single-particle electron microscopy (EM) and 3D reconstruction of negatively stained TtCsm complexes to gain structural information about this crRNP. Raw EM micrographs showed mono-dispersed, elongated particles with a length of ~220 Å in the largest direction (Figure 7A). Using the automated data collection program Leginon (Suloway et al., 2009) and the Appion image-processing pipeline (Lander et al., 2009), we recorded ~420 micrographs and extracted a stack of ~60,000 individual particle images. Reference-free 2D alignment and classification produced class averages with striking features, clearly

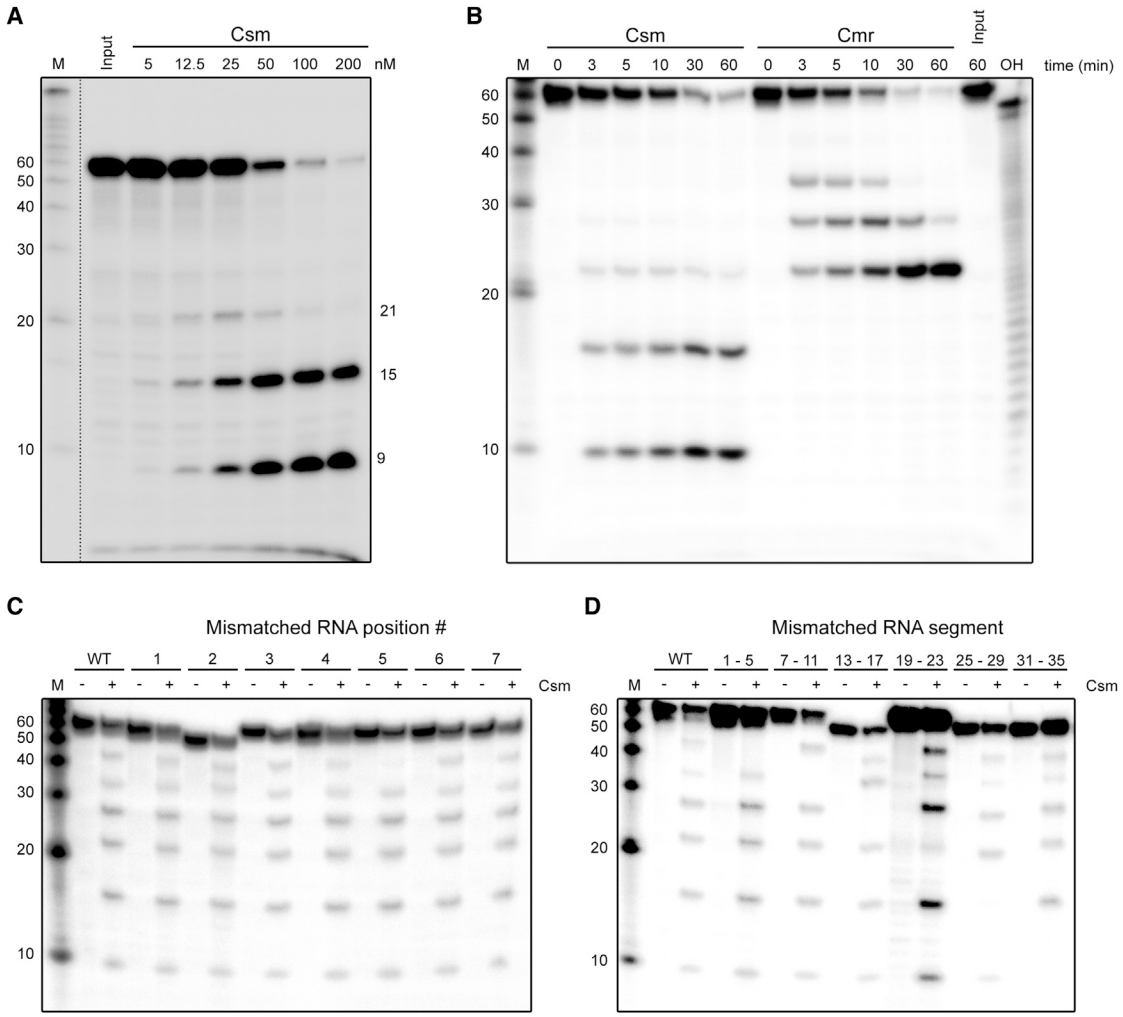


Figure 5. In Vitro Activity Assays with the TtCsm Complex

(A) A 5′-labeled ssRNA target complementary to crRNA 4.5 was incubated with different concentration of the purified, endogenous TtCsm complex in a buffer containing 2 mM Mg^{2+}. Samples were analyzed by denaturing PAGE (20% AA, 7 M urea), followed by phosphorimaging.

(B) The ssRNA target was incubated with 100 nM of the endogenous Csm or Cmr complex for the indicated amount of time. OH, alkaline hydrolysis ladder of the 50 nt RNA target.

(C) RNA targets (complementary to crRNA 4.5) with base-pair-disrupting mutations at the indicated positions (also see Figure 6A) were incubated in the absence ("–") or presence ("+") of TtCsm. In order to visualize more (transient) degradation products, the assay was performed with a lower (10 μM) Mg^{2+} concentration. "WT" refers to the unmodified "wild-type" target RNA.

(D) RNA targets (complementary to crRNA 4.5) with mutated, base-pair-disrupting regions at the indicated positions (also see Figure 6A) were incubated in the absence ("–") or presence ("+") of TtCsm in a buffer containing 10 μM Mg^{2+}. Additional Csm activity assays with different RNA or DNA substrates and different cofactors are provided in Figure S3. Discontinuous gel lanes are indicated by the dashed line.

showing a wormlike architecture of two intertwined filaments with a base resembling a foot (Figure 7B). Using the structure of *E. coli* Cascade (EMDB-5314) (Wiedenheft et al., 2011a) low-pass filtered to 60 Å as an initial model, we performed iterative projection-matching refinement to generate a final 3D electron density map at 17 Å (using the 0.5 Fourier Shell Correlation criterion) (Figure 7C; Figure S4). The 3D structure of TtCsm resembles a "sea worm" similar to TtCmr (Staals et al., 2013), composed of two intertwined filaments that terminate in a footlike base. We segmented the TtCsm structure on the basis of visual inspection of the map, our mass spectrometry analyses, and comparison with our previous segmentation of TtCmr. Of the two filaments, one is clearly larger and appears to be composed of identical, repeating subunits. This larger filament is likely composed of Csm3 on the basis of homology to the Cmr4 backbone subunit of TtCmr and our native mass spectrometry analyses (Figure 4). The smaller filament is likely composed of Csm2 subunits on the basis of a similar analysis. The head of the complex is most likely capped by Csm5 and the footlike base contains Csm1 at the toe on the basis of comparison with our previous Cmr structure and homology between these respective subunits (Figures S5A and S5B).

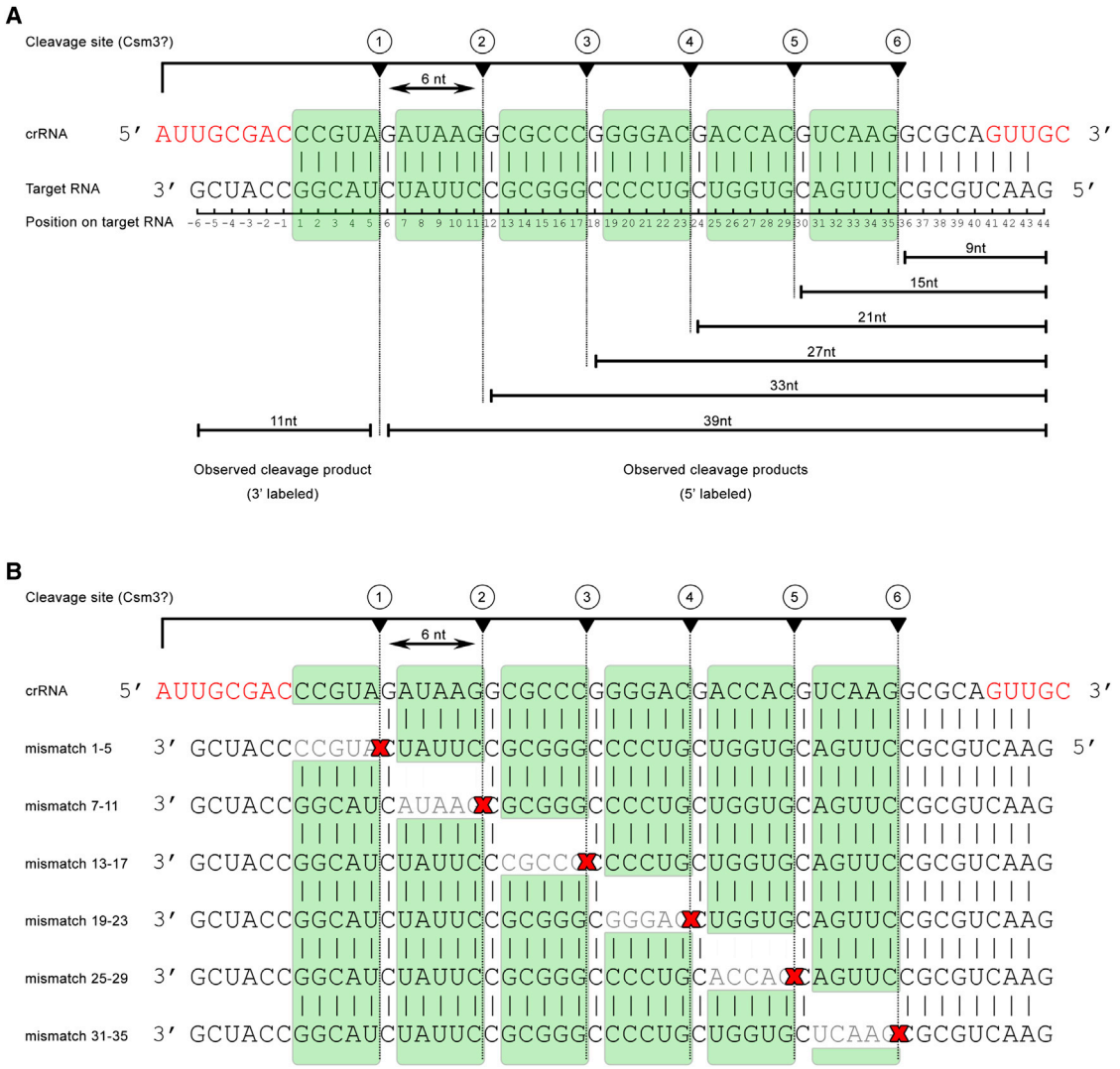


Figure 6. Model of Cleavage Activity of the TtCsm Complex

(A) Schematic representation of the cleavage activity of the TtCsm complex.

(B) Schematic representation of the impact of base-pair-disrupting mutations in regions of the target RNA on activity (also see Figure 5D). Cleavages observed in this study are indicated by dotted lines. Skipped cleavage sites are indicated with red crosses.

Nucleotides in red indicate repeat-derived sequences of the crRNA.

At the current resolution, it is difficult to unambiguously segment and dock atomic structures into the EM density; nevertheless, we constructed a model that is consistent with our biochemical results above. Using Cas7 from the *E. coli* Cascade crystal structure (Jackson et al., 2014), the Csm3 crystal structure from *M. kandleri* (35% identity) (Hrle et al., 2013), and the PHYRE automatic fold recognition server (Kelley and Sternberg, 2009), structures for TtCsm3 were generated that fit reasonably well into the larger of the two filaments, especially for the three Csm3 subunits near the head (Figure S5C). The three Csm3 subunits near the foot appear to have substantial heterogeneity and/or adopt a different helical geometry. We hypothesize that an additional Csm3 backbone subunit in a subpopulation of the purified TtCsm sample (isolated from *Thermus*) may contribute to the observed heterogeneity. Additionally, the crRNA-protein crosslinking experiments described above suggest a likely path of the crRNA along the docked TtCsm3 homology structures. In this model the crRNA would bind along the ~25 Å wide channel located between the Csm3 and Csm2 filaments and engage the conserved thumb domain of Csm3 (Figures S5D–S5H).

Interestingly, the two Csm3 structure predictions show handlike characteristics similar to those attributed to Cas7 in the recent crystal structures of Cascade (Jackson et al., 2014; Mulepati et al., 2014; Zhao et al., 2014) (Figures S5D–S5H). This thumb domain is particularly evident in the PHYRE model (Figure S5D) based on Cas7 (Jackson et al., 2014), whose crRNA binding regions are known from the structure of the intact

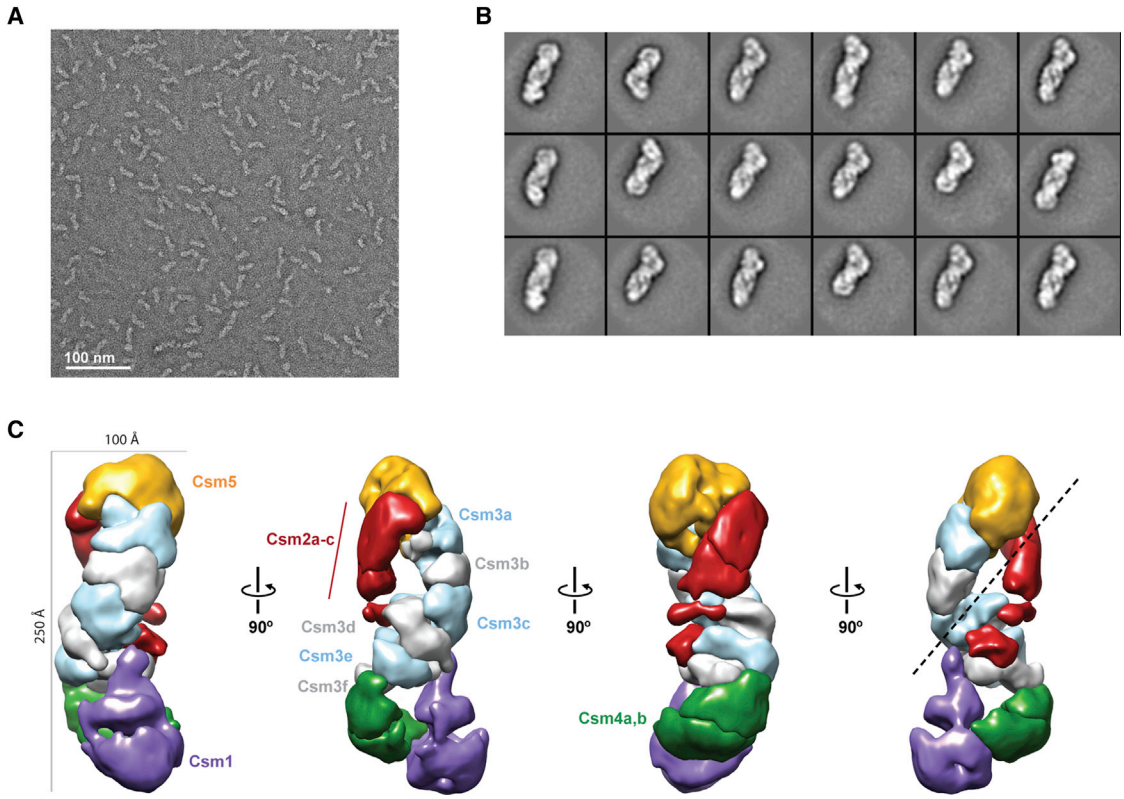


Figure 7. Molecular Architecture of the *T. thermophilus* Csm Complex

(A) Raw micrograph of negatively stained TtCsm complexes. Scale bar, 100 nm.

(B) Representative reference-free 2D class averages of TtCsm complexes. The width of the boxes is ~400 Å.

(C) Working segmentation of the TtCsm complex reconstruction at 17 Å resolution highlighting the "sea worm" architecture. Segmented regions are colored and labeled as Csm1 (purple), Csm2 (red), Csm3 (alternating light blue and gray), Csm4 (green), and Csm5 (orange).

Cascade. We generated a multiple sequence alignment for Csm3 using ClustalO and showed that most of the proposed crosslinked residues are highly conserved (Figure S6). Further structural studies will be necessary to verify the accuracy of our model.

DISCUSSION

By studying the native Type III-A Csm complex, purified from *T. thermophilus*, we have revealed several important features of this CRISPR-Cas system, including its composition, structure, and activity in vitro.

The TtCsm complex consists of five Csm proteins (Csm1–Csm5) and one crRNA of variable sizes. Our RNA-seq analysis revealed that the Csm-bound crRNAs contain an invariable 5′ handle, 5′-AUUGCGAC, which is consistent with the primary, Cas6-mediated cleavage of the pre-crRNA at this position (Carte et al., 2008). The presence of the tag in the mature crRNA may play a role in initiating RNP complex formation, with the tag being specifically recognized and bound by one of the protein subunits of the complex. In comparison with other CRISPR-Cas complexes, it is most likely that the Cas5-like Csm4 subunit might perform this function (van der Oost et al., 2014; Wiedenheft et al., 2011a). The fact that mature Csm-bound crRNAs have a variable 3′ end hints at a processing step for trimming these crRNAs that appears to be a typical feature for Type III CRISPR-Cas complexes (Hale et al., 2009; Hatoum-Aslan et al., 2011; Rouillon et al., 2013; Staals et al., 2013; Zhang et al., 2012). Interestingly, we found more extensive 3′ end processing of Csm-bound crRNAs with a noncanonical 5′ handle (e.g., crRNAs from CRISPR6 and CRISPR7). This indicates that proper crRNP complex assembly (including recognition of the 5′ tag sequence) is directly coupled to mature 3′ end crRNA formation and suggests that 3′ processing occurs when bound within the surveillance complex. Hence, failure to properly assemble the complex (in the case of a noncanonical tag sequence) would result in crRNAs with more exposed 3′ ends that are susceptible to 3′ trimming (Brendel et al., 2014). These observations are consistent with a model where the sizes of mature crRNAs are determined by the dimensions of the complex. Indeed, TtCsm crRNAs (45/53 nt) were somewhat longer than the TtCmr-bound ones (40/46 nt), because of a complex with a slightly more extended backbone and therefore more protection of their 3′ ends (Figure S7). Our model where Csm3 forms the backbone of the complex (model 1) is therefore in good agreement with the previously reported 6 nt binding periodicity of this protein (Hatoum-Aslan et al., 2013).

The crRNA content of the TtCsm complex showed a remarkable bias for particular spacers, both in terms of variety and abundance. These results were strikingly similar to our previous observations with the TtCmr complex (Staals et al., 2013) (Figure S1B), suggesting that crRNAs can be shared among complexes from different CRISPR-Cas subtypes. In stark contrast, *Sulfolobus solfataricus* Csm and Cmr complexes interact with crRNAs derived from different CRISPR arrays (Rouillon et al., 2013; Zhang et al., 2012). This sorting phenomenon might be accounted for by the differences in repeat specificity of these complexes or the different repertoire of Cas6 paralogues in these species and the handover mechanisms of these paralogues (Niewoehner et al., 2014; Sokolowski et al., 2014). It is possible that bias in spacer selection occurs at the level of primary processing by Cas6. Indeed, in the Cascade-associated Cas6e/6f, single turnover processing of pre-crRNAs results in delivery to the appropriate crRNP complexes (Niewoehner et al., 2014; Reeks et al., 2013; van der Oost et al., 2014).

Type III-A CRISPR-Cas systems have been implicated in providing protection against plasmid conjugation and transformation in vivo (Hatoum-Aslan et al., 2014; Marraffini and Sontheimer, 2008, 2010). The results of these studies were interpreted as evidence for a DNA-targeting mechanism, although no confirmation of this activity has yet been obtained in vitro. Interestingly, the in vitro analyses presented here reveal that (under the tested conditions) the TtCsm complex harbors in vitro RNase rather than DNase activity. TtCsm catalyzes the cleavage of complementary target RNAs with a 5′ ruler-like endoribonuclease mechanism similar to that of Type III-B systems (Hale et al., 2009; Staals et al., 2013). This ruler mechanism defines six cleavage sites at the target RNA, each separated by 6 nt distances. Analogous to the Type I Cascade complex, the target RNA is most likely base pairing with the crRNA guide along the backbone of the TtCsm, a model that is supported by the specificity of the TtCsm complex for complementary RNA targets and the extensive crRNA-Csm3 crosslinks observed in this study. We therefore hypothesize that the Cas7-like subunits, constituting the backbone of these Type III complexes (Csm3 and Cmr4 in Type III-A and Type III-B, respectively), harbor the active sites. Our in vitro activity assay with partially mismatching target RNAs showed that adjacent active sites were impaired when base pairing of the guide's upstream nucleotide(s) was disrupted. Nevertheless, these partially mismatching target RNAs were still degraded at the more distantly located active sites, which indicates that the TtCsm crRNA guides lack a defined seed sequence as is present in at least a subset of the Type I systems (Maier et al., 2013; Semenova et al., 2011; Wiedenheft et al., 2011b). Although a more detailed analysis on the boundaries of target recognition is still required, these results indicate that RNA targeting by Type III-A systems is quite flexible. This flexibility could also explain why target interrogation in Type III systems does not rely on a PAM (as in Type I systems), since target versus nontarget discrimination should be dispensable for RNA. Whether or not RNA targeting relies on other motifs outside of the protospacer region is an interesting task for further investigations.

The intriguing discrepancy between the apparent DNA-targeting activities of the Csm complex in vivo (Marraffini and Sontheimer, 2008) and its RNA targeting activity in vitro (this study) opens up the question of whether DNA, RNA, or both are the natural targets of the Type III-A system (and possibly of the III-B system). Several recent studies may provide pieces of the puzzle. (1) A Type III system of *Sulfolobus islandicus* has been reported to result in degradation of plasmid DNA. Interestingly, DNA interference appeared to be dependent on both transcription of the target sequence and on the presence of Csx1 (Csm6) (Deng et al., 2013). (2) Csx1/Csm6 are members of a highly variable protein family sharing a CARF domain (CRISPR-Cas-associated Rossmann fold) and are strongly associated with Type III CRISPR-Cas systems (Makarova et al., 2014). Furthermore, Csx1 of *Pyrococcus furiosus* has been demonstrated to associate with both dsDNA and dsRNA (Kim et al., 2013). (3) The large subunit of the Csm complex (Cas10/Csm1) of *Staphylococcus epidermidis* degrades single-stranded DNA and RNA in vitro (Ramia et al., 2014). (4) In *S. epidermidis,* deletion of the *csm*6 gene (encoding a Csx1 homolog) and mutations of conserved residues in the Palm polymerase domain of Cas10/Csm1 prevented CRISPR immunity in vivo, without affecting either complex formation or crRNA production, strongly suggesting their involvement in target degradation (Hatoum-Aslan et al., 2014). (5) Given the clustering of the gene encoding Csx1/Csm6 with the five genes encoding the Csm complex (Figure 1) (Makarova et al., 2014), it is possible the Type III-A system uses a similar (Csm6/Csx1-dependent and transcription-dependent) DNA-targeting mechanism as well. Indeed, during the revision of this manuscript, a new study by the Marraffini group showed that interference by the Type III-A Csm complex in *S. epidermidis* proceeds in a transcription-dependent fashion, which was shown to confer resistance against lytic viruses (Goldberg et al., 2014). These observations together with the in vitro RNase activities from this study strongly suggest a role of the Type III-A system in degrading transcriptionally active MGEs. Future in vivo and in vitro analyses are required to fully understand how this intriguing CRISPR-Cas variant functions to protect its host from MGE invasions.

EXPERIMENTAL PROCEDURES

Purification of the Csm Complex and Identification of the Csm Proteins

The TtCsm complex was from a $(His)_6$-tagged Csm complex expressing *T. thermophilus* HB8 strain, as described in detail in the Supplemental Experimental Procedures.

RNA-Seq Analysis

crRNAs were purified and sequenced essentially as described previously (Staals et al., 2013), the details of which can be read in the Supplemental Experimental Procedures.

In Vitro Activity Assays

In vitro activity assays were performed essentially as described previously (Staals et al., 2013), details are described in the Supplemental Experimental Procedures.

UV Crosslinking and Identification of crRNA-Protein Interactions by LC-MS/MS

Protein-RNA crosslinking was performed using UV irradiation at 254 nm and the crosslinked peptides were enriched as described previously (Kramer et al., 2014; Schmidt et al., 2012). The sample was analyzed by LC-MS/MS

essential according to Kramer et al. (2011). A more detailed description is provided in the Supplemental Experimental Procedures.

Native Mass Spectrometry

Native MS was performed as described in detail previously (van Duijn et al., 2012). Details about measurements of TtCsm subcomplexes and individual TtCsm proteins under denaturing conditions can be found in the Supplemental Experimental Procedures.

Single Particle Electron Microscopy and Analysis

TtCsm complexes diluted to ~25–50 nM were applied immediately to a glow-discharged continuous-carbon grid and then negatively stained with four consecutive droplets of 2% uranyl acetate. The sample was examined using a Technai-20 electron microscope equipped with a field emission gun and operated at 120 kV acceleration voltage. Image processing and 3D reconstruction were performed as described in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The EM-derived density map of the TtCsm complex has been deposited in the Electron Microscopy Data Bank under accession number EMD-6122. The RNA-seq data set of the Csm-bound crRNAs has been deposited in the European Nucleotide Archive database under accession number PRJEB7461.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2014.10.005.

AUTHOR CONTRIBUTIONS

A.S., K. Sakamoto, T.S., N.D., S.Y., and J.v.d.O. contributed to the initial design of this study. K. Sakamoto, S.Y., and A.S. contributed to the production and purification of the Csm complex. T.S. and N.D. contributed to the MS identification of the Csm proteins. R.H.J.S., J.J.K., and P.J.S. contributed to RNA-seq analysis. R.H.J.S., Y.Z., K.V., N.N., and M.V. contributed to biochemical analyses. D.W.T., J.E.K., E.N., and J.A.D. contributed to electron microscopy and 3D structure determination. K. Sharma and H.U. contributed to MS-based protein-RNA crosslinking. A.B. and A.J.R.H. contributed to MS-based stoichiometry analysis. R.H.J.S., Y.Z., D.W.T., K. Sharma, A.B., A.S., and J.v.d.O., with input from all, contributed to the writing of manuscript.

ACKNOWLEDGMENTS

The authors thank Virgis Siksnys (Vilnius) for sharing data before publication. We thank Aimi Osaki for construction of the recombinant *T. thermophilus* strain. R.H.J.S. was supported by the University of Otago's Division of Health Sciences Career Development postdoctoral fellowships. This work was supported by an ALW grant (820.02.003 to J.v.d.O.) from the Netherlands Organization for Scientific Research (NWO) and by JSPS KAKENHI grant no.25440013 (to A.S.). A.B. and A.J.R.H. were supported by the Netherlands Proteomics Centre. E.N. and J.A.D. are Howard Hughes Medical Institute Investigators.

Received: July 25, 2014
Revised: September 27, 2014
Accepted: October 2, 2014
Published: November 6, 2014

REFERENCES

Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S., and Shinkai, A. (2010). Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. J. Mol. Biol. *395*, 270–281.

Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, Ü. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. Nucleic Acids Res. *42*, 7884–7893.

Barrangou, R., and Marraffini, L.A. (2014). CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol. Cell *54*, 234–244.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248–254.

Brendel, J., Stoll, B., Lange, S.J., Sharma, K., Lenz, C., Stachler, A.E., Maier, L.K., Richter, H., Nickel, L., Schmitz, R.A., et al. (2014). A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in Haloferax volcanii. J. Biol. Chem. *289*, 7164–7177.

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 Dev. *22*, 3489–3496.

Carte, J., Christopher, R.T., Smith, J.T., Olson, S., Barrangou, R., Moineau, S., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M., and Terns, M.P. (2014). The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in *Streptococcus thermophilus*. Mol. Microbiol. *93*, 98–112.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607.

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*. Mol. Microbiol. *87*, 1088–1099.

Fineran, P.C., Gerritzen, M.J., Suárez-Diez, M., Künne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H., and Brouns, S.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proc. Natl. Acad. Sci. USA *111*, E1629–E1638.

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. USA *109*, E2579–E2586.

Gasiunas, G., Sinkunas, T., and Siksnys, V. (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. Cell. Mol. Life Sci. *71*, 449–465.

Goldberg, G.W., Jiang, W., Bikard, D., and Marraffini, L.A. (2014). Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. Nature. Published online August 31, 2014. http://dx.doi.org/10.1038/nature13637.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics *8*, 172.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput. Biol. *1*, e60.

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.

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M., and Terns, M.P. (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Mol. Cell *45*, 292–302.

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. USA *108*, 21218–21222.

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. J. Biol. Chem. *288*, 27888–27897.

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.

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 Biol. *10*, 1670–1678.

Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. Science *345*, 1473–1479.

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.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. *18*, 529–536.

Juranek, S., Eban, T., Altuvia, Y., Brown, M., Morozov, P., Tuschl, T., and Margalit, H. (2012). A genome-wide view of the expression and processing patterns of *Thermus thermophilus* HB8 CRISPR RNAs. RNA *18*, 783–794.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. *4*, 363–371.

Kim, Y.K., Kim, Y.G., and Oh, B.H. (2013). Crystal structure and nucleic acid-binding activity of the CRISPR-associated protein Csx1 of *Pyrococcus furiosus*. Proteins *81*, 261–270.

Koonin, E.V., and Makarova, K.S. (2013). CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. RNA Biol. *10*, 679–686.

Kramer, K., Hummel, P., Hsiao, H.-H., Luo, X., Wahl, M., and Urlaub, H. (2011). Mass-spectrometric analysis of proteins cross-linked to 4-thio-uracil- and 5-bromo-uracil-substituted RNA. Int. J. Mass Spectrom. *304*, 184–194.

Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.L., Hentze, M.W., Kohlbacher, O., and Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat. Methods *11*, 1064–1070.

Lander, G.C., Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.W., et al. (2009). Appion: an integrated, database-driven pipeline to facilitate EM image processing. J. Struct. Biol. *166*, 95–102.

Maier, L.K., Lange, S.J., Stoll, B., Haas, K.A., Fischer, S., Fischer, E., Duchardt-Ferner, E., Wöhnert, J., Backofen, R., and Marchfelder, A. (2013). Essential requirements for the detection and degradation of invaders by the Haloferax volcanii CRISPR/Cas system I-B. RNA Biol. *10*, 865–874.

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. (2011). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. *9*, 467–477.

Makarova, K.S., Anantharaman, V., Grishin, N.V., Koonin, E.V., and Aravind, L. (2014). CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. Front. Genet. *5*, 102.

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). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568–571.

Mulepati, S., Héroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science *345*, 1479–1484.

Niewoehner, O., Jinek, M., and Doudna, J.A. (2014). Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. Nucleic Acids Res. *42*, 1341–1353.

Ramia, N.F., Tang, L., Cocozaki, A.I., and Li, H. (2014). *Staphylococcus epidermidis* Csm1 is a 3′-5′ exonuclease. Nucleic Acids Res. *42*, 1129–1138.

Reeks, J., Naismith, J.H., and White, M.F. (2013). CRISPR interference: a structural perspective. Biochem. J. *453*, 155–166.

Richter, H., Zoephel, J., Schermuly, J., Maticzka, D., Backofen, R., and Randau, L. (2012). Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. Nucleic Acids Res. *40*, 9887–9896.

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.

Schmidt, C., Kramer, K., and Urlaub, H. (2012). Investigation of protein-RNA interactions by mass spectrometry—Techniques and applications. J. Proteomics *75*, 3478–3494.

Scholz, I., Lange, S.J., Hein, S., Hess, W.R., and Backofen, R. (2013). CRISPR-Cas systems in the cyanobacterium *Synechocystis* sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. PLoS ONE *8*, e56470.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc. Natl. Acad. Sci. USA *108*, 10098–10103.

Sokolowski, R.D., Graham, S., and White, M.F. (2014). Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system. Nucleic Acids Res. *42*, 6532–6541.

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.

Suloway, C., Shi, J., Cheng, A., Pulokas, J., Carragher, B., Potter, C.S., Zheng, S.Q., Agard, D.A., and Jensen, G.J. (2009). Fully automated, sequential tilt-series acquisition with Leginon. J. Struct. Biol. *167*, 11–18.

Terns, R.M., and Terns, M.P. (2014). CRISPR-based technologies: prokaryotic defense weapons repurposed. Trends Genet. *30*, 111–118.

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat. Rev. Microbiol. *12*, 479–492.

van Duijn, E., Barbu, I.M., Barendregt, A., Jore, M.M., Wiedenheft, B., Lundgren, M., Westra, E.R., Brouns, S.J., Doudna, J.A., van der Oost, J., and Heck, A.J. (2012). Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced shot-palindromic-repeats (CRISPR)-associated protein complexes from *Escherichia coli* and *Pseudomonas aeruginosa*. Mol. Cell. Proteomics *11*, 1430–1441.

Westra, E.R., van Erp, P.B., Künne, 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.

Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K., and Brouns, S.J. (2013). Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. PLoS Genet. *9*, e1003742.

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

Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J., Dickman, M.J., and Doudna, J.A. (2011b). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc. Natl. Acad. Sci. USA *108*, 10092–10097.

Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Res. *40*, 5569–5576.

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.

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. Nature Published online August 12, 2014. 10.1038/nature13733.

Molecular Cell, Volume *56*

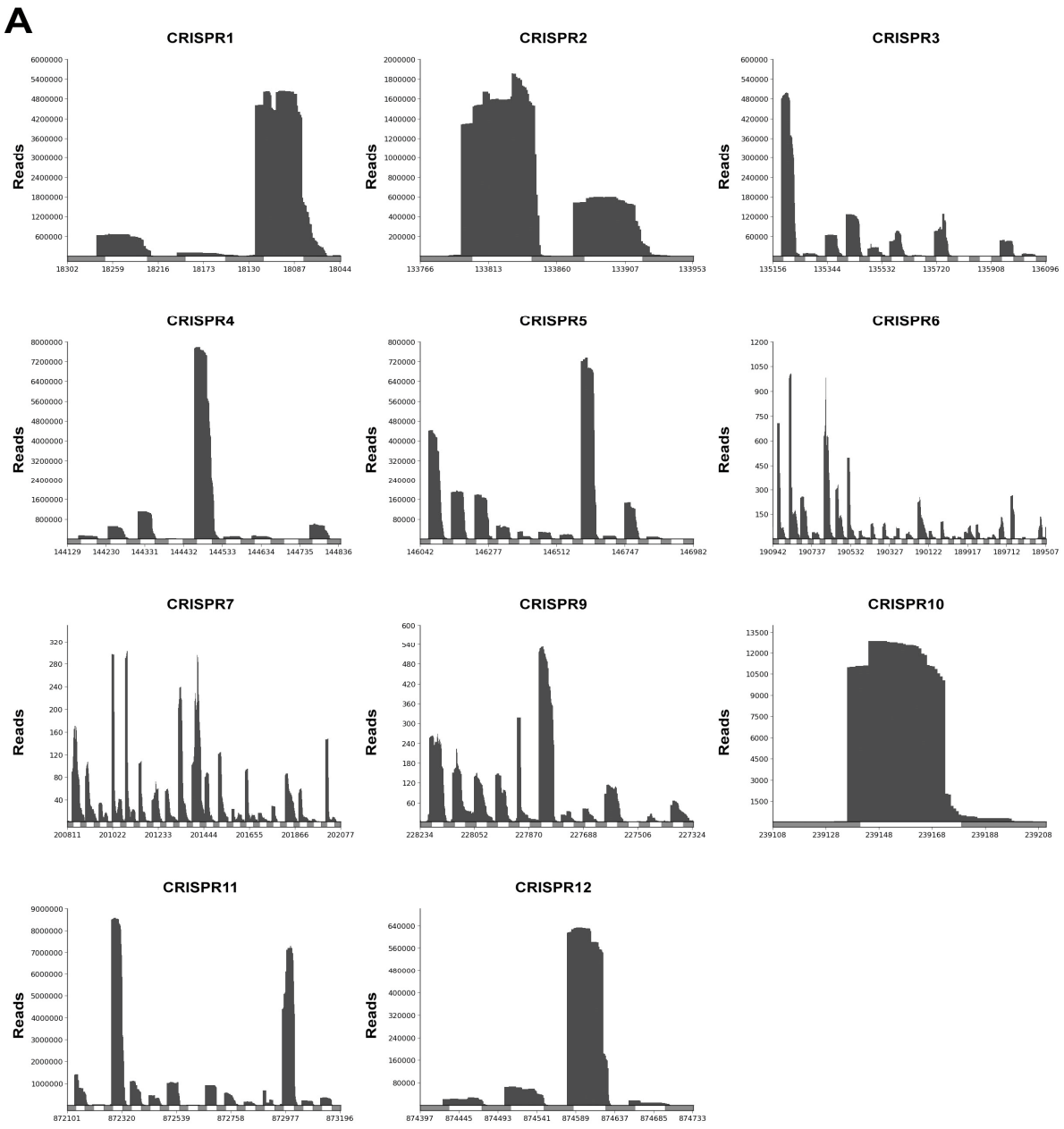
Supplemental Information

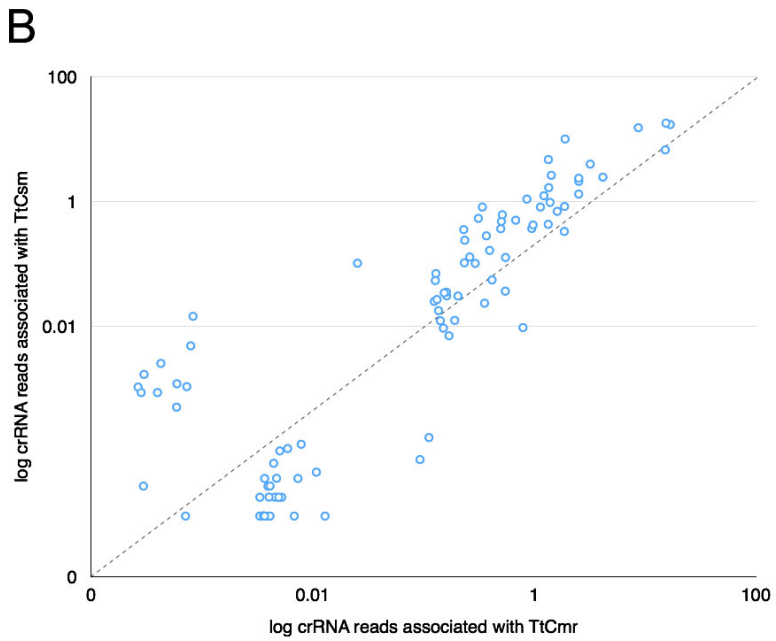
RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of *Thermus thermophilus*

Raymond H.J. Staals, Yifan Zhu, David W. Taylor, Jack E. Kornfeld, Kundan Sharma, Arjan Barendregt, Jasper J. Koehorst, Marnix Vlot, Nirajan Neupane, Koen Varossieau, Keiko Sakamoto, Takehiro Suzuki, Naoshi Dohmae, Shigeyuki Yokoyama, Peter J. Schaap, Henning Urlaub, Albert J.R. Heck, Eva Nogales, Jennifer A. Doudna, Akeo Shinkai, and John van der Oost

Supplemental Figures

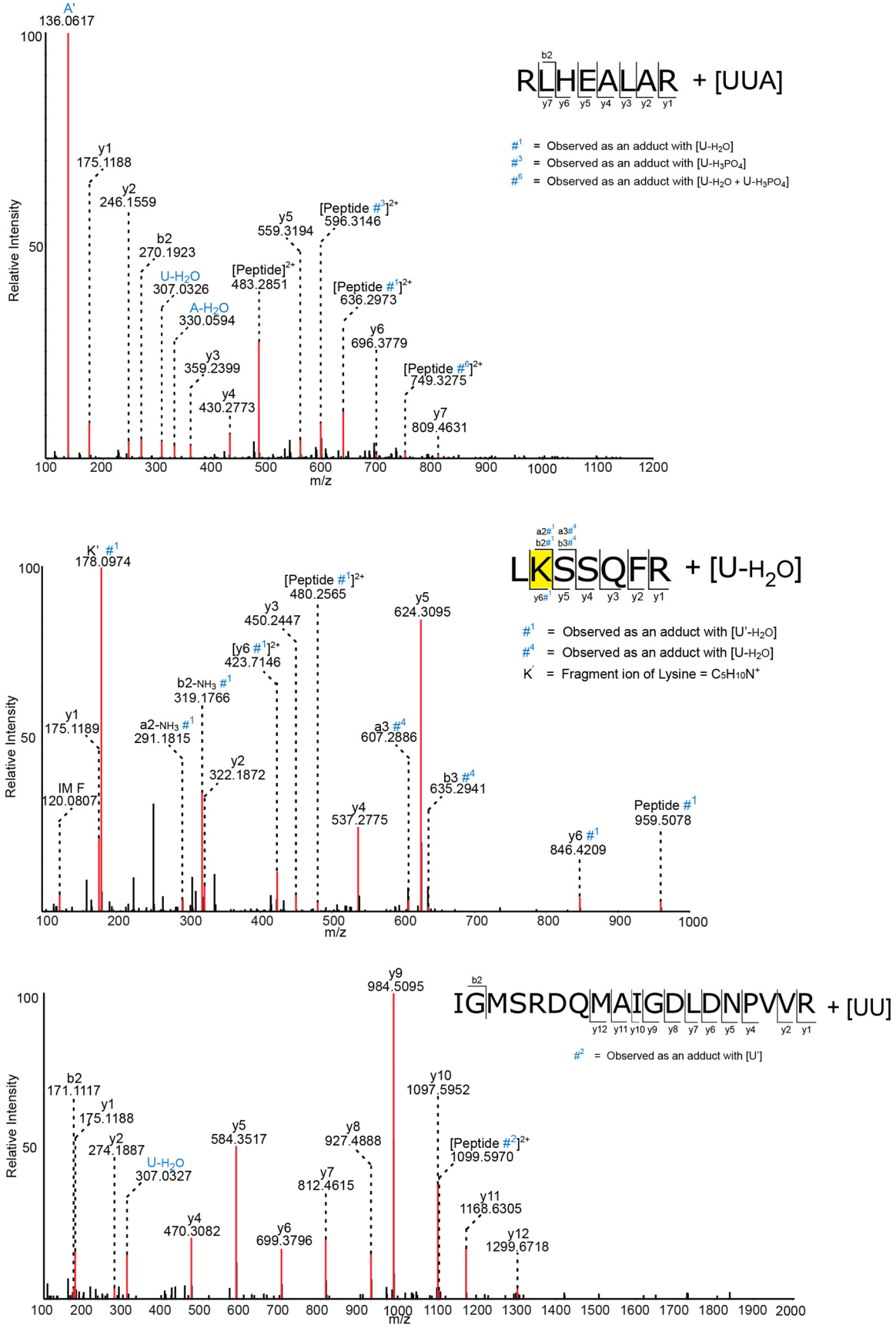
Figure S1 (related to Figure 3)

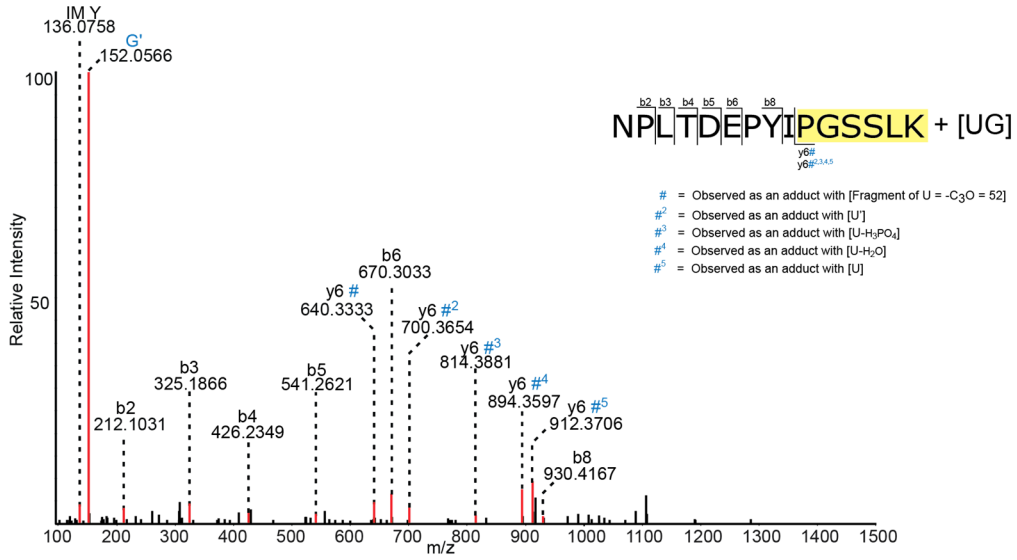




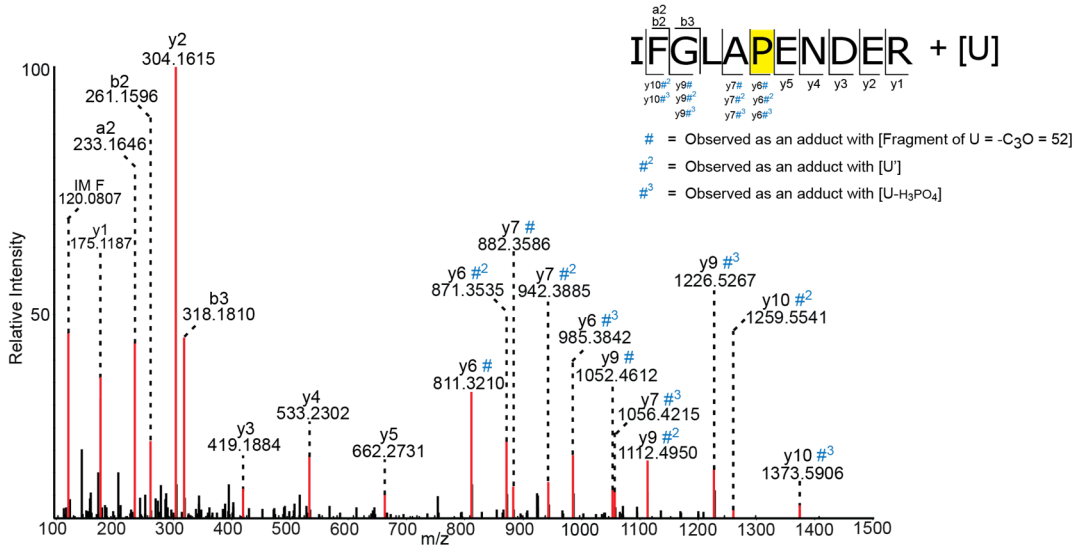
Mapping of the RNAseq data on the *T. thermophilus* CRISPR arrays and spacers and comparison of the crRNA content of the TtCsm and TtCmr complexes. (A) crRNAs were isolated from the TtCsm complex and analyzed by RNAseq. The resulting reads were mapped on the genome of *T. thermophilus* HB8. Depicted are the absolute number of reads (y-axis) mapping to the 11 different CRISPR arrays. The genomic locations are presented on the x-axis with the repeat and spacer sequences indicated in gray and white respectively. (B) Comparison of the crRNA content of the TtCsm and TtCmr complexes. The log number of reads of TtCmr-associated crRNAs (x-axis) is plotted against the log number of reads of TtCsm-associated crRNAs (y-axis). Each node represents a different crRNA which was normalized by the total amount of mapped reads.

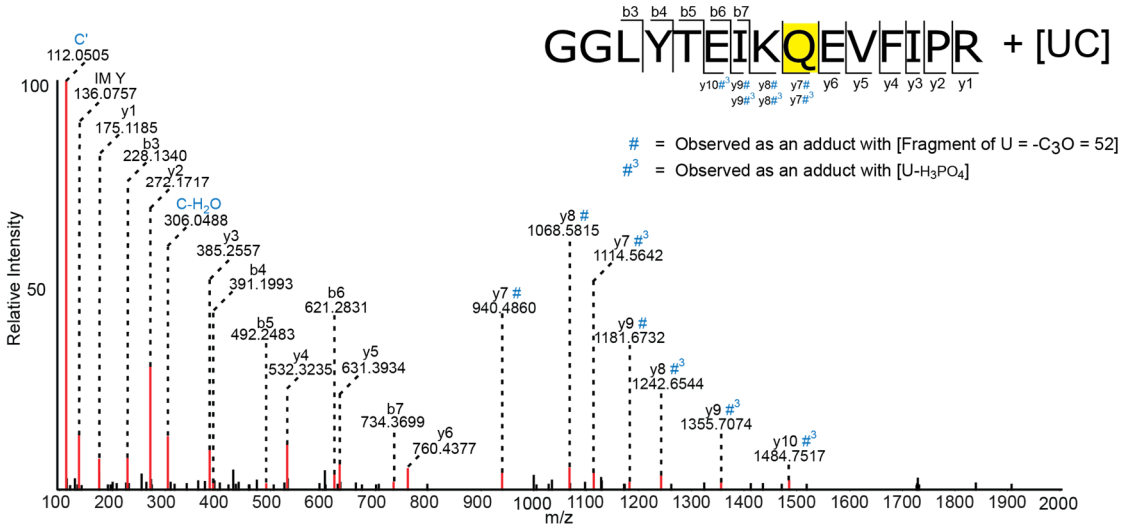
Figure S2 (related to Table 1)

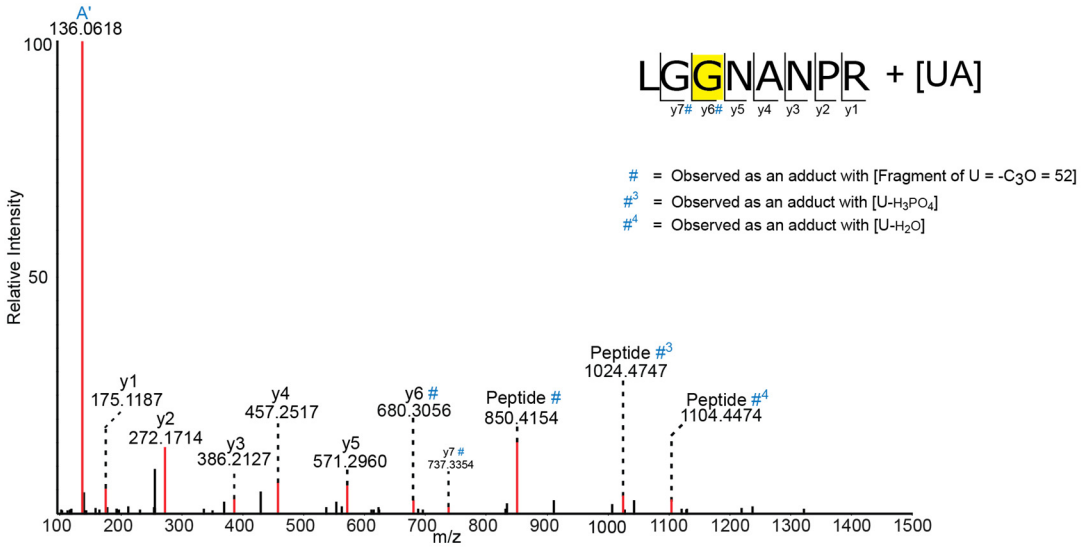


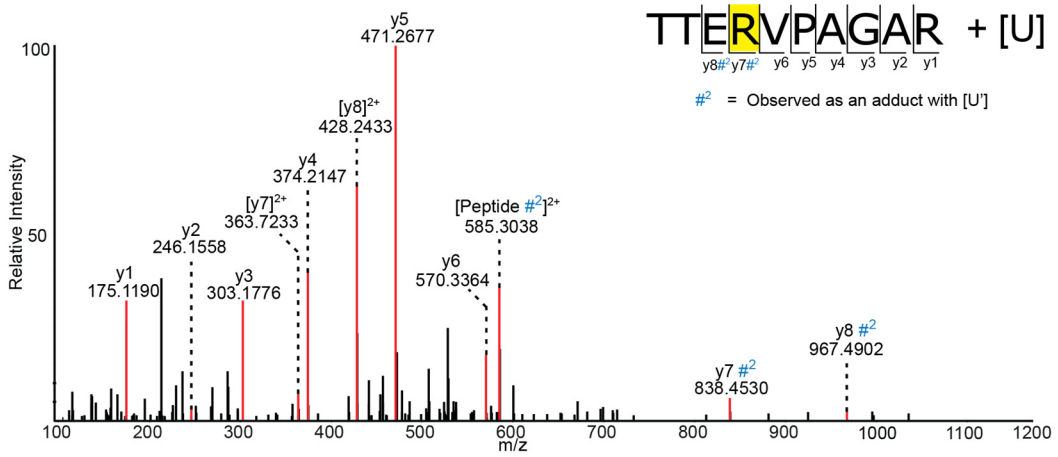


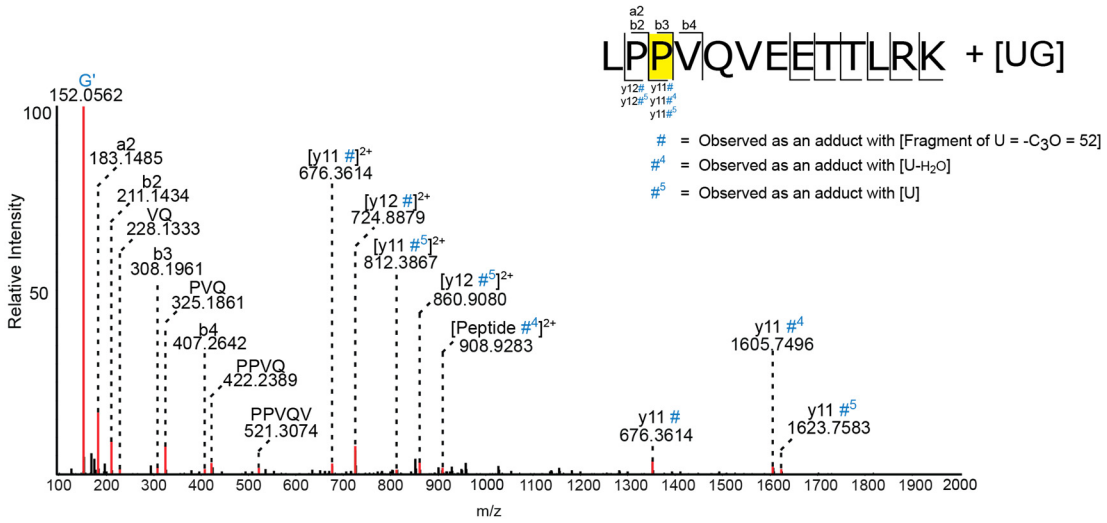
Exact crosslinked residue could not be identified, either of the Amino acids highlighted in yellow can be crosslink site.

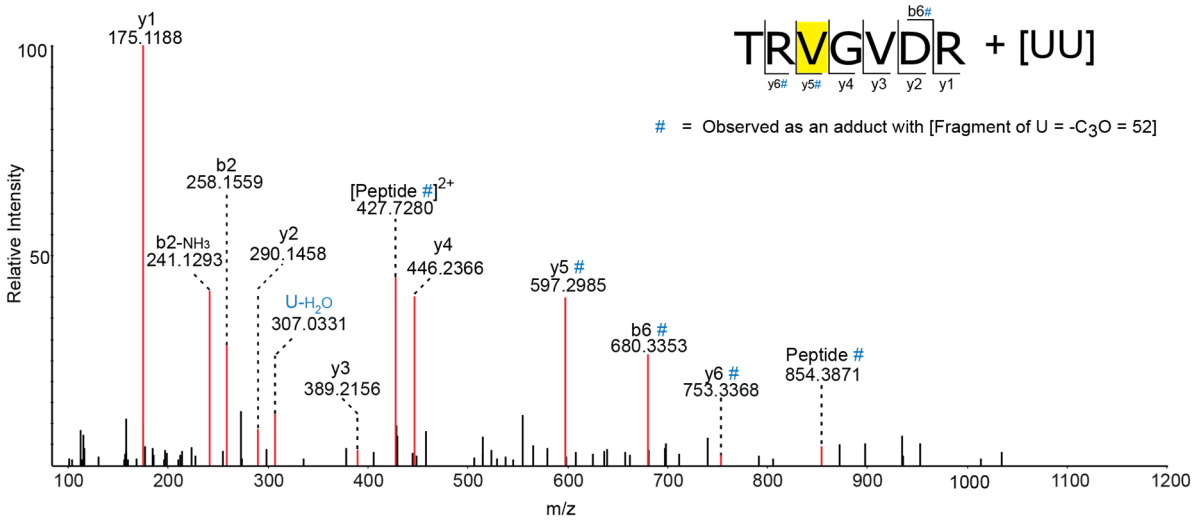


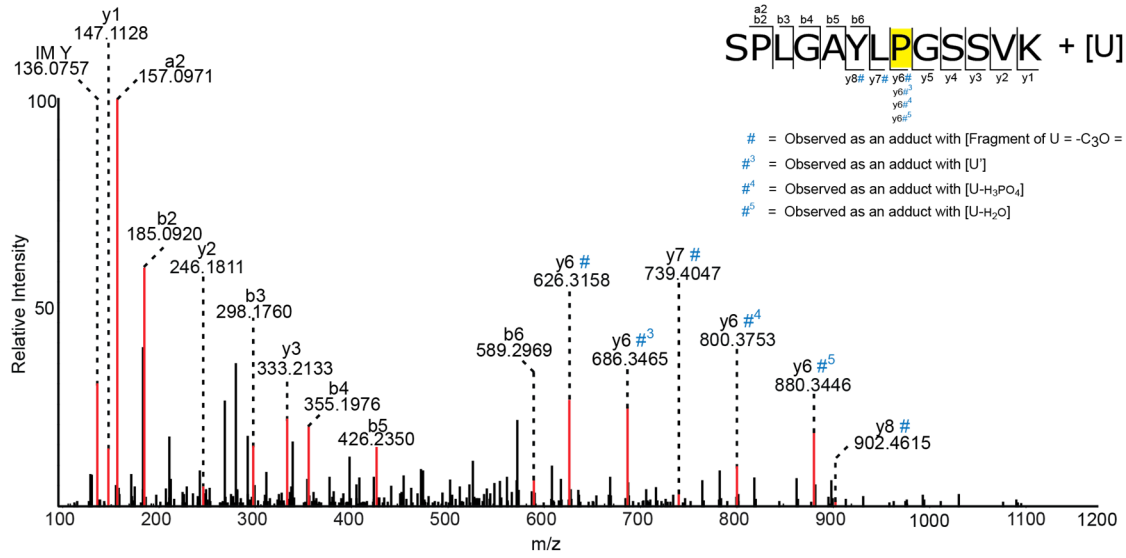


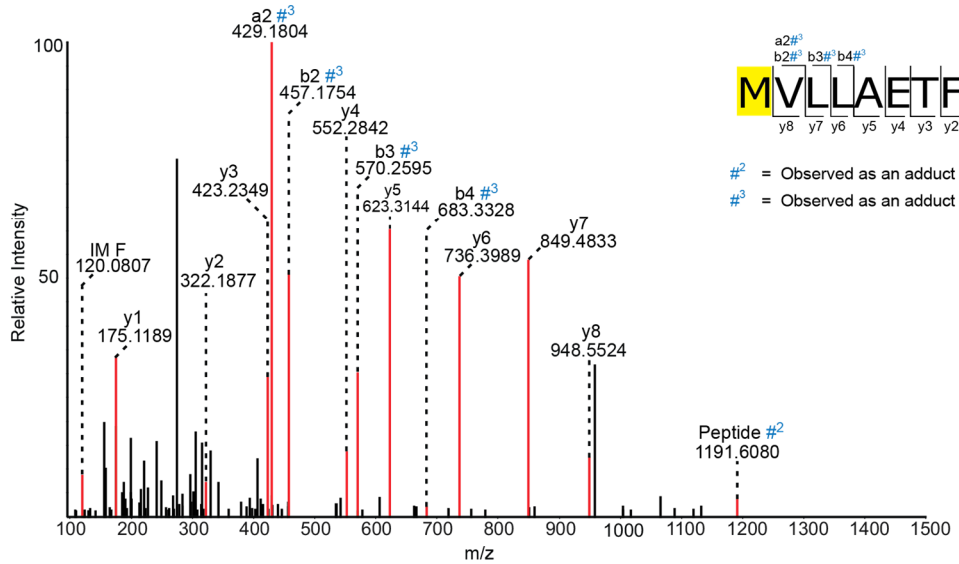

LGGNANPR + [UA]
y7# y6# y5 y4 y3 y2 y1
= Observed as an adduct with [Fragment of U = -C3O = 52]
#3 = Observed as an adduct with [U-H3PO4]
#4 = Observed as an adduct with [U-H2O]
A'
136.0618
y1
175.1187
y2
272.1714
y3
386.2127
y4
457.2517
y5
571.2960
y6 #
680.3056
y7 #
737.3354
Peptide #
850.4154
Peptide #3
1024.4747
Peptide #4
1104.4474
Relative Intensity
m/z


TTERVPAGAR + [U]
y8#2 y7#2 y6 y5 y4 y3 y2 y1
#2 = Observed as an adduct with [U']
y5
471.2677
[y8]2+
428.2433
y4
374.2147
[y7]2+
363.7233
y2
246.1558
y1
175.1190
y3
303.1776
[Peptide #2]2+
585.3038
y6
570.3364
y8 #2
967.4902
y7 #2
838.4530
Relative Intensity
m/z


LPPPVQVEETTLRK + [UG]
a2
b2
b3
b4
y12#
y12#5
y11#
y11#4
y11#5
= Observed as an adduct with [Fragment of U = -C3O = 52]
#4 = Observed as an adduct with [U-H2O]
#5 = Observed as an adduct with [U]
G'
152.0562
a2
183.1485
b2
211.1434
VQ
228.1333
b3
308.1961
PVQ
325.1861
b4
407.2642
PPVQ
422.2389
PPVQV
521.3074
[y11 #]2+
676.3614
[y12 #]2+
724.8879
[y11 #5]2+
812.3867
[y12 #5]2+
860.9080
[Peptide #4]2+
908.9283
y11 #
676.3614
y11 #4
1605.7496
y11 #5
1623.7583
Relative Intensity
m/z

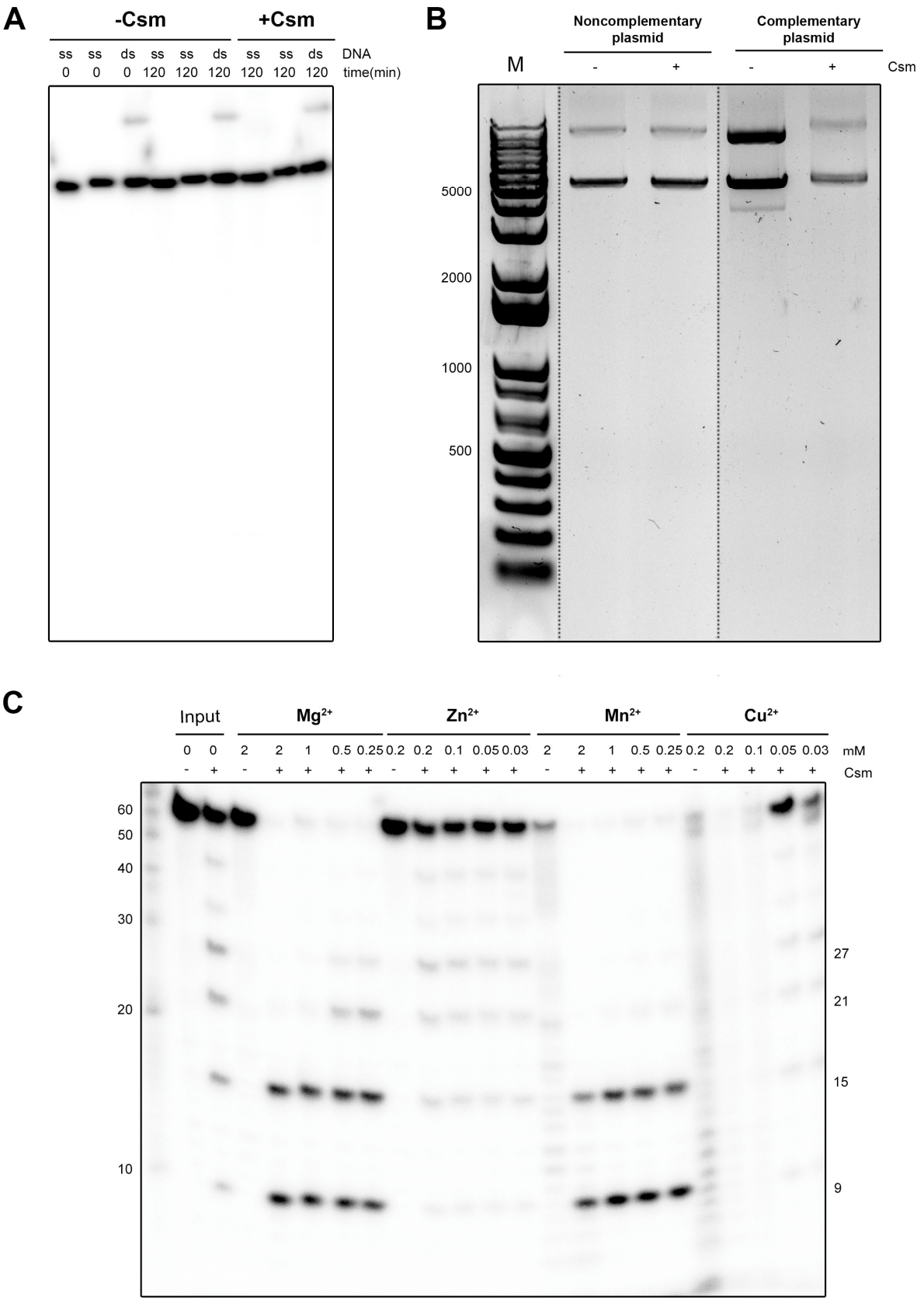

y1
175.1188
b2
258.1559
b2-NH3
241.1293
y2
290.1458
[Peptide #]2+
427.7280
y4
446.2366
U-H2O
307.0331
y3
389.2156
y5 #
597.2985
b6 #
680.3353
y6 #
753.3368
Peptide #
854.3871
Relative Intensity
m/z
TRVGVDR + [UU]
b6#
y6# y5# y4 y3 y2 y1
= Observed as an adduct with [Fragment of U = -C3O = 52]

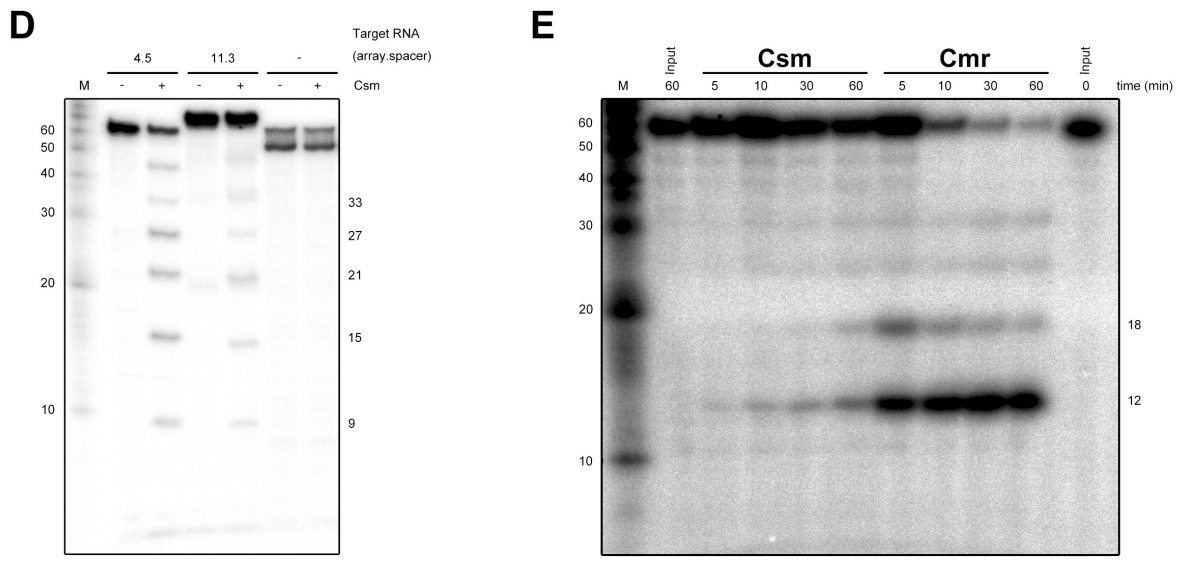

y1
147.1128
IM Y
136.0757
a2
157.0971
b2
185.0920
y2
246.1811
b3
298.1760
y3
333.2133
b4
355.1976
b5
426.2350
b6
589.2969
y6 #
626.3158
y7 #
739.4047
y6 #3
686.3465
y6 #4
800.3753
y6 #5
880.3446
y8 #
902.4615
Relative Intensity
m/z
SPLGAYLPGSSVK + [U]
a2 b2 b3 b4 b5 b6
y8# y7# y6# y5 y4 y3 y2 y1
y6#3
y6#4
y6#5
= Observed as an adduct with [Fragment of U = -C3O = 52]
#3 = Observed as an adduct with [U']
#4 = Observed as an adduct with [U-H3PO4]
#5 = Observed as an adduct with [U-H2O]


a2 #3
429.1804
b2 #3
457.1754
y4
552.2842
b3 #3
570.2595
y3
423.2349
y5
623.3144
b4 #3
683.3328
y6
736.3989
y7
849.4833
IM F
120.0807
y2
322.1877
y1
175.1189
y8
948.5524
Peptide #2
1191.6080
Relative Intensity
m/z
a2#3
b2#3 b3#3 b4#3
MVLLAETFR + [U]
y8 y7 y6 y5 y4 y3 y2 y1
#2 = Observed as an adduct with [U']
#3 = Observed as an adduct with [U-H3PO4]

MS/MS spectra of the protein-crRNA crosslinks identified within the TtCsm complex. For every peptide observed to be crosslinked (as described in Table 1) here we show the sample spectra of the peptide crosslinked with one of the RNA moieties. The peptide sequence and fragment ions are indicated on the top. The crosslinked residues are highlighted in yellow. The peptide is fragmented with the cleavage of amide bonds resulting in fragments retaining the amino-terminal (b –ions) and the carboxy-terminal (y – ions) respectively. Some of the b- and y- ions were observed with a mass shift of #, #^{1}, #^{2}, #^{3}, #^{4}, #^{5} and #^{6} corresponding to -C_3O (a fragment of Uracil), U'-H_2O, U', U-H_3PO_4, U-H_2O, U and U-H_2O + U-H_3PO_4 respectively. Of note, for cross-linked peptides derived from Csm1 (positions 371 – 378) and (positions 21-39) the cross-linked amino acid could not be identified due to a lack of a corresponding mass shift in the b- or y-type fragment ions of the peptide. IM: Immonium ions, U': U marker ion adduct of 112.0273 Da.

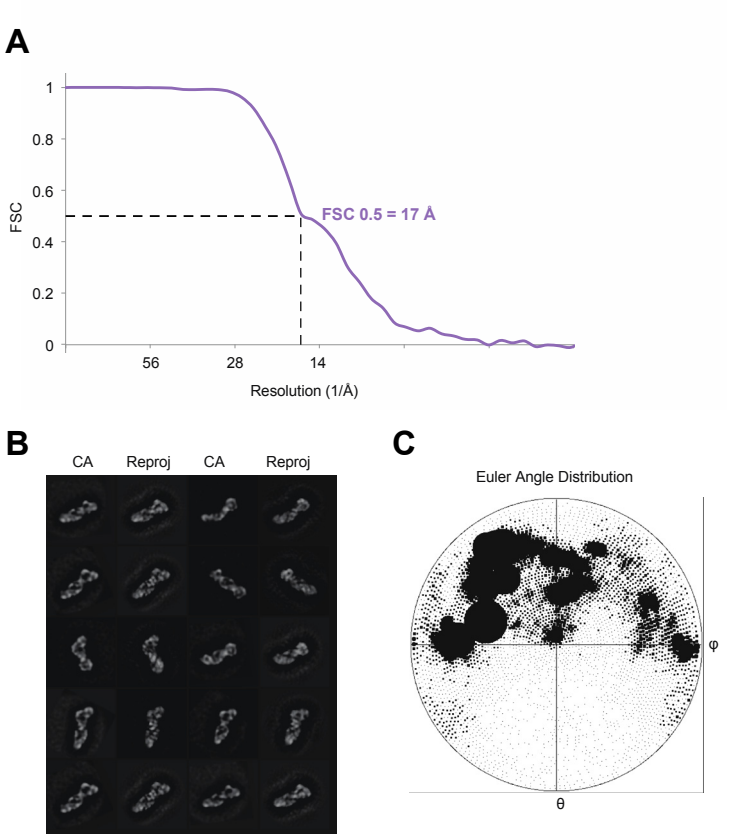
Figure S3 (related to Figure 5)





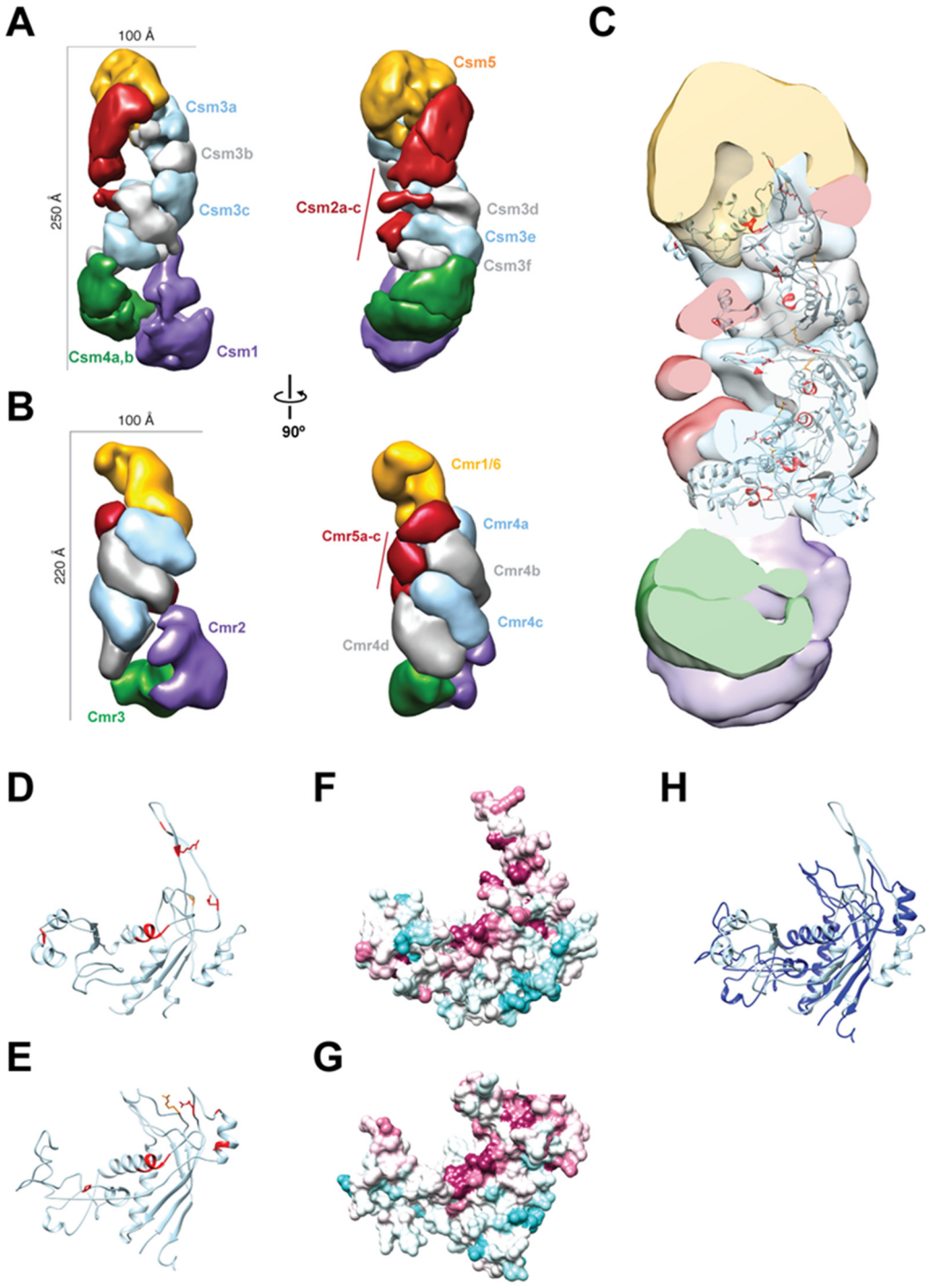
TtCsm *in vitro* activity assays with complementary ssDNA, dsDNA, plasmid DNA and RNA targets. (A) Denaturing gel analysis (20% AA, 7M Urea) of 5' radiolabeled ssDNA (ss) and dsDNA (ds) complementary to crRNA 4.5, which were incubated with ("+") or without ("-") the TtCsm complex for the indicated amount of time. (B) A plasmid was constructed by cloning dsDNA (complementary to crRNA 1.1) on a plasmid ("complementary plasmid"). This plasmid was incubated with ("+") or without ("-") the TtCsm complex and analyzed on a 0.8% agarose gel. The empty cloning vector, pCR2.1-TOPO, was used as a control ("noncomplementary plasmid"). (C) A 50 nt, 5' radiolabeled ssRNA substrate complementary to crRNA 4.5 was incubated with the TtCsm complex in a buffer containing different co-factors (Mg^{2+}, Mn^{2+}, Zn^{2+} and Cu^{2+}) followed by denaturing gel analysis (20% AA, 7M Urea). (D) Csm activity assay with 5' labeled ssRNA substrates complementary to crRNA 4.5 ("4.5") or crRNA 11.3 ("11.3"). Noncomplementary 50 and 60 nt ssRNAs ("-"), derived from the Decade marker bands ("M"), were tested in parallel as negative control. In order to visualize more (transient) degradation products, the assay was performed with a lower (10 µM) Mg^{2+}-concentration. (E) Csm and Cmr activity assays with a 3' labeled ssRNA substrate complementary to crRNA 4.5.

Figure S4 (related to Figure 7)



Architecture of TtCsm. (A) Fourier shell correlation curve indicates the reconstruction has a resolution of ~17 Å at the 0.5 cut-off criterion. (B) Comparison of reprojections of the TtCsm complex reconstruction (even columns, Reproj) with corresponding reference-free 2D class averages (odd columns, CA). Width of each box corresponds to 400 Å. (C) Euler angle distribution of the reconstruction. The size of the spot is proportional to the number of particles that belong to that specific view.

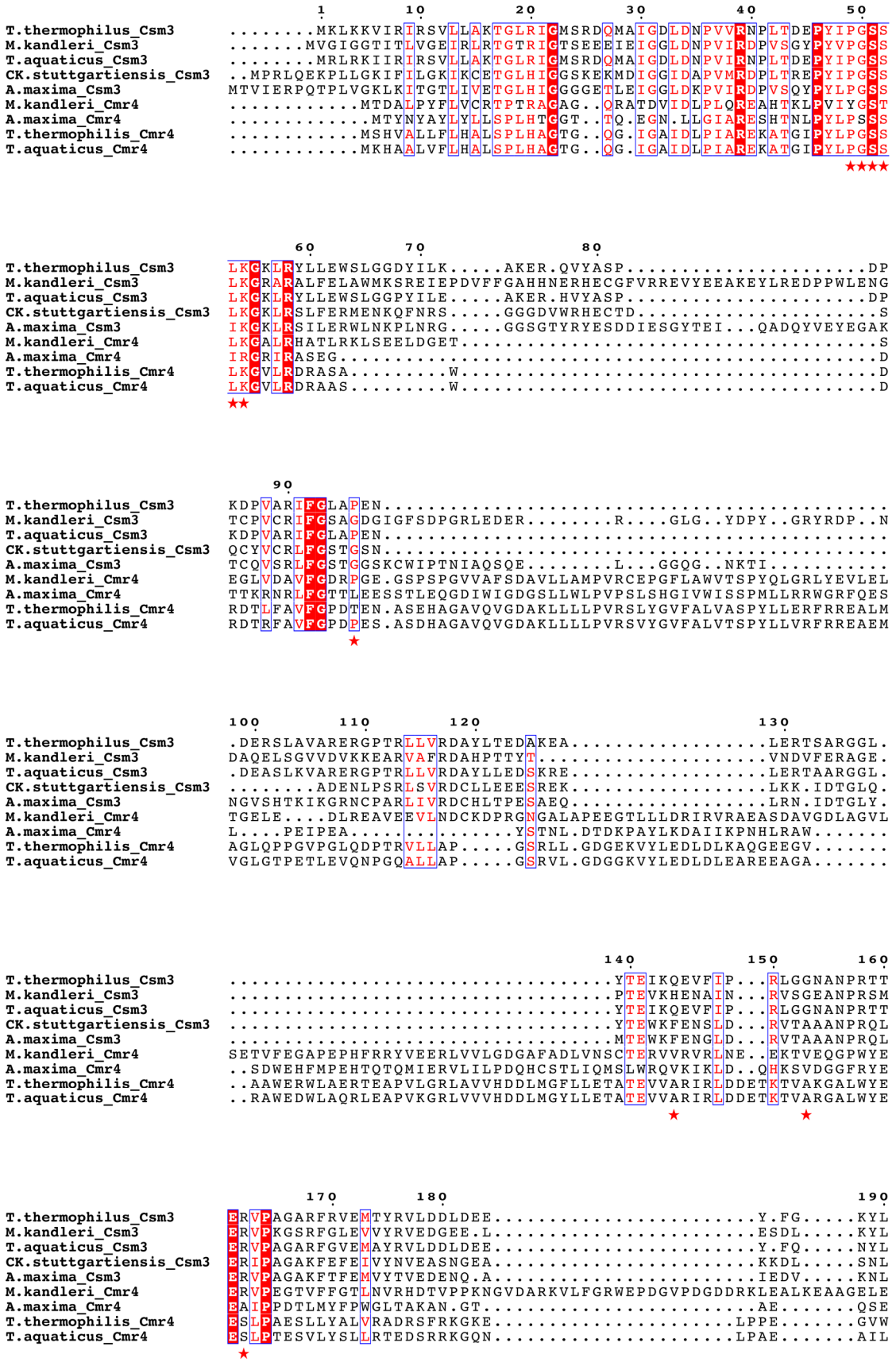
Figure S5 (related to Figure 7)



Comparison of TtCsm and TtCmr and path of crRNA along backbone of TtCsm. Segmentations of TtCsm (A) and TtCmr (B) showing the similarities in subunit organization. (C) Transparent surface of TtCsm reconstruction with Phyre models of Csm3 docked into the corresponding segments. Other subunits have been removed for clarity. Residues in Csm3 are color coded as follows: red, crRNA-protein cross-linking data from this study; green, inter-subunit cross-linking data from this study; crRNA

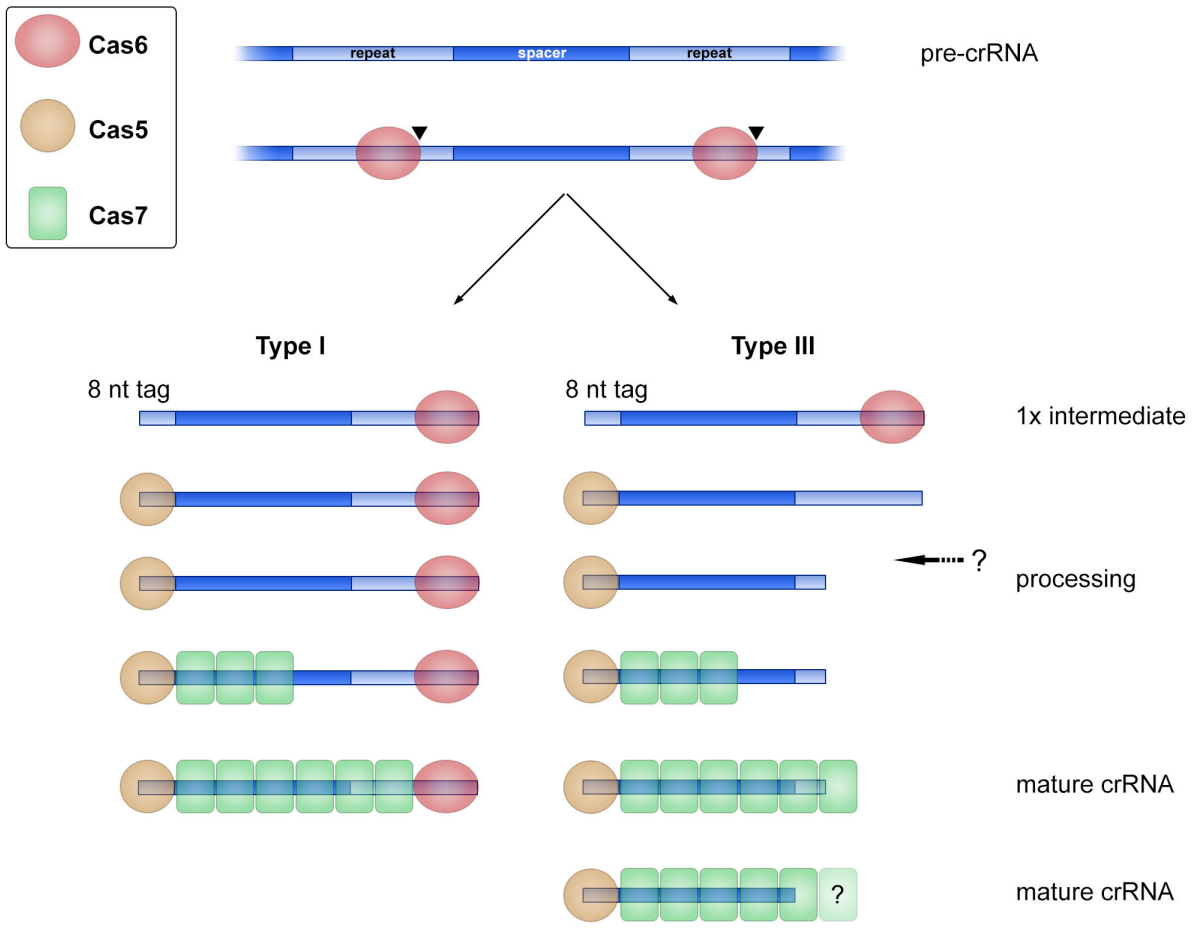
binding residues identified in (Hrle et al., 2013). (D) PHYRE structure prediction of Csm3 (based on *E. coli* Cas7 (Jackson et al., 2014)) with residues colored as in (C) and showing the crRNA-binding thumb of this subunit. (E) PHYRE structure prediction of Csm3 (based on *M. kandleri* Csm3 (Hrle et al., 2013)) with residues colored as in (C) and showing the crRNA-binding thumb of this subunit. (F, G) Surface conservation of the Csm3 structure prediction in (D) and (E), respectively. The surface is colored according to amino acid conservation among Csm3 proteins shown in Fig. S6 by the Consurf Server (Ashkenazy et al., 2010), where purple/red represents highly conserved residues, while white/light blue denotes the most variant residues.

Figure S6 (related to Figure 7)



Multiple sequence alignment of Csm3 and Cmr4. The primary sequence of Csm3 for *Thermus thermophilus* (UniProt: Q53W06), *Methanopyrus kandleri* (UniProt: Q8TVS2), *Thermus aquaticus* (UniProt: B7A9Y4), *Methanosarcina acetivorans* (UniProt: Q8TPH9), *Candidatus Kuenenia stuttgartiensis* (UniProt: Q1Q3H6) and *Arthrospira maxima* (UniProt: B5W7G0) and Cmr4 from *Methanopyrus kandleri* (UniProt: Q8TVT9), *Arthrospira maxima* (UniProt: B5W4P3), *Thermus thermophilus* (UniProt: Q53W06) and *Thermus aquaticus* (UniProt: B7A6X3) were aligned using Clustal Omega (Sievers et al., 2011). The alignment was generated using ESPript with default settings. White letters highlighted in red represent completely conserved residues. Residues with >70% conservation are shown as red letters on a white background with a blue frame. Residues that crosslinked to the crRNA are denoted with red stars.

Figure S7 (related to Figure 3)



Hypothetical model for RNP complex formation and 3' crRNA processing in Type I and III CRISPR-Cas systems. After transcription of the pre-crRNA, crRNA maturation is initiated by the Cas6-mediated endoribonucleolytic cleavages (black triangles) in the repeat sequence. The complex-bound Cas6 protein in Type I systems remains attached to the 3' end of the cleaved repeat sequence, while the 'standalone' Cas6 in Type III systems dissociates, exposing the 3' end of the crRNA for 3'-5' exonucleolytic trimming by an unknown nuclease. The 5' end of the crRNA is bound and protected by the Cas5 family of proteins.

Supplemental Tables

Table S1 (related to Figure 4)

Csm subunit	**Theoretical Mass (Da)**	**Experimental Mass (Da)**
Csm1	90,302.2	90,316.5*
Csm2	15,634.0	15,637,2 ± 0.8
Csm3	27,381.5	27,387.8 ± 2.5
Csm4	32,832.9	32,839.2 ± 2.1
Csm5	44,281.7	44,286.3 ± 1.7
crRNA	15,600.0	n.d.
Model 1 (1:3:6:2:1:1)	**Theoretical Mass (Da)**	**Experimental Mass (Da)**
Csm	427,040.7	426,998.1 ± 217.6
Csm – Csm5	382,759.0	381,896.2 ± 261.3
Model 2 (1:3:2:4:2:1)	**Theoretical Mass (Da)**	**Experimental Mass (Da)**
Csm	427,462.2	426,998.1 ± 217.6
Csm – Csm5	383,180.5	381,896.2 ± 261.3

*exact mass of Csm1 is determined only once.

Exact masses of individual TtCsm subunits (denaturing and tandem MS) and TtCsm complexes (native MS).

Table S2 (related to Figure 4)

Model 1 (proposed stoichiometry 1:3:6:2:1:1)

Mass of (sub) complexes in solution	Theoretical mass (Da)	Mass products (Da)	Annotation	Stoichiometry 1 2 3 4 5 crRNA
426,998.1	427,040.7		Csm	1 3 6 2 1 1
		404,464.8	Csm-?	1 3 6 2 1 1
		22,345.9	?	0 0 0 0 0 0
381,896.2	382,759.0		Csm-Csm5	1 3 6 2 0 1
		359,676.4	Csm-Csm5-?	1 3 6 2 0 1
		22,405.8	?	0 0 0 0 0 0
336,914.9	336,738.5		Csm-Csm1	0 3 6 2 1 1
318,728.8	318,658.1		Csm-2*Csm2-Csm4-Csm5	1 1 6 1 0 1
289,683.3	291,276.6		Csm-2*Csm2-Csm3-Csm4-Csm5	1 1 5 1 0 1
	258,443.7	256,450.3	Csm-2*Csm2-Csm3-2*Csm4-Csm5	1 1 5 0 0 1
	231062.2	228,950.9	Csm-2*Csm2-2*Csm3-2*Csm4-Csm5	1 1 4 0 0 1
	32,832.9	32,822.2	Csm4	0 0 0 1 0 0
	27,381.5	27,373.6	Csm3	0 0 1 0 0 0
273,803.6	275,642.6		Csm-3*Csm2-Csm3-Csm4-Csm5	1 0 5 1 0 1
	242,809.7	242,308.2	Csm-3*Csm2-Csm3-2*Csm4-Csm5	1 0 5 0 0 1
	215,428.2	213,391.0	Csm-3*Csm2-2*Csm3-2*Csm4-Csm5	1 0 4 0 0 1
	32,832.9	32,834.1	Csm4	0 0 0 1 0 0
260,064.9	259,623.9		Csm-Csm1-Csm4-Csm5	0 3 6 1 0 1
	226,791.0	226,631.5	Csm-Csm1-2*Csm4-Csm5	0 3 6 0 0 1
	199,409.5	199,171.1	Csm-Csm1-Csm3-2*Csm4-Csm5	0 3 5 0 0 1
	32,832.9	32,823.0	Csm4	0 0 0 1 0 0
	27,381.5	27,366.8	Csm3	0 0 1 0 0 0
244,252.1	243,989.9		Csm-Csm1-Csm2-Csm4-Csm5	0 2 6 1 0 1
	211,157.0	211,502.5	Csm-Csm1-Csm2-2*Csm4-Csm5	0 2 6 0 0 1
	32,832.9	32,828.4	Csm4	0 0 0 1 0 0
181,649.2	183,775.5		Csm-Csm1-Csm2-Csm3-2*Csm4-Csm5	0 2 5 0 0 1
90,952.1	89,883.5		Csm-Csm1-5*Csm3-2*Csm4-Csm5	0 3 1 0 0 1

Model 2 (proposed stoichiometry 1:3:2:4:2:1)

Mass of (sub) complexes in solution	Theoretical mass (Da)	Mass products (Da)	Annotation	Stoichiometry 1 2 3 4 5 crRNA
426,998.1	427,462.2		Csm	1 3 2 4 2 1
		404,464.8	Csm-?	1 3 2 4 2 1
		22,345.9	?	0 0 0 0 0 0
381,896.2	383,180.5		Csm-Csm5	1 3 2 4 1 1
		359,676.4	Csm-Csm5-?	1 3 2 4 1 1
		22,405.8	?	0 0 0 0 0 0
336,914.9	338,898.8		Csm-2*Csm5	1 3 2 4 0 1
318,728.8	318,780.9		Csm-Csm2-Csm3-2*Csm4	1 2 1 2 2 1
289,683.3	290,133.2		Csm-Csm3-2*Csm4-Csm5	1 3 1 2 1 1
	257,300.3	256,450.3	Csm-Csm3-3*Csm4-Csm5	1 3 1 1 1 1
	229,918.8	228,950.9	Csm-2*Csm3-3*Csm4-Csm5	1 3 0 1 1 1
	32,832.9	32,822.2	Csm4	0 0 0 1 0 0
	27,381.5	27,373.6	Csm3	0 0 1 0 0 0
273803.6	274,499.2		Csm-Csm2-Csm3-2*Csm4-Csm5	1 2 1 2 1 1
	241,666.3	242,308.2	Csm-Csm2-Csm3-3*Csm4-Csm5	1 2 1 1 1 1
	214,284.8	213,391.0	Csm-Csm2-2*Csm3-3*Csm4-Csm5	1 2 0 1 1 1
	32,832.9	32,834.1	Csm4	0 0 0 1 0 0
260,064.9	258,865.2		Csm-2*Csm2-Csm3-2*Csm4-Csm5	1 1 1 2 1 1
	226,032.3	226,631.5	Csm-2*Csm2-Csm3-3*Csm4-Csm5	1 1 1 1 1 1
	198,650.8	199,171.1	Csm-2*Csm2-2*Csm3-3*Csm4-Csm5	1 1 0 1 1 1
	32,832.9	32,823.0	Csm4	0 0 0 1 0 0
	27,381.5	27,366.8	Csm3	0 0 1 0 0 0
244,252.1	243,231.2		Csm-3*Csm2-Csm3-2*Csm4-Csm5	1 0 1 2 1 1
	210,398.3	211,502.5	Csm-3*Csm2-Csm3-3*Csm4-Csm5	1 0 1 1 1 1
	32,832.9	32,828.4	Csm4	0 0 0 1 0 0
229,238.5	229,918.8		Csm-2*Csm3-3*Csm4-Csm5	1 3 0 1 1 1
	197,085.9	196,373.6	Csm-2*Csm3-4*Csm4-Csm5	1 3 0 0 1 1
	32,832.9	32,833.7	Csm4	0 0 0 1 0 0
181,649.2	181,451.9		Csm-Csm2-2*Csm3-4*Csm4-Csm5	1 2 0 0 1 1
90,952.1	92,714.6		Csm-3*Csm2-2*Csm3-3*Csm4-Csm5	0 0 0 1 1 1

Overview of the experimental masses for all Csm (sub)complexes present in solution, matched against the 2 proposed stoichiometries. For each complex the theoretical mass (based on the protein amino acid sequence and estimated crRNA mass of 15,600 Da) and stoichiometric information is given. 1=Csm1, 2=Csm2, 3=Csm3, 4=Csm4, 5=Csm5 and minus (-) indicates the elimination of that subunit, n.d. is not determined.

Table S3 (related to Figures 2, 5 and 6)

oligo name	Sequence (5' to 3')	comments
P1; P2	AAGCTTGGACCTCTACCGCGACCCCTTCCGGGCGGT; TCTAGATCATCAGTGGTGGTGGTGGTGGTGGAGGGGCTCTAGCCTCCCCACCATC CAGCCTAAGG	Construction of the plasmid pUC-csm5h
P3; P4	CTGCAGCTCACCAGCGGCACCAAGGCCATGAGCGCG; GAATTCGGGCGAGGCCGTACACCCCCTCCTTAAGGG	
P5; P6	AAGCTTCCTGAAGGCCCGGGACTTCGCCCTTAAGGA; TCTAGATCATCAGTGGTGGTGGTGGTGGTGAAACCCCAGGGGGACGGGCTCCGGG GAAAGGGGGC	Construction of the plasmid pUC-csx1h
P7; P8	CTGCAGCCTTGACCTGGGATCACCGCCCCTCCCGGT; GAATTCTGAGGGTTTTTGAGGGCTTACACCGATAGA	
P9	GAACTGCGCCTTGACGTGGTCGTCCCCGGGCGCCTTATCTACGGCCATCG	Target DNA (complementary to crRNA 4.5)
P10	CGATGGCCGTAGATAAGGCGCCCGGGGACGACCACGTCAAGGCGCAGTTC	Reverse complement of P9, to generate dsDNA target
P11	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGGCCAUCG	Wildtype target RNA (complementary to crRNA 4.5)
P12	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGCCCAUCG	Mutated +1 target RNA (complementary to crRNA 4.5)
P13	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACCGCCAUCG	Mutated +2 target RNA (complementary to crRNA 4.5)
P14	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUAGGGCCAUCG	Mutated +3 target RNA (complementary to crRNA 4.5)
P15	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUUCGGCCAUCG	Mutated +4 target RNA (complementary to crRNA 4.5)
P16	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCAACGGCCAUCG	Mutated +5 target RNA (complementary to crRNA 4.5)
P17	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUGUACGGCCAUCG	Mutated +6 target RNA (complementary to crRNA 4.5)
P18	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAACUACGGCCAUCG	Mutated +7 target RNA (complementary to crRNA 4.5)
P19	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCAUGCCCCAUCG	Mutated 1 to 5 target RNA (complementary to crRNA 4.5)
P20	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCGAAUACUACGGCCAUCG	Mutated 7 to 11 target RNA (complementary to crRNA 4.5)
P21	GAACUGCGCCUUGACGUGGUCGUCCCCCCCGCCCUUAUCUACGGCCAUCG	Mutated 13 to 17 target RNA (complementary to crRNA 4.5)
P22	GAACUGCGCCUUGACGUGGUCCAGGGCGGGCGCCUUAUCUACGGCCAUCG	Mutated 19 to 23 target RNA (complementary to crRNA 4.5)
P23	GAACUGCGCCUUGACCACCACGUCCCCGGGCGCCUUAUCUACGGCCAUCG	Mutated 25 to 29 target RNA (complementary to crRNA 4.5)

P24	GAACUGCGCGAACUCGUGGUCGUCCCCGGGCGCCUUAUCUACGGCCAUCG	Mutated 31 to 35 target RNA (complementary to crRNA 4.5)

Oligonucleotides used in this study. Sequences in yellow indicate the base pair-disrupting mutations in the target RNAs used for the *in vitro* activity assays.

Supplementary Experimental Procedures

Construction and cultivation of the *T. thermophilus* HB8 strain producing the $(His)_6$-tagged protein

In order to produce the C-terminal $(His)_6$-tagged Csm5 in *T. thermophilus* HB8, the tag-coding sequence was inserted within the genome by homologous recombination. The plasmid pUC-csm5h, used for the homologous recombination, was constructed as follows. A DNA fragment (fragment 1; 570-bp *Hin*dIII-*Xba*I fragment) carrying the 3′-terminal coding region of *csm5* (positions 141,509 to 142,049 on the megaplasmid pTT27) followed by a $(His)_6$ tag, and another DNA fragment (fragment 2; 510-bp *Pst*I-*Eco*RI fragment) carrying the downstream region of *csm5* (positions 142,456 to 142,981 on the megaplasmid pTT27), were amplified by genomic PCR using the primers P1/P2 and P3/P4 (**Table S1**), respectively, and then cloned into pUC19 (*Hin*dIII-*Eco*RI sites) together with the thermostable kanamycin-resistance marker gene (Hashimoto et al., 2001) (1.1-kbp *Xba*I-*Pst*I fragment), to construct pUC-csm5h. The plasmid pUC-csx1h, used for insertion of the $(His)_6$-tag-coding sequence at the 3' of the *csx1* gene in *T. thermophilus* HB8, was constructed as follows. A DNA fragment (fragment 3; 560-bp *Hin*dIII-*Xba*I fragment) carrying the 3′-terminal coding region of *csx1* (positions 142,926 to 143,454 on the megaplasmid pTT27) followed by a $(His)_6$ tag, and another DNA fragment (fragment 4; 530-bp *Pst*I-*Eco*RI fragment) carrying the downstream region of *csx1* (positions 143,514 to 144,037 on the megaplasmid pTT27), were amplified by genomic PCR using the primers P5/P6 and P7/P8 (**Table S1**), respectively, and then cloned into pUC19 (*Hin*dIII-*Eco*RI sites) together with the thermostable kanamycin-resistance marker gene, to construct pUC-csx1h.

Plasmid pUC-csm5h or pUC-csx1h was introduced into the *T. thermophilus* HB8 strain, and kanamycin-resistant clone was obtained as described previously (Hashimoto et al., 2001). In the strain, the downstream region of the *csm5* gene on the genome (positions 142,050 to 142,455) or that of the *csx1* gene (position 143,455 to 143,514) is replaced by the $(His)_6$ tag and two stop codons, followed by the kanamycin-resistance marker gene. The *T. thermophilus* HB8 cells producing the $(His)_6$-tagged proteins were cultured at 70°C in a rich (TT) medium (Agari et al., 2008) until an A_{600} = 1.5 to 4.5 was attained.

Detailed description of the purification of the TtCsm complex and identification of the Csm proteins

The *T. thermophilus* HB8 cells producing the $(His)_6$-tagged Csm complex were resuspended in 20 mM Tris-HCl (pH 8.0), containing 50 mM NaCl and 0.1 mM phenylmethylsulfonyl fluoride, disrupted by sonication in ice water, and then ultracentrifuged (200,000 × *g*) for 1 h at 4°C. The supernatant was applied to a HisTrap HP column (GE Healthcare), pre-equilibrated with 20 mM Tris-HCl (pH 8.0), containing 0.15 M NaCl, and then the bound protein was eluted with a linear gradient of 0 to 0.5 M imidazole. The target fractions were collected, and desalted by fractionation on a HiPrep 26/10 desalting column (GE Healthcare). The sample was then applied to a RESOURCE Q column (GE Healthcare), pre-equilibrated with 20 mM Tris-HCl (pH 8.0), and the bound protein was eluted with a linear gradient of 0 to 0.5 M NaCl. The target fraction was collected and concentrated. The sample was then applied to a HiLoad 16/60 Superdex 200 pg column (GE Healthcare), pre-equilibrated with 20 mM Tris-HCl (pH 8.0) containing 0.15 M NaCl. The target fractions were collected, and desalted by fractionation on a HiPrep 26/10 desalting column, pre-equilibrated with 20 mM Tris-HCl (pH 8.0). The sample was then applied to HiTrap Heparin column (GE Healthcare), pre-equilibrated with 20 mM Tris-HCl (pH 8.0), and the bound protein was eluted with a linear gradient of 0 to 1 M NaCl. The target fractions were collected, and desalted by fractionation on a HiPrep 26/10 desalting column, pre-equilibrated with 10 mM sodium phosphate buffer (pH 7.0). The sample was then applied to the CHT2-1 column (Bio-Rad Laboratories, Inc.), pre-equilibrated with 10 mM sodium phosphate buffer (pH 7.0), and eluted with a linear gradient of 10 to 500 mM sodium phosphate buffer (pH 7.0).

The components of the complex were identified using a peptide mass fingerprinting method. Briefly, the purified complex was subjected to SDS-PAGE and visualized by staining with Coomassie Brilliant Blue R-250**.** Each protein band was excised and digested by in-gel digestion with L-(tosylamido-2-phenyl) ethyl chloromethyl ketone-treated trypsin. The digestion mixtures were mixed with α-cyano-4-hydroxycinnamic acid as a matrix and subjected to matrix-assisted laser desorption ionization time-of-flight mass spectrometry (Bruker Daltonics Inc., Germany, Ultraflex). The lists of observed monoisotopic peptide ion peaks were searched in the NCBI database using MASCOT (Matrix science Inc., Boston, MA).

RNAseq analysis

crRNAs were purified from the TtCsm complexes by phenol-chloroform-isoamyl alcohol (PCI) extraction followed by ethanol precipitation. crRNAs were phosphatase and T4 polynucleotide kinase (PNK) treated prior to library preparation using the Illumina TruSeq Small RNA Sample Preparation Kit. Different adapters ligated to the 5' and 3' ends of the crRNAs allowed for subsequent orientation of the sequencing reads. The ligated RNAs were then reverse transcribed and amplified by PCR. The resulting library was sequenced using 2 × 100 bp reads (Paired-End) on a HiSeq Illumina platform (Plateforme de Séquençage à Haut Débit Imagif, Gif-sur-Yvette, France). A total of 73,695,063 mate-paired reads were obtained and were aligned using blast (non-overlapping reads were removed: 17,604,430 reads). The adapter-stripped reads were mapped to the genome of *T. thermophilus* with Bowtie2 using the default settings (Langmead and Salzberg, 2012). Reads containing any insertions, deletions, mismatches or reads that mapped multiple times (e.g. the 8 nt repeat-derived sequences) with the reference genome were discarded, resulting in 52,823,733 (94.18%) mapped reads. Visualization was performed using Microsoft Excel and Matplotlib (Hunter, 2007).

UV-crosslinking and identification of crRNA-protein interactions by LC-MS/MS

Around 1 nmol of the TtCsm complex was resuspended in 100 µl of 20 mM Tris-HCl (pH 8.0) and 150 mM NaCl. The complex was incubated at 65°C for 10 min. The samples were then transferred to black polypropylene microplates (Greiner Bio-One) and irradiated at 254 nm for 10 min at room temperature as described previously (Kramer et al., 2011). The samples were ethanol precipitated and the pellet was dissolved in 4 M urea and 50 mM Tris-HCl pH 7.9. The final concentration of urea was then adjusted to 1 M with 50 mM Tris-HCl pH 7.9, and the RNA was hydrolysed using 1 µg RNase A and T1 (Ambion, Applied Biosystems) for 2 h at 52°C. Following RNA digestion, the sample was digested with trypsin (Promega) at 37°C overnight. The sample was desalted to remove non cross-linked RNA fragments using an in-house prepared C18 (Dr. Maisch GmbH) column, and the cross-linked peptides were enriched on an in-house prepared TiO_2 (GL Sciences) (Kramer et al., 2011). The samples were then dried and resuspended in 12 µl sample solvent (5% v/v ACN, 1% v/v FA) for mass spectrometry analysis. The sample was injected onto a nano-liquid chromatography system (Agilent 1100 series, Agilent Technologies) coupled with a LTQ-Orbitrap Velos instrument (Thermo Scientific) as described previously (Christian et al., 2014). Online ESI-MS was performed in data-dependent mode using a

TOP10 HCD method. All precursor ions as well as fragment ions were scanned in the Orbitrap, and the resulting spectra were measured with high accuracy (< 5 ppm) both in the MS and MS/MS level. Data analysis was done essentially as described previously (Christian et al., 2014), using a dedicated database search tool (Urlaub lab, unpublished data).

In vitro activity assays

All DNA and RNA substrates were purchased from Integrated DNA Technologies (IDT) or Eurogentec. A full list of all the oligonucleotides is provided in Table S3. 5' terminally labeled DNA or RNA substrates were generated with T4 polynucleotide kinase (PNK) and ^{32}P γ-ATP (Perkin Elmer), followed by denaturing gel purification (20% acrylamide, 7 M Urea). 3' terminal labeling of RNA was performed with T4 RNA Ligase 1 and ^{32}P pCp (Perkin Elmer) followed by denaturing gel purification. *In vitro* activity assays were performed by incubating the substrate with 100 nM of the Csm complex (unless indicated otherwise) at 65°C for 1 h in a buffer containing: 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 mM DTT, 1 mM ATP and 2 mM $MgCl_2$ (unless indicated otherwise). After incubation, an equal volume of formamide RNA loading buffer was added and incubated for 5 min at 95°C. Samples and 5' labeled ssDNA markers or ssRNA Decade Markers (Ambion) were analyzed by denaturing PAGE (20% acrylamide, 7M Urea) and visualized by autoradiography.

Native mass spectrometry

TtCsm was buffer exchange to 0.175 M ammonium acetate (pH 7.9) at 40°C, using five sequential steps on a centrifugal filter with a cut-off of 10 kDa (Sartorius). The TtCsm complex was kept at room temperature and sprayed at a concentration of 1 µM from borosilicate glass capillaries. A modified Exactive plus (EMR,Thermo Scientific, USA) (Rosati et al., 2012; Rose et al., 2012; Snijder et al., 2014) and modified quadrupole time-of-flight instrument (Waters, United Kingdom) adjusted for optimal performance in high mass detection was used (van den Heuvel et al., 2006). Exact mass measurements of the individual TtCsm proteins were acquired under denaturing conditions (10% formic acid, 50% / 50% ACN/Mq + 0.2% FA). TtCsm was heated to 65°C prior to buffer exchange (performed at 40°C). Although we attempted other organic modifiers, subcomplexes in solution were generated successfully by the addition of 30% DMSO or alternatively by acidifying the used buffer with acetic acid (to a pH of 3.6 - 4). Instrument settings for the modified Qtof were as follows: needle voltage ~1.3 kV, cone voltage

~175 V, source pressure 10 mbar. Xenon was used as the collision gas for tandem mass spectrometric analysis at a pressure of 2 x 10^{-2} mbar. The collision voltage was varied between 10–200 V. The voltages on the flatapoles and transport octapoles were manually tuned to enhance transmission of protein ions on a modified Exactive plus with capillary voltage between 1.2 – 1.4 kV. For the highly charged protein, Xenon was used in the HCD cell at a pressure of 5 x 10^{-10} mbar, with acceleration voltages between 5 – 100V to increase sensitivity, desolvation and dissociation. Both instruments were calibrated using a cluster of Caesium Iodide (25 mg/ml).

Single particle electron microscopy and analysis

Micrographs were recorded automatically using the MSI-Raster application within Leginon on a 4k x 4k Gatan CCD camera at a nominal magnification of ×80,000 (1.45 Å/pixel at the specimen level) with a randomly set defocus range (–0.5 to –1.3 um) and a dose of ~20 $e^{-}A^{-2}$. We used the Appion image-processing environment to automatically select ~60,000 TtCsm particles using FindEM (Roseman, 2004), with Type I-E Cascade class averages as templates. The contrast transfer function (CTF) was estimated using ACE2 (Mallick et al., 2005) within Appion. Micrographs were CTF corrected using ACE2, and the negatively stained TtCsm complexes were extracted using boxes of 288 × 288 pixels. These particles were subjected to reference-free alignment and classification using multivariate statistical analysis and multi-reference alignment in IMAGIC (Tang et al., 2007) into a total of ~300 classes.

We used the *E. coli* Cascade structure (Wiedenheft et al., 2011) low-pass filtered to 60 Å as an initial model for three-dimensional reconstruction using iterative projection matching refinement with libraries from the EMAN2 and SPARX software packages (Hohn et al., 2007; Tang et al., 2007) as described previously (Lander et al., 2009; Wiedenheft et al., 2011). The reconstruction showed structural features to 17 Å resolution (based on the 0.5 FSC criterion), with excellent agreement between reference-free 2D class averages and reprojections of the structure, and displayed a large distribution of Euler angles, despite some preferential orientations of the particles on the carbon film (**Figure S4A-C**). The reconstruction was segmented automatically using Segger (Pintilie and Chiu, 2012) in Chimera (Pettersen et al., 2004) based on the biochemical analyses and MS results. All atomic structures shown were generated using the PHYRE automatic fold recognition server (Kelley and Sternberg, 2009) and the amino acid sequence of the respective *T. thermophilus* protein.

Supplemental References

Agari, Y., Kashihara, A., Yokoyama, S., Kuramitsu, S., and Shinkai, A. (2008). Global gene expression mediated by Thermus thermophilus SdrP, a CRP/FNR family transcriptional regulator. Molecular microbiology *70*, 60-75.

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 research *38*, W529-533.

Christian, H., Hofele, R.V., Urlaub, H., and Ficner, R. (2014). Insights into the activation of the helicase Prp43 by biochemical studies and structural mass spectrometry. Nucleic acids research *42*, 1162-1179.

Hashimoto, Y., Yano, T., Kuramitsu, S., and Kagamiyama, H. (2001). Disruption of Thermus thermophilus genes by homologous recombination using a thermostable kanamycin-resistant marker. FEBS letters *506*, 231-234.

Hohn, M., Tang, G., Goodyear, G., Baldwin, P.R., Huang, Z., Penczek, P.A., Yang, C., Glaeser, R.M., Adams, P.D., and Ludtke, S.J. (2007). SPARX, a new environment for Cryo-EM image processing. Journal of structural biology *157*, 47-55.

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.

Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. pp. 90-95.

Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science *345*, 1473-1479.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols *4*, 363-371.

Kramer, K., Hummel, P., Hsiao, H.-H., Luo, X., Wahl, M., and Urlaub, H. (2011). Mass-spectrometric analysis of proteins cross-linked to 4-thio-uracil- and 5-bromo-uracil-substituted RNA. International Journal of Mass Spectrometry *304*, 184-194.

Lander, G.C., Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.W.*, et al.* (2009). Appion: an integrated, database-driven pipeline to facilitate EM image processing. Journal of structural biology *166*, 95-102.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

Mallick, S.P., Carragher, B., Potter, C.S., and Kriegman, D.J. (2005). ACE: automated CTF estimation. Ultramicroscopy *104*, 8-29.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry *25*, 1605-1612.

Pintilie, G., and Chiu, W. (2012). Comparison of Segger and other methods for segmentation and rigid-body docking of molecular components in cryo-EM density maps. Biopolymers *97*, 742-760.

Rosati, S., Rose, R.J., Thompson, N.J., van Duijn, E., Damoc, E., Denisov, E., Makarov, A., and Heck, A.J. (2012). Exploring an orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. Angewandte Chemie *51*, 12992-12996.

Rose, R.J., Damoc, E., Denisov, E., Makarov, A., and Heck, A.J. (2012). High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. Nature methods *9*, 1084-1086.

Roseman, A.M. (2004). FindEM--a fast, efficient program for automatic selection of particles from electron micrographs. Journal of structural biology *145*, 91-99.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J.*, et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology *7*, 539.

Snijder, J., van de Waterbeemd, M., Damoc, E., Denisov, E., Grinfeld, D., Bennett, A., Agbandje-McKenna, M., Makarov, A., and Heck, A.J. (2014). Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by Orbitrap mass spectrometry. Journal of the American Chemical Society *136*, 7295-7299.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. Journal of structural biology *157*, 38-46.

van den Heuvel, R.H., van Duijn, E., Mazon, H., Synowsky, S.A., Lorenzen, K., Versluis, C., Brouns, S.J., Langridge, D., van der Oost, J., Hoyes, J.*, et al.* (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. Analytical chemistry *78*, 7473-7483.

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.