Structural basis for CRISPR RNA-guided DNA recognition by Cascade

Jore, Matthijs M.; Lundgren, Magnus; van Duijn, Esther; Bultema, Jelle B.; Westra, Edze R.; Waghmare, Sakharam P.; Wiedenheft, Blake; Pul, Uemit; Wurm, Reinhild; Wagner, Rolf

Published in:
Nature Structural & Molecular Biology

DOI:
10.1038/nsmb.2019

Publication date:
2011

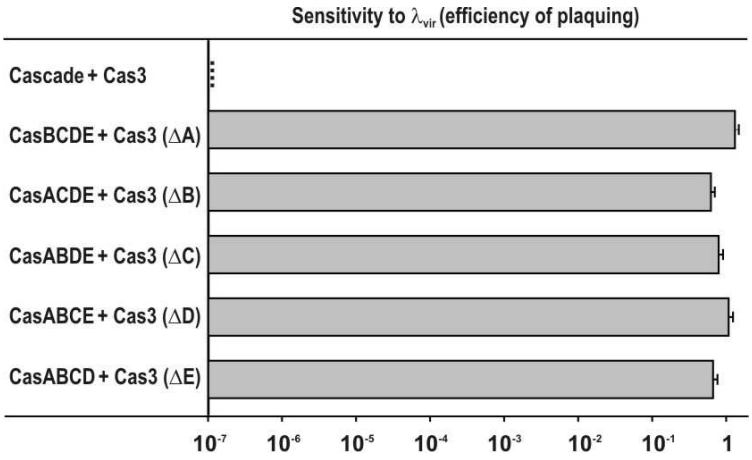
Supplementary information for

Structural basis for CRISPR RNA-guided DNA recognition by Cascade

Matthijs M. Jore, Magnus Lundgren, Esther van Duijn, Jelle B. Bultema, Edze R. Westra, Sakharam P. Waghmare, Blake Wiedenheft, Ümit Pul, Reinhild Wurm, Rolf Wagner, Marieke R. Beijer, Arjan Barendregt, Kaihong Zhou, Ambrosius P.L. Snijders, Mark J. Dickman, Jennifer A. Doudna, Egbert J. Boekema, Albert J. R. Heck, John van der Oost & Stan J.J. Brouns

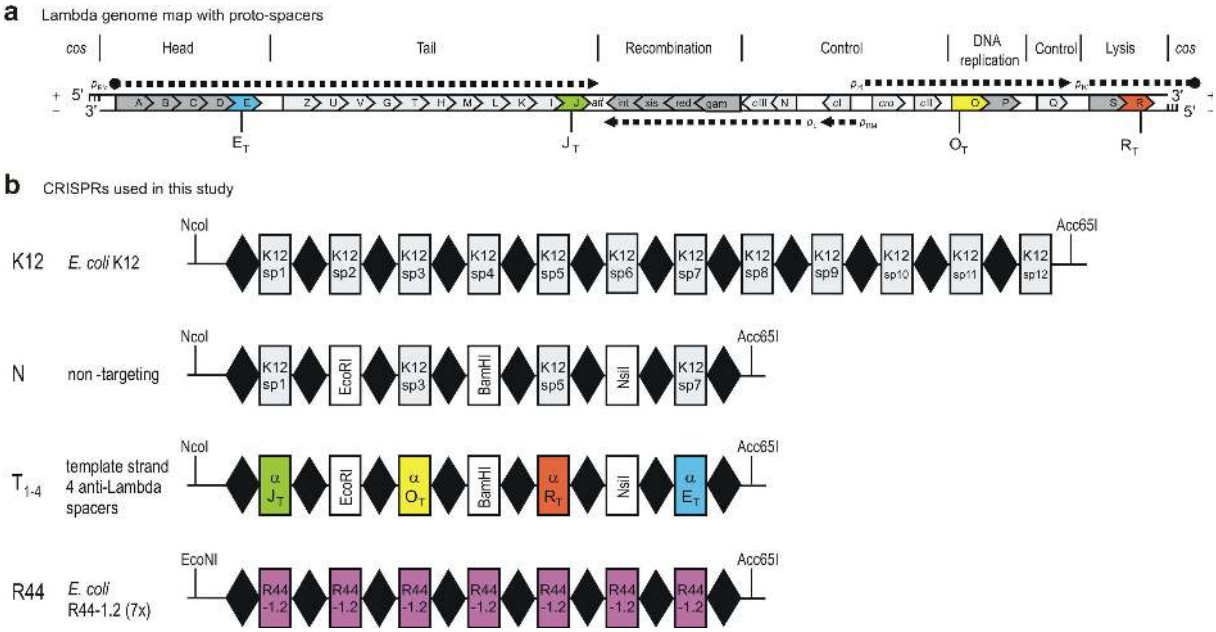
Nature Structural & Molecular Biology: doi:10.1038/nsmb.2019

Supplementary Fig. 1



All Cascade subunits are required for immunity. The efficiency of plaquing (EOP) was determined in *E. coli* BL21-AI containing complete Cascade, phage λ-targetting $T_{1\text{-}4}$ CRISPR (see Supplementary Fig. 2) and Cas3[1], and for strains expressing Cascade sub-complexes lacking one of the Cas proteins. Host strain *E. coli* BL21-AI does not contain *cas*-genes[2]. Error bars indicate one standard deviation.

Supplementary Fig. 2



CRISPRs used in this study. **(a)** Phage λ genome map indicating the main genes and transcripts (dotted arrows), and the positions of the protospacers on the coding or template strand. **(b)** Schematic overview of CRISPRs used in this study (repeats: diamonds, spacers: rectangles). The N CRISPR was designed as a non-targeting control containing the naturally occurring spacers 1, 3, 5, and 7 from *E. coli* K12, which have no homology to any known phage[1]. The number of plaque forming units obtained in the presence of this CRISPR was used to calculate the efficiency of plaquing (Supplementary Fig. 1). The uniform CRISPR R44 was designed based on a natural spacer at the second position of the CRISPR in *E. coli* R44 (ECOR44)[3,4]. This spacer was derived from the *upfA* gene of phage P1 and P7. The R44 CRISPR was designed such that the spacer was repeated seven times. The sequences of the CRISPRs are provided below.

K12 CRISPR, pWUR396

GCGTACCATGGCATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTTTTATGGGAAAAAATGCTTTAAGAACAA
ATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTCGCAGACGCGCGGCGATACGCTCACGC
AGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCCGCGC
CAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGAGTTCCCCGCGCCAGCGGGGATAAAC
CGTTTGGATCGGGTCTGGAATTTCTGAGCGGTCGCGAGTTCCCCGCGCCAGCGGGGATAAACCGCGAATCGCGCA
TACCCTGCGCGTCGCCGCCTGCGAGTTCCCCGCGCCAGCGGGGATAAACCGTCAGCTTTATAAATCCGGAGATAC
GGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGGACTCACCCCGAAAGAGATTGCCAGCCAGCTTGAGTTC
CCCGCGCCAGCGGGGATAAACCGCTGCTGGAGCTGGCTGCAAGGCAAGCCGCCCAGAGTTCCCCGCGCCAGCGGG
GATAAACCGGGGGGCGCATGACCGTAAACATTATCCCCCGGGAGTTCCCCGCGCCAGCGGGGATAAACCGGGAGT
TCAGACATAGGTGGAATGATGGACTACGAGTTCCCCGCGTTAGCGGGGATAAACCGCCCGGTAGCCAGGTTTGCA
ACGCCTGAACCGAGAGTTCCCCGCGCCAGCAGGGATAAACCGGCAACGACGGTGAGATTTCACGCCTGACGCTGG
GTACCGGACC

Non-targeting CRISPR (N), pWUR477

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTT
CGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGA
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGCGAATCGCGCATACCCTGCGCGTCGCCGCCTGGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGGACTCACCCCGAA
AGAGATTGCCAGCCAGCTTGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Template CRISPR (T1-4), pWUR478

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTGA
GTGTGATCGATGCCATCAGCGAAGGGCCGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGCAAGCAACAGGCAGGCGTGACAGCCAGCAAAC
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGTGGGATGCCTACCGCAAGCAGCTTGGCCTGAAGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGTGACAAGTCCACG
TATGACCCGACCGACGATAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

ECOR44-1.2 (7x), pWUR547

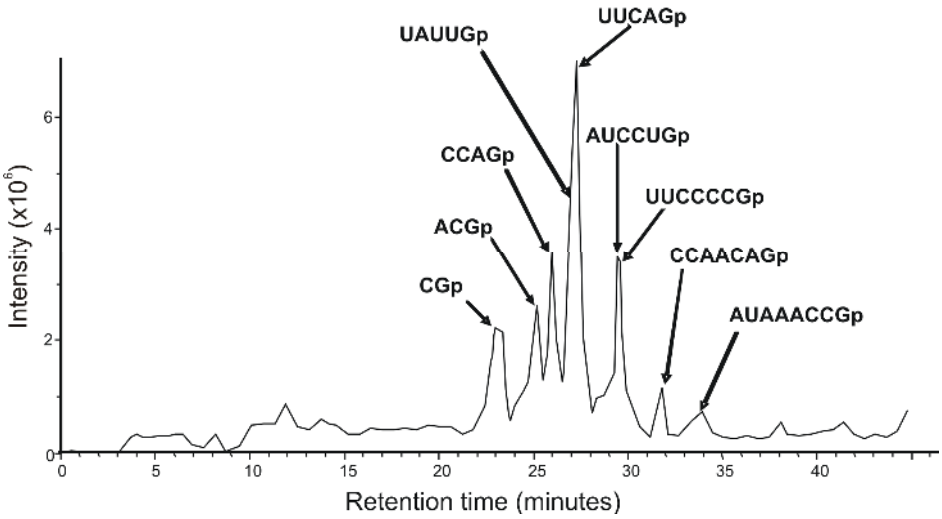
CCTGCATTAGGTAATACGACTCACTATAGGATAAACCGACGGTATTGTTCAGATCCTGGCTTGCCAACAGGAGTT
CCCCGCGCCAGCGGGGATAAACCGACGGTATTGTTCAGATCCTGGCTTGCCAACAGGAGTTCCCCGCGCCAGCGG
GGATAAACCGACGGTATTGTTCAGATCCTGGCTTGCCAACAGGAGTTCCCCGCGCCAGCGGGGATAAACCGACGG
TATTGTTCAGATCCTGGCTTGCCAACAGGAGTTCCCCGCGCCAGCGGGGATAAACCGACGGTATTGTTCAGATCC
TGGCTTGCCAACAGGAGTTCCCCGCGCCAGCGGGGATAAACCGACGGTATTGTTCAGATCCTGGCTTGCCAACAG
GAGTTCCCCGCGCCAGCGGGGATAAACCGACGGTATTGTTCAGATCCTGGCTTGCCAACAGGAGTTCCCCGCGCC
AGCGGGGATAAACCGGGTACC

Supplementary Fig. 3

a

5' OH – AUAAACCGACGGUAUUGUUCAGAUCCUGGCUUGCCAACAGGAGUUCCCCCGCGCCAGCGGGG – 2', 3'-P

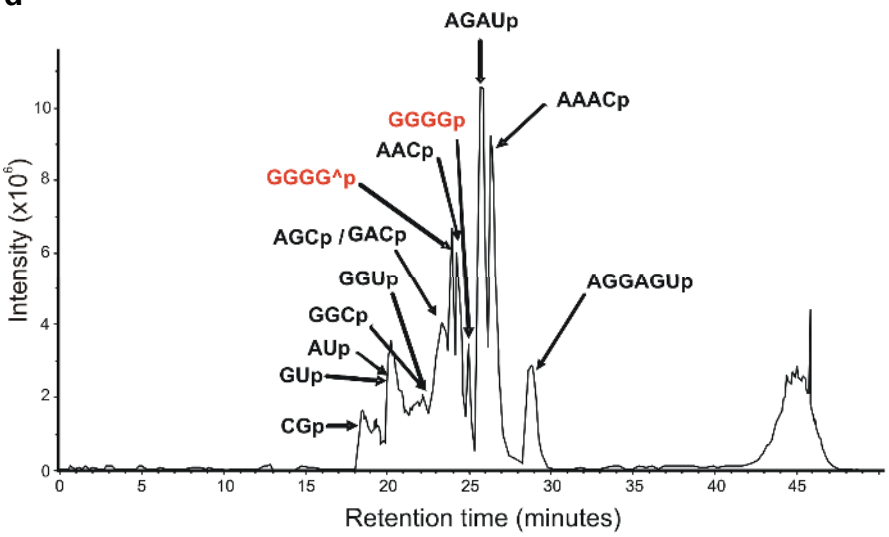
b



c

Sequence	Position	Theoretical mass (Da)	Experimental mass (Da)
AUAAACCGp	A1:G8	2596.597	2596.2
ACGp	A9:G11	997.617	997.1
UAUUGp	U13:G17	1610.94	1610.1
UUCAGp	U18:G22	1609.955	1609.3
AUCCUGp	A23:G28	1915.139	1914.9
CUUGp	C30:G33	1280.745	1280.2
CCAACAGp	C34:G40	2266.403	2266.1
AGp	A42:G43	692.433	692.1
UUCCCCGp	U44:G50	2196.298	2195.3
CCAGp	C53:G56	1302.801	1302.3

d

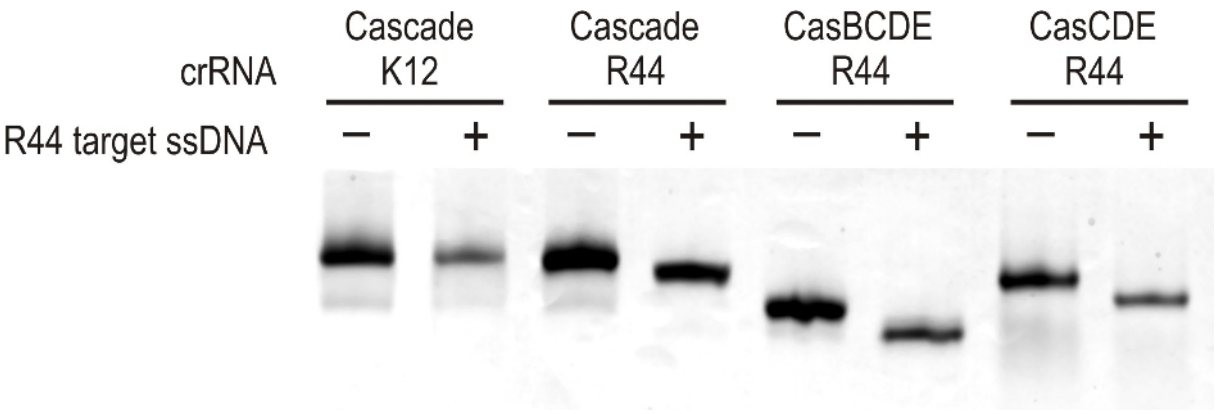


e

Sequence	Position	Theoretical mass (Da)	Experimental mass (Da)
AUp	A1:U2 / A14:U15	653.393	653.1
AAACp	A3:C6	1310.826	1310.4
GACp / AGCp	G8:C10 / A55:C57	997.617	997.1
GGUp	G11:U13	1014.602	1014.2
GUp	G17:U18	669.393	669.0
AGAUp	A21:U24	1327.811	1327.3
GGCp	G28:C30	1013.617	1013.4
AACp	A36:C38	981.617	981.2
AGGAGUp	A39:U44	2018.229	2017.7
GGGGp	G58:G61	1398.851	1398.4
GGGG^p	G58:G61	1380.836	1380.4

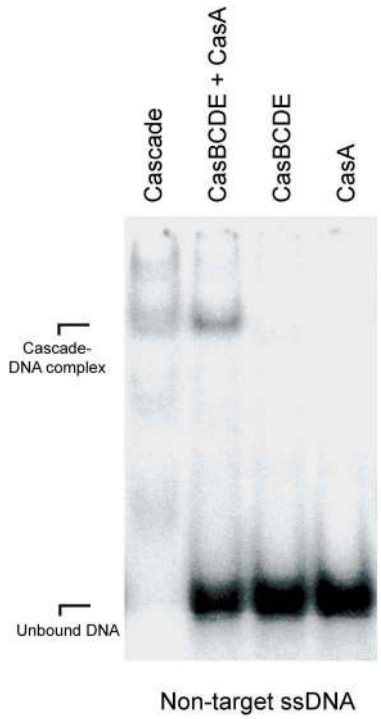
RNase digests of the crRNA. **(a)** Sequence of mature crRNA with the R44 spacer sequence shaded, and the hairpin underlined. **(b)** Base peak chromatogram of the RNaseT1 digest of mature crRNA. The predominant oligoribonucleotide peaks assigned to the mature crRNA are highlighted. **(c)** Summary of the identified oligoribonucleotides assigned to mature crRNA from the RNaseT1 digest. **(d)** Base peak chromatogram of the RNaseA digest of mature crRNA. The predominant oligoribonucleotide peaks assigned to the mature crRNA are highlighted. **(e)** Summary of the identified oligoribonucleotides assigned to mature crRNA from the RNaseA digest. ^ indicates cyclic 2',3'-phosphate. It is worth mentioning that the cyclic phosphate 3'-end is not a substrate for *E. coli* poly(A) RNA polymerase[5], and this explains why only shorter, apparently partly degraded crRNAs were cloned and sequenced previously[1].

Supplementary Fig. 4



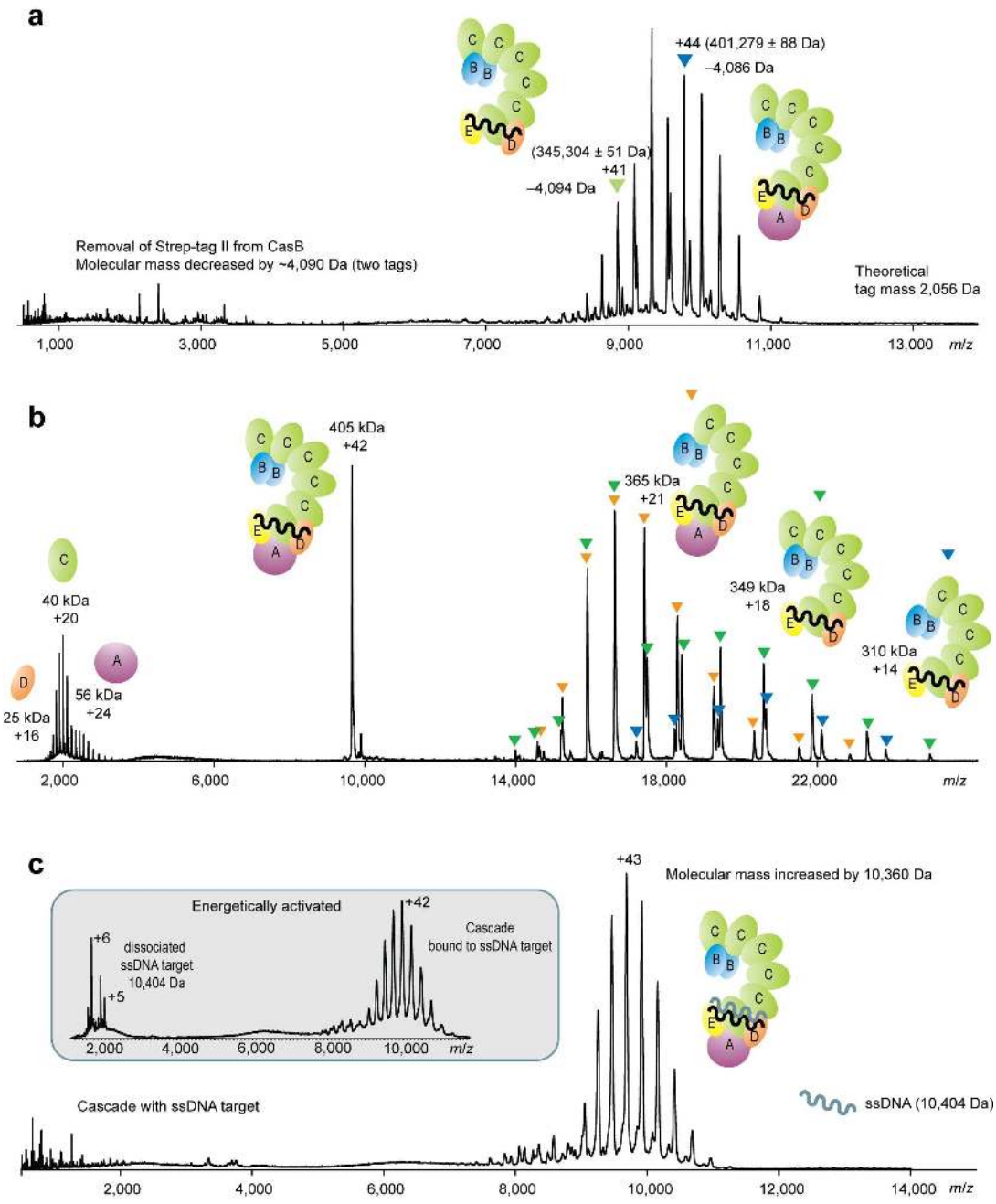
Native PAGE of Cascade-target DNA complexes. Coomassie-blue stained native PAGE analysis of Cascade and sub-complexes binding ssDNA oligonucleotides complementary to the R44 crRNA (BG3028, see Table S3). The gel shows the increased migration rates of Cascade complexes loaded with uniform R44 crRNA in the presence their complementary target ssDNA due to the additional charge negative charge of the ssDNA. By contrast, the migration rate of Cascade loaded with different crRNAs derived from the *E. coli* K12 CRISPR I array that does not contain the R44 spacer is not affected. In addition, differences in migration rates of the various complexes (Cascade, CasBCDE and CasCDE) are visible.

Supplementary Fig. 5



CasA complementation of non-specific DNA binding. Binding ability of uniform crRNA-loaded (R44) Cascade, CasBCDE + CasA, CasBCDE and CasA to non-target ssDNA is shown. While CasA does not exhibit non-specific DNA binding on by itself, it is able to restore the non-specific DNA binding of CasBCDE.

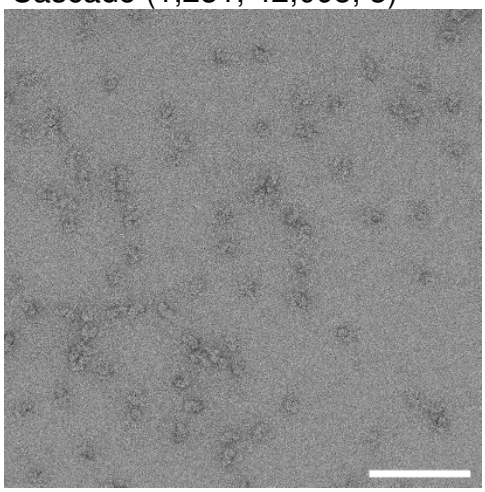
Supplementary Fig. 6



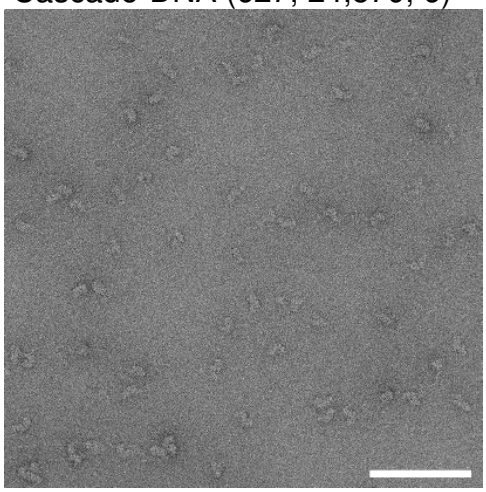
(a) Native mass spectrum of Cascade after treatment with HRV3C protease. A dominant species with a mass of 401,279 Da (blue triangle) was observed, confirming the presence of two copies of CasB in the intact assembly. Indicated by the green triangles is the complex lacking CasA. **(b)** Tandem mass spectrum of the 42+ ion of Cascade. Besides the dissociation of CasA (green) also CasC (orange) dissociated from the complex. The complex lacking CasA further expels a CasC subunit to form a 310 kDa Cascade sub-complex (blue). The low m/z region of the spectrum shows the dissociated CasA, CasC and CasD proteins. Overlapping peaks of two different complexes are indicated by two colors. **(c)** Native mass spectrum of Cascade bound to the ssDNA-probe. The mass of the complex increased by 10,201 Da, indicating the presence of one crRNA per Cascade. The inset shows the same spectrum after energetically activating the Cascade-ssDNA probe complex. The charge state distribution for the ssDNA probe is centered around 2,000 m/z.

Supplementary Fig. 7

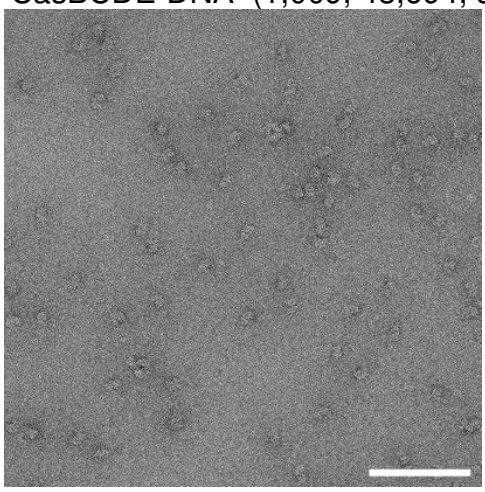
a Cascade (1,231, 42,098, 3)



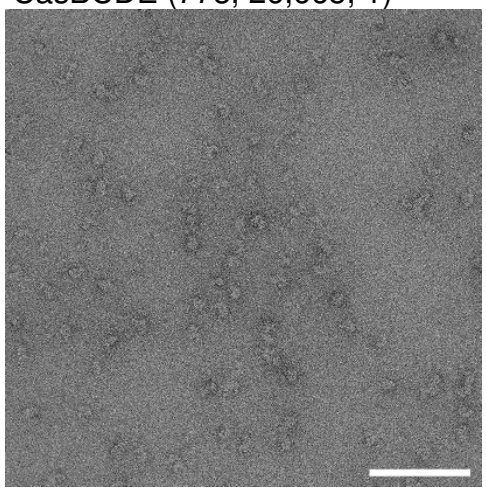
b Cascade-DNA (627, 24,579, 6)



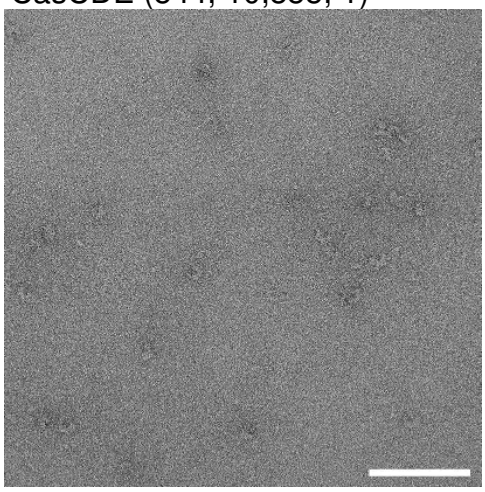
c CasBCDE-DNA (1,009, 43,694, 3)



d CasBCDE (773, 26,665, 1)

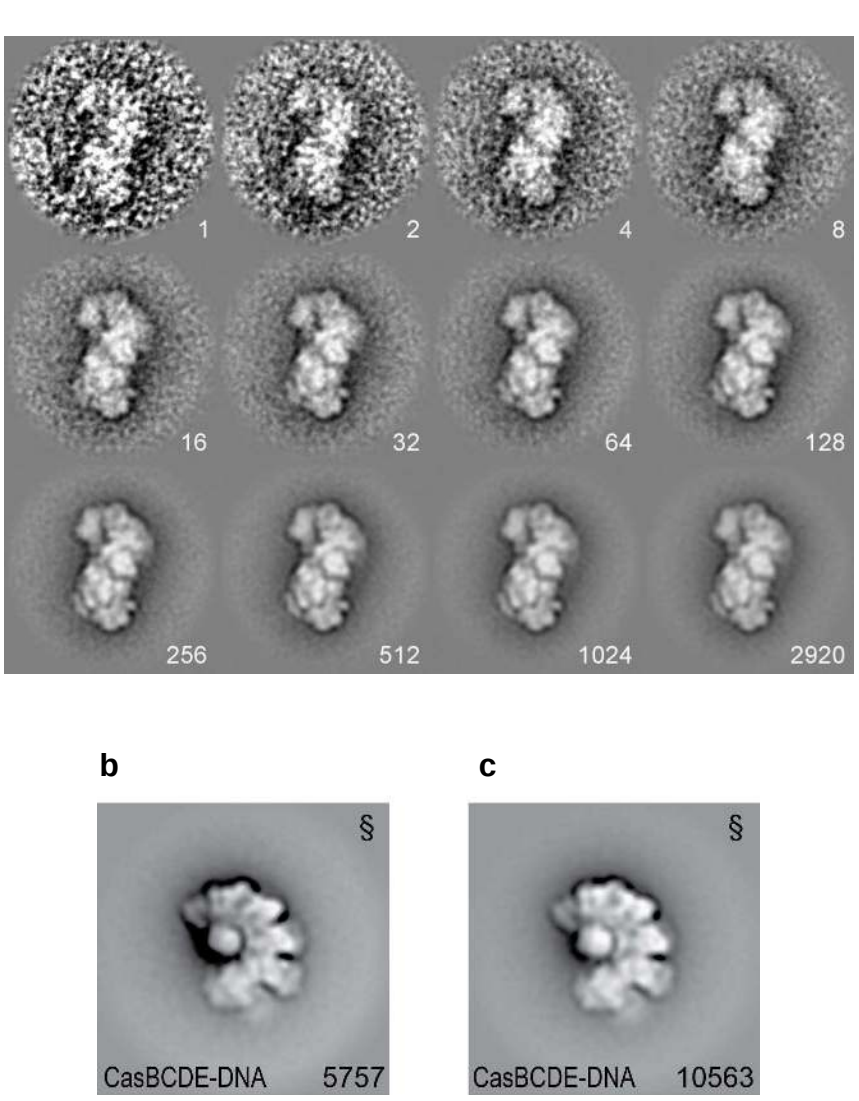


e CasCDE (544, 10,855, 1)



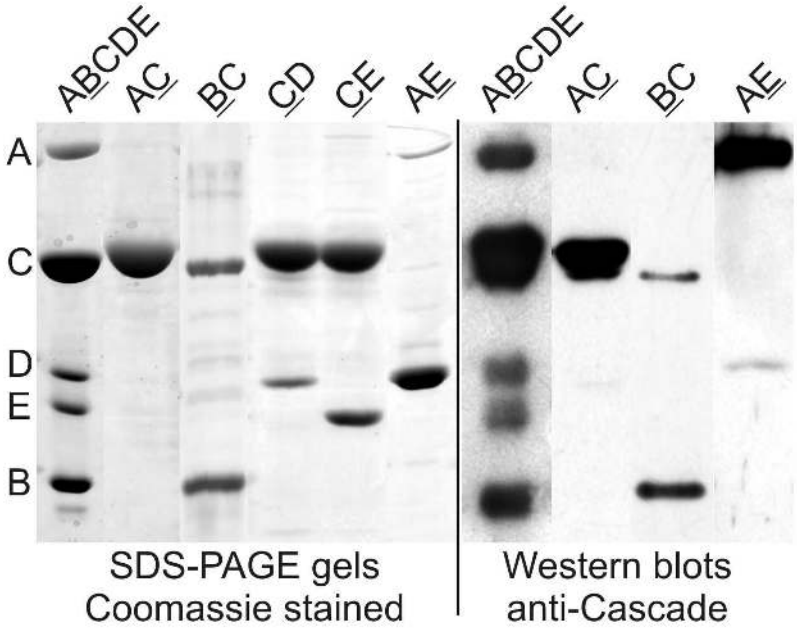
(a-e) Raw Electron Micrographs of Cascade, CasBCDE and CasCDE preparations in the presence or absence of target ssDNA. The number of raw micrographs used to generate the class averages in Fig. 6, the total number of particles analysed, and the number of particle classes are shown in brackets, respectively. The scale bar represents 100 nm.

Supplementary Fig. 8



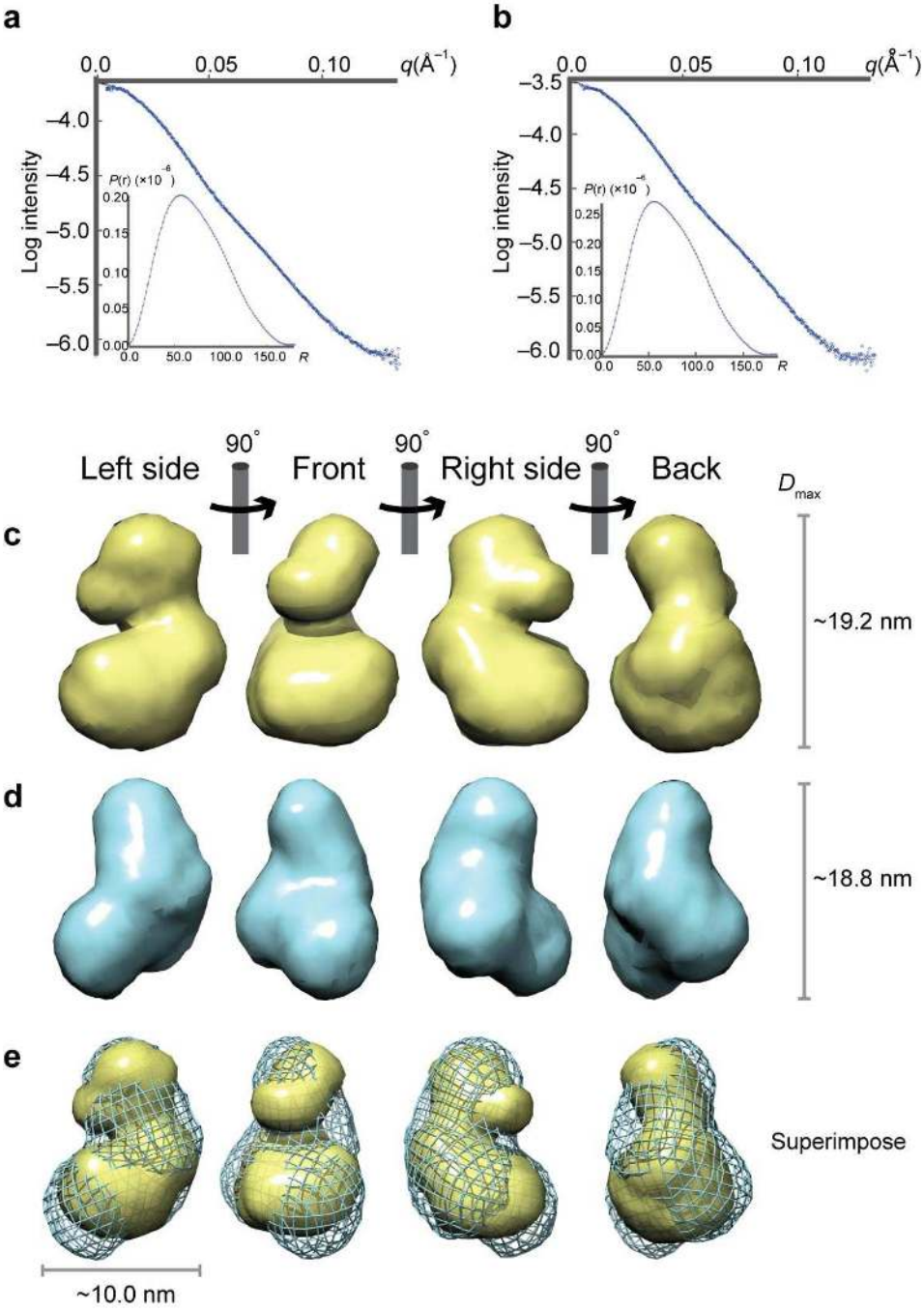
(a) Single particle EM of Cascade. After translational and rotational alignment of a data set with single particle projections, sums with increasing numbers of copies in equivalent positions show the gradual improvement in the signal-to-noise ratio. **(b)** and **(c)** The two remaining classes of CasBCDE-DNA particles that are not depicted in Fig. 6. The number of particle projections to create the average shown is depicted on the bottom-right corner of each image. The § sign indicates that the depicted particle is a mirrored view of the original to match the predominant orientation of the Cascade projection shown in Fig. 6A.

Supplementary Fig. 9



Interactions between individual Cascade subunits. Composite of Coomassie blue-stained SDS-polyacrylamide gels and Western blots of affinity purified tagged Cascade subunits (underlined) co-expressed with untagged Cascade subunits and pre-crRNA. All subunits directly interact with CasC except CasA, which interacts with CasE. Western blots were probed with anti-Cascade antibodies. Note that the N-terminal Strep-tagged Cascade subunits migrate slower through the gel than the untagged proteins.

Supplementary Fig. 10



Solution scattering model of Cascade obtained with Small-Angle X-ray Scattering. Scattering data for Cascade were collected at 10 keV (λ = 1.24 Å) from two protein concentrations, and include scattering vectors *(q)*, ranging from **(a)** 0.015 $Å^{-1}$ to 0.127 $Å^{-1}$ for Cascade and **(b)** 0.015 $Å^{-1}$ to 0.133 $Å^{-1}$ for Cascade bound to target DNA. Guinier approximations of each curve estimates the radius of gyration (R(g)) for Cascade and for Cascade bound to target DNA at 5.6 nm. The pair-distribution function (insert) indicates that the radius of gyration for both particles is ~5.6 nm. **(c)** *Ab initio* reconstructions of Cascade reveal a seahorse shaped complex, consistent with EM imaging. **(d)** DNA binding induces a conformational change in Cascade. **(e)** Superposition of the solution structures of Cascade without (yellow) and with target DNA (mesh). Images have been rendered using Chimera[6].

Supplementary Table 1

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

Cascade component	**Theoretical Mass (Da)**	**Experimental Mass (Da)**
CasA	55,972.4	55,972.2 ± 14.8
CasB with tag	21,260.4	21,261.5 ± 1.1
CasB without tag	19,204.0	19,201.9 ± 1.8
CasC	39,894.4	39,896.3 ± 1.3
CasD	25,208.9	25,210.4 ± 3.8
CasE	22,364.1	22,364.7 ± 1.1
crRNA	19,660.8	19,708[a]
Complex	**Theoretical Mass (Da)**	**Experimental Mass (Da)**
Cascade	405,095	405,365 ± 135
Cascade without CasA	349,122	349,399 ± 84
CasB dimer	42,521	42,524 ± 8

[a] The mass of crRNA was indirectly calculated via tandem mass spectrometry analysis of CasE-crRNA. Since only dissociation products for CasE were observed after collisional activation no standard deviation can be calculated for the mass of crRNA.

Supplementary Table 2

List of masses for all Cascade (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 a crRNA mass of 19,662 Da) and stoichiometric information is given. (A=CasA, B=CasB, C=CasC, D=CasD, E=CasE, minus (-) indicates the lacking subunit, n.d.is not determined.

Mass of (sub) complexes in solution (Da)	Mass products (Da)	Theoretical mass (Da)	Annotation	Stoichiometry A B C D E crRNA
405,256		405,095	Cascade	1 2 6 1 1 1
	365,316	365,200	Cascade-CasC	1 2 5 1 1 1
	349,384	349,122	Cascade-CasA	0 2 6 1 1 1
	55,950	55,972	CasA	1 0 0 0 0 0
	39,941	39,894	CasC	0 0 1 0 0 0
349,333		349,122	Cascade-CasA	0 2 6 1 1 1
	324,389	323,914	Cascade-CasA-CasD	0 2 6 0 1 1
	309,644	309,228	Cascade-CasA-CasC	0 2 5 1 1 1
	270,031	269,334	Cascade-CasA-2·CasC	0 2 4 1 1 1
	39,946	39,894	CasC	0 0 1 0 0 0
	25,231	25,209	CasD	0 0 0 1 0 0
306,932		306,602	Cascade-CasA-2·CasB	0 0 6 1 1 1
	281,915	281,393	Cascade-CasA-2·CasB-CasD	0 0 6 0 1 1
	267,215	266,707	Cascade-CasA-2·CasB-CasC	0 0 5 1 1 1
	227,267	226,813	Cascade-CasA-2·CasB-2·CasC	0 0 4 1 1 1
	39,935	39,894	CasC	0 0 1 0 0 0
	25,250	25,209	CasD	0 0 0 1 0 0
267,076		266,707	Cascade-CasA-2·CasB-CasC	0 0 5 1 1 1
	241,503	241,498	Cascade-CasA-2·CasB-CasC-CasD	0 0 5 0 1 1
	226,673	226,813	Cascade-CasA-2·CasB-2·CasC	0 0 4 1 1 1
	39,928	39,894	CasC	0 0 1 0 0 0
	25,227	25,209	CasD	0 0 0 1 0 0
227,127		226,813	Cascade-CasA-2·CasB-2·CasC	0 0 5 1 1 1
	202,019	201,604	Cascade-CasA-2·CasB-2·CasC-CasD	0 0 4 0 1 1
	187,241	186,918	Cascade-CasA-2·CasB-3·CasC	0 0 3 1 1 1
	39,946	39,894	CasC	0 0 1 0 0 0
187,352		186,918	Cascade-CasA-2·CasB-3·CasC	0 0 3 1 1 1
	161,882	161,710	Cascade-CasA-2·CasB-3·CasC-CasD	0 0 3 0 1 1
	147,087	147,024	Cascade-CasA-2·CasB-4·CasC	0 0 2 1 1 1
	39,910	39,894	CasC	0 0 1 0 0 0
	25,218	25,209	CasD	0 0 0 1 0 0
147,210		147,024	Cascade-CasA-2·CasB-4·CasC	0 0 2 1 1 1
	121,869	121,815	Cascade-CasA-2·CasB-4·CasC-CasD	0 0 2 0 1 1
	n.d.	107,130	Cascade-CasA-2·CasB-5·CasC	0 0 1 1 1 1
	39,916	39,894	CasC	0 0 1 0 0 0
	25,219	25,209	CasD	0 0 0 1 0 0
107,431		107,130	Cascade-CasA-2·CasB-5·CasC	0 0 1 1 1 1
	84,696	84,766	Cascade-CasA-2·CasB-5·CasC-CasE	0 0 1 1 0 1
	22,375	22,364	CasE	0 0 0 0 1 0

Supplementary Table 3

Strains, plasmids and primers used in this study

Strains	Description	Source
E.coli BL21(DE3)	F⁻ ompT gal dcm lon hsdSB(rB -mB -) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Novagen
E.coli BL21-AI	F⁻ ompT gal dcm lon hsdSB(rB -mB -) araB::T7RNAP-tetA	Invitrogen
E.coli NEB5α	fhuA2 _(argF-lacZ)U169 phoA glnV44 Φ80_ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	New England Biolabs
E.coli DH5α	F⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(rK -mK +), λ–	

Plasmids	Description and order of genes (5'-3')	Restriction sites	Primers	Source
pET-52b(+)	T7 RNA polymerase based expression vector, Amp^R			Novagen
pRSF-1b	T7 RNA polymerase based expression vector, Kan^R			Novagen
pCDF-1b	T7 RNA polymerase based expression vector, Str^R			Novagen
pACYCduet-1	T7 RNA polymerase based expression vector, Cam^R			Novagen
pWUR381	*cas3* in pET-52b with both Strep-tag II (N-term) and His_{10}-tag (C-term)	BamHI/NotI	BG2243 + BG2244	This study
pWUR388	*casA* in pET-52b with both Strep-tag II (N-term) and His_{10}-tag (C-term)			[1]
pWUR396	*E. coli* K12 CRISPR in pACYCDuet-1, see Fig S2			[1]
pWUR397	*cas3* in pRSF-1b, no tags			[1]
pWUR400	*casA-casB-casC-casD-casE* in pCDF-1b, no tags			[1]
pWUR401	*casB-casC-casD-casE* in pCDF-1b, no tags			[1]
pWUR402	*casC-casD-casE* in pCDF-1b, no tags			[1]
pWUR403	*casD-casE* in pCDF-1b, no tags			[1]
pWUR404	*casE* in pCDF-1b, no tags			[1]
pWUR405	*casA-casB-casC-casD* in pRSF-1b, no tags			[1]
pWUR406	*casA-casB-casC* in pRSF-1b, no tags			[1]
pWUR407	*casA-casB* in pRSF-1b, no tags			[1]
pWUR408	*casA* in pRSF-1b, no tags			[1]
pWUR413	*casD* in pRSF-1b, no tags	NcoI/NotI	BG2466 + BG2482	This study
pWUR477	non targeting CRISPR in pACYCDuet-1 (N), see Fig S2			[1]
pWUR478	template CRISPR in pACYCDuet-1 (T_{1-4}), see Fig S2			[1]
pWUR479	coding CRISPR in pACYCDuet-1 (C_{1-4})			[1]
pWUR480	*casB* with Strep-tag II (N-term)-*casC-casD* in pET52b			[1]
pWUR482	*casE* in pET52-1b, with Strep-tag II (N-term)	BamHI/NotI	BG2586 + BG2253	This study
pWUR514	*casB* with Strep-tag II (N-term)-*casC-casD-CasE* in pET52b	Acc65I/NotI	BG2573 + BG2586	This study
pWUR547	*E. coli* R44 CRISPR, 7x spacer nr. 2, in pACYCDuet-1, see Fig S2	EcoNI/Acc65I		[4], Geneart, Germany
pWUR553	*casB* in pET52-1b, with Strep-tag II (N-term)	Acc65I/NotI	BG2573 + BG2484	This study
pWUR554	*casC* in pCDF-1b, no tags	NcoI/NotI	BG2465 + BG2483	This study
pWUR555	*casC* in pET52-1b, with Strep-tag II (N-term)	BamHI/NotI	BG2249 + BG2483	This study
pWUR613	pUC19 containing R44-protospacer on a 350 bp phage P7 amplicon	BamHI/HindIII	BG3297 + BG3298	This study

Experiment	Primer	Sequence (5'-3')	Description
Cloning	BG2243	GCGCGGGATCCTATGGAACCTTTTAAATATATATGCC	*cas3* + BamHI (fw)
	BG2244	GGCCCGCGGCCGCTTTGGGATTTGCAGGGATGACT	*cas3* + NotI (rv)
	BG2249	GCGCGGGATCCTATGTCTAACTTTATCAATATTCATGT	*casC* + BamHI (fw)
	BG2253	GCGCGGGATCCTATGTATCTCAGTAAAGTCATCATTG	*casE* + BamHI (fw)
	BG2316	GCGCGGGTACCAGATGAGATCTTATTTGATCTTGCGG	*casD* + Acc65I (fw)
	BG2465	GCGCGCCATGGCTATGTCTAACTTTATCAATATTCATGT	*casC* + NcoI (fw)
	BG2466	GCGCGCCATGGCTATGAGATCTTATTTGATCTTGCGG	*casD* + NcoI (fw)
	BG2482	GGCCCGCGGCCGCTTACTGAGATACATCCATACCTCC	*casD* + NotI (rv) + stopcodon
	BG2483	GGCCCGCGGCCGCTCACGCCTCGCCATTATTACGA	*casC* +NotI (rv) + stopcodon
	BG2484	GGCCCGCGGCCGCTTACGCATTTTTGTTTGTGGTCAAT	*casB* + NotI (rv) + stopcodon
	BG2573	GCGCGGGTACCAGATGGCTGATGAAATTGATGCAATG	*casB* + Acc65I (fw)
	BG2586	GGCCCGCGGCCGCCTGCAGTCACAGTGGAGCCAAAGATAGCAAG	*casE* + NotI (rv) + stopcodon
	BG3048	GCGCGGAATTCATGAGATCTTATTTGATCTTGCGG	*casD* + EcoRI (fw)
	BG3049	GGCCCCTGCAGTTACTGAGATACATCCATACCTCC	*casD* + PstI + stopcodon (rv)
	BG3297	GGCCCGGATCCCCAATACAGCGATCAACGCCTTC	phage P7 fragment + BamHI (fw)
	BG3298	CGCGCAAGCTTCGATGCTGATGGGGATTAACAG	phage P7 fragment + HindIII (rv)
EMSA	BG3009	AAGCTGATCGGCAAGCTCGAAAGCACG**ACGGTATTGTTCAGATCCTGGCTTGCCAACAG**TG ATTTGCTCAATTTTGTAGATTGAAG	Target with R44 protospacer, and flanking sequences[1]
	BG3010	CTTCAATCTACAAAATTGAGCAAATCA**CTGTTGGCAAGCCAGGATCTGAACAATACCGT**CG TGCTTTCGAGCTTGCCGATCAGCTT	Target with R44 protospacer, and flanking sequences[2]
	BG3011	AAGCTGATCGGCAAGCTCGAAAGCACG**TTGACTTGGATCACCGACCTTGGCGATTACGA**TG ATTTGCTCAATTTTGTAGATTGAAG	Target with scrambled R44 protospacer and flanking sequences[1]
	BG3012	CTTCAATCTACAAAATTGAGCAAATCA**TCGTAATCGCCAAGGTCGGTGATCCAAGTCAA**CG TGCTTTCGAGCTTGCCGATCAGCTT	Target with scrambled R44 protospacer and flanking sequences[2]
	BG3032	**ACGGTATTGTTCAGATCCTGGCTTGCCAACAG**	Target with R44 protospacer, without flanking sequences[1]
	BG3033	**CTGTTGGCAAGCCAGGATCTGAACAATACCGT**	Target with R44 protospacer, without flanking sequences[2]
In vitro transcription (to generate RNA products for EMSA)	BG3030	CCATGGTAATACGACTCACTATAGGGCTTCAATCTACAAAATTGAGCAA	T7 promoter primer to generate a template for in vitro transcription [3]
	BG3031	AAGCTGATCGGCAAGCTCGAAAGC	Primer to generate a template for in vitro transcription[3]
	BG3079	CCATGGTAATACGACTCACTATAGGGAAGCTGATCGGCAAGCTCGAAAG	T7 promoter primer to generate a template for in vitro transcription [3]
	BG3080	CTTCAATCTACAAAATTGAG	Primer to generate a template for in vitro transcription[3]
ESI-MS Native PAGE EM	BG3028	biotin-TEG-CTGTTGGCAAGCCAGGATCTGAACAATACCGT	R44 protospacer (A)

[1] DNA sequence corresponds to crRNA sequence, [2] DNA sequence complementary to crRNA sequence, [3] These primers were used to prepare material for *in vitro* transcription by PCR by adding on a T7 promoter, using either BG3009-BG3010 or BG3011-BG3012 as template. Restriction sites are underlined, Proto-spacer sequences are in bold.

Supplementary Methods

Mass Spectrometry analysis of RNase digests. RNaseT1 and RNaseA digests of mature crRNA was performed using an ESI-MS/MS using a HCT Esquire Quadruple Ion Trap (Bruker Daltonics) coupled to a Dionex. The oligoribonucleotide mixture was separated on a PepMap C-18 RP capillary column (300 µm × 150 mm I.D., Dionex, UK) at 50 °C using a gradient condition starting at 20% buffer D (0.4 M 1,1,1,3,3,3,-Hexafluoro-2-propanol (Sigma-Aldrich) adjusted with triethylamine (TEA) to pH 7.0, 0.1 mM TEAA, and 50% methanol (v/v) (Fisher)) and extending to 35% D in 20 min at a flow rate of 2 µl /min. The mass spectrometer was set select a mass range of 250−1500 m/z and the capillary voltage was kept at -3650 V. Oligoribonucleotides with -2 to -4 charge states were selected for tandem mass spectrometry using collision induced dissociation. The theoretical masses of the crRNA and predicted digests were determined using the Mongo Oligo Mass Calculator.

Electrophoretic Mobility Shift Assays. DNA targets were gel-purified long oligonucleotides (Isogen Life Sciences or Biolegio), listed in Supplementary Table 3. The oligonucleotides were end-labeled using γ^{32}P-ATP (PerkinElmer) and T4 kinase (Fermentas). Double-stranded DNA targets were prepared by annealing complementary oligonucleotides and digesting remaining ssDNA with Exonuclease I (Fermentas). Labeled RNA targets were *in vitro* transcribed using T7 Maxiscript or T7 Mega Shortscript kits (Ambion) with α^{32}P-CTP (PerkinElmer) and removing template by DNaseI (Fermentas) digestion. Double stranded RNA targets were prepared by annealing complementary RNAs and digesting surplus ssRNA with RNaseT1 (Fermentas), followed by phenol extraction.

Plasmid mobility shift assays were performed using plasmid pWUR613 (Supplementary Table 3) containing the R44 protospacer. The fragment containing the protospacer was PCR-amplified from bacteriophage P7 genomic DNA using primers BG3297 and BG3298 (Supplementary Table 3). Plasmid (0.4 µg) and Cascade were mixed in a 1:10 molar ratio and incubated at 37 °C for 30 minutes. Cascade proteins were then removed by proteinase K treatment (Fluka) (0.15 U, 15 min, 37 °C) and PCI extraction. RNA-DNA complexes were then treated with RNaseH (Promega) (2 U, 1 h, 37 °C).

Electron Microscopy data analysis. Approximately 400,000 single particles were selected and extracted from 17,000 electron micrographs. Single particle data sets were analyzed with

GRIP software[7] using multi-reference alignments and no-reference alignments, multivariate statistical analysis, and hierarchical ascendant classification. The final two-dimensional projection maps were calculated from the best resolved classes by summing the best 5–20% of the projections based on the correlation coefficient determined in the alignment step.

Western blotting. The purified proteins transferred to nitrocellulose membrane (Perkin-Elmer), and incubated for 1 h in Phosphate Buffered Saline (PBS) blocking buffer pH 7.5, supplemented with 0.1% Tween 20 and 5% milk powder. The membranes were incubated with anti-Cascade serum raised in rabbits (Eurogentec, Belgium) (1:1000) in blocking buffer for 12 h, followed by three 15 min wash steps in PBS supplemented with 0.1% Tween 20. The membrane was then incubated with HRP-conjugated anti-rabbit antibodies (GE) (1:5000) in wash buffer for 1 h, followed by three 15 min wash steps. Photographic film detection (KODAK) of the signal was performed using ECL-plus substrate (GE).

Small-Angle X-ray Scattering. Data were collected at the Advanced Light Source (Lawrence Berkeley National Laboratory) on beamline 7.3.3. Solution scattering of Cascade was collected at room temperature (~22 °C) using at least two different concentrations (between 1 mg/ml and 18 mg/ml) in a 20 µl sample at 10 keV (λ = 1.24 Å). The sample-to-detector distance was set to 3056.69 mm resulting in scattering vectors (q) ranging from 0.012 $Å^{-1}$ to 0.127 $Å^{-1}$. Scatter plots for the low and high concentrations were merged and background subtracted using PRIMUS[8]. One-dimensional scatter curves were transformed and distance distribution functions P(r) were calculated using GNOM[9]. The pair-distribution function P(r), the frequency of interatomic vector lengths within the scattering particle, was calculated from the entire scattering curve I(q), by an indirect Fourier transform using GNOM. Ten independent models were generated using a simulated annealing method in DAMMIF[10]. *Ab initio* reconstructions were aligned, filtered and averaged based on occupancy using DAMAVER[11]. The SAXS bead models were converted to volumetric format using the pdb2vol convolution kernel in the Situs software package[12]. Volumes were calculated using VOIDOO[13].

References

1. Brouns, S.J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-4 (2008).
2. Barrick, J.E. et al. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243-7 (2009).
3. Ochman, H. & Selander, R.K. Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* **157**, 690-3 (1984).
4. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174-82 (2005).
5. Zaug, A.J., Linger, J. & Cech, T.R. Method for determining RNA 3' ends and application to human telomerase RNA. *Nucleic Acids Res* **24**, 532-3 (1996).
6. Goddard, T.D., Huang, C.C. & Ferrin, T.E. Software extensions to UCSF Chimera for interactive visualization of large molecular assemblies. *Structure* **13**, 473-482 (2005).
7. van Heel, M. et al. Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics* **33**, 307-369 (2000).
8. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**, 1277-1282 (2003).
9. Svergun, D.I. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *Journal of Applied Crystallography* **25**, 495-503 (1992).
10. Franke, D. & Svergun, D.I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **42**, 342-346 (2009).
11. Volkov, V.V. & Svergun, D.I. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **36**, 860-864 (2003).
12. Wriggers, W., Milligan, R.A. & McCammon, J.A. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *Journal of Structural Biology* **125**, 185-195 (1999).
13. Kleywegt, G.J. & Jones, T.A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D Biol Crystallogr* **50**, 178-85 (1994).