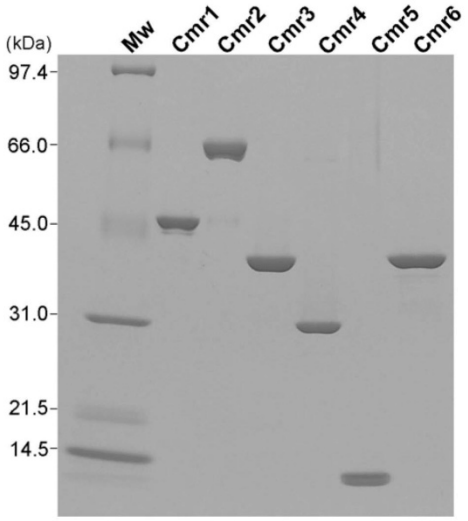
Supplemental Information

Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*

Raymond H.J. Staals, Yoshihiro Agari, Saori Maki-Yonekura, Yifan Zhu, David W. Taylor, Esther van Duijn, Arjan Barendregt, Marnix Vlot, Jasper J. Koehorst, Keiko Sakamoto, Akiko Masuda, Naoshi Dohmae, Peter J. Schaap, Jennifer A. Doudna, Albert J. Heck, Koji Yonekura, John van der Oost, Akeo Shinkai

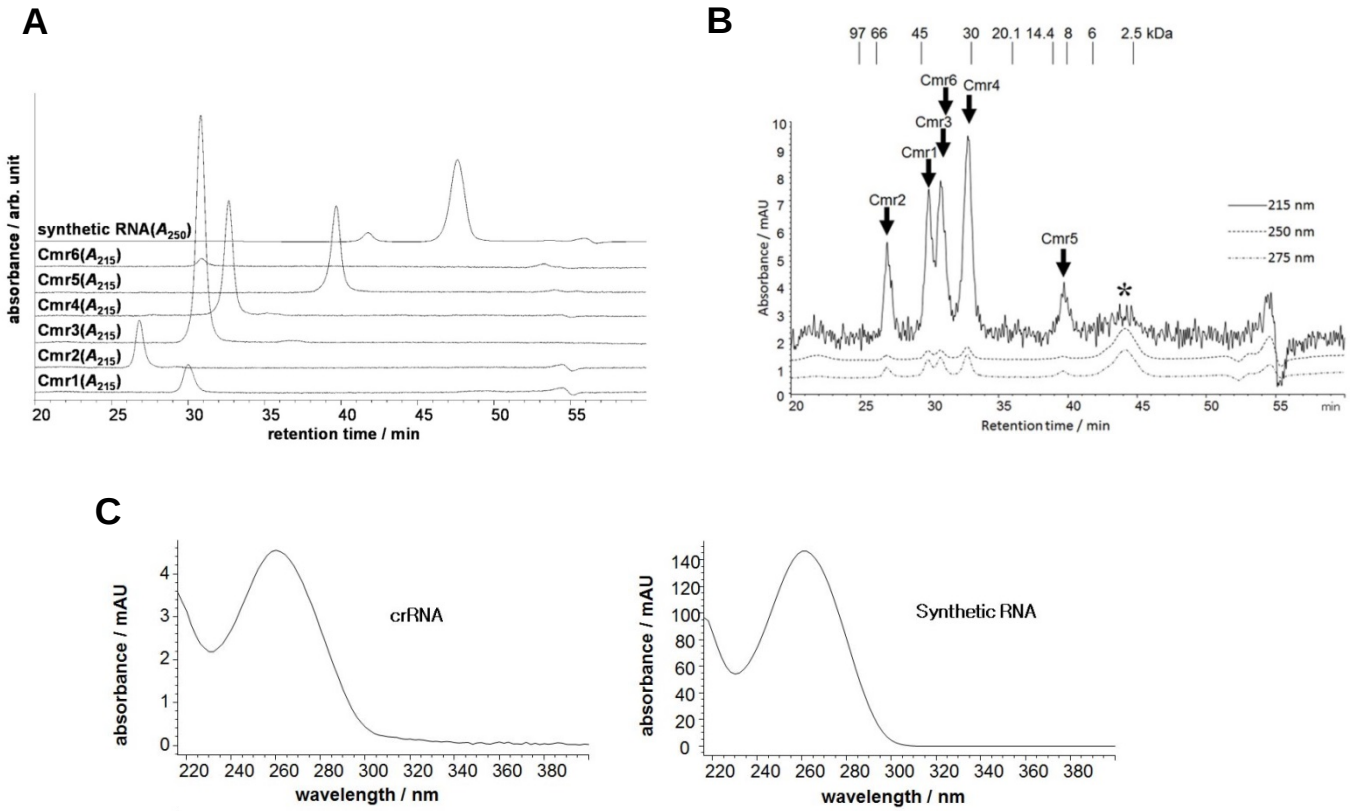
Supplemental figures

Fig. S1 (related to Fig. 2)



SDS-polyacrylamide gel analysis of the *T. thermophilus* Cmr proteins. The proteins were expressed in *E. coli*, and the purified proteins (each 1 μg) were analyzed on a pre-cast 5 to 20% polyacrylamide gel (ATTO Corp.), and the gel was stained with Coomassie Brilliant Blue R-250. Mw, molecular-weight markers. The protein concentration used was determined by the Bradford method (Bradford, 1976), using bovine serum albumin as a standard.

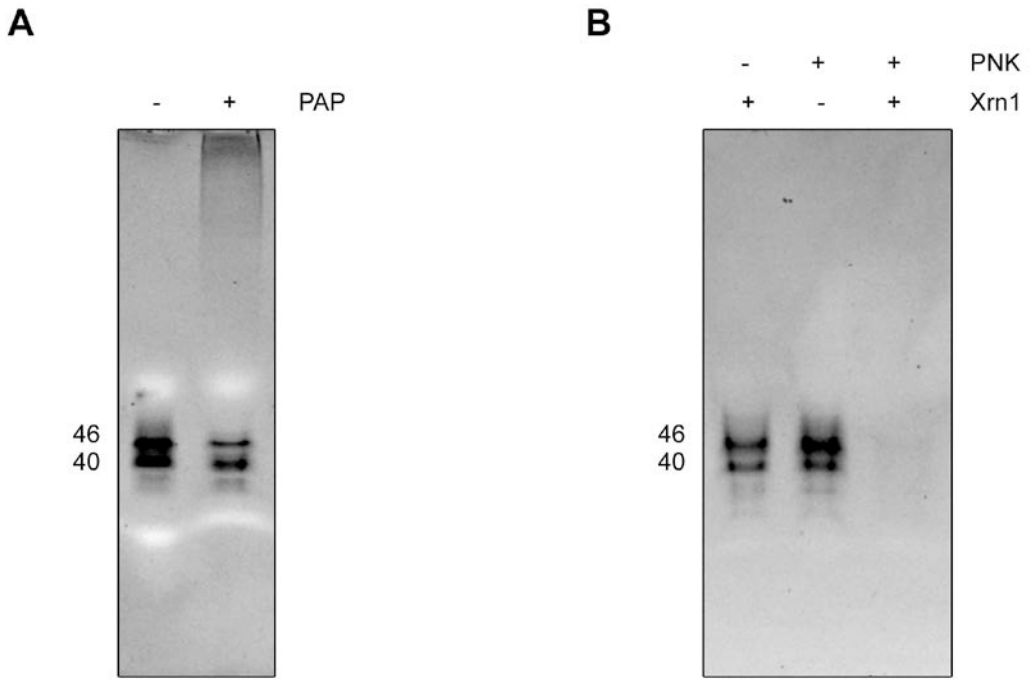
Fig. S2 (related to Fig. 2)



Subunit composition analysis of the Cmr complex. (A) Chromatograms of the recombinant Cmr proteins and a synthetic RNA. One µl (0.6 to 5.8 µg) of the recombinant Cmr proteins and 2 µl (~1.4 µg) of the synthetic RNA (7 kDa) were subjected to the GPC, in the presence of 6 M guanidine-HCl. The elution profiles were recorded by monitoring the absorbances at 215 and 250 nm for the recombinant Cmr proteins and the synthetic RNA, respectively. (B) Representative chromatogram of a GPC analysis of the Cmr complex, under denaturing conditions. A one µl aliquot (0.75 µg protein) of the Cmr complex, purified from *T. thermophilus* (Fig. 2), was subjected to the GPC in the presence of 6 M guanidine-HCl. The peak for each subunit, identified from the profile of each recombinant Cmr protein (Fig. S2A), is indicated by an arrow. The retention times of the molecular weight markers are also shown by the vertical lines, with

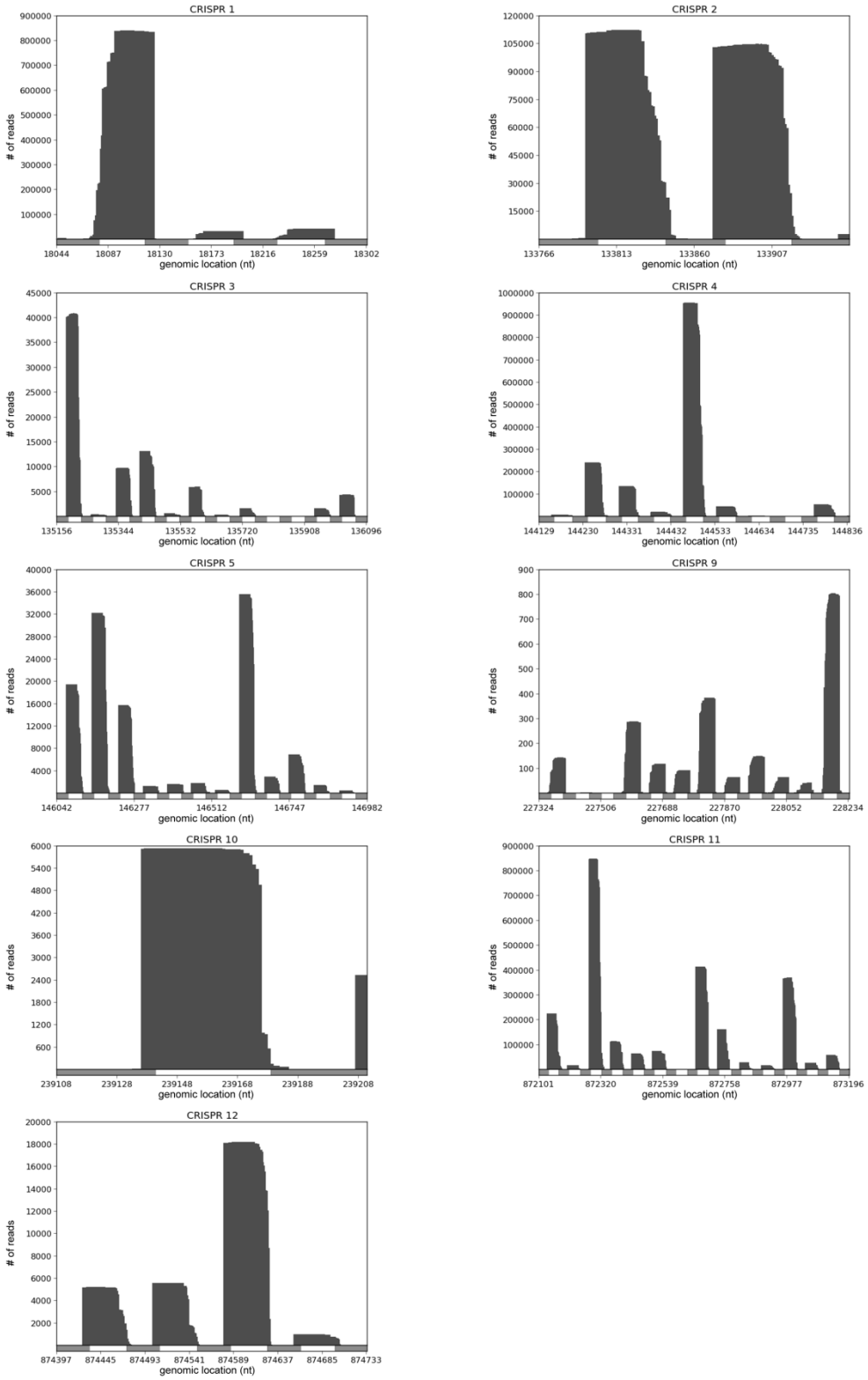
molecular weights in kDa. The elution profile of the sample was detected by measuring the absorbances at 215, 250, and 275 nm. The peak labeled with an asterisk is from crRNA. (C) The UV spectra of the crRNA derived from the Cmr complex and a synthetic RNA. Left panel: The UV spectrum of the elution peak at 43.8 min from the GPC chromatography (Fig. S2B). Right panel: The UV spectrum of a synthetic RNA (Fig. S2A).

Fig. S3 (related to Fig. 3)



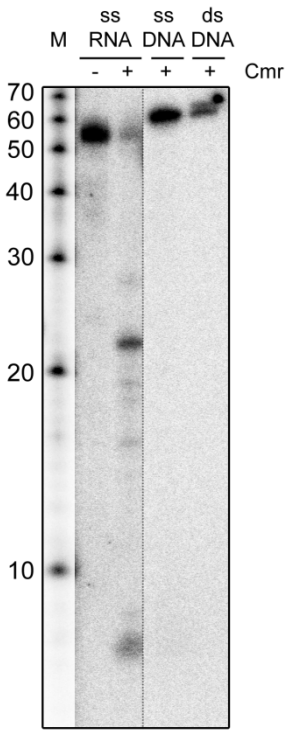
Denaturing gel of the 3’ and 5’ end characterization of the Cmr-bound crRNAs. crRNAs from the Cmr complex were isolated by PCI extraction. (A) The 3’ end of the crRNA was successfully 3’ polyadenylated with *E. coli* Poly(A) Polymerase (PAP). (B) The isolated crRNAs were incubated with T4 polynucleotide kinase (PNK), the 5’ phosphate-dependent exoribonuclease Xrn1 or both. These results demonstrated that the isolated crRNAs could be used for adapter ligation preceding to the deep sequencing analysis.

Fig. S4 (related to Fig. 3)



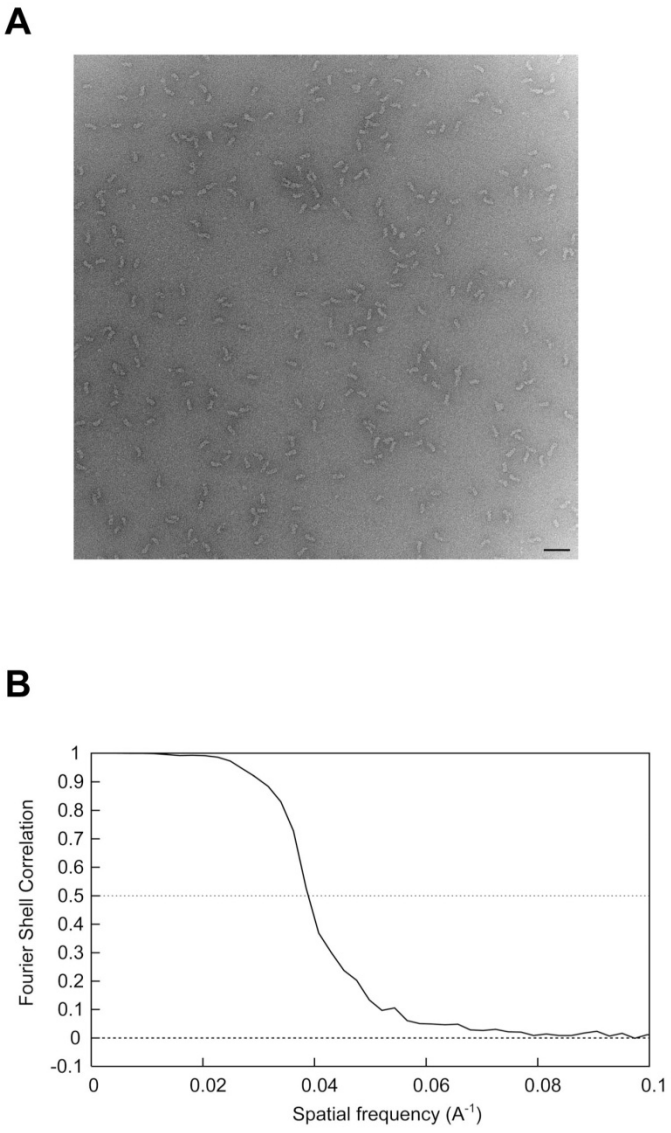
Distribution of the deep-sequencing reads over the different *T. thermophilus* CRISPR arrays. crRNAs isolated from the TtCmr complex were send for deep sequencing analysis. The resulting reads were mapped on the genome of *T. thermophilus*. This overview displays all the reads that were found to map to any of the 11 different CRISPR arrays. On the horizontal axis, the repeat sequences are indicated in gray, while the spacer sequences are indicated in white.

Fig. S5 (related to Fig. 5)



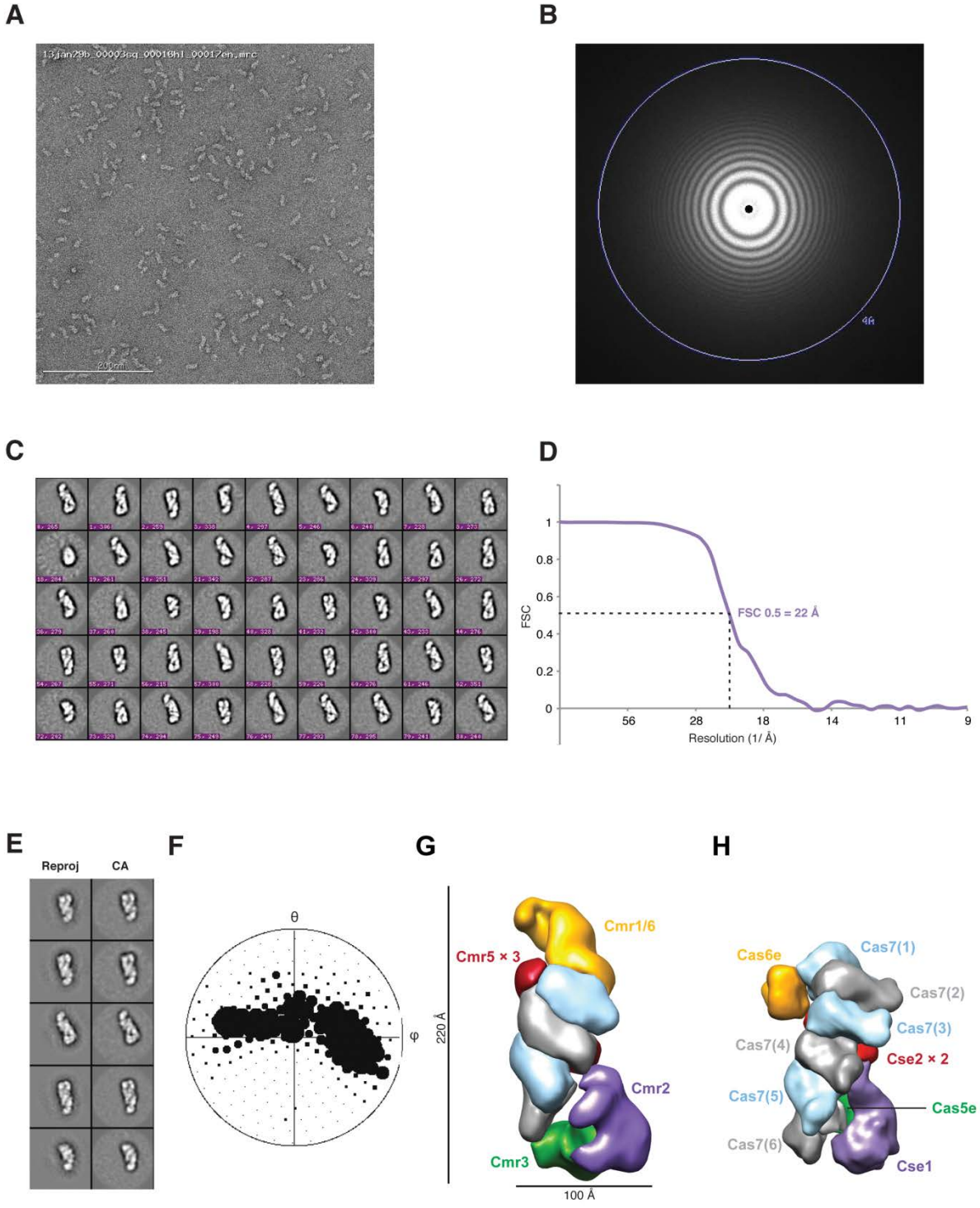
Cmr activity assay with ssRNA, ssDNA and dsDNA substrates. An internally labeled RNA substrate complementary to CRISPR-4.5 was incubated without (-) and with (+) the endogenous Cmr complex for 2 hours in a buffer containing 2 mM Mg^{2+}. In parallel, a 5' end labeled ssDNA (complementary to CRISPR-4.5) and a dsDNA substrate were incubated with (+) assayed as well. Samples were analyzed by denaturing polyacrylamide gel electrophoresis, followed by phosphorimaging. Noncontiguous lanes from the same gel are indicated with dotted lines.

Fig. S6 (related to Fig. 7)



EM images and three-dimensional reconstruction of the Cmr complex. (A) Representative raw images of negatively stained particles of the Cmr complexes. The scale bar represents 500 Å. (B) Fourier shell correlation curve, indicating a resolution of ~26 Å with a 0.5 cut-off criterion.

Fig. S7 (related to Fig. 7)



Architecture of Cmr complex determined using automated molecular microscopy. (A,B) Representative micrograph and FFT (fast Fourier transform) of the micrograph of negatively stained Cmr complexes acquired using LEGINON, respectively. Scale bar indicates 200 nm. (C) Reference-free 2D class averages of the Cmr complex. Width of each box corresponds to 400 Å. (D) Fourier shell correlation curve indicates the reconstruction has a resolution of ~22 Å at the 0.5 cut-off criterion. (E) Comparison of reprojections of the Cmr complex reconstruction (odd column) with corresponding reference-free 2D class averages (even column). Width of each box corresponds to 400 Å. (F) Euler angle distribution of the reconstruction. The size of the spot is proportional to the number of particles that belong to that specific view. (G,H) Comparison of the three-dimensional structures of (G) the *T. thermophilus* Cmr complex (this study) and (H) the *E. coli* Cascade complex (Wiedenheft et al., 2011).

Table S1 (related to Fig. 4)

CMR component	**Theoretical Mass (Da)**	**Experimental Mass (Da)**
CMR 1	44,485.9	44,478.8 ± 4.9
CMR 2	64,819.3	64,844.3 ± 7.7
CMR 3	39,726.7	39,737.5 ± 3.7
CMR 4	30,626.3	30,635.5 ± 6.8
CMR 5	12,960.9	12,965.3 ± 1.7
CMR 6	38,536.1	38,563.7 ± 3.6
crRNA	15,081.0	n.d.
Complex	**Theoretical Mass (Da)**	**Experimental Mass (Da)**
CMR	364,036.9	364,610.3 ± 46
CMR – CMR1	319,551.0	319,059.6 ± 40

Exact masses of individual Cmr protein subunits of Cmr. The masses of the Cmr (sub-) complexes observed under native mass spectrometry conditions are also listed.

Table S2 (related to Fig. 4)

Mass of (sub) complexes in solution (Da)	Theoretical mass (Da)	Mass dissociated products (Da)	Annotation	Stoichiometry 1 2 3 4 5 6 crRNA
364,610.3	364,036.9		CMR	1 1 1 4 3 1 1
	319,551.0	319,553.5	CMR-CMR1	0 1 1 4 3 1 1
	44,485.9	44,478.8	CMR1	1 0 0 0 0 0 0
304,755.4	304,470.0		CMR-CMR1-crRNA	0 1 1 4 3 1 0
	291,509.1	291,268.8	CMR-CMR1-CMR5-crRNA	0 1 1 4 2 1 0
	265,933.9	266,177.0	CMR-CMR1-CMR6-crRNA	0 1 1 4 3 0 0
	38,536.1	38,546.7	CMR6	0 0 0 0 0 1 0
	12,960.9	12,965.1	CMR5	0 0 0 0 1 0 0
266,226.4	265,933.9		CMR-CMR1-CMR6-crRNA	0 1 1 4 3 0 0
	252,973.0	253,806.6	CMR-CMR1-CMR5-CMR6-crRNA	0 1 1 4 2 0 0
	235,307.6	236,078.4	CMR-CMR1-CMR4-CMR6-crRNA	0 1 1 3 3 0 0
	12,960.9	12,961.8	CMR5	0 0 0 0 1 0 0
	30,626.3	30,630.3	CMR4	0 0 0 1 0 0 0
200,092.7	199,924.0		CMR-CMR1-CMR2-CMR3-crRNA	0 0 0 4 3 1 0
	186,963.1	186,991.0	CMR-CMR1-CMR2-CMR3-CMR5-crRNA	0 0 0 4 2 1 0
	169,297.7	169,378.3	CMR-CMR1-CMR2-CMR3-CMR4-crRNA	0 0 0 3 3 1 0
	12,960.9	12,968.5	CMR5	0 0 0 0 1 0 0
	30,626.3	30,630.0	CMR4	0 0 0 1 0 0 0
169,438.4	169,297.7		CMR-CMR1-CMr2-CMR3-CMR4-crRNA	0 0 0 3 3 1 0
161,519.9	161,387.9		CMR-CMR1-CMR2-CMR3-CMR6-crRNA	0 0 0 4 3 0 0
	148,427.0	148,521.7	CMR-CMR1-CMR2-CMR3-CMR5-CMR6-crRNA	0 0 0 4 2 0 0
	130,761.6	130,839.9	CMR-CMR1-CMR2-CMR3-CMR4-CMR6-crRNA	0 0 0 3 3 0 0
	12,960.9	12,960.4	CMR5	0 0 0 0 1 0 0
	30,626.3	30,627.4	CMR4	0 0 0 1 0 0 0
125,813.9	125,710.5		CMR-CMR1-CMR2-CMR3-2*CMR4-CMR5-crRNA	0 0 0 2 2 1 0
104,601.0	104,546.0		CMR-CMR1-4*CMR4-3*CMR5-CMR6-crRNA	0 1 1 0 0 0 0
	64,819.3	64,815.8	CMR2	0 1 0 0 0 0 0

List of masses for all Cmr (sub)complexes present in solution, and their dissociated products that are formed in the gas phase after collisional activation during tandem mass spectrometry experiments. In addition for each complex the theoretical mass (based on amino acid sequence and crRNA mass of 15,081 Da) and stoichiometric information is given. 1=CMR1, 2=CMR2, 3=CMR3, 4=CMR5, 5=CMR5, 6=CMR6, minus (-) indicates the lacking subunit, n.d. is not determined.

Table S3 (related to Fig. 1-7)

oligo name	sequence (5′ to 3′)	comments
P1; P2	AAGCTTACTATGGGAGCGTCAAAGCGCCCCCCGCGG; TCTAGACTATCAGTGGTGGTGGTGGTGGGCCGAGAGGCCCAGCTTGGCGAGAACCCCTTCCA	construction of the plasmid pUC-cmr6h
P3; P4	CTGCAGCCGGGCGCAGACCTTCTTTATCCTGGACTT; GAATTCCCGCCTCCTTGGCCCACTCCAGCAGGGCCA	
P5; P6	ATACATATGCGTAGGGCGTTTCGTGAGGCG; TATAGATCTTTATTACAAGCTTTTCACGAAGGCCAAGACG	cloning of *cmr1*
P7; P8	ATATCATATGGAACACCTTCTCGCTATCGC; GGCCTCCTTAAGGACCGCCTCCAGGCCCGC	cloning of *cmr2*
P9; P10	CGGTCCTTAAGGAGGCCCTGGAGGAGCTTG; GAAGCTCGTACGCGGCCCTGGAGGGCACCT	
P11; P12	GGCCGCGTACGAGCTTCTGGCCCTGGTCCG; ATATAGATCTTTATTAATGAGCATCCCACACCTCCCGGCTCT	
P13; P14	ATATCATATGTTGATTGAGCCCAGGGATCC; CAGGGCCACGTGGGTGCGCGTCTCCGGGAC	cloning of *cmr3*
P15; P16	GCACCCACGTGGCCCTGGACCCCGCGGCCC; ATATAGATCTTTATTATGCCTCCTCCCAGCGCCTAAGCTTCC	
P17; P18	ATATCATATGAGCCACGTTGCGCTTCTTTTC; ATATGGATCCTTATTACCTCCCCACCACGATCCGGCA	cloning of *cmr4*
P19; P20	ATATCATATGCGTACCCGCTCGCAGGTGT; ATATGGATCCTTATTACTCCTCAATGAGGGCCTGGGCG	cloning of *cmr5*
P21; P22	GGAATTCATATGGGCCGCAGATCCGCGCTTG; GGAATTGCTAGCTTATTAGGCCGAGAGGCCCAGCTTG	cloning of *cmr6*
P23; P24	ATTTAGGTGACACTATA**G**AACTGCGCCTTGACGTGGTCGTCCCCGGGCGCCTTATCTACGGCCATCG; CGATGGCCGTAGATAAGGCGCCCGGGGACGACCACGTCAAGGCGCAGTTCTATAGTGTCACCTAAAT	*In-vitro* transcription of the CRISPR-4.5 target
P25; P26	ATTTAGGTGACACTATA**G**AACTGGGCGCCCGAGGTCAACGGCCCACTACCGGGCCTCGGGACCATCG; CGATGGTCCCGAGGCCCGGTAGTGGGCCGTTGACCTCGGGCGCCCAGTTCTATAGTGTCACCTAAAT	*In-vitro* transcription of the CRISPR-4.2 target
P27; P28	ATTTAGGTGACACTATA**G**AACTGAAGAACATGCCGAAGCCGTCGTCGTTGGACTGCCACATCCATCG; CGATGGATGTGGCAGTCCAACGACGACGGCTTCGGCATGTTCTTCAGTTCTATAGTGTCACCTAAAT	*In-vitro* transcription of the CRISPR-2.1 target
P29; P30	ATTTAGGTGACACTATA**G**AACTGGTGTCTGGGGCTAGCGGCATTCCCTTTGGGCGGGTTGTCCATCG; CGATGGACAACCCGCCCAAAGGGAATGCCGCTAGCCCCAGACACCAGTTCTATAGTGTCACCTAAAT	*In-vitro* transcription of the CRISPR-3.2 target
P31	GAACUGCGCCUUGACGUGGUCGUCCCCGGGCGCCUUAUCUACGGCCAUCG	Synthetic RNA target (CRISPR-4.5)
P32	AUUGCGACCCGUAGAUAAGGCGCCCGGGGACGACCACGUCAAG	Synthetic 43 nt crRNA (CRISPR-4.5)
P33	AUUGCGACCCGUAGAUAAGGCGCCCGGGGACGACCACGUCAAGGCG	Synthetic 46 nt crRNA (CRISPR-4.5)

Oligonucleotides used in this study. Sequences marked in yellow indicate the presence of a SP6 promoter in oligonucleotides that were used for *in-vitro* transcription. Bold lettering represents the first nucleotide incorporated into RNA during transcription. Underlined are the sequence with complementarity its cognate crRNA.

Supplemental Experimental Procedures

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

In order to produce the C-terminal $(His)_6$-tagged Cmr6 in *T. thermophilus* HB8, the tag-coding sequence was inserted within the genome by means of homologous recombination. The plasmid pUC-cmr6h, used for the homologous recombination, was constructed as follows. A DNA fragment (fragment 1; 530-bp *Hin*dIII-*Xba*I fragment) carrying the 3′-terminal coding region of *cmr6* (positions 157,391 to 157,893 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 *cmr6* (positions 157,941 to 158,435 on the megaplasmid pTT27), were amplified by genomic PCR using the primers P1/P2 and P3/P4 (Table S3), 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-cmr6h. The plasmid was introduced into the *T. thermophilus* HB8 strain, and a kanamycin-resistant clone was obtained as described previously (Hashimoto et al., 2001). In this strain, the downstream region of the *cmr6* gene on the genome (positions 157,894 to 157,940) 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 Cmr complex, constructed as described above, were cultured at 70°C in three liters of rich (TT) medium (Agari et al., 2008) until an A_{600} = 2 to 3 was attained. *T. thermophilus* HB8 cells infected by ΦYS40 phage (provided by Dr. Masatada Tamakoshi) at a multiplicity of infection of ~1 were also prepared, as described previously (Agari et al., 2010).

Detailed description of the purification of the Cmr complex

The *T. thermophilus* HB8 cells producing the $(His)_6$-tagged Cmr complex were resuspended in 100 ml of 20 mM Tris-HCl (pH 8.0), containing 50 mM NaCl, 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 20 mM imidazole, and then the bound protein was eluted with a linear gradient of 20 to 500 mM 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 sample was then applied to a HiLoad 16/60 Superdex 200 pg (GE Healthcare) column, pre-equilibrated with 20 mM Tris-HCl (pH 8.0) containing 0.15 M NaCl. The target fractions were collected, and concentrated with a Vivaspin 20 concentrator (30,000 Da molecular weight cut-off, Sartorius).

Expression and purification of recombinant Cmr proteins

The *cmr1*, *cmr2*, *cmr3*, *cmr4*, *cmr5*, and *cmr6* genes were each amplified by genomic PCR, using the primers P5/P6, P7 to P12, P13 to P16, P17/P18, P19/P20, and P21/P22, respectively. Each recombinant protein was expressed in *E. coli* by means of the pET expression system (Merck). **Cmr1:** The *cmr1* gene was cloned under the control of the T7 promoter of pET-21a(+), and then the plasmid was introduced into the *E. coli* BL21(DE3) strain. The recombinant cells were cultured at 37°C in two liters of Luria–Bertani (LB) broth, containing 50 μg ampicillin ml^{-1} and supplement A (0.05% glucose, 0.2% lactose monohydrate, 0.5%(v/v) glycerol, 50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, and 1 mM $MgSO_4$), for 16 h. The cells were resuspended

in 60 ml of 20 mM Tris-HCl (pH 8.0), containing 50 mM NaCl and 0.1 ml of RQ1 RNase-free DNase (Promega), and were disrupted by sonication in ice water. The same volume of buffer, preheated at 70°C, was added to the cell lysate. This mixture was incubated for 10 min at 70°C, and then ultracentrifuged (200,000 × *g*) for 1 h at 4°C. The supernatant was applied to a RESOURCE PHE column (GE Healthcare), pre-equilibrated with 50 mM sodium phosphate buffer (pH 7.0) containing 1.35 M ammonium sulfate, and the flow-through fraction was collected. Ammonium sulfate was added to the sample at a final concentration of 1.5 M, and then the sample was applied to a RESOURCE PHE column (GE Healthcare), pre-equilibrated with the same buffer. The bound protein was eluted with a linear gradient of 1.5 to 0 M ammonium sulfate. The target fractions were collected, and dialyzed against 10 mM sodium phosphate buffer (pH 7.0), containing 0.15 M NaCl. The sample was then applied to a HiTrap heparin column (GE Healthcare), pre-equilibrated with 10 mM Tris-HCl (pH 8.0) containing 0.15 M NaCl, and the bound protein was eluted with a linear gradient of 0.15 to 1 M NaCl. The target fractions were collected, and dialyzed against 10 mM MES-NaOH (pH 6.0) containing 0.1 M NaCl. The sample was then applied to a RESOURCE S column (GE Healthcare), pre-equilibrated with the same buffer, and the bound protein was eluted with a linear gradient of 0.1 to 0.5 M NaCl. The target fractions were collected, and dialyzed against 20 mM Tris-HCl (pH 8.0), containing 0.15 M NaCl. The target fractions were collected, and concentrated with a Vivaspin 10 concentrator (10,000 Da molecular weight cut-off, Sartorius). **Cmr2:** The *cmr2* gene was cloned under the control of the T7 promoter of pET-11a, and then the plasmid was introduced into the *E. coli* BL21(DE3) strain. The recombinant cells were cultured at 37°C in three liters of Luria–Bertani (LB) broth, containing 50 μg ampicillin ml^{-1}, 30 μg chloramphenicol ml^{-1}, and supplement A, for 16 h. The cells were resuspended in 60 ml of 20

mM Tris-HCl (pH 8.0), containing 50 mM NaCl, 5 mM β-mercaptoethanol, and 0.1 mM phenylmethylsulfonyl fluoride, and were disrupted by sonication in ice water. The same volume of buffer, preheated at 65°C, was added to the cell lysate. This mixture was incubated for 10 min at 65°C, and then 0.95 ml of protease inhibitor cocktail (Nacalai Tesque) was added to the mixture, which was then ultracentrifuged (200,000 × *g*) for 1 h at 4°C. The supernatant was applied to a RESOURCE ISO column (GE Healthcare), pre-equilibrated with 50 mM sodium phosphate buffer (pH 7.0) containing 1.5 M ammonium sulfate, and then the bound protein was eluted with a linear gradient of 1.5 to 0 M ammonium sulfate. 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 fractions were collected, and desalted by fractionation on a HiPrep 26/10 desalting column. The sample was then applied to a 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 applied to a HiLoad 16/60 Superdex 200 pg (GE Healthcare) column, pre-equilibrated with 20 mM Tris-HCl (pH 8.0), containing 0.15 M NaCl. **Cmr3:** The *cmr3* gene was cloned under the control of the T7 promoter of pET-11a, and then the plasmid was introduced into the *E. coli* BL21(DE3) strain. The recombinant cells were cultured at 37°C in two liters of Luria–Bertani (LB) broth, containing 50 μg ampicillin ml^{-1} and supplement A, for 16 h. The cells were resuspended in 60 ml of 20 mM Tris-HCl (pH 8.0), containing 0.2 M NaCl and 0.6 ml of protease inhibitor cocktail (Nacalai Tesque), and were disrupted by sonication in ice water. The same volume of buffer, preheated at 65°C, was added to the cell lysate. This mixture was incubated for 10 min at 65°C, and then

ultracentrifuged (200,000 × *g*) for 1 h at 4°C. The supernatant was applied to a RESOURCE PHE column (GE Healthcare), pre-equilibrated with 50 mM sodium phosphate buffer (pH 7.0) containing 1.5 M ammonium sulfate, and then the bound protein was eluted with a linear gradient of 1.5 to 0 M ammonium sulfate. The target fractions were collected, and desalted by dialysis against 20 mM Tris-HCl (pH 8.0), containing 0.1 M NaCl. The sample was applied to a HiTrap heparin column pre-equilibrated with the same buffer, and then the bound protein was eluted with a linear gradient of 0.1 to 0.4 M NaCl. The target fractions were collected, and dialyzed against 20 mM Tris-HCl (pH 8.0), containing 50 mM NaCl. The sample was applied to a RESOURCE S column (GE Healthcare) pre-equilibrated with the same buffer, and then the bound protein was eluted with a linear gradient of 50 to 200 mM NaCl. The target fractions were collected, and dialyzed against 20 mM Tris-HCl (pH 8.0), containing 0.15 M NaCl. **Cmr4:** The *cmr4* gene was cloned under the control of the T7 promoter of pET-11a, and then the plasmid was introduced into the *E. coli* Rosetta2(DE3) strain. The recombinant cells were cultured at 37°C in six liters of Luria–Bertani (LB) broth, containing 50 μg ampicillin ml^{-1}, 30 μg chloramphenicol ml^{-1}, and supplement A, for 16 h. The cells were resuspended in 120 ml of 20 mM Tris-HCl (pH 8.0), containing 50 mM NaCl, and were disrupted by sonication in ice water. The same volume of buffer, preheated at 75°C, was added to the cell lysate. This mixture was incubated for 50 min at 75°C, and then ultracentrifuged (200,000 × *g*) for 1 h at 4°C. The supernatant was applied to a RESOURCE ISO column (GE Healthcare), pre-equilibrated with 50 mM sodium phosphate buffer (pH 7.0) containing 1.05 M ammonium sulfate, and then the bound protein was eluted with a linear gradient of 1.05 to 0 M ammonium sulfate. 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 fractions were collected, and applied to a HiLoad 16/60 Superdex 75 pg (GE Healthcare) column, pre-equilibrated with 20 mM Tris-HCl (pH 8.0) containing 0.15 M NaCl. **Cmr5:** The *cmr5* gene was cloned under the control of the T7 promoter of pET-11a, and then the plasmid was introduced into the *E. coli* BL21(DE3) strain. The recombinant cells were cultured at 37°C in four liters of Luria–Bertani (LB) broth, containing 50 μg ampicillin ml^{-1}, for 16 h. The cells were resuspended in 60 ml of 20 mM Tris-HCl (pH 8.0), containing 50 mM NaCl, and were disrupted by sonication in ice water. The same volume of buffer, preheated at 70°C, was added to the cell lysate. This mixture was incubated for 10 min at 70°C, and then ultracentrifuged (200,000 × *g*) for 1 h at 4°C. The supernatant was applied to a Super Q Toyopearl 650M column (Tosoh), pre-equilibrated with 20 mM Tris-HCl (pH 8.0), and the flow-through fraction was collected. This fraction was applied to a Resource S column (GE Healthcare), pre-equilibrated with 20 mM MES-NaOH buffer (pH 6.0), and the bound protein was eluted with a linear gradient of 0 to 0.6 M NaCl. The target fractions were collected, and applied to a HiTrap heparin column (GE Healthcare), pre-equilibrated with 20 mM MES-NaOH buffer (pH 6.0), and the bound protein was eluted with a linear gradient of 0 to 2 M NaCl. The target fractions were collected and applied to a HiLoad 16/60 Superdex 75pg column (GE Healthcare), pre-equilibrated with 20 mM MES-NaOH buffer (pH 6.0), containing 0.5 M NaCl. The target fractions were collected and concentrated with a Centriprep concentrator (3,000 Da molecular-weight cutoff, Millipore), and dithiothreitol was added to the sample to a final concentration of 1 mM. **Cmr6:** The *cmr6* gene was cloned under the control of the T7 promoter of pET-HisTEV vector, with the tobacco etch virus (TEV) protease-recognition site, Glu-Asn-Leu-Tyr-Phe-Gln-Gly, instead of the thrombin-recognition site of the pET-15b vector (Merck),

and then the plasmid was introduced into the *E. coli* Rosetta2(DE3) strain. The recombinant cells were cultured at 37°C in six liters of Luria–Bertani (LB) broth, containing 50 μg ampicillin ml^{-1}, 30 μg chloramphenicol ml^{-1}, and supplement A, for 16 h. The cells were resuspended in 120 ml of 20 mM Hepes-NaOH (pH 7.5), containing 0.5 M NaCl, and were disrupted by sonication in ice water. The sample was ultracentrifuged (200,000 × *g*) for 1 h at 4°C, and the supernatant was applied to a HisTrap HP column (GE Healthcare), pre-equilibrated with 20 mM Hepes-NaOH (pH 7.5), containing 0.5 M NaCl and 20 mM imidazole, and the bound protein was eluted with a linear gradient of 20 to 500 mM imidazole. The target fractions were collected, and treated with TEV protease at 30°C for 1 h. The sample was applied to a HisTrap HP column (GE Healthcare), pre-equilibrated with 20 mM Tris-HCl (pH 8.0), containing 0.5 M NaCl and 20 mM imidazole. The flow-through fraction was collected and dialyzed against 20 mM Hepes-NaOH (pH 7.5), containing 0.5 M NaCl. The sample was concentrated with an Amicon Ultra-15 concentrator (10,000 Da molecular-weight cutoff, Millipore).

Identification of the Cmr subunits

The protein spot was excised from the SDS-polyacrylamide gel, after the gel was washed twice with distilled water. The excised protein spot was further washed a few times with 0.1 M ammonium bicarbonate buffer (pH 8.0) and 100% acetonitrile, alternately, and then completely dried with a speed vacuum drier. Twenty ng of trypsin (Trypsin Gold, Mass Spec Grade, Promega) were applied to the gel piece, followed by rehydration for 30 min on ice. After this incubation, nine μl of buffer (40 mM ammonium bicarbonate, pH 8.0 and 10% acetonitrile) were added to the solution, and the mixture was incubated at 37°C for 16 h. The tryptic peptide fragments were injected into the pre-column, at a flow rate of 10 μl min^{-1} using the autosampler

of the EASY-nLC (Proxeon), and the pre-column was washed with solution A (0.1% formic acid in double distilled water) for 10 min. The desalted peptides were subsequently separated on an analytical column, at a flow rate of 200 nl min^{-1}, with a linear gradient of 5 to 50% solution B (0.1% formic acid in acetonitrile) over 60 min. The peptide separated on the C_{18} column was introduced into the micrOTOF-QII mass spectrometer with nanoelectrospray ionization in the positive mode. The capillary voltage was –4.5 kV, and the drying gas temperature was 200°C. The collision energy of the quadrupole for MS/MS fragmentation of the precursor peptides, using Ar gas, was set at 20 to 40 eV. The fragmentation was performed automatically for the three most abundant peptide ions. The data were processed with DataAnalysis 3.4 (Bruker Daltonics). The peptides were identified using the Mascot database search (http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS).

GPC-HPLC under denaturing conditions

The GPC-HPLC was performed on tandem columns of SuperSW3000 (4.6 × 300 mm, Tosoh Corp.), using a Gilson model 307 pump and an Agilent 1100 series diode array detector. Elution was conducted with 10 mM phosphate buffer (pH 6.0), containing 6 M guanidine-HCl, at a flow rate of 0.15 ml min^{-1}, and the effluent was monitored at 215, 250, and 275 nm. The columns were calibrated with reduced and carboxyl-methylated proteins (phosphorylase b, serum albumin, ovalbumin, carbonic anhydrase, trypsin inhibitor, lactalbumin; LMW marker proteins kit, GE Healthcare) and peptides derived from CNBr-cleaved horse heart myoglobin.

Single particle electron microscopy and analysis with automated electron microscopy

Cmr complex samples were diluted to 25 nM and immediately applied to glow-discharged continuous carbon grids. After 1 min, grids were stained consecutively in 6 droplets of 2% (w/v) uranyl acetate solution and excess stain removed by gentle blotting with filter paper. The sample was examined using a Technai-20 transmission electron microscope equipped with a field-emission gun and operated at 120 keV. The Cmr complex was imaged at a nominal magnification of ×80,000 (1.45 Å/pixel at the specimen level). Images were recorded automatically using the MSI-Raster application in LEGINON (Suloway et al., 2005) on a Gatan 4k × 4k CCD camera with an exposure dose of 20 $e^{-}Å^{-2}$ and a randomly set focus ranging from −0.5 μm to −1.3 μm. All particle pre-processing was performed using the Appion image-processing pipeline (Lander et al., 2009). The contrast transfer function (CTF) of each micrograph was estimated and particles were selected concurrently with data collection using ACE2 (Mallick et al., 2005) and a template-based particle picker (Roseman, 2004) respectively. Micrograph phases were corrected using ACE2, and the negatively stained Cmr complexes were extracted using a 288 × 288-pixel box size. The particle stacks were binned by a factor of 2 for processing and particles were normalized to remove pixels whose values were above or below 4.5-σ of the mean pixel value using XMIPP (Sorzano et al., 2004). These particles were subjected to reference-free alignment and classification using multivariate statistical analysis and multi-reference alignment in IMAGIC (van Heel et al., 1996) to a total of ~400 classes. Particles belonging to bad class averages were discarded and a new stack of the remaining particles was created for the three-dimensional reconstruction. The first reconstruction (Fig. 7A) low-pass filtered to 70 Å resolution served as the 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 (Lander et al., 2009; Wiedenheft et

al., 2011). The reconstruction showed structural features to 22 Å resolution (based on the 0.5 FSC criterion), showed 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 (Fig. S7D-F). Interestingly, we also used the previous *E. coli* Cascade structure (Wiedenheft et al., 2011) low-pass filtered to 70 Å as an initial model and obtained a reconstruction identical to the final structure presented (Fig. 7D). The reconstruction at 22 Å resolution was segmented automatically using Segger (Pintilie et al., 2010) in Chimera (Pettersen et al., 2004) based on the MS and biochemical analyses. The atomic coordinates of *P. furiosus* Cmr2-Cmr3 (PDB 4H4K) was docked into the map using Fit-in-Map in Chimera by searching for the optimal cross-correlation coefficient between the low-pass filtered crystal structure (at 22 Å resolution) and the EM map.

crRNA isolation and 3' and 5' end characterization

Cmr-bound crRNAs were isolated from the endogenous Cmr complex by phenol-chloroform isoamyl alcohol (PCI) extraction and ethanol precipitation. For the 5' end characterization, crRNAs were phosphorylated for 30 min at 37°C using T4 Polynucleotide Kinase, followed by heat-inactivation for 20 min at 65°C. After the addition of 100 mM NaCl, the crRNAs were subjected to digestion by Xrn-1 (NEB) for 30 min at 37°C. For the 3' end characterization, crRNAs were 3' polyadenylated for 30 min at 37°C with *E. coli* Poly(A) Polymerase (NEB), according to the manufacturer's recommendations. Samples were separated on a 20% polyacrylamide denaturing gel, containing 7 M urea and visualized by ethidium bromide staining.

Purification of TTHB231 (Cas6) protein

The *cas6* gene was cloned under the control of the T7 promoter of pET-11a, and then the plasmid was introduced into the *E. coli* Rosetta(DE3) strain. The recombinant cells were cultured at 37°C in six liters of Luria–Bertani (LB) broth, containing 50 µg ampicillin ml^{-1} ,30 µg chloramphenicol ml^{-1}, and supplement A (0.05% glucose, 0.2% lactose monohydrate, 0.5%(v/v) glycerol, 50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, and 1 mM $MgSO_4$), for 16 h. The cells were resuspended in 60 ml of 20 mM Tris-HCl (pH 8.0), containing 50 mM NaCl, and were disrupted by sonication in ice water. The same volume of buffer, preheated at 70°C, was added to the cell lysate. This mixture was incubated for 10 min at 70°C, and then ultracentrifuged (200,000 × *g*) for 1 h at 4°C. The supernatant was applied to a RESOURCE PHE column (GE Healthcare), pre-equilibrated with 50 mM sodium phosphate buffer (pH 7.0) containing 1.2 M ammonium sulfate, and the bound protein was eluted with a linear gradient of 1.2 to 0 M ammonium sulfate. The target fractions were collected, and desalted by fractionation on a HiPrep 26/10 desalting column (GE Healthcare) pre-equilibrated with 20 mM MES (pH 6.0) containing 150 mM NaCl. The sample was then applied to a HiTrap heparin column (GE Healthcare), pre-equilibrated with the same buffer, 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 (GE Healthcare) pre-equilibrated with 10 mM NaPi (pH 7.0) containing 0.5 M NaCl. The sample was then applied to the hydroxyapatite CHT10-I column (BioRad Laboratories, Inc.) pre-equilibrated with the same buffer, and flowthrough fraction was collected. The sample were applied to a HiLoad 16/60 Superdex 75 pg (GE Healthcare) column, pre-equilibrated with 20 mM Tris-HCl (pH 8.0), containing 0.15 M

NaCl. The target fractions were collected, and concentrated with a Vivaspin 10 concentrator (10,000 Da molecular weight cut-off, Sartorius).

Other methods

SDS-PAGE analysis was performed, according to the method of Laemmli (Laemmli, 1970). Blue native-PAGE analysis was performed using a Native-PAGE Bis-Tris Gel System (Life Technologies Corp.), according to the manufacturer's instructions.

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. Mol. Microbiol. *70*, 60-75.

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.

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.

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 Lett. *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. J. Struct. Biol. *157*, 47-55.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

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.

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. J. Comput. Chem. *25*, 1605-1612.

Pintilie, G.D., Zhang, J., Goddard, T.D., Chiu, W., and Gossard, D.C. (2010). Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. J. Struct. Biol. *170*, 427-438.

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

Sorzano, C.O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., Carazo, J.M., and Pascual-Montano, A. (2004). XMIPP: a new generation of an open-source image processing package for electron microscopy. J. Struct. Biol. *148*, 194-204.

Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., and Carragher, B. (2005). Automated molecular microscopy: the new Leginon system. J. Struct. Biol. *151*, 41-60.

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. J. Struct. Biol. *157*, 38-46.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J. Struct. Biol. *116*, 17-24.

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.