University of Groningen

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, Sakham P.; Wiedenheft, Blake; Pul, Uemit; Wurm, Reinhild; Wagner, Rolf

Published in:
Nature Structural & Molecular Biology

DOI:
[10.1038/nsmb.2019](https://doi.org/10.1038/nsmb.2019)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P. L., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J. R., van der Oost, J., ... Pul, Ü. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Structural & Molecular Biology*, 18(5), 529-541. <https://doi.org/10.1038/nsmb.2019>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Structural basis for CRISPR RNA-guided DNA recognition by Cascade

Matthijs M Jore^{1,11}, Magnus Lundgren^{1,10,11}, Esther van Duijn^{2,3,11}, Jelle B Bultema^{4,11}, Edze R Westra¹, Saktham P Waghmare⁵, Blake Wiedenheft^{6–8}, Ümit Pul⁹, Reinhild Wurm⁹, Rolf Wagner⁹, Marieke R Beijer¹, Arjan Barendregt^{2,3}, Kaihong Zhou^{6–8}, Ambrosius P L Snijders^{5,10}, Mark J Dickman⁵, Jennifer A Doudna^{6–8}, Egbert J Boekema⁴, Albert J R Heck^{2,3}, John van der Oost¹ & Stan J J Brouns¹

The CRISPR (clustered regularly interspaced short palindromic repeats) immune system in prokaryotes uses small guide RNAs to neutralize invading viruses and plasmids. In *Escherichia coli*, immunity depends on a ribonucleoprotein complex called Cascade. Here we present the composition and low-resolution structure of Cascade and show how it recognizes double-stranded DNA (dsDNA) targets in a sequence-specific manner. Cascade is a 405-kDa complex comprising five functionally essential CRISPR-associated (Cas) proteins (CasA₁B₂C₆D₁E₁) and a 61-nucleotide CRISPR RNA (crRNA) with 5'-hydroxyl and 2',3'-cyclic phosphate termini. The crRNA guides Cascade to dsDNA target sequences by forming base pairs with the complementary DNA strand while displacing the noncomplementary strand to form an R-loop. Cascade recognizes target DNA without consuming ATP, which suggests that continuous invader DNA surveillance takes place without energy investment. The structure of Cascade shows an unusual seahorse shape that undergoes conformational changes when it binds target DNA.

The constant pressure of invading viruses and conjugative plasmids has shaped the evolution of host defense systems in prokaryotes. The widely distributed CRISPR immune system represents a recently discovered adaptive and inheritable defense strategy^{1–5}. The system consists of repeats that are interspersed with unique sequences called spacers, which are derived from viral and plasmid DNA^{6–8}.

The Cas protein machinery is encoded by gene clusters that are located close to the CRISPR locus⁹. Multiple types of *cas* gene sets have been recognized^{10,11} that correlate with specific families of repeat sequences¹².

The mechanism by which CRISPR and Cas induce immunity has been divided into three stages. In the first stage—CRISPR adaptation—the host encounters an invader and integrates a random fragment of foreign DNA nondirectionally into the CRISPR locus as a new spacer, resulting in resistance to foreign genetic elements carrying this sequence^{13–15}. The metal-dependent DNase Cas1 might be involved in generating small DNA fragments that are used as precursors for CRISPR adaptation¹⁶.

The second stage—CRISPR expression—involves transcription and translation of the *cas* genes and transcription of the CRISPR, yielding a precursor of the CRISPR RNA (pre-crRNA). The pre-crRNA is cleaved in the repeat regions^{17–19} by specific Cas endonucleases: CasE from *E. coli*²⁰, Cas6

from *Pyrococcus furiosus*²¹ and Csy4 from *Pseudomonas aeruginosa*²². CasE-generated mature crRNAs remain bound to a protein complex called Cascade (CRISPR-associated complex for antiviral defense), which consists of five Cas proteins of the Cse-type (CasABCDE; **Fig. 1a**). Cas6-generated crRNAs end up in the Cmr complex (Cmr1–Cmr6) and are further trimmed at the 3' end from around 67 to 39 or 45 nucleotides. The guide RNA-loaded Cmr complex cleaves single-stranded target RNA sequence-specifically²³. In *E. coli*, the third stage—CRISPR interference—requires not only Cascade loaded with anti-invader crRNA, but also the participation of the predicted nuclease-helicase Cas3. Because crRNAs that are complementary to either strand of the phage DNA provide resistance, it has been proposed that Cascade is a crRNA-guided complex that targets DNA rather than mRNA²⁰. *In vivo* experiments in *Staphylococcus epidermidis* involving a conjugative plasmid proved that the Csm-type CRISPR-Cas system targets DNA²⁴. In *Streptococcus thermophilus*, bacteriophage resistance and plasmid loss result from site-specific cleavage of viral and plasmid DNA within the spacer-targeted sequence, and this process requires the *csn1-like* gene¹⁵.

Although viral or plasmid DNA seems to be the target molecule in these three bacterial systems, there is no biochemical evidence

¹Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Wageningen, The Netherlands. ²Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and the Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands. ³The Netherlands Proteomics Center, Utrecht, The Netherlands. ⁴Department of Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, The Netherlands. ⁵ChELSI Institute, Department of Chemical and Process Engineering, University of Sheffield, Sheffield, UK. ⁶Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, California, USA. ⁷Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California. ⁸Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA. ⁹Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany. ¹⁰Present addresses: Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden (M.L.); MRC Clinical Sciences Centre, Imperial College London, London, UK (A.P.L.S.). ¹¹These authors contributed equally to this work. Correspondence should be addressed to J.v.d.O. (john.vanderoot@wur.nl).

Received 6 May 2010; accepted 24 January 2011; published online 3 April 2011; doi:10.1038/nsmb.2019

Figure 1 Core complexes of Cascade retain crRNA. (a) Schematic diagram of the CRISPR-Cas locus in *E. coli* K12 containing *cas3* (*ycgB*), *casA* (*cse1*, *ycgI*), *casB* (*cse2*, *ycgK*), *casC* (*cse4*, *ycgJ*), *casD* (*cas5e*, *ycgI*), *casE* (*cse3*, *ycgH*), *cas1* (*ygb7*) and *cas2* (*ygbF*)^{9,10}. (b) Coomassie blue-stained SDS-polyacrylamide gel showing StrepTactin-purified Cascade, CasBCDE and CasCDE. Protein marker sizes in kDa. Asterisk, Strep-tagged subunits. (c) Ethidium bromide-stained denaturing PAA-gel showing nucleic acids isolated from purified Cascade (sub)complexes. RNA marker sizes in nucleotides. (d) RNase A or DNase I treatment of Cascade-bound nucleic acids. (e) Size exclusion elution profiles of CasCDE, CasBCDE and Cascade before and after DNase I treatment.

to explain how Cas protein complexes recognize their target DNA molecules. Here we show that the molecular basis of the specificity of CRISPR immune system relies on the formation of R-loops: that is, the formation of Watson-Crick base pairs between the crRNA-spacer sequence and the complementary strand of double-stranded target DNA. We also present a structural model of Cascade that provides insight into the architecture of Cas protein complexes.

RESULTS

Core subcomplexes of Cascade

The *E. coli* K12 CRISPR-Cas system (Cse-subtype) consists of a gene cluster that includes *cas3*, the Cascade genes (*casABCDE*), *cas1*, *cas2* and a downstream CRISPR locus²⁰ (Fig. 1a). To investigate the role of the individual Cascade subunits, we first tested whether each subunit was required for antiviral defense. Viral plaque assays with Cas3 and Cascade lacking one type of protein subunit showed that all protein components of Cascade are indispensable for the virus-resistant phenotype of *E. coli* (Supplementary Fig. 1).

We systematically overproduced and purified different combinations of Cascade subunits and checked for the presence of mature crRNA. We found that CasA or CasAB could be omitted without affecting the apparent stoichiometry of the remaining subunits or the mature crRNA (Fig. 1b,c). Cascade, unlike CasBCDE and CasCDE, always copurified with large nucleic acid molecules (>300 nucleotides; Fig. 1c). Removal of the Cas proteins followed by nuclease treatments showed that RNase A only hydrolyzed the crRNA, whereas DNase I removed the long nucleic acids, thereby identifying the copurified nucleic acid as DNA (Fig. 1d). Size exclusion chromatography of the three types of complex revealed that the majority of CasBCDE and CasCDE were present in a single form, whereas Cascade showed a substantial void peak in addition to a discrete peak at ~11 ml (Fig. 1e). DNase I treatment before gel filtration eliminated the void peak without disrupting the discrete Cascade peak, again indicating the presence of Cascade-bound DNA (Fig. 1e).

Architecture of crRNA

The characteristics of the mature crRNA species have been accurately determined by subjecting mature crRNAs isolated from Cascade to denaturing RNA chromatography^{25,26} and electrospray ionization mass spectrometry (ESI-MS).

To simplify the analysis, we obtained a uniform crRNA preparation by coexpressing Cascade with a designed CRISPR containing eight repeats and seven identical spacers (denoted R44 CRISPR (Supplementary Fig. 2)). This setup resulted in a Cascade preparation in which each molecule was loaded with the same crRNA. We used chromatography to show the purity and homogeneity of the crRNA in Cascade (Fig. 2a), and the retention time was consistent with an approximate length of 60 nucleotides. The ESI-MS spectra indicated that the crRNA had a molecular weight of 19,660.80 Da (Fig. 2b), which corresponds well to an expected molecular weight of 19,660.82 Da for a 61-nucleotide crRNA resulting from single CasE endoRNase cleavage events in each repeat.

ESI-MS/MS analysis of crRNA treated with RNase T1 and RNase A resulted in a number of oligoribonucleotide digests that could be assigned to the mature crRNA sequence (Supplementary Fig. 3) and were consistent with a cleavage site 5' of the terminal base of the hairpin²⁰. The molecular weight analysis of the crRNA indicated a 5'-hydroxyl group and a 2',3'-cyclic phosphate terminus. We confirmed the latter by acid treatment of the crRNA, which showed a mass shift of 18 Da, corresponding to the hydrolysis of the cyclic phosphate to a 2' or 3' phosphate (Fig. 2c). Mature crRNA is 61 nucleotides long and contains the 32-nucleotide spacer sequence, flanked by repeat-derived sequences on either end: 8 bases at the 5' terminus (5' handle) and 21 bases forming a hairpin with a tetranucleotide loop at the 3' terminus (3' handle; Fig. 2d).

Figure 2 Architecture of crRNA. (a) Ion-pair reversed-phase HPLC purification of mature R44 crRNA at 75 °C. (b) Multiple-charged ESI-MS spectrum of the purified mature crRNA. (c) Enhanced view of the -18 charged species before (top) and after (bottom) acid treatment indicating hydrolysis of the 2',3'-cyclic phosphate. (d) Diagram of mature crRNA derived from the R44 CRISPR.

Target DNA recognition

The observation that DNA copurified with Cascade prompted us to analyze the DNA-binding behavior of Cascade in detail (Fig. 3). Electrophoretic mobility shift assays (EMSAs) showed that Cascade could bind single-stranded (ss) DNA containing a sequence complementary to the spacer sequence of the crRNA (Fig. 3a,c and Supplementary Fig. 4). Cascade also bound double-stranded target DNA (Fig. 3b,d) without the need for, or enhancement by, additional cofactors such as divalent metal ions or ATP (data not shown). The dissociation constants (K_d) of Cascade for single- and double-stranded target DNA were 8 and 790 nM, respectively. Cascade also bound weakly to nontarget DNA (Fig. 3c,d). Competitor DNA had little effect on preformed Cascade–target DNA complexes, indicating that the interaction between Cascade and complementary DNA substrates is stable. At high competitor concentrations a proportion of Cascade lacking CasA bound target DNA, as was evident from the faster migration rate of the CasBCDE–DNA complex (Fig. 3a,c and Supplementary Fig. 4).

Cascade subcomplexes that lacked CasA (CasBCDE and CasCDE) showed only sequence-specific binding to ssDNA and dsDNA targets, and were not affected by the addition of competitor DNA (Fig. 3e–h). The addition of purified CasA to CasBCDE preparations restored Cascade-like nonspecific DNA binding (Supplementary Fig. 5), but neither CasA alone (Supplementary Fig. 5) nor a combination of CasA and CasCDE (not shown) showed nonspecific DNA binding. Competition assays between Cascade and CasBCDE showed that a ratio of 1:25 was required to distribute the ssDNA target equally between the types of complexes (Fig. 4a). This difference was more pronounced when we used dsDNA target probes (data not shown). Furthermore, much less dsDNA target was shifted by CasBCDE than by Cascade under equivalent conditions (Fig. 3d,f), indicating again that CasA enhanced target DNA localization.

It seems that Cascade recognizes 32-nucleotide target sequences (referred to as protospacers) in dsDNA molecules by

formation of base pairs between the crRNA spacer sequence and the complementary DNA strand. Analysis of long dsDNA targets (protospacer with 27-base-pair (bp) flanks) showed that both strands shifted (Fig. 4b), probably owing to base-paired flanking regions. Only the complementary strand shifted when we used short dsDNA targets (corresponding to the protospacer), which suggests that the formation of base pairs between crRNA and the complementary strand displaced the noncomplementary strand (Fig. 4c).

We demonstrated that the noncomplementary strand had been displaced by performing enzymatic and chemical footprint analyses specific for ssDNA²⁷ using Cascade loaded with targeting (R44) and nontargeting (K12) crRNA. Endonuclease P1 footprints showed that an 18-base region (G27–C44) of the nontargeted strand corresponding to about half of the protospacer was susceptible to cleavage after binding by Cascade (Fig. 4d). In line with this observation, six thymines and one adenine in the same region of the nontargeted strand were also sensitive to permanganate modification and subsequent piperidine cleavage (Fig. 4e). These results indicate that only the first half of the nontargeted strand of the protospacer was exposed in the Cascade–DNA complex and the second half of the protospacer was shielded. The targeted strand was sensitive to P1 in a region that overlapped the protospacer adjacent motif (PAM)²⁸ (Fig. 4d), which may be indicative of a distortion in

Figure 3 Target recognition by Cascade. (a,b) Effect of the type of crRNA bound. Cascade was loaded with either targeting crRNA (R44 CRISPR, Supplementary Fig. 2) or nontargeting crRNA (K12 CRISPR). The binding of these two types of Cascade complex to one type of probe is shown. DNA probes are 86-nucleotide ssDNA or dsDNA sequences containing the R44 protospacer (32 nucleotides) flanked by 27 nucleotides on either end. (c–h) Effect of uniform crRNA-loaded (sub)complexes (R44 CRISPR) on the binding of single- or double-stranded target and nontarget DNA. Nontarget DNA probes contain a scrambled R44 protospacer sequence. (i,j) Effect of uniform crRNA-loaded Cascade (R44 CRISPR) on the binding of target and nontarget ssRNA and dsRNA. (a–h) Labeled probe concentration 1 nM; DNA competitor concentration 2,500, 500, 50, 5 and 0.5 ng μl^{-1} (highest concentration not used for CasCDE); protein concentration 200–300 nM except in i,j where the Cascade concentrations were 200, 100, 50, 25 and 12.5 nM.

DNA conformation resulting from Cascade-induced strand opening further downstream. Permanganate assays also indicated that a single base of the targeted strand of the protospacer (T33) was sensitive to modification (Fig. 4e), suggesting that this base is unpaired in the hybrid crRNA–DNA duplex. Exonuclease III footprints showed that the sequences that flank the protospacer are double-stranded and that

Cascade protects a region of around 9 bases including the PAM on one flank and around 14 bases on the non-PAM flank (Fig. 4f).

To show that crRNA and DNA form base pairs, we used a 3-kb plasmid containing the target sequence. Specific binding of R44–Cascade to the supercoiled plasmid resulted in a mobility shift of the plasmid (Fig. 4g). The retardation of the plasmid was partially relieved by proteolytic

Figure 4 R-loop formation by Cascade. **(a)** Competition assay between R44–crRNA–loaded Cascade and CasBCDE for R44 ssDNA target. Total protein concentration was 500 nM in each reaction, and the Cascade:CasBCDE ratio was 1:0, 100:1, 10:1, 1:1, 1:10, 1:100 and 0:1. DNA competitor concentration was $1 \mu\text{g } \mu\text{l}^{-1}$. **(b,c)** Effect of labeling the complementary or noncomplementary strand of a target dsDNA with 27-bp flanks **(b)** or without protospacer-flanking sequences **(c)**. Cascade concentrations were 1,500, 300, 60 and 12.5 nM. **(d,e)** Mapping of ssDNA regions in the Cascade–target DNA complex using nuclease P1 and KMnO_4 . Sensitive regions are indicated by dashed lines and the protospacer by a solid line according to the G+A sequencing lanes of each strand. Cascade loaded with K12-derived crRNA was used as a control. **(f)** Exonuclease III mapping of accessible dsDNA regions upstream and downstream of the Cascade–DNA complex. The borders of the Cascade-protected regions are indicated by arrows. **(g)** Detection of the R-loop in a target plasmid. Agarose gel indicating the mobility of the different plasmid forms (SC, supercoiled; L, linear; OC, open circular) and the mobility shifts caused by R44–Cascade and R44–crRNA binding. **(h)** Schematic diagram of the R-loop formed in crRNA-guided dsDNA recognition by Cascade. Regions sensitive to nuclease P1 and KMnO_4 are indicated by hash and asterisk signs, respectively.

Figure 5 Subunit composition of Cascade. (a) Native nano-ESI mass spectrum of Cascade. Two charge state distributions are present at high m/z values, corresponding to complexes of 405 kDa (purple) and 349 kDa (pink). The charge state distribution in red corresponds to the CasB dimer. (b) Cascade (sub)complexes analyzed by native mass spectrometry. The subcomplexes were formed in solution after adding 5% 2-propanol to the buffer solution containing Cascade.

removal of the Cascade protein subunits, leaving only the crRNA bound to the target DNA (Fig. 4g). These nucleic acid complexes produce a slight shift in gel electrophoretic mobility of the plasmid²⁹. Subsequent RNase H treatment restored the original gel mobility of the plasmid, indicating that Watson-Crick base-pairing occurred between the crRNA and one strand of the plasmid DNA (Fig. 4g). The specific binding of Cascade thus led to the formation of an R-loop (Fig. 4h): a structure in which the spacer part of the crRNA was base paired with one DNA strand of a duplex, leaving the displaced strand single stranded³⁰.

We did not observe Cascade-mediated cleavage of target DNA under the conditions tested. We could detect binding to complementary ssRNA (Fig. 3i), but this did not result in cleavage of target RNA. In contrast to dsDNA, there was no binding to dsRNA (Fig. 3j). In addition, there was very little nonspecific interaction between Cascade and ss- or dsRNA (Fig. 3i,j).

Subunit stoichiometry

To understand the structural basis of the interaction between Cascade and target DNA, we determined the composition of the Cascade protein assembly using mass spectrometric analyses. Denaturing and tandem MS analyses resulted in accurate mass measurements for each component of Cascade (Supplementary Table 1). The measured masses of CasA, CasB and CasD were consistent with the expected values, and the masses of CasC and CasE agreed with the primary amino acid sequence lacking the N-terminal methionine. A complex composed of one copy of each Cascade component (CasA₁B₁C₁D₁E₁-crRNA₁) would have a mass of 184 kDa. However, analysis of the intact assembly by native MS³¹ showed two major charge state distributions, corresponding to masses of $405,365 \pm 135$ Da and $349,399 \pm 84$ Da (Fig. 5a and Supplementary Table 1). A third, low-intensity charge state distribution around 2,200 m/z was observed with a mass of $42,524 \pm 8$ Da, which is close to the theoretical molecular weight of a CasB dimer (42,521 Da). Proteolytic removal of the affinity tag on CasB unambiguously confirmed the presence of two CasB copies (Supplementary Fig. 6a).

In agreement with the mass difference of 55,966 Da, the two major complexes of 405 and 349 kDa probably resemble the intact Cascade and a Cascade subcomplex lacking CasA. Tandem MS experiments on intact Cascade ions revealed that CasA was the first subunit to be expelled from the complex under collisional activation conditions, again indicative of the fact that at least one CasA copy is located at the periphery of Cascade with a rather low affinity (Supplementary Fig. 6b). In addition to the elimination of CasA, we also observed the loss of CasC. Selection and activation of the 349-kDa Cascade subcomplex showed the loss of one CasD subunit and up to two CasC subunits. Cascade could not be further disrupted by tandem MS, hampering the full assignment of its stoichiometry. Therefore, we devised a number of alternative strategies.

As Cascade was loaded with only R44 crRNA, we could determine the number of bound crRNA molecules by adding a complementary ssDNA probe. The total molecular weight of Cascade increased by the mass of a single ssDNA probe, indicating the presence of one accessible crRNA (Supplementary Fig. 6c).

For further characterization we used a combined approach of in solution and in gas-phase dissociation of Cascade^{32,33}. When we added a low percentage of 2-propanol, Cascade was partially disrupted, resulting in a variety of subcomplexes (Fig. 5b and Supplementary Table 2). In addition to intact Cascade (405 kDa), we also detected Cascade lacking CasA (349 kDa), and seven additional subcomplexes, the largest of which was a 307-kDa species. The difference between the 349 kDa and 307 kDa subcomplexes is 42,442 Da and probably reflects the loss of the CasB dimer (Fig. 5a).

The apparent consecutive loss of five masses of around 40 kDa each (from 307 kDa down to 107 kDa) was directly evident from this catalog of subcomplexes (Fig. 5b). This implies that at least five CasC subunits were present in intact Cascade. We confirmed this hypothesis by subjecting each of these subcomplexes to tandem MS (Supplementary Table 2). In addition to CasD, CasE and one crRNA molecule, this analysis revealed the presence of a sixth CasC subunit in the 107-kDa subcomplex (Fig. 5b). Combining all MS data resulted in a Cascade stoichiometry of CasA₁B₂C₆D₁E₁-crRNA₁. The theoretical mass of this complex is consistent with the experimental mass ($405,095$ Da versus $405,365 \pm 135$ Da), band intensities on protein gels (Fig. 1b) and elution profiles on a calibrated size exclusion column (Fig. 1e). Similar analyses of the purified subcomplexes revealed masses and compositions consistent with Cascade: 349 kDa for CasB₂C₆D₁E₁-crRNA₁ and 324 kDa for CasC₆D₁E₁-crRNA₁ (data not shown).

Structure of Cascade

We gained insight into the structural organization of Cascade by single-particle EM (Fig. 6). The Cascade complex is an elongated particle with no discernible symmetry and with no apparent structural resemblance to known protein complexes. The particle has approximate dimensions of 10×20 nm and resembles a seahorse with a curled-up tail (Fig. 6a–c,s). The particle has a striking indentation on one side that gives rise to the head and neck features of the seahorse. Although we obtained only three very similar types of projection map for Cascade (Fig. 6a–c and Supplementary Figs. 7,8), we observed six more diverse

Figure 6 EM structure of Cascade. (a–c,s) Three Cascade projections showing an elongated, seahorse-shaped particle with dimensions 20 × 10 nm. (d–i) Six Cascade projections bound to target ssDNA. The arrow indicates an anticipated rotation along the vertical axis. The six regularly arranged CasC subunits are indicated by asterisks. (j–l) Difference map (l) of Cascade (j) and Cascade with target ssDNA bound (k) indicating morphological changes upon DNA binding in mainly the head and back areas. (m–o) Difference map (o) of Cascade (m) and CasBCDE (n) with target ssDNAs bound, showing the location of the CasA subunit. (p–r) Difference map (r) of CasBCDE (p) and CasCDE (q), showing the location of the CasB subunits. (s) Enlargement of a defining the seahorse-like morphological features of Cascade. A typical indentation that contributes to the head feature of the seahorse is indicated with a hash sign. 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 particle orientation in a. Raw electron micrographs and an overview of the particle analysis method are given in **Supplementary Figures 7 and 8**, respectively. Scale bar, 10 nm. (t) Structural model of Cascade.

groups of projections for Cascade with target ssDNA bound (**Fig. 6d–i**). The apparent loss of adsorption orientation bias upon binding of target DNA could be due to altered surface properties of the complex resulting from DNA binding or might reflect some conformational heterogeneity. However, some projections showed clear similarity to Cascade projections in the DNA-free form (for example, **Fig. 6a,f**), which allowed us to compare the DNA-bound and DNA-free states of Cascade. Difference map analysis revealed relatively small changes throughout the Cascade particle and more pronounced changes in the head and back of the seahorse-shaped morphology of Cascade comprising the CasB and CasC subunits (see below; **Fig. 6j–l**). The observed change suggests that Cascade undergoes conformational changes when it binds target DNA.

The availability of the stable subcomplexes Cascade, CasBCDE and CasCDE and knowledge of their subunit compositions allowed us to investigate subunit localization in the Cascade structure. Some of the projections of Cascade with target ssDNA showed a regularly shaped and evenly spaced feature with sharp edges that spanned the torso of the complex (**Fig. 6i**). This repeated feature traversed the spine of this structure and was consistent with the six copies of CasC. We determined the position of CasA from difference maps between target DNA-bound forms of Cascade and one of three classes of the CasBCDE particle (**Fig. 6m–o** and **Supplementary Fig. 8b,c**). The analysis revealed that CasA was located in the curled-up tail of the seahorse (**Fig. 6o**). We identified the position of CasB using difference maps between the CasBCDE and CasCDE complexes (**Fig. 6p–r**). Both of these subcomplexes yielded only one particle class average with strong resemblance to each other. Apart from the six better-resolved CasC subunits in these projections (**Fig. 6p,q**), the difference map showed one region compatible with two CasB copies (**Fig. 6r**), consistent with the CasB homodimer observed with MS (**Fig. 5a**). The CasB dimer contributes the nose to the seahorse-shaped morphology of Cascade (**Fig. 6r,s**).

In addition to the position of some of the subunits in the Cascade structure, we also investigated the topological constraints of Cascade. Protein pulldown experiments between CasC and one of the remaining Cascade components showed that CasB, CasD and CasE formed stable two-component complexes with CasC, whereas CasA did not (**Supplementary Fig. 9**). The association of this subunit with the CasC backbone was mediated by CasE (**Supplementary Figs. 7,9**).

The combined observations result in a structural model in which the positions of CasA, CasB and CasC can be assigned with confidence,

contributing the curled-up tail, nose and torso of the seahorse, respectively (**Fig. 6t**). The model predicts that CasD, CasE and at least part of the crRNA are located in the main body of Cascade at the tail end of the CasC spine, close to CasA. Alternatively, the crRNA could also be bound along the spine of the CasC backbone and in this way define the rather unusual number and arrangement of CasC subunits by the length of the crRNA.

Ab initio small-angle X-ray scattering (SAXS)-based reconstruction of Cascade and Cascade bound to target DNA resulted in models (**Supplementary Fig. 10**) with a shape and size similar to those of the particle observed by EM (**Fig. 6**). The SAXS model of Cascade reveals an asymmetric seahorse-shaped particle with a relatively narrow head compared to its main body. When it bound target DNA, the volume of Cascade increased from 1,024 to 1,235 nm³ and the morphology of the particle changed into a shape with fewer prominent features. The changes include the loss of the indentation on one side which gives rise to the nose, head and neck features of the seahorse (**Supplementary Fig. 10c,d**). These structural changes in Cascade involve regions of the complex assigned to the CasB and CasC subunits. Although both EM and SAXS showed that Cascade undergoes a conformational change when it binds target DNA, the degree of the conformational change appeared to be more modest for EM. It is worth noting that the changes

in the shape of Cascade are not a consequence of DNA-induced changes to the subunit stoichiometry of the complex (**Supplementary Fig. 6c**), but may be indicative of a ligand-induced conformational change.

DISCUSSION

Composition of Cascade

Cascade is a ribonucleoprotein complex that has a key role in CRISPR-based defense against mobile genetic elements. The backbone of the complex consists of six CasC (Cse4) subunits, which belong to COG1857 and are present in several Cas subtypes¹¹. The CasC backbone forms a minimal protein core in conjunction with single copies of the widespread CasD subunit (Cas5e), the pre-crRNA endoRNase CasE (Cse3) and the mature crRNA. This minimal core is expanded by two proteins that are found only in the Cse subtype: a dimer of CasB (Cse2), a strictly α -helical protein with a conserved basic surface patch³⁴, and one copy of CasA (Cse1), the largest and most loosely attached subunit of the five.

The crRNA is remarkably stable when bound by Cascade or the CasBCDE and CasCDE subcomplexes, indicating that it is tightly associated with the CasCDE core of the protein complex and well shielded from cellular RNase activities. However, it is sufficiently exposed to allow base pairing with complementary nucleic acids. The crRNA is characterized by the presence of a 5'-hydroxyl group, as was shown for Cas6-generated crRNA²¹. Eukaryotic small interfering RNA (siRNA) and microRNA (miRNA), by contrast, need to be 5'-phosphorylated in order to bind to Argonaute and serve as a guide for the RNA-induced silencing complex (RISC)³⁵. *E. coli* crRNAs are unmodified, unlike, for example, plant siRNAs and miRNAs or vertebrate piwi-interacting RNAs, which are methylated at the 2'-hydroxyl group of the 3'-terminal ribonucleotide to prevent uridylation and associated destabilization^{36,37}. The cyclic 2',3'-phosphate of crRNA results from a metal ion-independent endonuclease activity on the 3'-side of the phosphodiester bond of pre-crRNA²¹. These initial pre-crRNA endonuclease cleavage products are the mature form of the crRNA in *E. coli*, in contrast to crRNAs from *P. furiosus*, *S. solfataricus* and *S. epidermidis*, which are trimmed at the 3' end^{19,23,38}. However, the 5' end of crRNAs has a constant length of eight nucleotides and is derived from the repeat^{20,21,24} (**Fig. 2d**). Differential complementarity of this 5' handle of the crRNA with a protospacer flanking sequence allows discrimination between self DNA (the CRISPR) and non-self DNA (the target) in the *S. epidermidis* Csm-type Cas system³⁸. In other CRISPR-Cas systems, the regions that flank the protospacer (opposite to the 5' or 3' handle of the crRNA) often harbor short sequence motifs that are known as protospacer adjacent motifs (PAMs)^{8,14,28,39}. The relevance of these short nucleotide sequences was originally shown in *S. thermophilus* by sequencing phages that had overcome host immunity by mutating a single nucleotide of the motif³⁹.

Target DNA recognition by R-loop formation

There is considerable evidence that crRNAs directly target invader DNA in the *E. coli* (Cse), *S. epidermidis* (Csm) and *S. thermophilus* (Csn) model systems^{15,20,24}. Binding studies showed that Cascade enables Watson-Crick base-pairing of the crRNA with the complementary strand of double-stranded protospacer DNA without supplemented cofactors such as ATP. This characteristic of Cascade is physiologically important, as most cellular and invader DNA is double stranded. The ATP-independence of this scanning process is in line with the absence of homology between the Cascade subunits and classical ATP-consuming helicases⁴⁰. It allows cells to continuously patrol the cytoplasm for nucleic acids that match the crRNA, without making major energy investments. In addition to sequence-specific DNA recognition, Cascade also interacts nonspecifically with DNA.

This nonspecific mode of interaction is mediated by CasA and results in enhanced sequence-specific DNA localization by Cascade. The sequence-specific interaction gives rise to the formation of an R-loop³⁰—a structure that is known to prime replication of bacterial plasmids⁴¹, bacterial chromosomes⁴² and mitochondrial DNA⁴³. Moreover, R-loop formation has been associated with recombination events at immunoglobulin class switch sequences⁴⁴. A sliding 8-bp R-loop occurs in elongating RNA polymerase complexes during transcription⁴⁵. We propose that the formation of such an R-loop by Cascade marks the DNA for subsequent interference processes catalyzed by other Cas proteins, such as Cas3.

Further analysis showed that although Cascade had little nonspecific affinity for RNA, it could also bind complementary ssRNA. Base pairing of the crRNA with mRNA occurs when the coding strand of an invading nucleic acid is targeted. This could titrate out unpaired Cascade-bound crRNAs, reducing the effectiveness of the immune response when the coding strand is targeted²⁰. Cascade does not recognize a protospacer within a dsRNA molecule, indicating that the ability to locate protospacers in a nucleic acid duplex is restricted to dsDNA. A distinct Cas protein complex of the Cmr-subtype from *P. furiosus* cleaves ssRNA complementary to the bound guide RNA²³, akin to RISC in eukaryotes. Hydrolysis of the target RNA takes place 14 bases from the 3' end of the guide RNA, suggesting that target RNA is cleaved by a ruler mechanism. In *S. thermophilus*, dsDNA breaks have been identified at specific sites within the protospacer region of CRISPR-targeted bacteriophage and plasmid DNA¹⁵. Inactivation of the *csn1-like* gene abolished DNA cleavage, suggesting that this protein has a key role in target DNA cleavage in the Csn subtype.

Structural basis of target DNA recognition

Cascade has an unusual seahorse-shaped architecture. The CasC subunits are arranged in an arch, representing the torso of the Cascade particle. CasC assembles into this stable hexameric arrangement only in the presence of CasD, CasE and crRNA. Although we could not determine the location of these three components in the structure by direct methods, mass spectrometry showed the existence of a 107-kDa subcomplex containing single copies of CasC, CasD, CasE and crRNA (**Fig. 5b**). This subassembly of Cascade implies that these four components are close to each other, or that they remain associated by stable binding to the crRNA. Further disassembly of Cascade revealed a CasE-crRNA particle, indicating that CasE holds on tightly to the crRNA after pre-crRNA cleavage. Despite the fact that the location of the crRNA remains unclear, we expect that the 61-nucleotide molecule is bound by a substantial part of Cascade. The length of the 32-nucleotide spacer sequence base paired to its target ssDNA should be about 8 nm. With the 5' and 3' handles, the crRNA could be up to 10 nm long. In an unpaired state, the crRNA probably experiences some conformational flexibility, whereas in a DNA-base-paired state, the conformation of the spacer sequence is more constrained by the formation of the RNA-DNA hybrid helix, which in contrast to the B-type dsDNA helix is close to A-type⁴⁶. This constraint might induce the observed conformational change of the particle and may facilitate recruitment of Cas3. It is tempting to speculate that the HD (histidine, aspartate) domain of Cas3, which carries out DNase activity on ssDNA and dsDNA⁴⁷, cleaves the invading DNA interacting with Cascade.

In conclusion, Cascade is a widespread, unusual ribonucleoprotein complex that is capable of sequence-specific recognition of double-stranded invader DNA archived in the CRISPR blacklist. The direct attack on invader DNA rather than its derived RNA transcripts might allow prokaryotes to neutralize invasive selfish DNA elements effectively.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/nsmb/>.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank M.H. Lai for experimental contributions, and E. Schaible, P. Zwart and M. Bokhove for technical support and for assistance with post processing of SAXS data. This work was financially supported by The Netherlands Organisation for Scientific Research (NWO) Vici grant to J.v.d.O. (865.05.001), Veni grants to S.J.J.B. (863.08.014) and E.v.D. (700.58.402), NWO TOP grant to E.J.B., and UK Engineering and Physical Sciences Research Council and Biotechnology and Biological Sciences Research Council grants to M.J.D. M.L. was financially supported by the Wenner-Gren Foundation, E.R.W. by Spinoza resources awarded to W.M. de Vos, A.P.L.S. by the Research Councils UK, B.W. by the Life Sciences Research Foundation of HHMI and Ü.P. by the Deutsche Forschungsgemeinschaft PU 435/1-1. We thank The Netherlands Proteomics Center for financial support.

AUTHOR CONTRIBUTIONS

M.M.J., M.L., E.R.W., M.R.B., J.J.B. and J.v.d.O. purified Cascade and performed binding assays and plaque assays. E.v.D., A.B. and A.J.R.H. conducted protein MS experiments. J.B.B. and E.J.B. performed EM. S.P.W., A.P.L.S. and M.J.D. conducted RNA MS experiments. B.W., K.Z. and J.A.D. performed SAXS. Ü.P., R. Wurm and R. Wagner did the footprint analyses. All the authors analyzed the data, and S.J.J.B., M.M.J. and J.v.d.O. wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/nsmb/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M. & Brouns, S.J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* **34**, 401–407 (2009).
- Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).
- Karginov, F.V. & Hannon, G.J. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell* **37**, 7–19 (2010).
- Marraffini, L.A. & Sontheimer, E.J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* **11**, 181–190 (2010).
- Deveau, H., Garneau, J.E. & Moineau, S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* **64**, 475–493 (2010).
- 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**, 174–182 (2005).
- Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).
- Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S.D. Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
- Jansen, R., Embden, J.D., Gastra, W. & Schouls, L.M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
- Haft, D.H., Selengut, J., Mongodin, E.F. & Nelson, K.E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, e60 (2005).
- Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I. & Koonin, E.V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**, 7 (2006).
- Kunin, V., Sorek, R. & Hugenholtz, P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8**, R61 (2007).
- Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
- Horvath, P. *et al.* Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1401–1412 (2008).
- Garneau, J.E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
- Wiedenheft, B. *et al.* Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **17**, 904–912 (2009).
- Tang, T.H. *et al.* Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. USA* **99**, 7536–7541 (2002).
- Hale, C., Kleppe, K., Terns, R.M. & Terns, M.P. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* **14**, 2572–2579 (2008).
- Lillestøl, R.K. *et al.* CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol. Microbiol.* **72**, 259–272 (2009).
- Brouns, S.J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
- Carte, J., Wang, R., Li, H., Terns, R.M. & Terns, M.P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496 (2008).
- Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J.A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355–1358 (2010).
- Hale, C.R. *et al.* RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945–956 (2009).
- Marraffini, L.A. & Sontheimer, E.J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).
- Dickman, M.J. & Hornby, D.P. Enrichment and analysis of RNA centered on ion pair reverse phase methodology. *RNA* **12**, 691–696 (2006).
- Waghmare, S.P., Pousinis, P., Hornby, D.P. & Dickman, M.J. Studying the mechanism of RNA separations using RNA chromatography and its application in the analysis of ribosomal RNA and RNA:RNA interactions. *J. Chromatogr. A* **1216**, 1377–1382 (2009).
- Raghavan, S.C., Tsai, A., Hsieh, C.L. & Lieber, M.R. Analysis of non-B DNA structure at chromosomal sites in the mammalian genome. *Methods Enzymol.* **409**, 301–316 (2006).
- Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
- Wilson-Sall, T. & Hsieh, T.S. Preferential cleavage of plasmid-based R-loops and D-loops by *Drosophila* topoisomerase IIbeta. *Proc. Natl. Acad. Sci. USA* **99**, 7974–7979 (2002).
- Thomas, M., White, R.L. & Davis, R.W. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc. Natl. Acad. Sci. USA* **73**, 2294–2298 (1976).
- Heck, A.J.R. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **5**, 927–933 (2008).
- Lorenzen, K., Vannini, A., Cramer, P. & Heck, A.J. Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* **15**, 1237–1245 (2007).
- Zhou, M. *et al.* Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *Proc. Natl. Acad. Sci. USA* **105**, 18139–18144 (2008).
- Agari, Y., Yokoyama, S., Kuramitsu, S. & Shinkai, A. X-ray crystal structure of a CRISPR-associated protein, Cse2, from *Thermus thermophilus* HB8. *Proteins* **73**, 1063–1067 (2008).
- Ma, J.B. *et al.* Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434**, 666–670 (2005).
- Li, J., Yang, Z., Yu, B., Liu, J. & Chen, X. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr. Biol.* **15**, 1501–1507 (2005).
- Houwings, S. *et al.* A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**, 69–82 (2007).
- Marraffini, L.A. & Sontheimer, E.J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571 (2010).
- Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
- Singleton, M.R., Dillingham, M.S. & Wigley, D.B. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50 (2007).
- Itoh, T. & Tomizawa, J. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proc. Natl. Acad. Sci. USA* **77**, 2450–2454 (1980).
- Baker, T.A. & Kornberg, A. Transcriptional activation of initiation of replication from the *E. coli* chromosomal origin: an RNA-DNA hybrid near oriC. *Cell* **55**, 113–123 (1988).
- Lee, D.Y. & Clayton, D.A. Initiation of mitochondrial DNA replication by transcription and R-loop processing. *J. Biol. Chem.* **273**, 30614–30621 (1998).
- Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E. & Lieber, M.R. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* **4**, 442–451 (2003).
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
- Noy, A., Perez, A., Marquez, M., Luque, F.J. & Orozco, M. Structure, recognition properties, and flexibility of the DNA:RNA hybrid. *J. Am. Chem. Soc.* **127**, 4910–4920 (2005).
- Han, D. & Krauss, G. Characterization of the endonuclease SS02001 from *Sulfolobus solfataricus* P2. *FEBS Lett.* **583**, 771–776 (2009).

ONLINE METHODS

Protein production and purification. Cascade complexes were produced and purified as described²⁰ using the expression plasmids listed in **Supplementary Table 3**. Size exclusion chromatography (Superdex 200 HR 10/30 (GE)) was performed using 20 mM Tris-HCl (pH 8.0), 0.1 M NaCl, 1 mM dithiothreitol. Copurified nucleic acids were isolated by extraction using an equal volume of phenol:chloroform:isoamylalcohol (PCI) (25:24:1, v/v/v), pH 8.0 (Fluka), and incubated with DNase I (Invitrogen) supplemented with 2.5 mM MgCl₂ or RNase A (Fermentas) for 10 min at 37 °C.

HPLC purification of crRNA. The crRNA was analyzed by ion-pair reverse-phase HPLC on an Agilent 1100 HPLC with UV_{260nm} detector (Agilent) equipped with a DNasep column 50 mm × 4.6 mm I.D. (Transgenomic) and the following chromatographic buffer conditions: 0.1 M triethylammonium acetate (TEAA) (pH 7.0) (buffer A; Fluka); and buffer A with 25% acetonitrile (v/v) (buffer B; Fisher). The crRNA was obtained by injecting purified intact Cascade at 75 °C using a linear gradient starting at 15% buffer B and extending to 60% buffer B in 12.5 min, followed by a linear extension to 100% B over 2 min at a flow rate of 1.0 ml min⁻¹. Hydrolysis of the cyclic phosphate terminus was performed by incubating the HPLC-purified crRNA in a final concentration of 0.1 M HCl at 4 °C for 1 h.

ESI-MS analysis of crRNA. Electrospray ionization MS was performed in negative mode using an UHR-TOF mass spectrometer (maXis) or an HCT Ultra PTM Discovery instrument (both Bruker Daltonics), coupled to an online capillary liquid chromatography system (Ultimate 3000, Dionex). RNA separations were performed using a monolithic (PS-DVB) capillary column (200 μm × 50 mm I.D., Dionex). The chromatography was performed using the following buffer conditions: 0.4 M 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP, Sigma-Aldrich) adjusted with triethylamine (TEA) to pH 7.0 and 0.1 mM TEAA (buffer C); and buffer C with 50% methanol (v/v) (buffer D; Fisher). RNA analysis was performed at 50 °C with 20% buffer D, extending to 40% buffer D in 5 min followed by a linear extension to 60% buffer D over 8 min at a flow rate of 2 μl min⁻¹.

Protein and Native Mass spectrometry. Cascade was analyzed in 0.15 M ammonium acetate (pH 8.0) at a protein concentration of 5 μM obtained by five sequential concentration and dilution steps at 4 °C using a centrifugal filter with a cut-off of 10 kDa (Millipore). Proteins were sprayed from borosilicate glass capillaries and analyzed on a LCT electrospray time-of-flight or modified quadrupole time-of-flight instruments (both from Waters) adjusted for optimal performance in high mass detection^{48,49}. Exact mass measurements of the individual Cas proteins were acquired under denaturing conditions (50% acetonitrile, 50% MQ, 0.1% formic acid). Subcomplexes in solution were generated by the addition of 2-propanol to the spray solution to a final concentration of 5% (v/v). Instrument settings were as follows: needle voltage ~1.2 kV, cone voltage ~175 V, source pressure 9 mbar. Xenon was used as the collision gas for tandem mass spectrometric analysis at a pressure of 1.5 10⁻² mbar. The collision voltage was 10–200 V.

Electrophoretic mobility shift assay. EMSAs were performed by incubating Cascade, CasBCDE or CasCDE with 1 nM ³²P-labeled DNA or RNA in 50 mM Tris-Cl, pH 7.5, 100 mM NaCl (see **Supplementary Methods**). Salmon sperm DNA (Invitrogen) was used as competitor. The EMSA reactions were incubated at 37 °C for 20–30 min before electrophoresis on 5% polyacrylamide gels. Target

DNA binding and cleavage activity of Cascade was tested in the presence of 1–10 mM calcium, magnesium or manganese ions.

Chemical and enzymatic probing analyses. Chemical and enzymatic probing assays were performed using dsDNA probes prepared by annealing of the complementary oligonucleotides BG3009 and BG3010 (**Supplementary Table 3**). We annealed 2.5 pmol of 5'-end-labeled oligonucleotide to 6 pmol of its unlabeled complementary oligonucleotide to ensure saturation of the labeled oligonucleotide.

Chemical probing with KMnO₄ was performed by incubating 0.01 nM DNA with 3 μM R44-Cascade or K12-Cascade in 50 mM Tris-Cl, pH 7.5, 100 mM NaCl in a total volume of 30 μl, followed by incubation at 30 °C for 30 min. The samples were diluted by adding equal amounts of Aquadest (Millipore) before addition of 4 μl of 160 mM KMnO₄. Permanganate modification was allowed for 2 min at 30 °C. The reaction was stopped by the addition of 4.8 μl β-mercaptoethanol and 5.3 μl 0.5 M EDTA, followed by phenol-chloroform extraction and precipitation with ethanol. The pellets were dissolved in 50 μl of 10% piperidine and incubated at 90 °C for 30 min. After lyophilization, the pellets were washed twice with 30 μl Aquadest and lyophilized. Finally, the pellets were dissolved in 50 μl of 0.3 M sodium acetate and precipitated with ethanol. The cleavage products were separated on a denaturing 10% polyacrylamide gel together with G+A-sequencing of oligonucleotides⁵⁰ and visualized by autoradiography.

For probing with P1 nuclease (Sigma) 30 μl of the Cascade-DNA complexes was diluted by addition of 30 μl 60 mM sodium acetate, pH 5.3, followed by incubation with 0.09 units of P1 nuclease at 37 °C for 30 min, PCI extraction and ethanol precipitation. For exonuclease III footprinting of Cascade-bound DNA complexes, the binding reactions were diluted with equal amounts of Aquadest and incubated with 60 U of Exo. III (Promega) at 37 °C for 20 min. The cleavage reactions were stopped by addition of EDTA to a final concentration of 20 mM, followed by PCI extraction and ethanol precipitation.

Electron microscopy. Protein samples were negatively stained with 2% uranyl acetate on glow-discharged carbon-coated copper grids. R44 target ssDNA (BG3028, **Supplementary Table 3**) was added to Cascade subcomplexes loaded with R44 crRNA in a twofold excess. EM was performed on a Philips CM120 equipped with a LaB6 tip (120 kV). Images were recorded with a Gatan 4000 SP 4K slow-scan CCD camera (130,000×) at a pixel size of 0.23 nm at the specimen level and analyzed with GRACE software⁵¹. Single-particle projections were selected from micrographs mainly by reference-based automated particle selection incorporated into GRIP software⁵².

48. Tahallah, N., Pinkse, M., Maier, C.S. & Heck, A.J. The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Commun. Mass Spectrom.* **15**, 596–601 (2001).
49. van den Heuvel, R.H. *et al.* Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* **78**, 7473–7483 (2006).
50. Maxam, A.M. & Gilbert, W. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499–560 (1980).
51. Oostergetel, G.T., Keegstra, W. & Brissos, A. Automation of specimen selection and data acquisition for protein electron crystallography. *Ultramicroscopy* **74**, 47–59 (1998).
52. van Heel, M. *et al.* Single-particle electron cryo-microscopy: towards atomic resolution. *Q. Rev. Biophys.* **33**, 307–369 (2000).