LETTER

doi:10.1038/nature10402

Structures of the RNA-guided surveillance complex from a bacterial immune system

Blake Wiedenheft[1,2]*, Gabriel C. Lander[3]*, Kaihong Zhou[1,2], Matthijs M. Jore[4], Stan J. J. Brouns[4], John van der Oost[4], Jennifer A. Doudna[1,2,5,6] & Eva Nogales[1,2,3]

Bacteria and archaea acquire resistance to viruses and plasmids by integrating short fragments of foreign DNA into clustered regularly interspaced short palindromic repeats (CRISPRs). These repetitive loci maintain a genetic record of all prior encounters with foreign transgressors[1–6]. CRISPRs are transcribed and the long primary transcript is processed into a library of short CRISPR-derived RNAs (crRNAs) that contain a unique sequence complementary to a foreign nucleic-acid challenger[7–12]. In *Escherichia coli*, crRNAs are incorporated into a multisubunit surveillance complex called Cascade (CRISPR-associated complex for antiviral defence), which is required for protection against bacteriophages[13,14]. Here we use cryo-electron microscopy to determine the subnanometre structures of Cascade before and after binding to a target sequence. These structures reveal a sea-horse-shaped architecture in which the crRNA is displayed along a helical arrangement of protein subunits that protect the crRNA from degradation while maintaining its availability for base pairing. Cascade engages invading nucleic acids through high-affinity base-pairing interactions near the 5′ end of the crRNA. Base pairing extends along the crRNA, resulting in a series of short helical segments that trigger a concerted conformational change. This conformational rearrangement may serve as a signal that recruits a *trans*-acting nuclease (Cas3) for destruction of invading nucleic-acid sequences.

The CRISPR RNA-guided adaptive immune system in *Escherichia coli* K12 consists of eight *cas* genes and a downstream CRISPR locus (Fig. 1a). Cascade is a 405-kDa ribonucleoprotein complex composed of 11 subunits of five functionally essential Cas proteins (one CasA protein, two CasB proteins, six CasC proteins, one CasD protein and one CasE protein) and a 61-nucleotide crRNA[13,14]. Previous structural and biochemical studies have determined the composition and general morphology of the Cascade complex[14]. However, the arrangement of subunits and the mechanism of target recognition remain largely unknown. Using single-particle cryo-electron microscopy (cryo-EM), we determined the structure of Cascade at a resolution of ~8 Å (Fig. 1, Supplementary Fig. 1 and Supplementary Movie 1). This structure provides a detailed description of the subunit organization with sufficient resolution to observe secondary structure elements within each of the 11 protein components and the crRNA.

Overall, Cascade has a sea-horse-shaped architecture with a helical backbone, as suggested by two-dimensional electron microscopy[14]. The backbone is capped at its ends by two prominent features representing the 'head' (CasE) and 'tail' (CasA) of the sea-horse anatomy. The resolution of the cryo-EM reconstruction, together with crystal structures of two individual subunits and previously established subunit stoichiometries, allowed us to delineate the molecular boundaries of all the individual components. The resulting model provides detailed insight into Cascade organization (Fig. 1, Supplementary Fig. 2 and Supplementary Movie 2).

The backbone of Cascade consists of six copies of CasC organized in a helical stack. Integral to the spine, the crRNA lies in a groove on the concave surface of the CasC helix. The extended conformation of the crRNA explains its importance for complex assembly and suggests that is has a structural role as a template for CasC subunit association.

The crRNA is anchored at both ends of the Cascade complex by specific protein–RNA interactions that can be seen in the cryo-EM structure. CasE is the endoribonuclease that specifically binds to a stem–loop in the CRISPR transcript, and cleavage results in a 61-nucleotide crRNA[13,14]. A co-crystal structure of the CasE homologue from *Thermus thermophilus* in complex with the crRNA stem–loop fits with high fidelity into the head structure of the complex[9,11] (Supplementary Fig. 2). After CRISPR cleavage, CasE remains bound to the 3′ end of the mature crRNA and the RNA stem–loop protrudes like a 'beak' from the head of the Cascade complex (Fig. 1 and Supplementary Fig. 2). The crRNA loops around the base of CasE and extends ~45 nucleotides along the binding groove in the helical backbone. The 5′ end of the crRNA terminates within the tail of the complex, forming a hook-like structure in a pocket between CasC6 (C6, the sixth CasC subunit) and CasA (Fig. 1b). CasD is adjacent to this interface, sits at the midpoint of CasA and makes extensive contacts with the neighbouring domain of C6 (Fig. 1).

The two CasB subunits form an elongated dimer positioned along the inner surface of the crRNA–CasC spine, connecting the head (CasE) and tail (CasA) of the Cascade complex. CasB1 sits next to the head, and has limited interactions with CasE and the first two CasC subunits (C1 and C2). The CasB2 subunit makes similar contacts with CasC subunits C3 and C4. The extended conformation of the CasB dimer creates a deep cleft that cradles the 3′ half of the crRNA spacer sequence (Fig. 1).

The three-dimensional structure of Cascade reveals how six identical CasC polypeptides assemble into an asymmetric helix that is programmed to terminate at C6. The first five CasC subunits are structurally similar, forming a right-handed helix with a pitch of 135 Å (Fig. 2a). However, this symmetry is perturbed between C5 and C6 owing to the interaction of C6 with the hook-like structure at the 5′ end of the crRNA (Fig. 2b). This interaction results in a ~160° rotation of the distal domain of C6 (Fig. 2). This rotation flips the distal domain out of the vertical axis of the CasC helix, breaking the helical arrangement. CasD and CasA stabilize this distinct structure and orientation of C6. The flipped-out conformation of the distal domain in C6 results in a larger gap between the C5 and C6 subunits, exposing a segment of crRNA in the 5′ region of the spacer sequence.

Cascade engages invading nucleic acids with high affinity when they bear sequence complementary to the 5′ end of the crRNA spacer sequence[15,16] (the 'seed' sequence: nucleotides 1–5, 7–8). Bacteriophages containing a single-nucleotide mutation in the seed region escape Cascade-mediated immunity, whereas point mutations outside the

[1]Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. [2]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. [3]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [4]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands. [5]Department of Chemistry, University of California, Berkeley, California 94720, USA. [6]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.
*These authors contributed equally to this work.

©2011 Macmillan Publishers Limited. All rights reserved

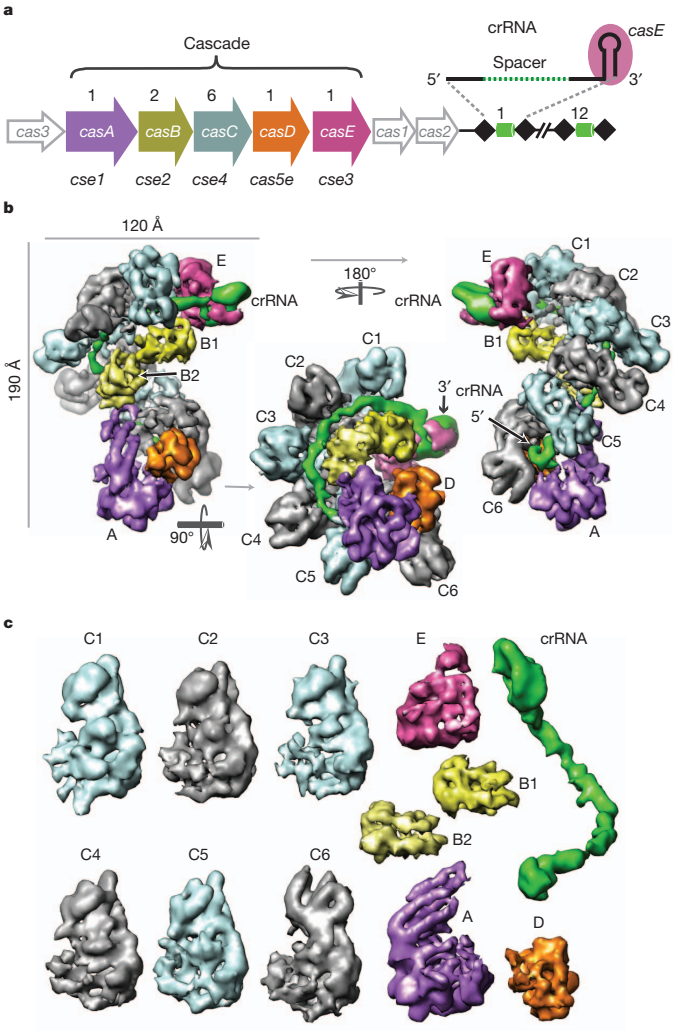

Figure 1 | Structure of the Cascade complex from *E. coli*. **a** The CRISPR system in *E. coli* K12 (Cse-type) consists of eight *cas* genes and a downstream CRISPR locus. *casA* to *casE* are members of large gene families, referred to as *cse1*, *cse2*, *cse4*, *cas5e* and *cse3*, respectively[28,29]. The CRISPR consists of a series of 29-nucleotide repeats (black diamonds) separated by 32-nucleotide spacer sequences (green cylinders). CasE (magenta) is an endoribonuclease that specifically binds to a stable stem–loop in the CRISPR RNA repeat and cleaves 8 nucleotides away from the spacer sequence in the 5′ direction[9,11,14]. **b**, Cascade assembles into a sea-horse-shaped architecture where the crRNA (green) is positioned along a helical arrangement of six CasC subunits (C1–6). The helical spine is capped at its ends by two prominent features that represent the head (E, CasE) and tail (A, CasA) of the sea-horse anatomy. D, CasD. **c**, Cascade consists of unequal numbers of Cas proteins and a crRNA ($CasA_1B_2C_6D_1E_1crRNA_1$). The first five CasC subunits (C1–5) are structurally similar, whereas CasC6 is distinct. B1, CasB1; B2, CasB2.

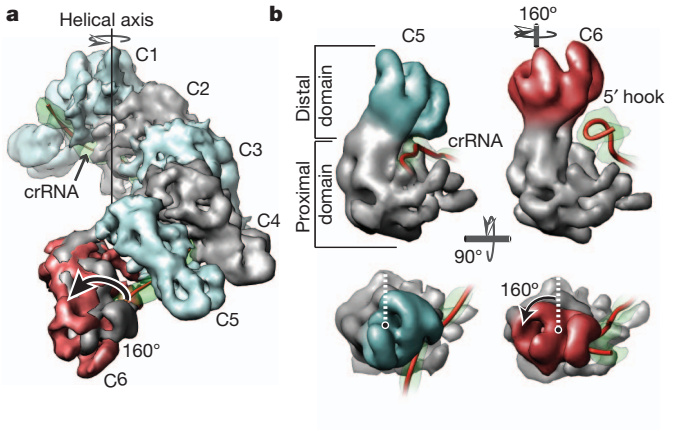

Figure 2 | Programmed capping of the CasC helix. **a**, C1–5 form a right-handed helix with a pitch of 135 Å. The two domains of each CasC subunit are referred to as proximal and distal, relative to the helical axis of the CasC subunits. The different conformation of C6 (red) relative to the other CasC subunits interrupts the helical symmetry (black arrow). **b**, The crRNA is positioned along a contiguous groove on the concave surface of the C1–5 helix. The 5′ end of the crRNA forms a hook-like structure that interacts with C6. This interaction correlates with the distinct conformation of C6 and the termination of the helix. Although the proximal domains of C5 and C6 have the same orientation, the distal domain of C6 is rotated by ~160° relative to the other CasC subunits (black arrow). The centre of rotation is indicated by a black dot.

seed region do not facilitate escape[15]. Despite these observations, little is known about the mechanism of target binding and the extent to which base pairing occurs between the crRNA and a target sequence. To address these issues, we determined the structure of Cascade bound to a 32-nucleotide single-stranded RNA (ssRNA) that is complementary to the spacer sequence of the crRNA. Although Cascade is thought to target DNA, it has also been shown to bind ssRNA targets with high affinity. *In vitro*, Cascade makes specific and nonspecific interactions with double-stranded DNA substrates but interacts with RNA in a strictly sequence-specific fashion[13,14]. We chose a ssRNA substrate to achieve maximal target site occupancy and sample homogeneity. Notably, RNA and DNA targets induce similar structural changes in the Cascade complex as detected by partial proteolysis (Supplementary Fig. 3).

The ~9 Å target-bound structure maintains the sea-horse morphology observed for the unbound complex, in which the CasC subunits can be superimposed on the unbound structure (Fig. 3a and Supplementary Fig. 4). However, examination of the other subunits reveals several significant differences that occur on target binding (Fig. 3b). The width of the crRNA density approximately doubles along the entire length of the spacer sequence, suggestive of duplex formation. Strikingly, however, the crRNA and target RNA strands do not form one contiguous double-stranded helix. Instead, we observe density consistent with five short duplex segments, each accommodating four or five base pairs of double-stranded RNA (Fig. 3b–e). The helical segments are connected by short (1–2-nucleotide) non-helical regions that seem to be the contact sites for individual CasC subunits (Fig. 3e).

In addition to changes in the RNA, we observed a concerted conformational change in the locations and orientations of CasE, CasB and CasA. CasE remains bound to the 3′ crRNA stem–loop, and target binding results in a clockwise rotation (~15°) consistent with a shortening of the crRNA spacer (Fig. 3b). This motion is coupled with movement of the CasB dimer, which forms a protein bridge between the head (CasE) and the tail (CasA) of the complex (Supplementary Movies 3 and 4). The two CasB subunits move ~17 Å along the crRNA binding groove, towards the tail. CasB2 interacts with a four-helix bundle in CasA, inducing a ~30° rotation of CasA. This rotation is centred around CasD, which functions as a hinge that connects CasA to C6. The distinct orientation of C6 relative to C1–5 is conserved in the target-bound Cascade complex (Fig. 3c, d). However, duplex formation on target binding seems to alter the interaction between C6 and the 5′ hook. Base pairing in the crRNA spacer is concomitant with a disruption of the hook-like structure at the 5′ end of the crRNA and results in a decrease in resolvable density for the distal domain of C6 (Fig. 3c, d and Supplementary Movie 5).

The target-bound structure of Cascade reveals segments of density along the length of the crRNA spacer that accommodate short regions of double-stranded helix. This structural observation indicates that the entire spacer sequence is available for base pairing to a complementary target sequence. However, previous genetic and biochemical assays have identified a preferred high-affinity binding site in the 5′ seed region of the crRNA that is essential for phage protection[15]. To test the relative binding affinities of discrete regions of the crRNA, we

©2011 Macmillan Publishers Limited. All rights reserved

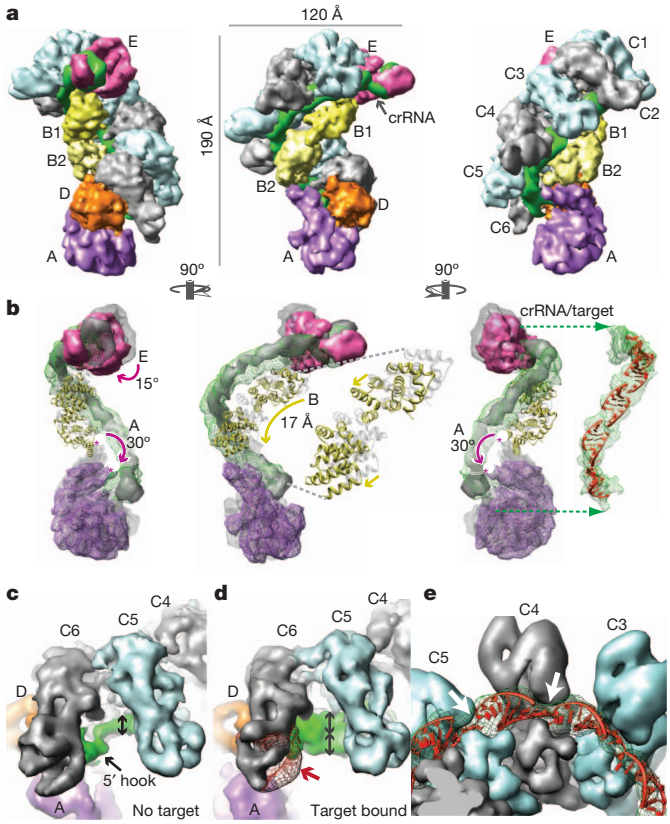

Figure 3 | Target binding triggers a concerted conformational change. **a**, Structure of Cascade bound to a 32-nucleotide target RNA complementary to the crRNA spacer sequence, at a resolution of ~9 Å. **b**, Removing the CasC subunits reveals significant structural differences between the unbound and target-bound structures. CasE, CasA and the crRNA from the unbound structure are shown as grey volumes. The same subunits from the target bound complex are shown in magenta, green mesh and purple mesh, respectively. Models of the CasB crystal structure are shown docked into the cryo-EM map before (grey) and after (yellow) target binding. Four or five base pairs of double-stranded RNA (red) fit with high fidelity into the crRNA density. On target binding, CasE rotates by ~15° (magenta arrow), both CasB subunits move ~17 Å along the concave surface of the CasC backbone (yellow arrow) and CasA rotates by ~30° (purple arrow). The purple asterisks indicate the positions of the four-helix bundle on CasA before and after target binding. **c**, In the unbound structure, CasC6 is rotated out of the CasC helix, exposing the 5′ region of the crRNA (double-headed arrow indicates density for the single-stranded crRNA). **d**, Additional density corresponding to the target nucleic acid is clearly visible in the target-bound complex (double-headed arrows). Base pairing in the crRNA spacer disrupts the hook-like structure at the 5′ end of the crRNA, and a difference map reveals a significant loss of resolvable density in the distal domain of C6 following target binding (red mesh and arrow). **e**, The short segments of double-stranded helices are connected by short, non-helical regions located at pinch points of the CasC subunits (white arrows).

designed a series of 16-nucleotide target DNAs that tile across the crRNA in 8-nucleotide steps (Supplementary Fig. 5). Using native gel mobility shift assays, we observed high-affinity interactions for targets that include the seed region, and that binding affinities decrease with increasing steps in the 3′ direction (Supplementary Fig. 5). This indicates that each portion of the crRNA spacer is accessible for target binding, but that the unique structural context of the 5′ region of the crRNA results in a higher binding affinity for this region. A high-affinity seed binding site, of approximately the same length, has also been observed in other gene silencing systems[16,17]. In eukaryotes, Argonaute proteins enhance target recognition by pre-ordering the microRNA seed sequence in a helical configuration, and we speculate that Cascade may use a similar mechanism[17].

Overall, our data suggest a model in which Cascade-mediated surveillance initially relies on high-affinity binding to the seed region of

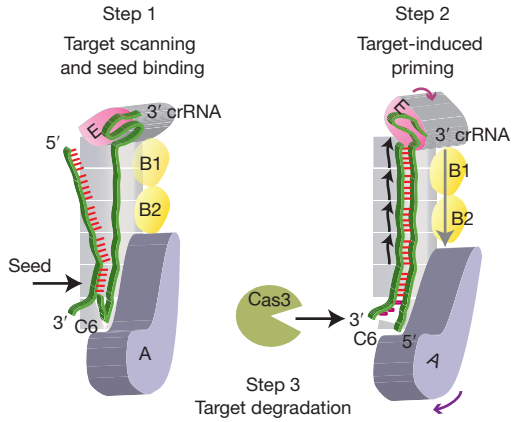

Figure 4 | A model for pathogen surveillance and signalling by Cascade. Efficient surveillance and detection of invading nucleic acids is mediated by base pairing in the seed sequence (nucleotides 1–5, 7–8) of the crRNA[15]. Duplex formation may proceed in the 3′ direction (four curved black arrows), resulting in a series of short helical duplexes that shorten the crRNA; this in turn causes a concerted conformational change in CasA, CasB and CasE (coloured arrows) that coincides with a disruption of the 5′ hook and results in a decrease in resolvable density for the distal domain of C6. Cascade binds single-stranded and double-stranded substrates. Here we depict the target as a single strand for simplicity.

the crRNA (Fig. 4). Following a seed match, duplex formation then proceeds in the 3′ direction along the length of the crRNA in increments of four or five base pairs. These helical segments reduce the overall length of the crRNA, triggering the concerted conformational change that may serve as a signal to recruit Cas3 for target destruction (Fig. 4 and Supplementary Movies 3, 4 and 5).

METHODS SUMMARY

Cascade preparation. Proteins of the CRISPR system in *E. coli* (CasA–E) and the synthetic CRISPR RNA were co-expressed in *E. coli* BL21(DE3). Cascade was affinity-purified on Strep-Tactin Superflow Plus resin (Qiagen) using an amino-terminal Strep-II tag on the CasB subunit. The Strep-II peptide was removed from CasB by cleavage with PreScission protease, and the complex was further purified by gel filtration. The purified protein was used for native gel mobility shift assays using standard methods.

Cryo-EM and image analysis. Purified complexes were applied to glow-discharged C-flats (Protochips Inc.), blotted and plunged into liquid ethane. Data were acquired using a Tecnai F20 Twin transmission electron microscope at $20\,e^{-}$ Å^{-2} on a Gatan 4,000 × 4,000-pixel charge-coupled-device camera using the LEGINON data collection software[18]. Data preprocessing was performed using functionalities within the APPION electron microscopy processing environment[19].

The contrast transfer function (CTF) of each image was estimated during data collection using ACE2 and CTFFIND[20]. Particles were initially selected using a difference-of-Gaussians particle picker, extracted using a box size of 280 × 280 pixels, and classified two dimensionally using the IMAGIC package[21–23]. The resulting reference-free class averages were used for template-based automatic particle selection[24]. Three-dimensional maps were calculated using an iterative projection-matching approach with libraries from the EMAN2 and SPARX software packages[25,26]. Volume segmentation, docking and visualization of molecular models were performed using CHIMERA[27]

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 May; accepted 27 July 2011.

1. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
2. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).
3. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320,** 1047–1050 (2008).
4. Bolotin, A., Ouinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).

©2011 Macmillan Publishers Limited. All rights reserved

5. Jansen, R., van Embden, J. D. A., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–1575 (2002).
6. 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).
7. 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).
8. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).
9. Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M. & MacMillan, A. M. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nature Struct. Mol. Biol.* **18,** 688–692 (2011).
10. 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).
11. Sashital, D. G., Jinek, M. & Doudna, J. A. An RNA-induced conformational change required for CRISPR RNA cleavage by the endonuclease Cse3. *Nature Struct. Mol. Biol.* **18,** 680–687 (2011).
12. Wang, R., Preamplume, G., Terns, M. P., Terns, R. M. & Li, H. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure* **19,** 257–264 (2011).
13. Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321,** 960–964 (2008).
14. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Struct. Mol. Biol.* **18,** 529–536 (2011).
15. Semenova, E. *et al.* A crRNA seed sequence governs CRISPR interference. *Proc. Natl Acad. Sci. USA* **108,** 10098–10103 (2011).
16. Wiedenheft, B. *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl Acad. Sci. USA* **108,** 10092–10097 (2011).
17. Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. & Barford, D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol. Cell* **33,** 204–214 (2009).
18. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151,** 41–60 (2005).
19. Lander, G. C. *et al.* Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166,** 95–102 (2009).
20. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
21. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166,** 205–213 (2009).
22. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128,** 82–97 (1999).
23. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116,** 17–24 (1996).
24. Roseman, A. M. FindEM–a fast, efficient program for automatic selection of particles from electron micrographs. *J. Struct. Biol.* **145,** 91–99 (2004).
25. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).
26. Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157,** 47–55 (2007).
27. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157,** 281–287 (2007).
28. 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).
29. 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).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the Doudna and Nogales lab members for their reading of this manuscript, and to P. Grob, S. Hill, R. Hall and T. Houweling for technical support. This project was funded by a National Science Foundation grant to J.A.D., a Veni grant to S.J.J.B. (863.08.014) and a NWO Vici grant to J.v.d.O. (865.05.001). G.C.L. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation. B.W. is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation. E.N. and J.A.D. are Howard Hughes Medical Institute investigators.

Author Contributions M.M.J., S.J.J.B. and J.v.d.O. designed expression constructs. B.W. and K.Z. purified samples. B.W., K.Z., M.M.J. and S.J.J.B. performed assays. B.W. and G.C.L. carried out the cryo-electron microscopy. G.C.L. performed the electron microscopy processing and segmentation analysis. All authors contributed to experimental design, data analysis and manuscript preparation.

Author Information The cryo-electron microscopy density maps for Cascade and Cascade bound to a 32-nucleotide target have been deposited at the Electron Microscopy Data Bank under accession numbers 5314 and 5315, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A.D. (doudna@berkeley.edu) and E.N. (enogales@lbl.gov).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Cascade preparation. The Cas proteins (CasA–E) and the synthetic CRISPR RNA were co-expressed in *E. coli* BL21(DE3) cells that were induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside at $D_{600\,nm} = 0.5$ in overnight cultures grown at 16 °C (refs 13, 14). Cells from the overnight expression cultures were collected by centrifugation (5,000*g* for 10 min). The cell pellet was resuspended in lysis buffer (100 mM Tris, pH 8.0, 300 mM KCl, 1 mM EDTA, 1 mM tris(2-carboxyethyl) phosphine hydrochloride (TCEP) and 5% glycerol), supplemented with protease inhibitors (Roche), and the slurry was sonicated on ice for 2 min in 10-s bursts. The lysate was clarified by centrifugation (22,000*g* for 20 min) and the complex was affinity-purified on Strep-Tactin Superflow Plus resin (Qiagen) using an N-terminal Strep-II tag on the CasB subunit. The complex was eluted from the resin in 50 ml lysis buffer containing 2.5 mM desthiobiotin. The Strep-II peptide was removed from CasB by cleavage with PreScission protease during dialysis at 4 °C overnight against gel filtration buffer (25 mM Hepes, pH 7.5, 100 mM KCl, 1 mM TCEP). The liberated Strep-II tag was removed using a second Strep-Tactin Superflow Plus column (Qiagen). The protein was concentrated (Amicon) for further purification on a Superose 6 size-exclusion column (GE Healthcare) equilibrated in gel filtration buffer. The target-bound complex was prepared by adding fivefold molar excess of an oligoribonucleotide complementary to the crRNA. The mixture was incubated at 37 °C for 15 min. The unbound oligoribonucleotide was separated from the target-bound complex on a Superdex 200 size-exclusion column (GE Healthcare).

Cryo-electron microscopy. Preservation of nucleoprotein complexes in vitreous ice was performed in the same manner for both unbound and target-bound specimens. Aliquots (4 µl) of purified sample (~1.2 mg ml^{-1}) were placed onto C-flats (Protochips Inc.) that had been just glow-discharged in a nitrogen atmosphere for 60 s using an Edwards carbon evaporator. Grids were loaded into an FEI Vitrobot whose incubation chamber maintained an environment of 4 °C and 100% humidity. The grids were blotted for 3 s using a blotting offset of −1, and were then plunged into liquid ethane and stored in liquid nitrogen until being loaded into the electron microscope. Data were acquired using a Tecnai F20 Twin transmission electron microscope operating at 120 keV at a nominal magnification of ×100,000 (1.15 Å at the specimen level) using low-dose exposures (~20 e^- Å^{-2}) with a randomly set focus ranging from −0.8 to −2.5 µm. A total of 2,370 images of unbound Cascade and 1,406 images of target-bound Cascade were automatically recorded on a Gatan 4,000 × 4,000-pixel charge-coupled-device camera (15-µm pixel size) using the LEGINON data collection software[18].

Single-particle pre-processing. All data preprocessing leading to three-dimensional reconstruction was performed using functionalities within the APPION processing environment[19]. Concurrent with data collection, carbon edges were manually masked from the acquired images, and particles were initially extracted automatically using a difference-of-Gaussians particle picker[21]. The contrast transfer function (CTF) was additionally estimated automatically during data collection using both the ACE2 program and the CTFFIND program[20]. Particle image stacks were generated by extracting selected particles with a box size of 288 × 288 (performed with the 'batchboxer' program[22]) from images whose estimated CTF confidence value was greater than 80%. The stack was reduced by a factor of four, and reference-free, two-dimensional classification was performed using iterative multivariate statistical analysis and multireference alignment analysis (MSA–MRA) within the IMAGIC software package[23]. The resulting class averages showing detailed structural information at a high signal-to-noise ratio were selected for use as templates for template-based automatic particle selection using FINDEM[24], resulting in a total of 498,137 and 389,166 particle selections for the unbound and target-bound particles, respectively. Particles were extracted in the same manner as previously described, and reference-free, two-dimensional classifications were again performed with the MSA–MRA methodology. The resulting gallery of 5,000 class averages was manually curated to remove ice contamination, false positives and damaged Cascade complexes. Another round of MSA–MRA was performed on the resulting 'cleaned' stack, and the resulting class averages were again inspected to remove false or damaged particle selections. Only particles contained in the final set of class averages were re-extracted from phase-flipped micrographs to generate the final stack for the data sets. The particle image stacks, which contained 275,573 and 176,090 particles for the unbound and target-bound complexes, respectively, were binned by a factor of two to a pixel size of 2.3 Å for three-dimensional reconstructions.

Initial models for three-dimensional reconstruction were determined using a low-resolution SAXS reconstruction[30]. The SAXS reconstruction was low-pass-filtered to a resolution of 60 Å and forward-projected at an angular increment of 15°, and a multireference alignment was performed using the final 5,000 reference-free class averages of each of the Cascade complexes. The aligned class averages were back-projected to generate a new density model, which was then used for another iteration of projection matching. Ten iterations of projection matching at an angular increment of 15° were performed using the EMAN reconstruction software to arrive at the unbound and target-bound Cascade densities, which were used as starting points for refinement using single particles.

Three-dimensional reconstruction and analysis. The unbound and target-bound data sets were processed separately, each using their corresponding initial model. Three-dimensional refinements of the starting densities were performed using an iterative projection-matching approach with libraries from the EMAN2 and SPARX software packages[25,26]. Projection matching began at an angular increment of 25°, progressing down to 0.8° over the course of dozens of iterations. The reconstruction algorithm dictated that the reconstruction was only allowed to proceed to the next smaller angular increment once >95% of the particles had a pixel error of less than one pixel. The resolution was estimated by splitting the data set into two separate halves and calculating the Fourier shell correlation between the resulting volumes. The density was conservatively low-pass-filtered to this estimated resolution before proceeding to the next iteration. The estimated resolutions based on the Fourier shell correlation for the unbound Cascade density were 8.8 Å at a correlation of 0.5 and 7.7 Å at a correlation of 0.143. The estimated resolutions based on the Fourier shell correlation for the target-bound Cascade density were 9.2 Å at a correlation of 0.5 and 8.0 Å at a correlation of 0.143.

To dampen predominant low-resolution amplitudes, the density Fourier amplitudes of the two final reconstructed densities were adjusted to match experimental one-dimensional SAXS curves using the SPIDER software package[30]. Segmentation of the densities was performed manually using the 'volume tracer' tool of the UCSF CHIMERA visualization software[27]. UCSF CHIMERA was also used for rigid-body docking of crystal structures into the segmented densities, as well as for generation of all surface renderings of cryo-EM densities. To assess the difference in position of CasC6 relative to the other CasC subunits, helical models of CasC and the RNA were generated by using two components of the iterative helical real-space reconstruction method[31]. Cryo-EM density corresponding to C1–5, as well as their associated nucleic-acid densities, were first segmented from the asymmetric reconstructions. A rough estimate of the axial rise and rotation of the subunits was determined manually, and were used as initial parameters for the 'hsearch_lorentz' program, which determined the true axial parameters. The 'himpose_long' program was then used to impose the helical symmetry on the segmented density, generating the helical structure.

Electrophoretic mobility shift assays. Binding assays were performed by incubating Cascade with 5′ ^{32}P-labelled single-stranded DNAs. Each reaction included 25 mM HEPES, pH 7.5, 100 mM KCl, 1 mM TCEP, 1% glycerol, 1 mM $MgCl_2$ and 1 mg ml^{-1} transfer RNA. All reactions were incubated for 15 min at 37 °C before electrophoresis on 6% polyacrylamide gels. Gels were dried and exposed using phosphor storage screens, scanned with a phosphorimager (GE Healthcare) and quantified using KALEIDAGRAPH (Synergy software).

Limited proteolysis. Preparations of Cascade were annealed to ssRNA or single-stranded DNA substrates complementary to the spacer sequence, spacer plus the 5′ handle (self) or spacer with a protospacer-adjacent motif. Limited proteolysis was performed at room temperature (25 °C) in a total reaction volume of 100 µl. Each reaction mixture contained 30 µM trypsin (Sigma), 3.7 µM Cascade, 25 mM Hepes, ph 7.5, 100 mM KCL, 5% glycerol and 1 mM TCEP. Aliquots (20 µl) of the reaction were sampled at each time point and added directly to ×5 SDS-loading buffer at 95 °C for 5 min. Reaction products were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis using 12% gels.

30. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
31. Egelman, E. H. Single-particle reconstruction from EM images of helical filaments. *Curr. Opin. Struct. Biol.* **17,** 556–561 (2007).

©2011 Macmillan Publishers Limited. All rights reserved