Molecular Cell, Volume *56*

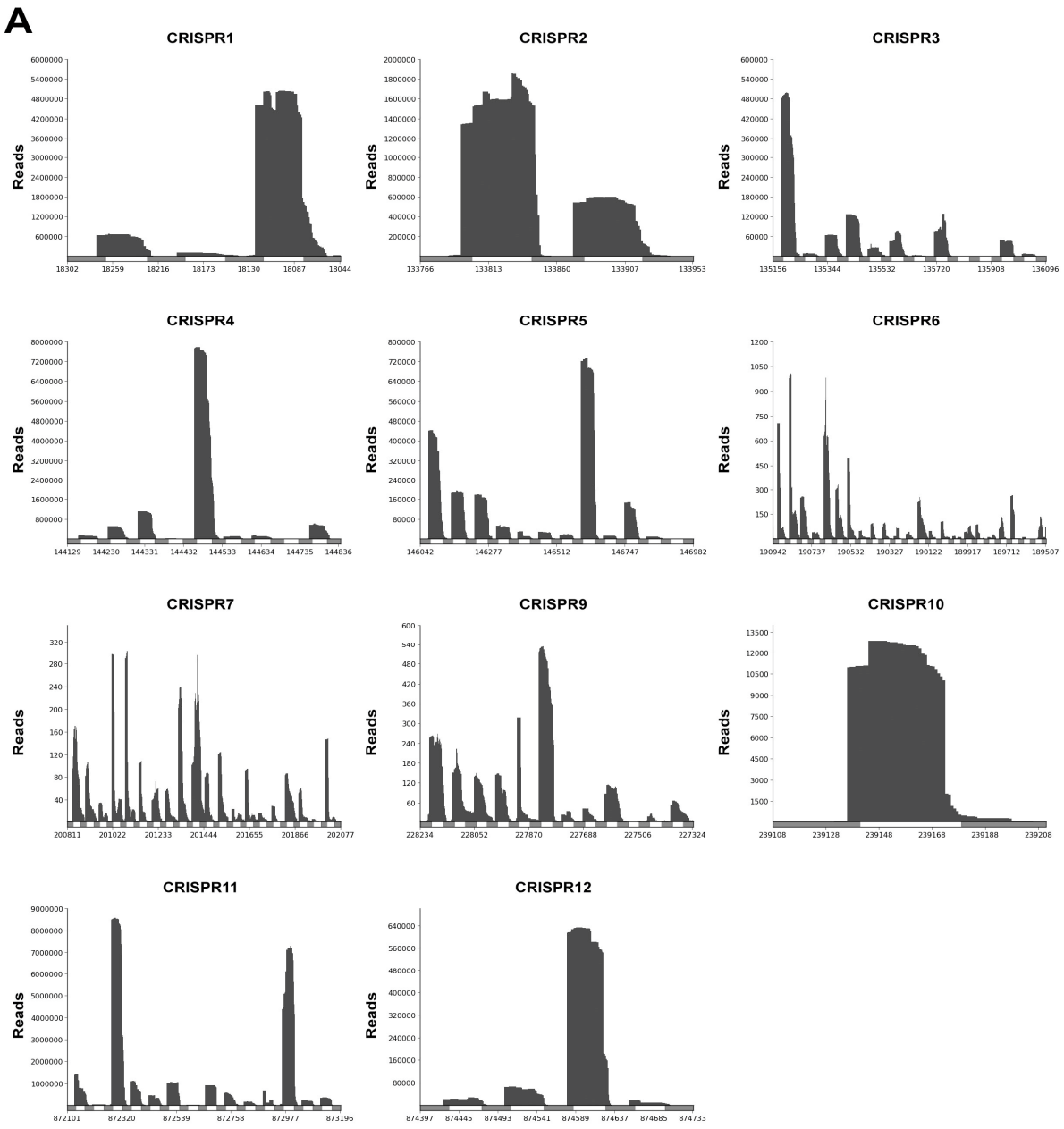
Supplemental Information

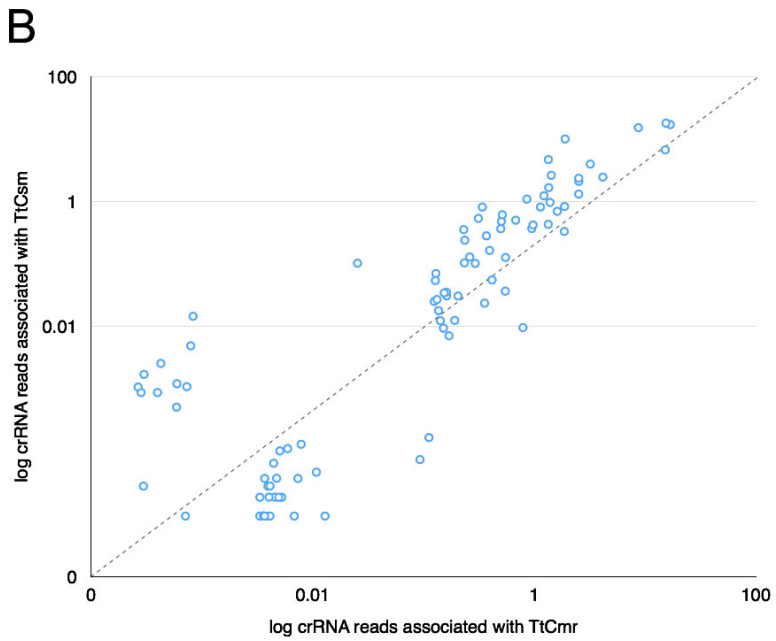
RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of *Thermus thermophilus*

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

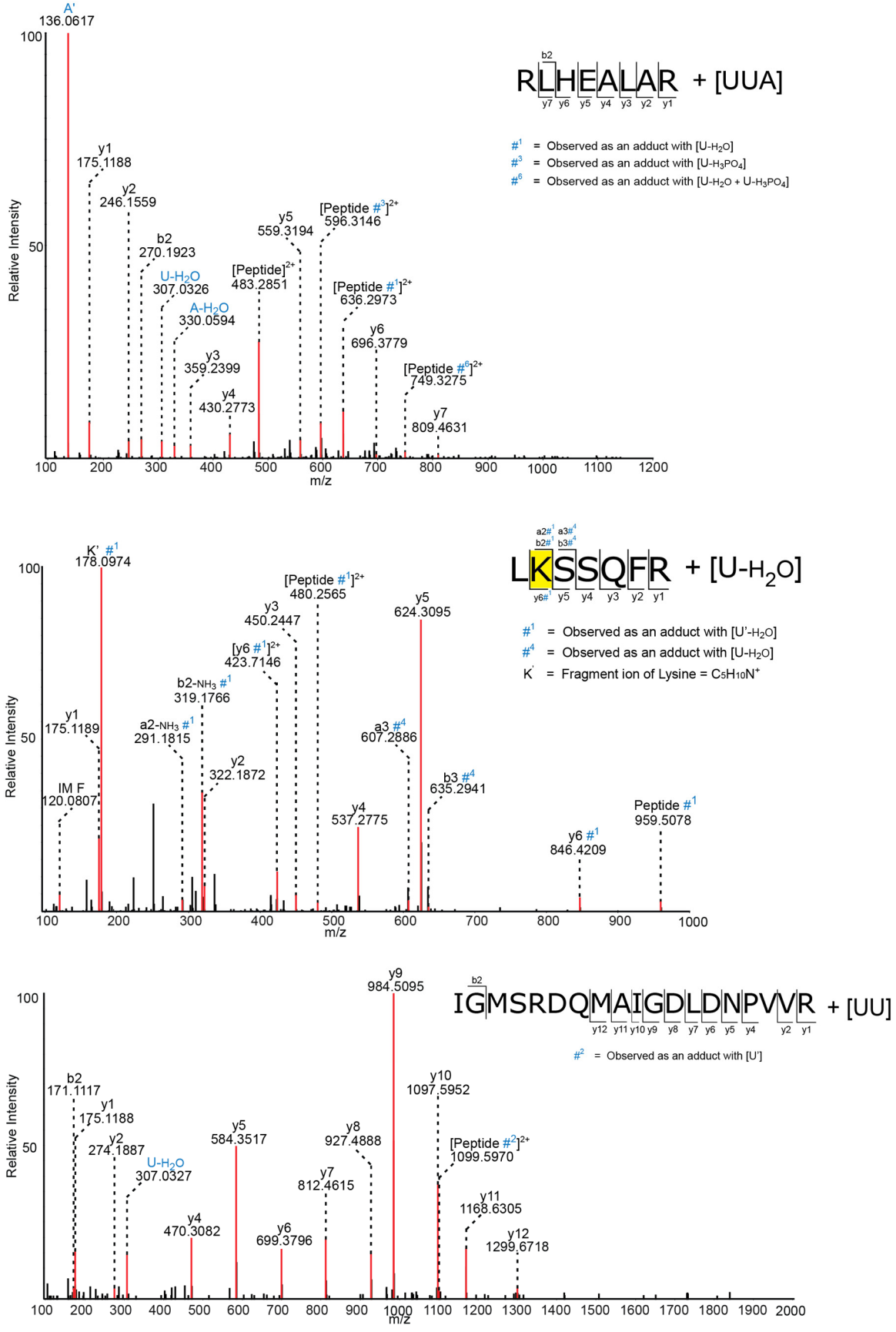
Figure S1 (related to Figure 3)

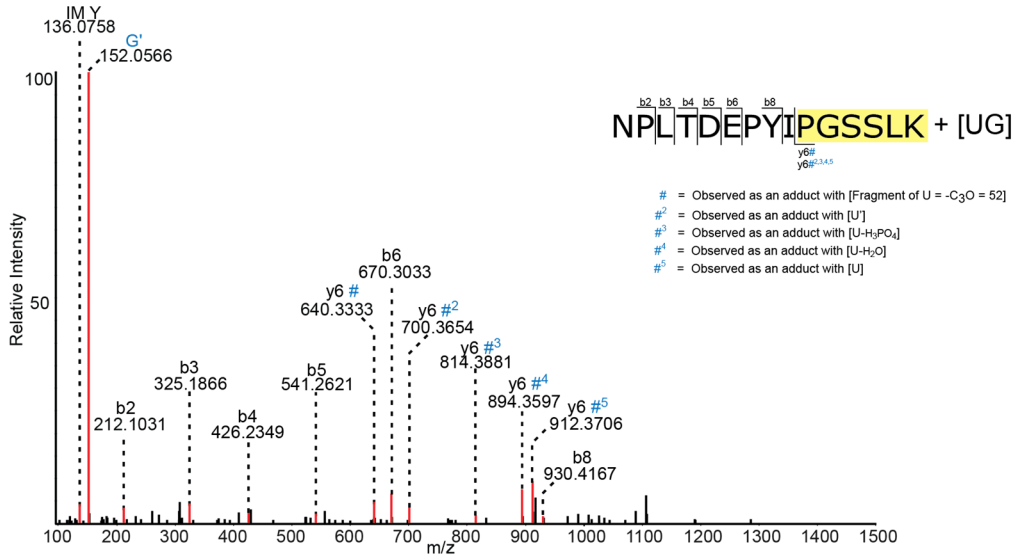




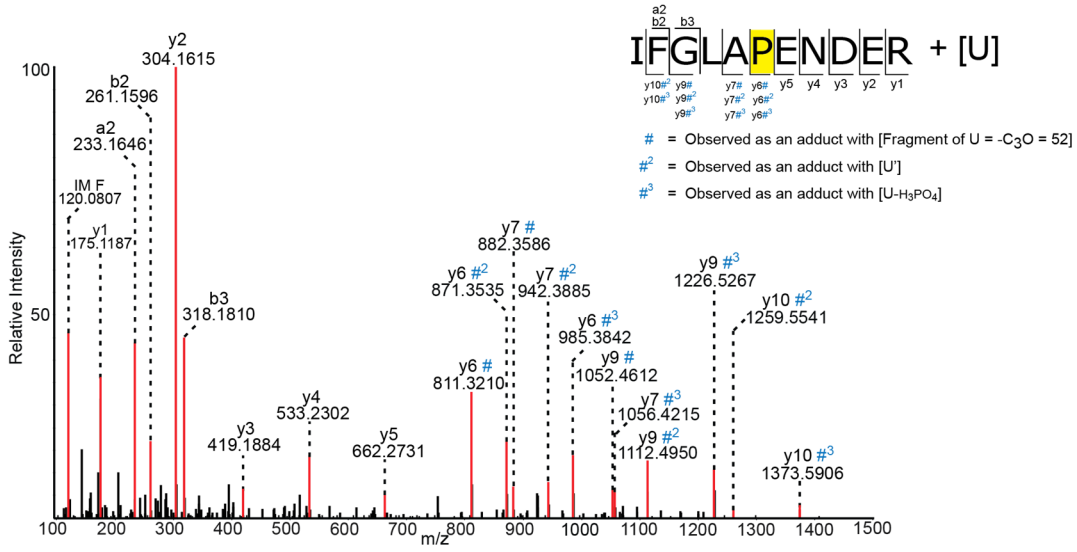
Mapping of the RNAseq data on the *T. thermophilus* CRISPR arrays and spacers and comparison of the crRNA content of the TtCsm and TtCmr complexes. (A) crRNAs were isolated from the TtCsm complex and analyzed by RNAseq. The resulting reads were mapped on the genome of *T. thermophilus* HB8. Depicted 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.

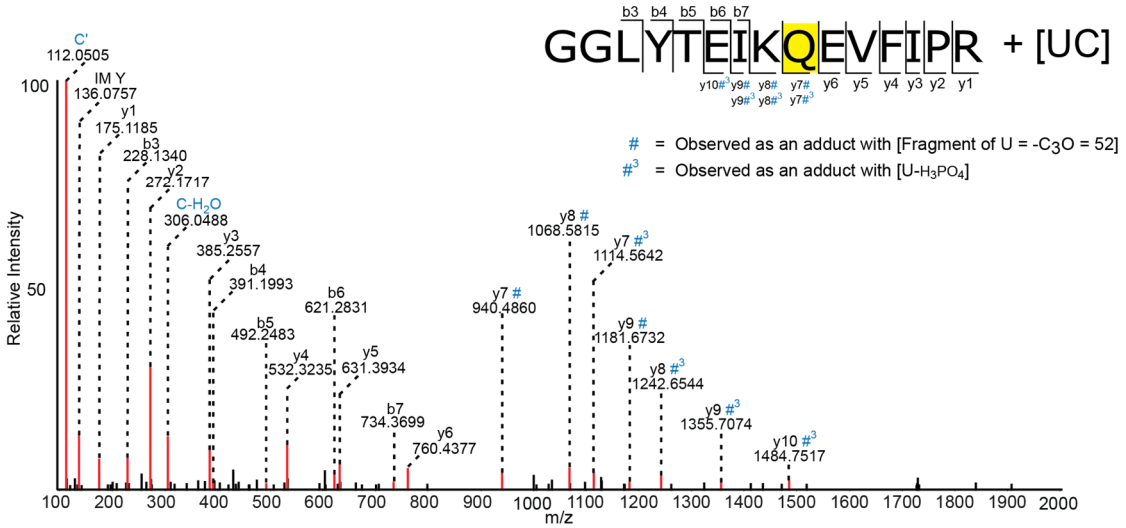
Figure S2 (related to Table 1)

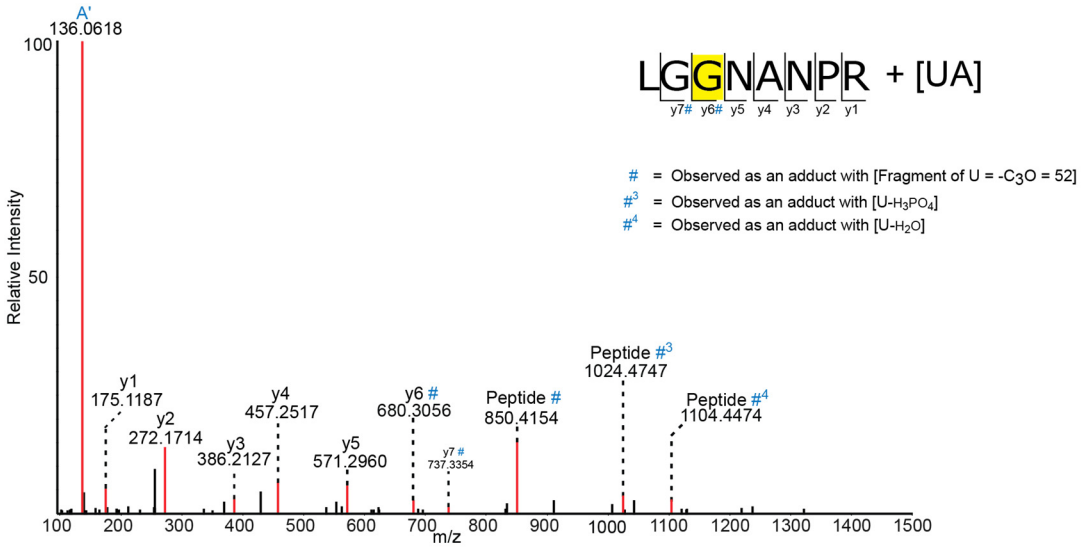


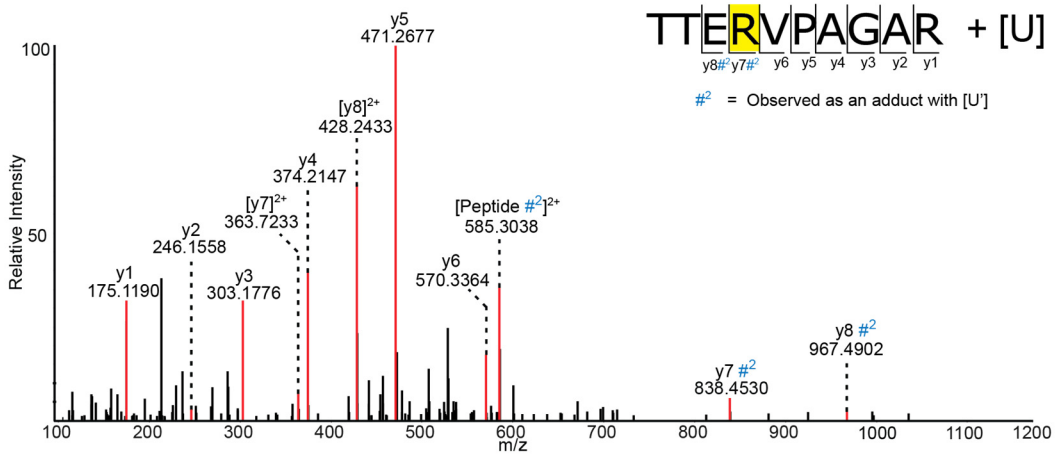


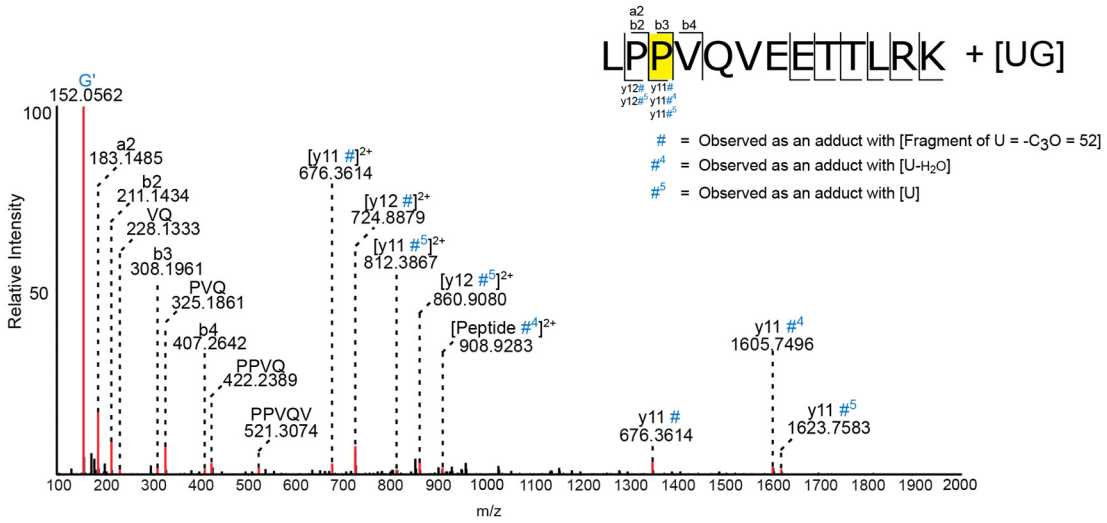
Exact crosslinked residue could not be identified, either of the Amino acids highlighted in yellow can be crosslink site.

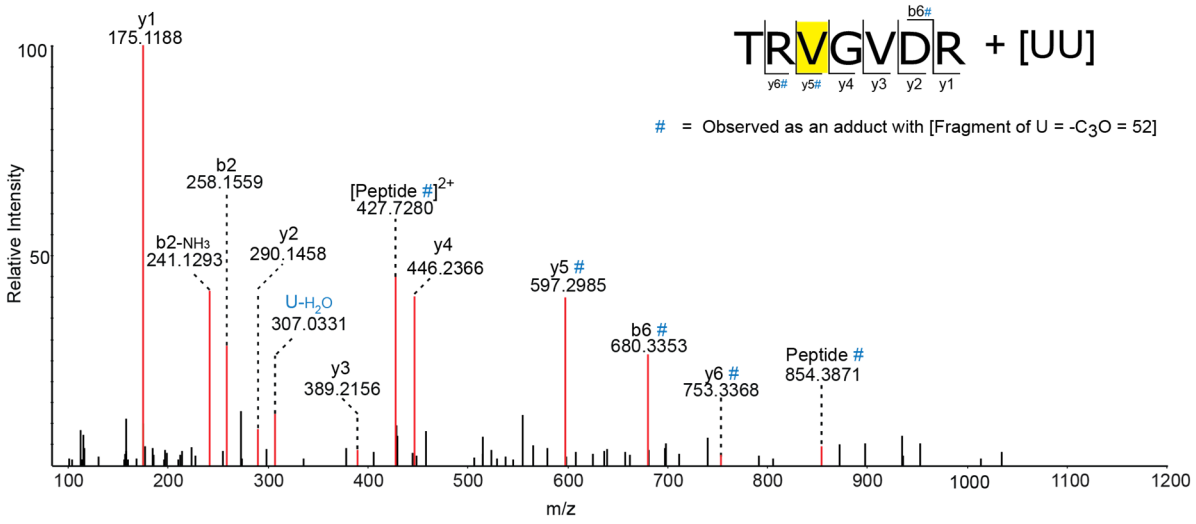


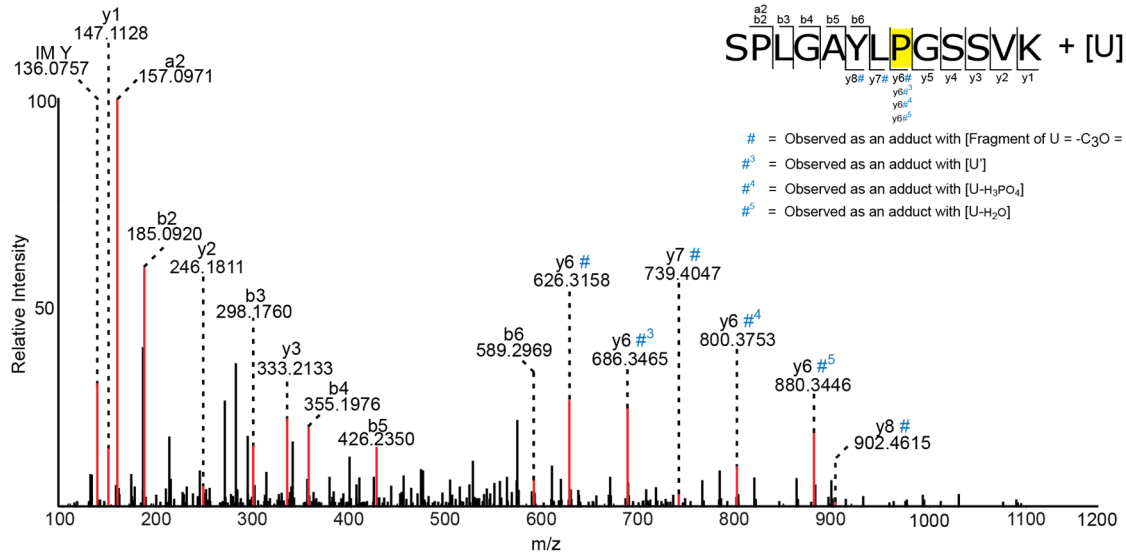


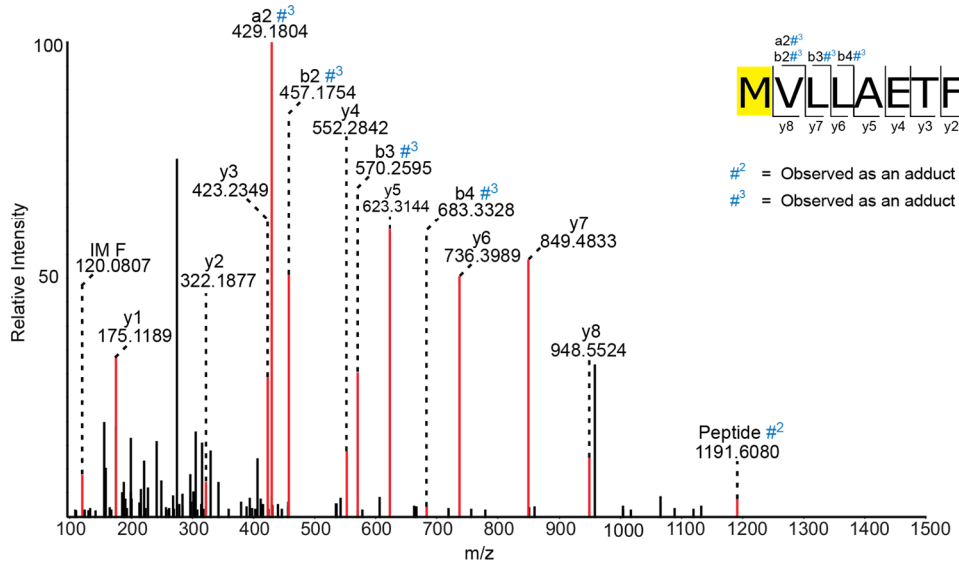

LGGNANPR + [UA]
y7# y6# y5 y4 y3 y2 y1
= Observed as an adduct with [Fragment of U = -C3O = 52]
#3 = Observed as an adduct with [U-H3PO4]
#4 = Observed as an adduct with [U-H2O]
A'
136.0618
y1
175.1187
y2
272.1714
y3
386.2127
y4
457.2517
y5
571.2960
y6 #
680.3056
y7 #
737.3354
Peptide #
850.4154
Peptide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

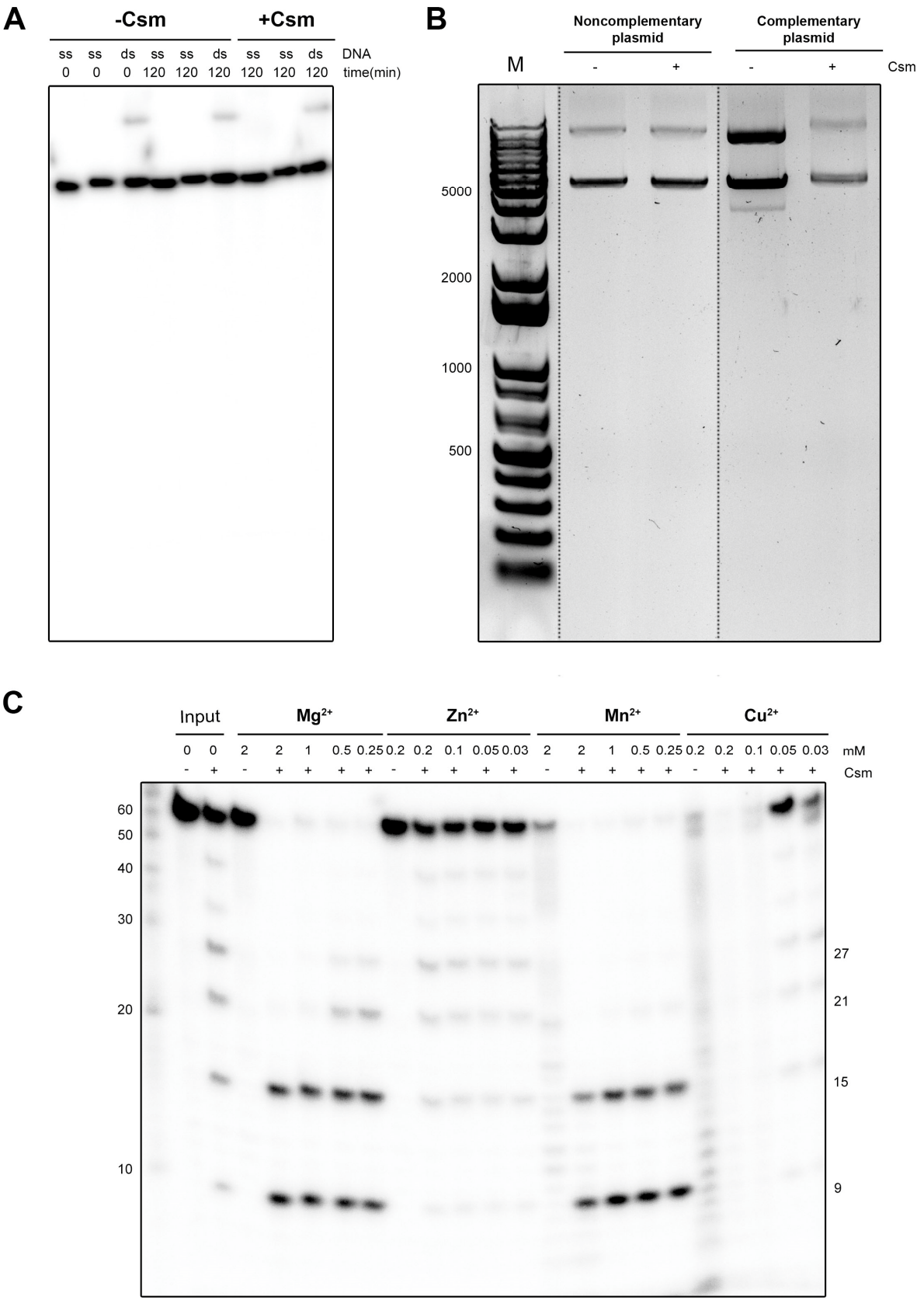

y1
175.1188
b2
258.1559
b2-NH3
241.1293
y2
290.1458
[Peptide #]2+
427.7280
y4
446.2366
U-H2O
307.0331
y3
389.2156
y5 #
597.2985
b6 #
680.3353
y6 #
753.3368
Peptide #
854.3871
Relative Intensity
m/z
TRVGVDR + [UU]
b6#
y6# y5# y4 y3 y2 y1
= Observed as an adduct with [Fragment of U = -C3O = 52]

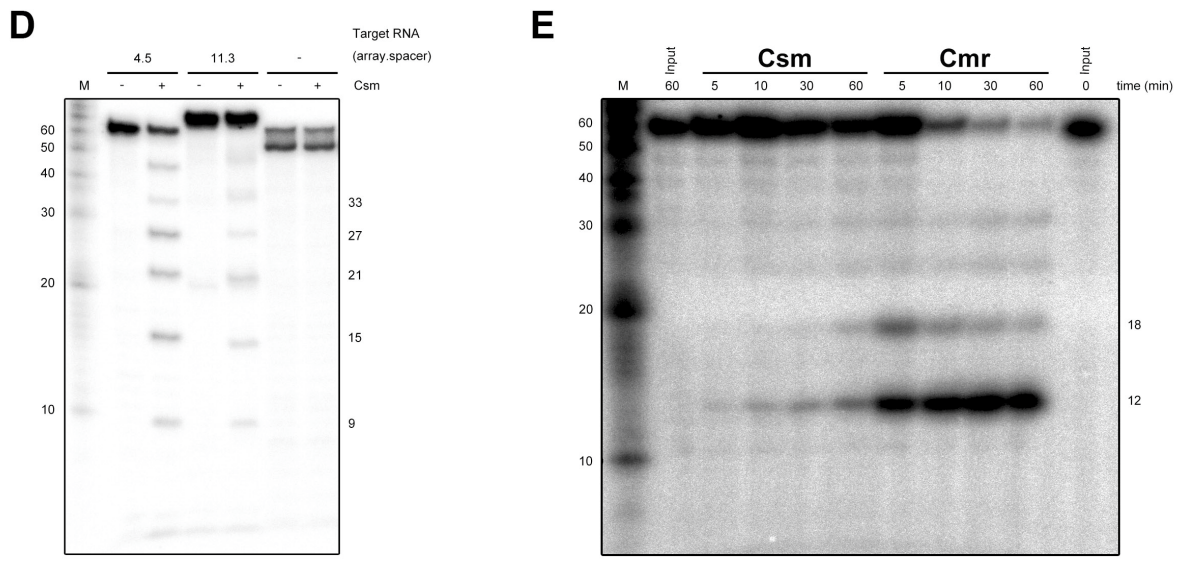

y1
147.1128
IM Y
136.0757
a2
157.0971
b2
185.0920
y2
246.1811
b3
298.1760
y3
333.2133
b4
355.1976
b5
426.2350
b6
589.2969
y6 #
626.3158
y7 #
739.4047
y6 #3
686.3465
y6 #4
800.3753
y6 #5
880.3446
y8 #
902.4615
Relative Intensity
m/z
SPLGAYLPGSSVK + [U]
a2 b2 b3 b4 b5 b6
y8# y7# y6# y5 y4 y3 y2 y1
y6#3
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.

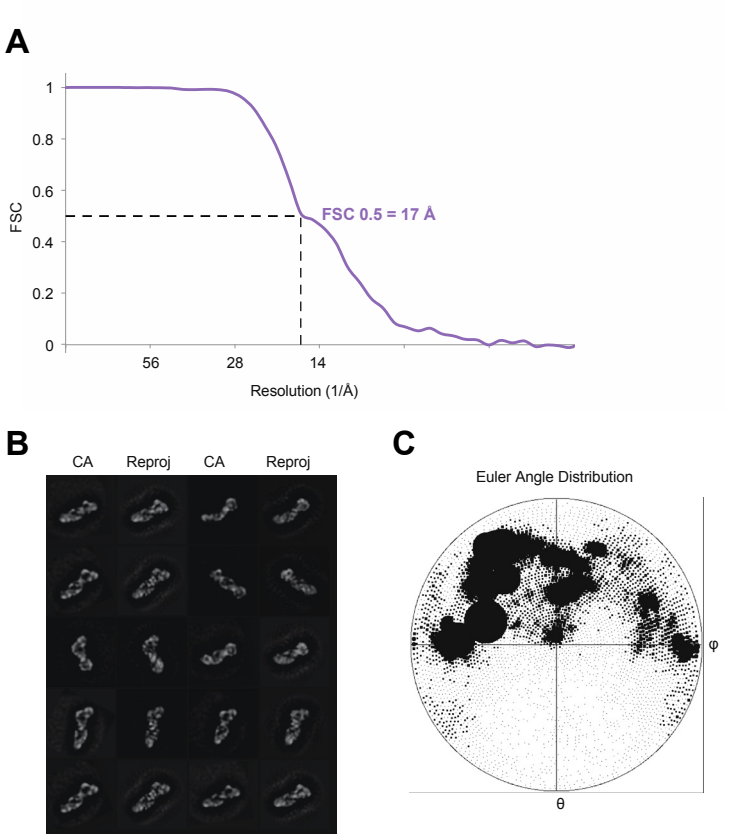
Figure S3 (related to Figure 5)





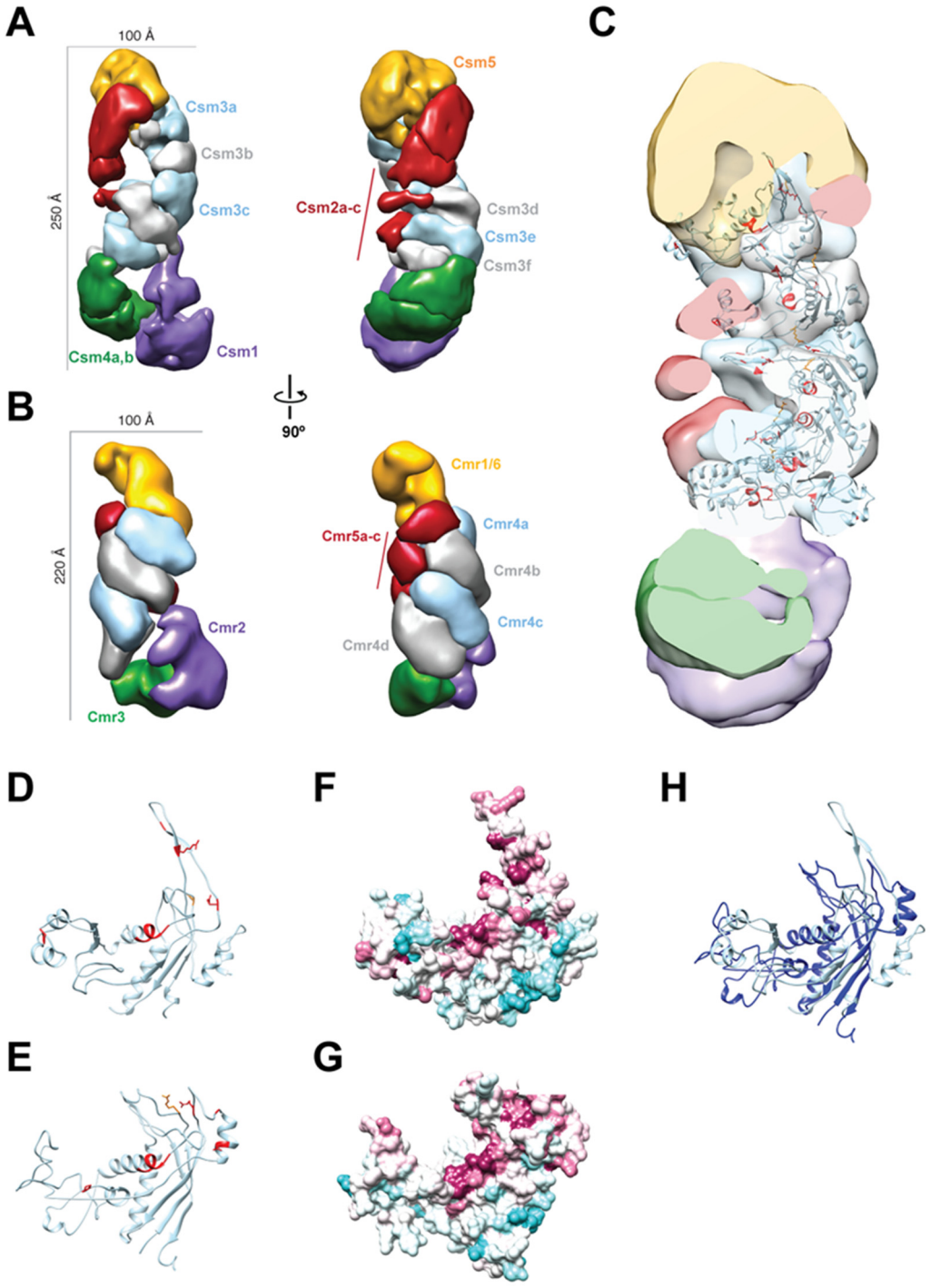
TtCsm *in vitro* activity assays with complementary ssDNA, dsDNA, plasmid DNA and RNA targets. (A) Denaturing gel analysis (20% AA, 7M Urea) of 5' radiolabeled ssDNA (ss) and dsDNA (ds) complementary to crRNA 4.5, which were incubated with ("+") or without ("-") the TtCsm complex for the indicated amount 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.

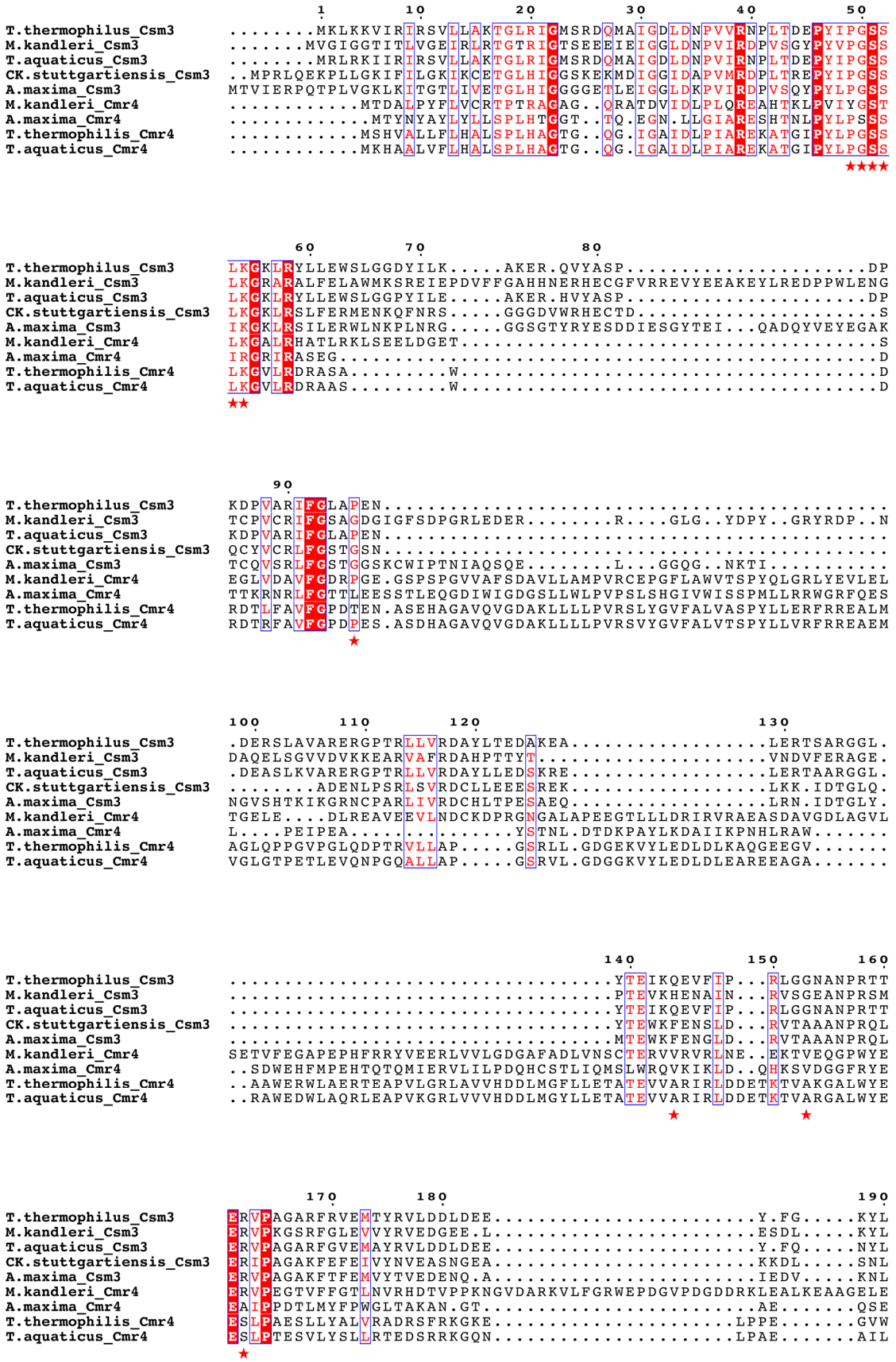
Figure S5 (related to Figure 7)



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

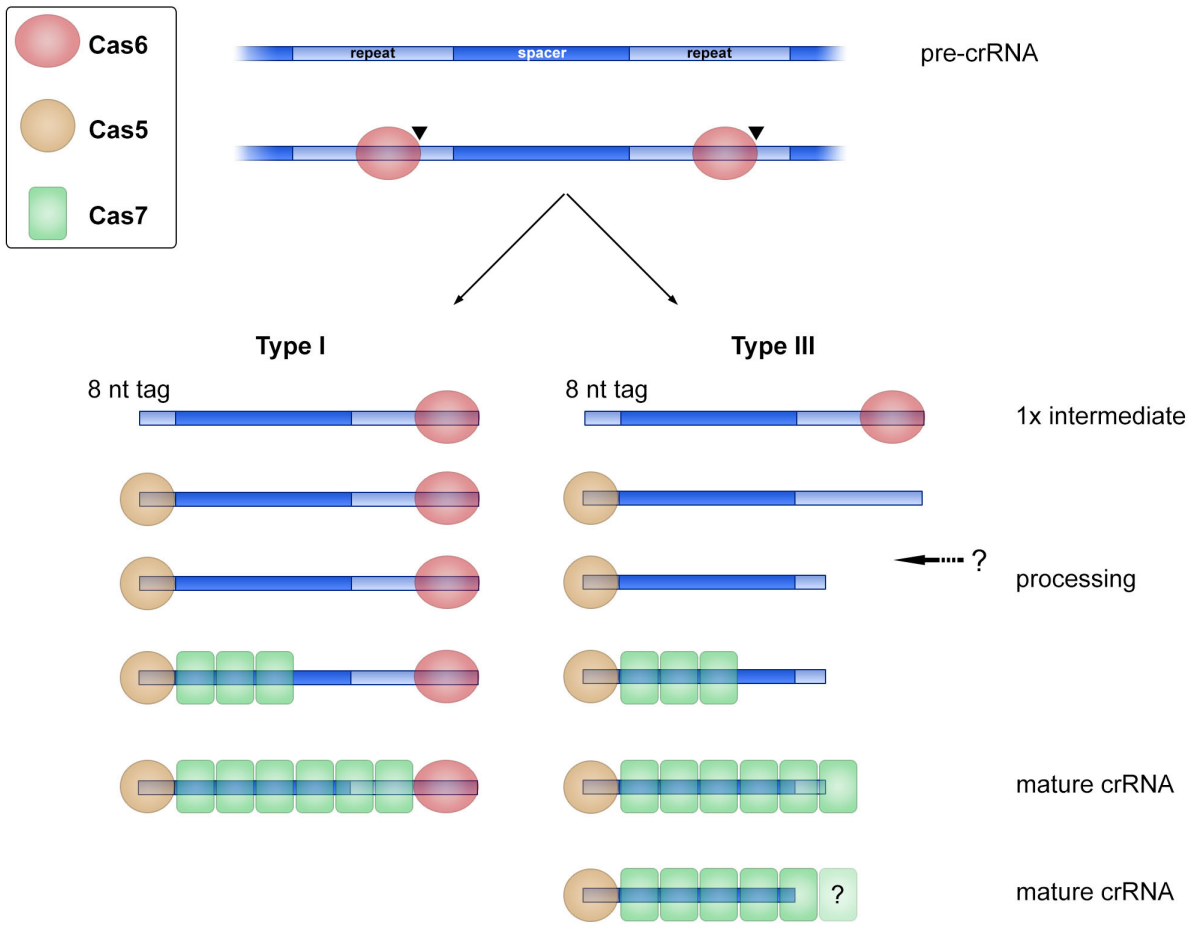
binding residues identified in (Hrle et al., 2013). (D) PHYRE structure prediction of Csm3 (based on *E. coli* Cas7 (Jackson et al., 2014)) with residues colored as in (C) and showing the crRNA-binding thumb of this subunit. (E) PHYRE structure prediction of Csm3 (based on *M. kandleri* Csm3 (Hrle et 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.