

In the format provided by the authors and unedited.

# Type IV CRISPR RNA processing and effector complex formation in *Aromatoleum aromaticum*

Ahsen Özcan<sup>1</sup>, Patrick Pausch <sup>2,3</sup>, Andreas Linden<sup>4,5</sup>, Alexander Wulf<sup>5</sup>, Karola Schühle<sup>6</sup>, Johann Heider<sup>2,6</sup>, Henning Urlaub<sup>4,5</sup>, Thomas Heimerl<sup>2</sup>, Gert Bange<sup>2,3</sup> and Lennart Randau <sup>1,2\*</sup>

---

<sup>1</sup>Max Planck Institute for Terrestrial Microbiology, Marburg, Germany. <sup>2</sup>LOEWE Center for Synthetic Microbiology (Synmikro), Philipps University Marburg, Marburg, Germany. <sup>3</sup>Faculty of Chemistry, Philipps-University-Marburg, Marburg, Germany. <sup>4</sup>Bioanalytics Research Group, Department of Clinical Chemistry, University Medical Centre, Göttingen, Germany. <sup>5</sup>Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. <sup>6</sup>Laboratory for Microbiology, Department of Biology, Philipps-University Marburg, Marburg, Germany.  
\*e-mail: [lennart.randau@mpi-marburg.mpg.de](mailto:lennart.randau@mpi-marburg.mpg.de)

**Supplementary information for**

**Type IV CRISPR RNA processing and effector complex formation in  
*Aromatoleum aromaticum***

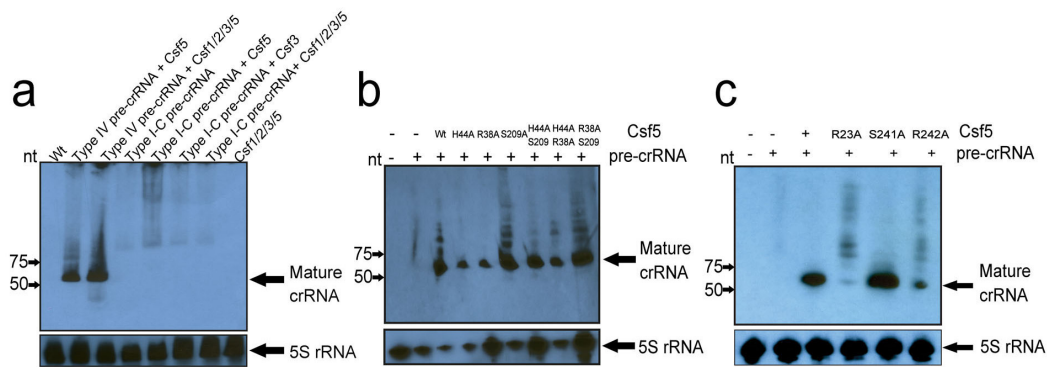
Ahsen Özcan, Patrick Pausch, Andreas Linden, Alexander Wulf, Karola Schühle,  
Johann Heider, Henning Urlaub, Thomas Heimerl, Gert Bange, Lennart Randau

**Content:**

**Supplementary Figures 1 – 6**

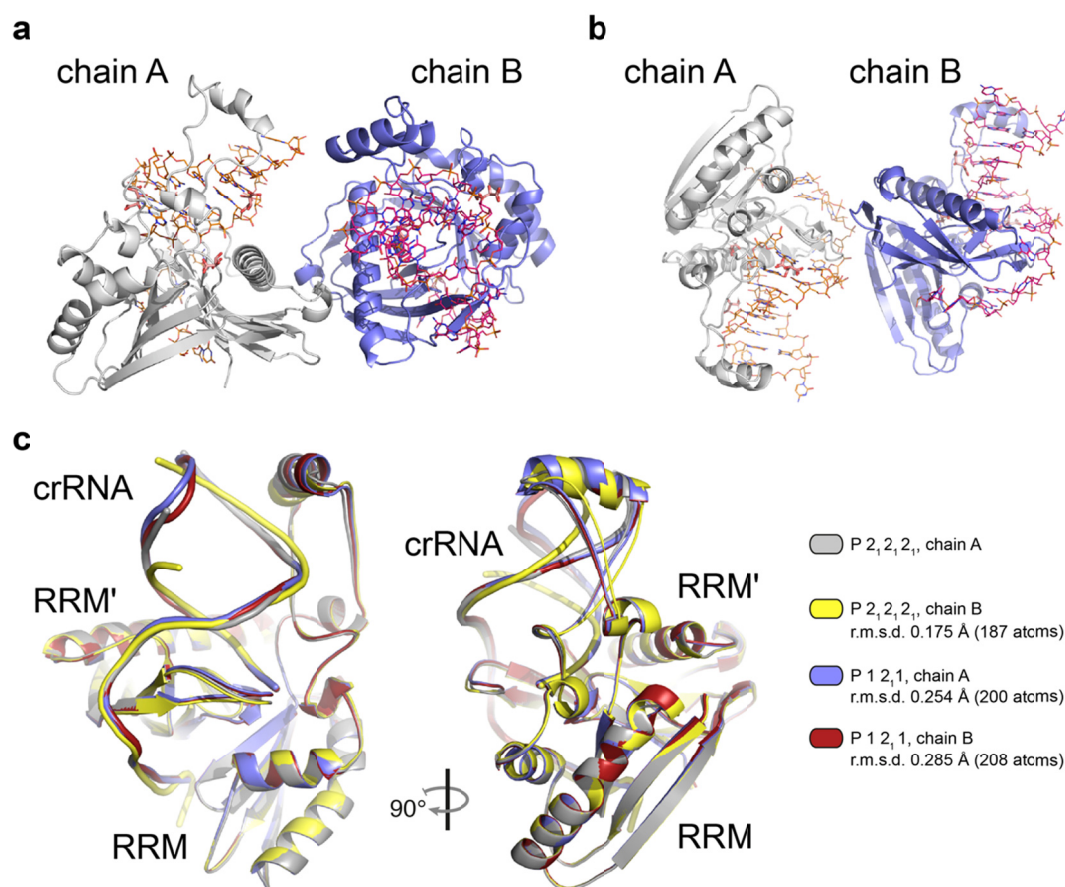
**Supplementary Tables 1-2, 4-5**

## Supplementary Figure 1



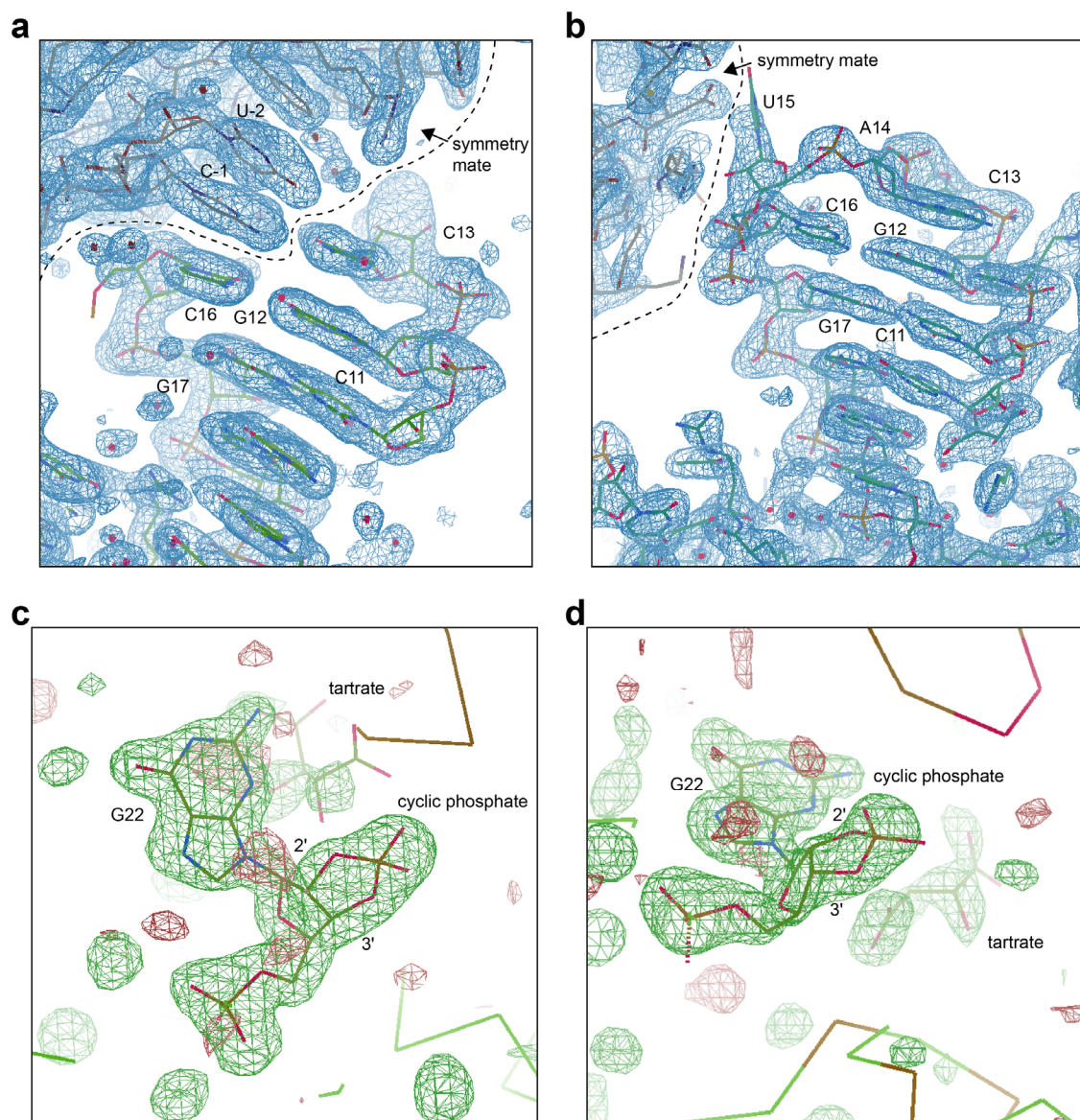
**Supplementary Figure 1, Csf5 crRNA maturation activity.** **a**, Indicated Cas proteins and a Type IV repeat-spacer-repeat-array were produced in *E. coli* BL21-AI (Wild type (Wt)). Northern blots were performed with a probe against the spacer. Repeat-spacer-repeat transcripts with identical spacers were produced either with Type IV or Type I-C repeats in *E. coli* BL21-AI together with the indicated Cas proteins. Mature crRNAs were only detected with Type IV repeats together with Csf5 or a crRNP complex (Csf1, Csf2, Csf3 and Csf5). **b**, H44A, R38A and S209A variants Csf5 were screened for crRNA maturation. **c**, R23A, S241A and R242A variants of Csf5 were assayed for crRNA production. The results in **a – c** are representatives of three experiments.

## Supplementary Figure 2



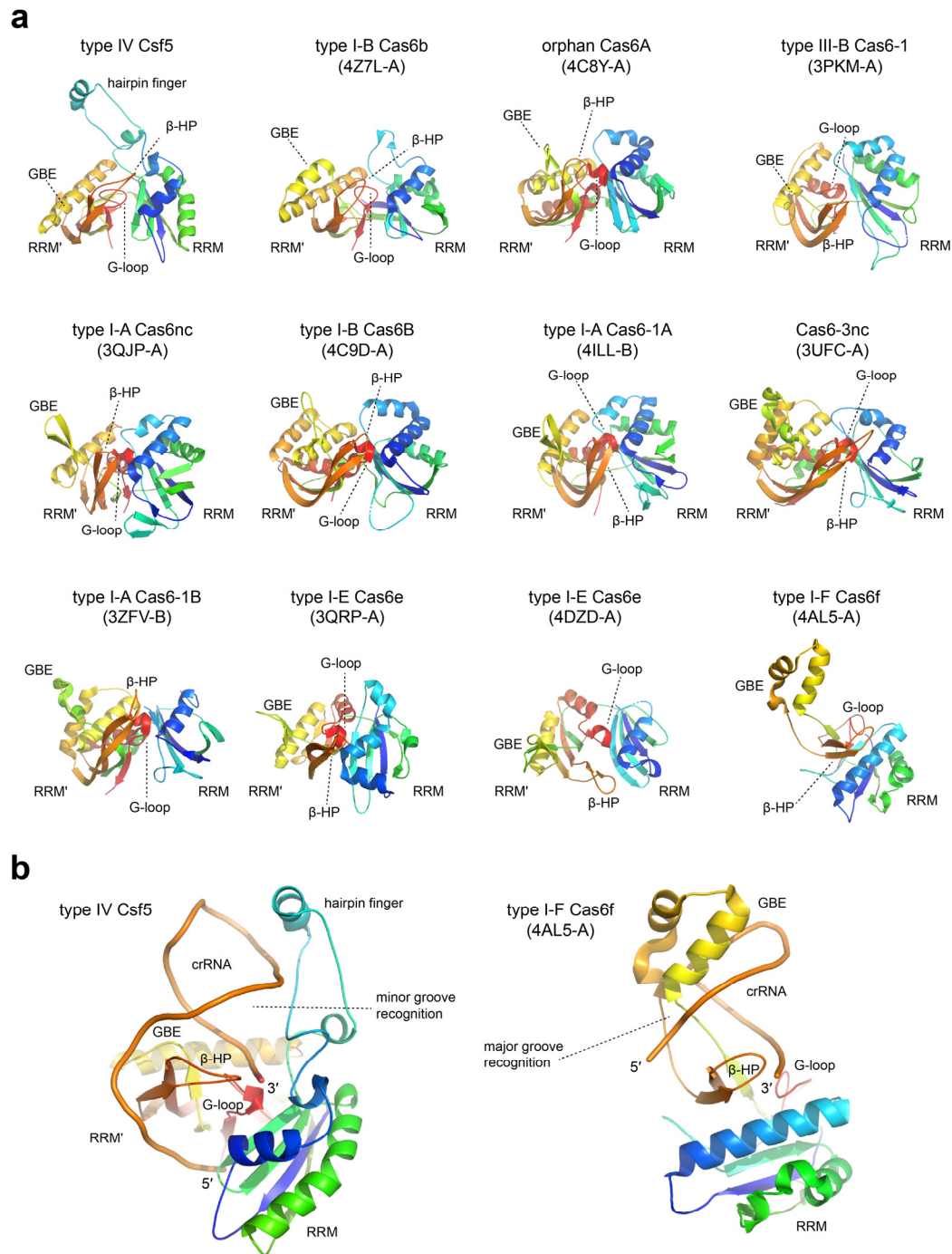
**Supplementary Figure 2, Crystal structures of Csf5. a and b,** Asymmetric unit content of Csf5 associated to the degraded (a) and intact hairpin (b). Csf5 chains are labeled according to their identity. **c,** superimposition of the individual Csf5 and crRNA structures of the P 2<sub>1</sub> 2<sub>1</sub> 2<sub>1</sub> (degraded hairpin) and P 1 2<sub>1</sub> 1 (intact hairpin) symmetry datasets. Coloring is according to the color scheme shown in the right panel. The r.m.s.d. values against chain A of the P 2<sub>1</sub> 2<sub>1</sub> 2<sub>1</sub> structure are shown in the right panel.

### Supplementary Figure 3



**Supplementary Figure 3, Electron density maps. a and b**, refined  $2F_o-F_c$  maps at  $1.1 \sigma$  (blue mesh) around the degraded (a) and intact (b) crRNA 3'-repeat tag hairpins. Nucleotides are labeled according to their identity. Symmetry mate regions are indicated by dotted lines and are indicated by arrows. **c and d**, unrefined  $F_o-F_c$  difference map at  $3 \sigma$  observed after molecular replacement of Csf5 associated with the degraded hairpin (c, chain A; d, chain B). Green mesh indicates positive density and red mesh indicates negative density. Shown is a close-up of the active site.

## Supplementary Figure 4



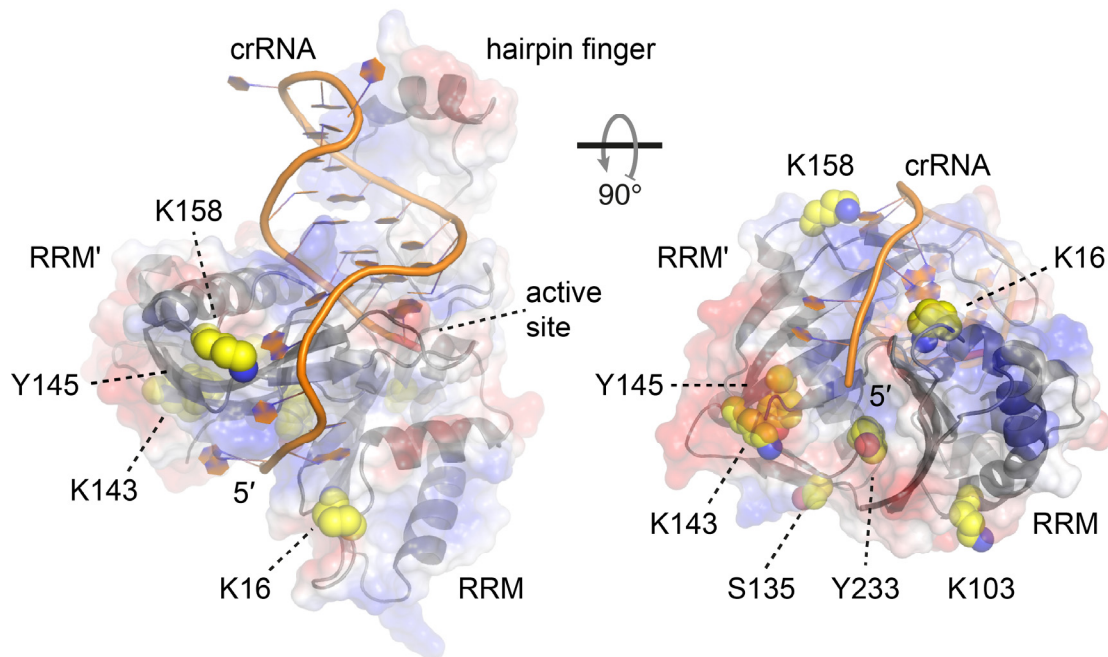
### Supplementary Figure 4, Comparison of Csf5 to other Cas6 family proteins.

**a**, side-by-side presentation of Csf5 and the Cas6 family proteins listed in Supplementary Table 2. The structures are shown in cartoon representation and are rainbow colored from the N-terminus in blue to the C-terminus in red. Comparison with an exemplary set of diverse Cas6 homologs revealed that Csf5 shares the conserved RRM and RRM' domains, as well as the G-loop,  $\beta$ -HP and GBE elements with other Cas6 structures. This conserved core fold

compares best with the type I-B Cas6b homolog from *Methanococcus maripaludis* (PDB-ID: 4Z7L, 33) with a Z-score of 17.7.

**b**, structures of Csf5 (left) and Cas6f (right) associated with the recognized crRNA hairpin (orange backbone). The  $\alpha$ -helical finger domain of Csf5 that interacts with the crRNA repeat hairpin is not found in other Cas6 proteins. However, the most distant Cas6 homolog, i.e. type I-F Cas6f with a Z-score of 3.2 (PDB-ID: 4AL5, 34), was found to feature a similar structure that forms extensive contacts with the crRNA repeat hairpin. In contrast to Csf5, the Cas6f hairpin interacting element is evolved from the C-terminal RRM' domain GBE structure and contacts the major groove of the crRNA hairpin.

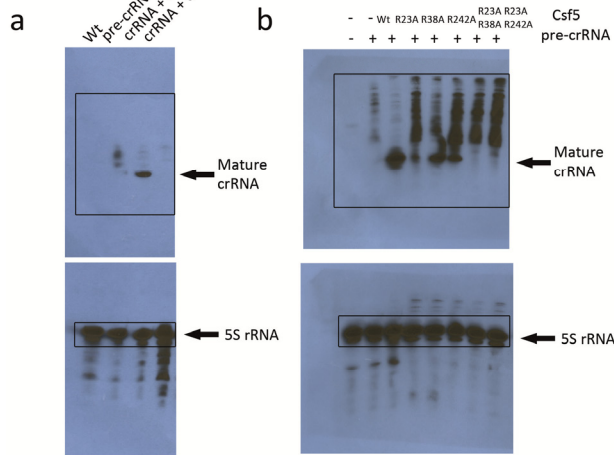
### Supplementary Figure 5



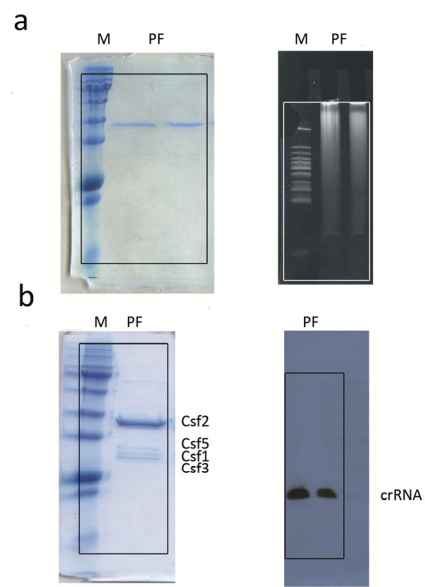
**Supplementary Figure 5, Csf5 – crRNA interactions.** Amino acids that were identified to UV crosslink with crRNA (Supplementary Table 3) are indicated by space-filling (yellow) and projected onto the Csf5 crystal structure.

**Supplementary Figure 6**

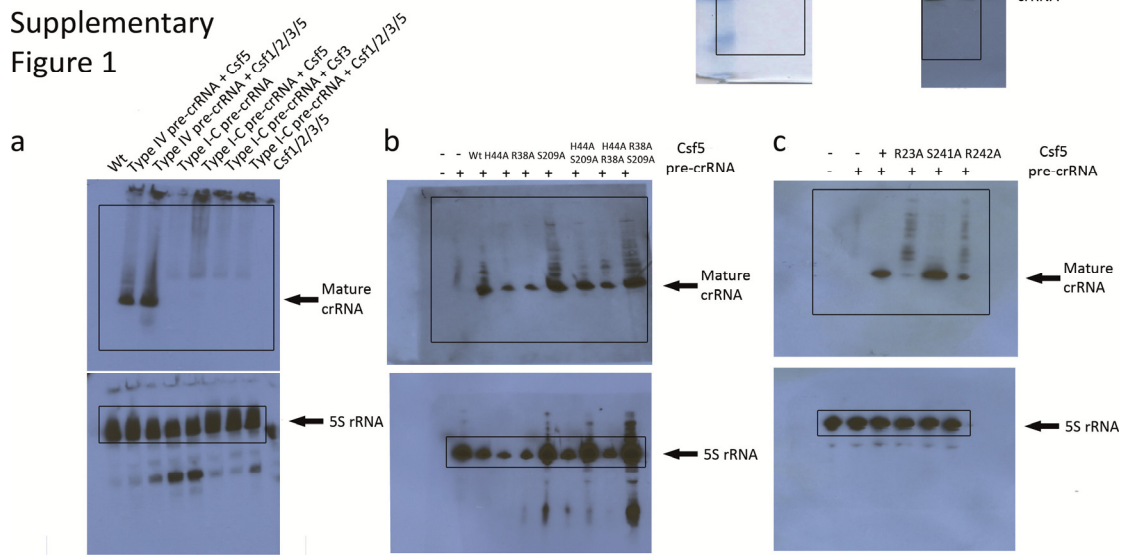
**Figure 1**



**Figure 4**



**Supplementary Figure 1**



**Supplementary Figure 6, Uncropped gels, Uncropped gels of indicated panels for Figure 2, Figure 4 and Supplementary Figure 1 are provided. The cropped areas are boxed.**



## Supplementary Tables

**Supplementary Table 1, Data collection and refinement statistics (MR and SAD)**

	Csf5 intact hairpin	Csf5 degraded hairpin	Csf5 VIL derivative
<b>Data collection</b>			
Method	MR	MR	SAD
Wavelength (Å)	0.966	0.972	1.750
Space group	P 1 2 <sub>1</sub> 1	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	62.85, 84.24, 71.74	85.56, 100.99, 101.93	85.540, 99.028, 101.559
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90, 100.35, 90	90, 90, 90	90, 90, 90
Resolution (Å)	43.76 - 2.29 (2.37 - 2.29)	45.25 - 1.75 (1.81 - 1.75)	45.19 - 3.19 (3.41 - 3.19)
Unique reflections	32620 (2997)	89451 (8838)	14803 (2554)
<i>R</i> <sub>merge</sub>	0.074 (0.718)	0.048 (1.127)	0.234 (1.364)
<i>R</i> <sub>pim</sub>	0.050 (0.506)	0.019 (0.437)	0.077 (0.449)
<i>I</i> / <i>σI</i>	10.83 (1.45)	20.32 (1.57)	7.7 (1.6)
CC1/2	0.998 (0.715)	0.999 (0.731)	0.993 (0.76)
Completeness (%)	98.29 (91.17)	99.88 (99.83)	99.3 (97.1)
Redundancy	3.1 (2.8)	7.4 (7.5)	9.7 (9.6)
Anom. completeness	-	-	98.0 (96.0)
Anom. redundancy	-	-	5.1 (5.0)
<b>Refinement</b>			
Resolution (Å)	43.76 - 2.29 (2.37 - 2.29)	45.25 - 1.75 (1.81 - 1.75)	
No. reflections			
total	32603 (2995)	89435 (8833)	
<i>R</i> <sub>free</sub>	1638 (154)	4426 (443)	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	21.8 / 25.1	21.2 / 23.7	
No. atoms	5145	5347	
Protein	4934	4939	
Ligand/ion	101	99	
Water	110	309	
Protein residues	500	500	
Ramachandran			
Favored (%)	98.19	99.4	
Allowed (%)	1.81	0.6	
Outliers (%)	0.00	0.00	
Rotamer outliers (%)	2.67	1.21	
<i>B</i> -factors			
Protein	51.02	39.33	
Ligand/ion	37.11	45.67	
Water	47.14	55.43	
R.m.s. deviations			
Bond lengths (Å)	0.005	0.008	
Bond angles (°)	1.12	1.25	

\*For each structure, the data was derived from a single crystal. \*Values in parentheses are for highest-resolution shell.

**Supplementary Table 2, Comparison of Csf5 and Cas6 structures**

<b>organism</b>	<b>type</b>	<b>PDB-ID</b>	<b>Z-score</b>	<b>r.m.s.d.*</b>	<b>length of alignment</b>	<b>length of chain</b>	<b>identity (%)</b>
<i>Methanococcus maripaludis</i>	type I-B Cas6b	4Z7L-A	17.7	3.0	196	218	13
<i>Thermus thermophilus</i>	orphan Cas6A	3C8Y-A	11.8	3.5	181	238	17
<i>Pyrococcus furiosus</i>	type III-B Cas6-1	3PKM-A	11.2	3.9	180	233	10
<i>Pyrococcus horikoshii</i>	type I-A Cas6nc	3QJP-A	10.1	3.9	174	240	9
<i>Thermus thermophilus</i>	type I-B(?) Cas6B	4C9D-A	10.0	4.1	181	262	9
<i>Sulfolobus solfataricus</i>	type I-A Cas6-1A	4ILL-B	9.7	3.6	170	277	10
<i>Pyrococcus furiosus</i>	Cas6nc-3nc	3UFC-A	9.6	3.9	172	243	9
<i>Thermus thermophilus</i>	type I-A Cas6-1B	3ZFV-B	8.4	4.0	162	258	12
<i>Thermus thermophilus</i>	type I-E Cas6e	3QRP-A	8.0	4.1	164	214	12
<i>Escherichia coli</i>	type I-E Cas6e	4DZD-A	6.3	4.1	133	193	11
<i>Pseudomonas aeruginosa</i>	type I-F Cas6f	4AL5-A	3.2	3.6	88	189	9

**Supplementary Table 2, Selected Cas6 homologs for comparison to Csf5.** \*r.m.s.d. (root mean square deviation). The structure of Csf5 was compared to all structures deposited in the PDB via the DALI protein structure comparison server. The DALI search confirmed that Csf5 is a structural homolog of the Cas6 family. The indicated Cas6 homologs were selected from all publicly available Cas6 structures (PDB database, [www.rcsb.org](http://www.rcsb.org)).

**Supplementary Table 4**

Peptide	Protein	Adduct
kTNPYDIAR	Csf3 (Cas5)	U-H2O
GLDDP <b>k</b> R	Csf5 (Cas6)	U-H2O
GL <b>Tk</b> TAENTTQVASR	Csf5 (Cas6)	U-H2O
FALPAG <b>k</b>	Csf5 (Cas6)	U-H2O
ALPLSNGSEAVESGD <b>Gy</b> K	Csf2 (Cas7)	U
AS <b>k</b> HVYM(ox)GLFGGGTR	Csf2 (Cas7)	U-H2O
GVGVELEQHTL <b>s</b> IK	Csf5 (Cas6)	U
GLLL <b>c</b> ALR	Csf2 (Cas7)	U
<b>c</b> ELAALTVPSLMEFFTGK	Csf2 (Cas7)	U
R <b>y</b> FVR	Csf5 (Cas6)	CU
<b>h</b> lvRVDDVTR	Csf2 (Cas7)	U
<b>Gy</b> IEQHGHEPPPTTEEQR	Csf5 (Cas6)	U
IIADH <b>p</b> AMSTGR	Csf5 (Cas6)	U
GLDDP <b>k</b> R	Csf5 (Cas6)	U-H2O
LTEV <b>r</b> HIVR	Csf2 (Cas7)	U
<b>m</b> (Ox)GQQHLLR	Csf5 (Cas6)	U-H2O
IADEIA <b>k</b> AAEK	Csf1 (LS)	U-H2O
LTAPM(ox)HVADSIQYSLDD <b>Ky</b> VVR	Csf2 (Cas7)	U
<b>Gy</b> WFAGNLTSR	Csf5 (Cas6)	AU
V <b>k</b> IPYFPANDLR	Csf2 (Cas7)	U-H2O
<b>Yk</b> PLLEAATLPAIQIVTQR	Csf5 (Cas6)	U-H2O
<b>k</b> SDLANVVSQSIAGTELYVR	Csf2 (Cas7)	U-H2O

**Supplementary table 4, Type IV crRNA interacting Cas protein residues** Recombinant Type IV crRNP complexes were subjected to ultraviolet light crosslinking, followed by mass spectrometry. Specific amino acid residues (red) of the Cas proteins were found to interact with the indicated crRNA bases (adduct).

**Supplementary Table 5**

<b>Name</b>	<b>Sequence (5'-3')</b>
Type IV mini-CRISPR array F	GATCCTAATACGACTCACTATAGGGGTGTTCCCCGCGCATCGCGGGGGTTGAAGCGTCAG GTCTGCAACAAAGATCAACCCTACTCGGTGTTCCCCGCGCATCGCGGGGGTTGAAGA
Type IV mini-CRISPR array R	AGCTTCTTCAACCCCCGCGATGCGCGGGGAACACCGAGTAGGGTTGATCTTTGTTGCAGA CCTGACGCTTCAACCCCCGCGATGCGCGGGGAACACCCCTATAGTGAGTCGTATTAG
Type I-C mini-CRISPR array F	GAT CCT AAT ACG ACT CAC TAT AGG GGC ATC GCC CCT CGG TGA CGG GGG GCG TGG ATT GAA ACC GTC AGG TCT GCA ACA AAG ATC AAC
Type I-C mini-CRISPR array R	AAT CCA CGC CCC CCG TCA CCG AGG GGC GAT GCC CCT ATA GTG AGT CGT ATT AG
Type I-C mini-CRISPR array F2	CCT ACT CGG CAT CGC CCC TCG GTG ACG GGG GGC GTG GAT TGA AAC A
Type I-C mini-CRISPR array R2	AGC TTG TTT CAA TCC ACG CCC CCC GTC ACC GAG GGG CGA TGC CGA GTA GGG TTG ATC TTT GTT GCA GAC CTG ACG GTT TC
5S rRNA compl. sequence	DIG - GGGGTCAGGTGGGACCACCGCGCTACGGCCGCC
Type IV compl. spacer	DIG - CGTCAGGTCTGCAACAAAGATCAACCCTACTCG
Csf5 H44A F	CGG GTG ATC GCA GCA CCG GTC TGC GGA TCA C
Csf5 H44A R	GTG ATC CGC AGA CCG GTG CTG CGA TCA CCC G
Csf5 S209A F	GGT CGC ACC GGT GGC GGT CAC CAG ACG C
Csf5 S209A R	GCG TCT GGT GAC CGC CAC CGG TGC GAC C
Csf5 R38A F	CCG GTC TGC GGA TCA GCG CTA AAG AAC AGC GG
Csf5 R38A R	CCG CTG TTC TTT AGC GCT GAT CCG CAG ACC GG
Csf5 R23A F	CGC CAG CGC CTC AGC CAG ATC ATT CGG C
Csf5 R23A R	GCC GAA TGA TCT GGC TGA GGC GCT GGC G
Csf5 S241A F	GAC CGT AGC CAC GGG CGG TCA GGT TAC CCG
Csf5 S241A R	CGG GTA ACC TGA CCG CCC GTG GCT ACG GTC
Csf5 R242A F	ATA CGA CCG TAG CCA GCG CTG GTC AGG TTA CC
Csf5 R242A R	GGT AAC CTG ACC AGC GCT GGC TAC GGT CGT AT

**Supplementary Table 5, Oligonucleotide list**