LETTER

doi:10.1038/nature13733

Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*

Hongtu Zhao[1,2], Gang Sheng[1], Jiuyu Wang[1], Min Wang[1], Gabor Bunkoczi[3], Weimin Gong[1], Zhiyi Wei[4] & Yanli Wang[1]

Clustered regularly interspaced short palindromic repeats (CRISPR) together with CRISPR-associated (Cas) proteins form the CRISPR/Cas system to defend against foreign nucleic acids of bacterial and archaeal origin[1–9]. In the I–E subtype CRISPR/Cas system, eleven subunits from five Cas proteins ($CasA_1B_2C_6D_1E_1$) assemble along a CRISPR RNA (crRNA) to form the Cascade complex[10–13]. Here we report on the 3.05 Å crystal structure of the 405-kilodalton *Escherichia coli* Cascade complex that provides molecular details beyond those available from earlier lower-resolution cryo-electron microscopy structures. The bound 61-nucleotide crRNA spans the entire 11-protein subunit-containing complex, where it interacts with all six CasC subunits (named CasC1–6), with its 5′ and 3′ terminal repeats anchored by CasD and CasE, respectively. The crRNA spacer region is positioned along a continuous groove on the concave surface generated by the aligned CasC1–6 subunits. The five long β-hairpins that project from individual CasC2–6 subunits extend across the crRNA, with each β-hairpin inserting into the gap between the last stacked base and its adjacent splayed counterpart, and positioned within the groove of the preceding CasC subunit. Therefore, instead of continuously stacking, the crRNA spacer region is divided into five equal fragments, with each fragment containing five stacked bases flanked by one flipped-out base. Each of those crRNA spacer fragments interacts with CasC in a similar fashion. Furthermore, our structure explains why the seed sequence, with its outward-directed bases, has a critical role in target DNA recognition. In conclusion, our structure of the Cascade complex provides novel molecular details of protein–protein and protein–RNA alignments and interactions required for generation of a complex mediating RNA-guided immune surveillance.

Our 3.05 Å crystal structure of the *E. coli* Cascade complex exhibits a sea-horse-shaped architecture, similar to previous cryo-electron microscopy structures[10]. Instead of forming a simple, compact structure, the 11 subunits are assembled into two structural layers. Six CasC proteins, together with CasD and CasE, tightly pack to form the outer layer (Fig. 1b–d and Extended Data Fig. 1). Within this outer arch-like configuration, the six CasC proteins labelled C1 to C6 are assembled into a symmetry-related helical alignment defined by a ~40° rotation, together with a vertical shift of 25 Å, per subunit. CasE and CasA cap the two ends of the CasC helical arrangement by associating with CasC1 and CasC6, respectively. The inner layer, consisting of the CasA and the CasB dimer, is connected to the outer layer mainly via CasA–CasD interactions as well as multiple relatively weak CasC contact sites at their lateral regions (Fig. 1c). We observed all 61 nucleotides of the crRNA, with the 32-nucleotide spacer shown in red and the two flanking repeat segments (also called 5′- and 3′-handle) shown in orange in Fig. 1. The crRNA is properly positioned along the long outer layer (Extended Data Fig. 2). Interestingly, each subunit in the outer layer contains a long β-hairpin protruding from its structured core (Fig. 1d).

The CasE endonuclease subunit at the head end of the Cascade complex adopts a two-ferredoxin fold, similar to the RNA-bound structures of *Thermus thermophilus* Cse3[14,15] and *Pseudomonas aeruginosa* Csy4[16]. The 3′-end of the crRNA adopts a stem loop-fold, with the RNA stem anchored within the positively charged cleft of the carboxy-terminal domain of CasE (Fig. 2a, b and Extended Data Fig. 3a). Its β6–β7 hairpin, a universal structural feature of the CasE family, contains several positively charged residues that contact the crRNA backbone. The 3′-end repeat of crRNA is recognized by the side chains of $R111^{CasE}$ and $R113^{CasE}$, which insert into the major groove of the RNA stem and contact the bases of G(+10) and G(+19) (Extended Data Fig. 3b, c). With the following continuous turn, the 3′ terminal repeat surrounds the β6–β7 hairpin, a feature not observed previously, and is probably responsible for proper anchoring of the crRNA in the Cascade complex (Fig. 2b). At the opposite face of the crRNA stem-binding surface, a V-shaped cleft is formed between the two domains of CasE. There, a specific loop from CasC1 residues 197-208 inserts into this V-shape cleft of CasE via hydrophobic interactions (Fig. 2c and Extended Data Fig. 3d).

Once the crRNA is Cascade-bound, the 8-nucleotide repeat of its 5′-end takes on a hook-like shape, with a sharp turn at the G1-G(−1) and G(−1)-C(−2) steps, preceded by continuous stacking within the C(−3)-A(−5) steps, and three residues pointing into a different direction at each base (Fig. 2d). The 5′-handle is buried into a pocket formed by CasD, CasC6 and CasA in a sequence-specific manner (Fig. 2d–g). Residues C(−2), C(−3), A(−5) and U(−7) interact with CasD and CasC6 sequence specifically. The base of G(−1) stacks onto the side-chain of $L89^{CasD}$. $R108^{CasD}$, which is located within the β5–β6 hairpin, forms hydrogen bonds with the base of C(−2) (Fig. 2e). The CRISPR/Cas system cannot launch a defence against the invading plasmid once $R108^{CasD}$ is substituted by Ala (Extended Data Fig. 4a). $F129^{CasA}$ and $K177^{CasC6}$ form respective hydrogen bonds with the bases of C(−3) and A(−5), which are further stabilized by stacking with the side chains of $F129^{CasA}$ and $R206^{CasD}$, respectively (Fig. 2e and f). The 5′-terminal residues A(−6), U(−7) and A(−8) are sandwiched by amino acid residues; that is, A(−6) between $F208^{CasD}$ and $P19^{CasD}$, U(−7) between $Y145^{CasD}$ and the peptide plane of residues $38–39^{CasD}$ on the α1-helix, and A(−8) between $Y142^{CasD}$ and $D179^{CasC6}$-$R48^{CasD}$ pair. Additionally, the 2′ hydroxyl and phosphate groups of the 5′-handle provide hydrogen bonds and ionic bonds with CasD (Fig. 2f, g and Extended Data Fig. 2).

A specific β5–β6 hairpin of CasD, which contacts CasA and CasC6, has a critical role in 5′-handle binding. This hairpin extends across the RNA backbone to CasC6 through the gap between G(−1) and C(−2), in similar fashion to that of CasC2–6 (discussed below). In addition, a loop formed by residues $120–135^{CasA}$ inserts into the cavity between the β-hairpin and the main body of CasD (Fig. 2h and Extended Data Fig. 4c), and a loop of CasC6 (residues 164–179) inserts into the opposite groove (Extended Data Fig. 4d), indicating a role of the β-hairpin for proper assembly of CasA and CasC6 into the Cascade complex. In the crRNA-free structure, this arm is highly flexible[17], suggesting mutual stabilization between CasD and bound crRNA, which is supported by the fact that upon replacement of the β-hairpin by a non-functional $(GGS)_4$ linker, the Cascade complex cannot be assembled properly and Cascade-mediated immunity was abolished (Extended Data Fig. 4a, b, e, f).

[1]Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. [2]University of Chinese Academy of Sciences, Beijing 100049, China. [3]Cambridge Institute for Medical Research, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0XY, UK. [4]Department of Biology, South University of Science and Technology of China, Shenzhen, 518055, China.

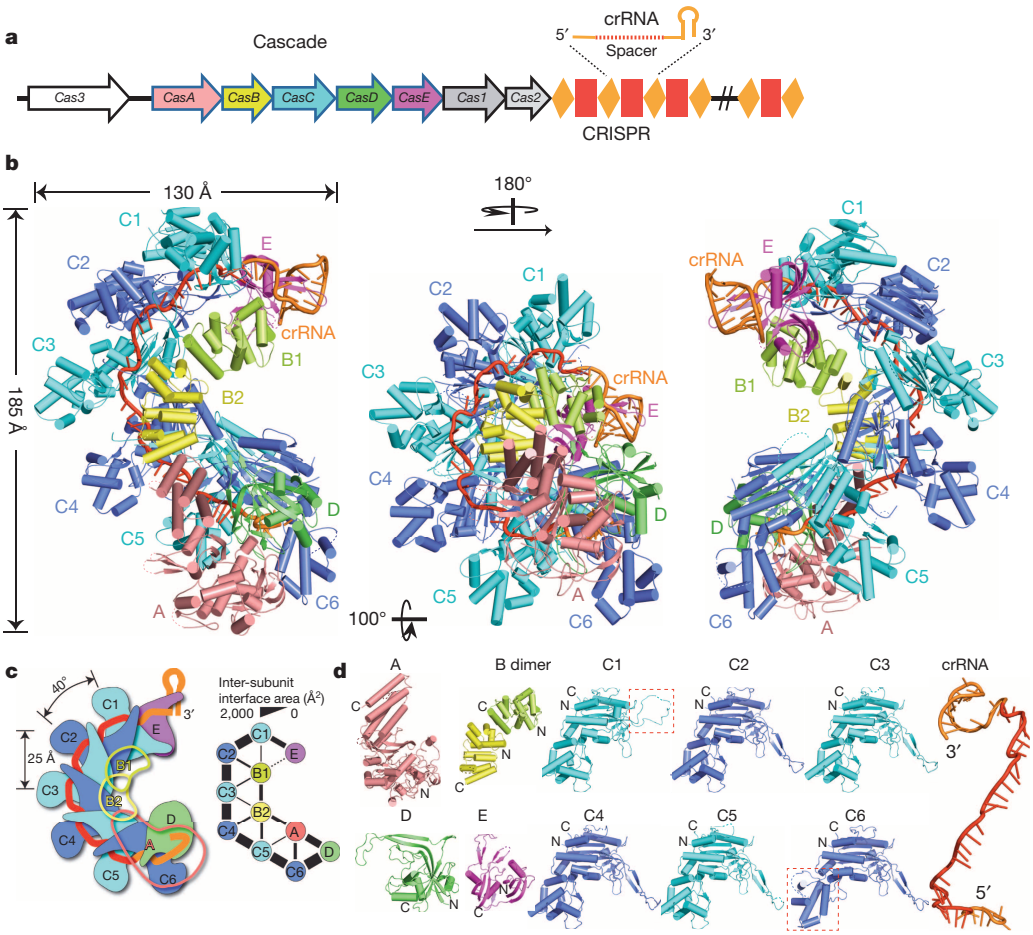


Figure 1 | Crystal structure of the Cascade complex from *E. coli*. **a**, The I–E subtype CRISPR system in *E. coli* (K12) consists of eight Cas proteins and CRISPR locus. **b**, Overall structure of the Cascade complex. **c**, Cartoon scheme of Cascade showing the structural uniqueness in both the crRNA recognition and the subunit organization. All subunits in the outer layer contain a long hairpin protruding from the structural core for stapling the crRNA spacer, for specific binding of the stem-loop structure, and for the outer layer organization. The sophisticated protein interaction network in the Cascade organization is shown. **d**, Individual components are shown.

Six CasC subunits form the backbone of the Cascade complex. The proximal (inner side of the helix) and distal domains (outer side) of each CasC adopt a concave shape (Fig. 3). Except for CasC1, all CasC proximal domains contain long β-hairpins (Fig. 1d), which are flexible in the RNA-free strucures[18,19], and stabilized upon crRNA–Cascade assembly. The β-hairpin of each CasC extends up into the groove of its preceding CasC subunit, where it is locked by interactions of the β-hairpin residues D204^{CasC}, D205^{CasC} and L206^{CasC} with the surrounding groove residues (Fig. 3a, b and Extended Data Fig. 5a–c). Consistently, mutating D204-L206^{CasC} to Ala abolishes Cascade-mediated immunity

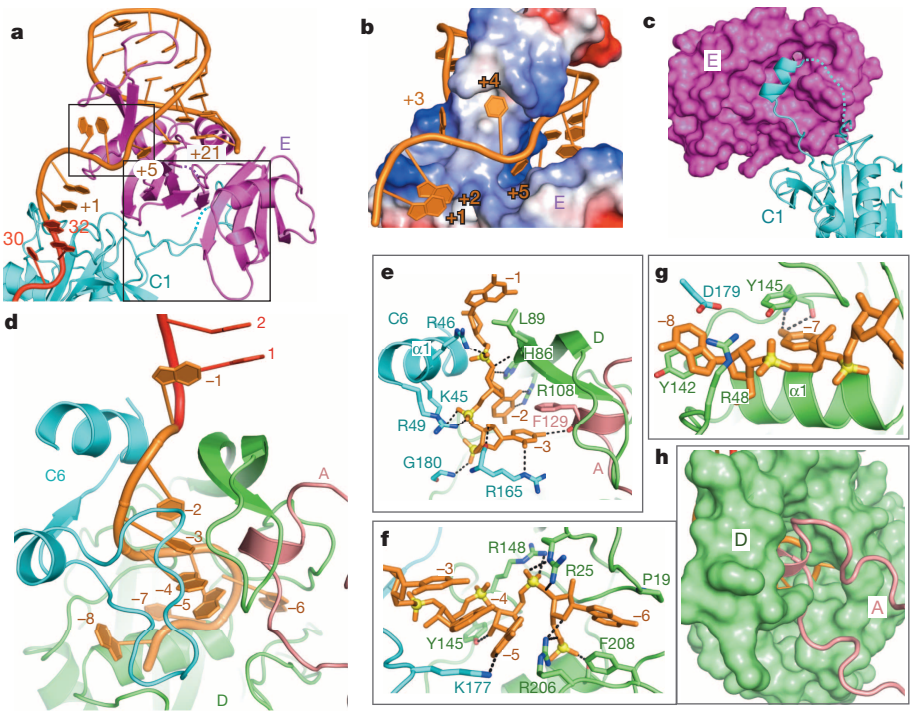


Figure 2 | The two ends of the crRNA are anchored. **a**, The 3′- repeat segment winds the β6–β7 hairpin of CasE. **b**, Positioning of non-helical residues (+1)-(+5). U(+5) is inserted into the cleft between two hairpins. U(+4) and G(+3) is inserted into the concave surface. A(+2) makes a break from G(+3) and stacks with G(+1). **c**, An α-helix from CasC1 inserts into a cleft of CasE. **d**, The 5′-end repeat is buried by CasD, CasC6 and CasA. **e**–**g**, Expanded views of the interactions between the Cas proteins and the 5′-repeat segment. **h**, A loop of CasA is inserted into the cavity of CasD.

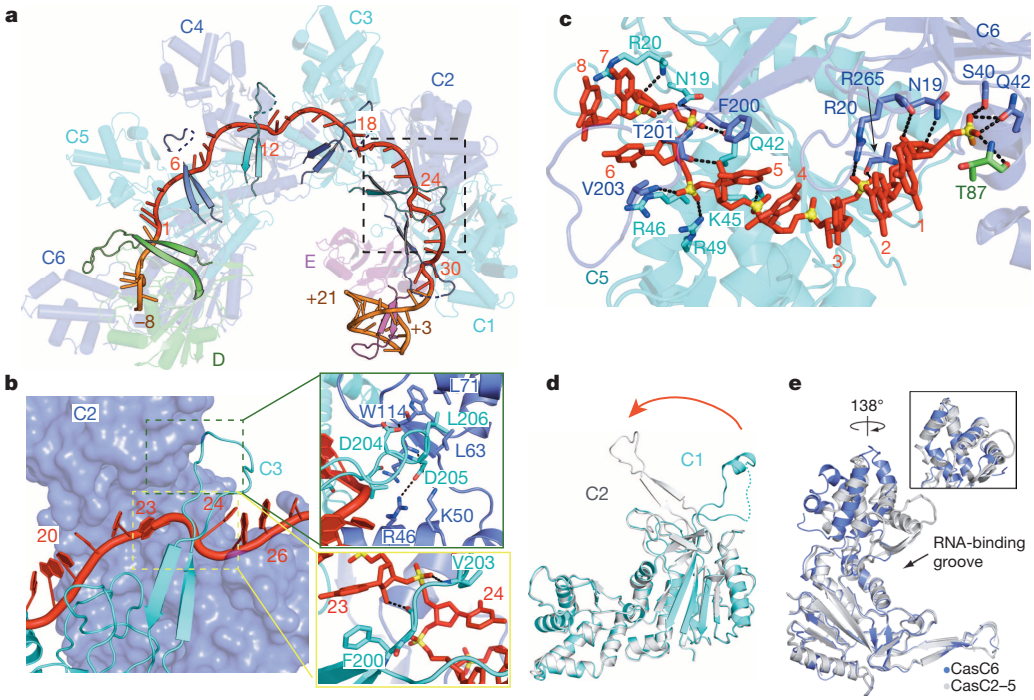


Figure 3 | The spacer fragment is positioned in the continuous groove on the concave surface of CasC1–6. **a**, The spacer fragment in red is not continuously stacked with five kinks present at regular intervals, namely at positions of the 6th, 12th, 18th, 24th and 30th nucleotide. Seven important β-hairpin arms from CasD, CasE and CasC2–6 subunits are highlighted. **b**, A β-hairpin of CasC3 inserts into the concave surface of CasC2. **c**, The interaction between the crRNA seed segment and CasC5–6. **d**, **e**, Structural comparison between CasC1 (cyan) and CasC2 (grey) (**d**), and between CasC5 (grey) and CasC6 (blue) (**e**).

(Extended Data Fig. 5d). Interestingly, because CasC6 is the last CasC subunit of the helical arrangement, it employs a CasD β-hairpin, even though CasD shares no similarity in overall folding with CasC.

Our structural analysis revealed an intriguing feature of a kink every 6 nucleotides within the crRNA spacer fragment. The 32-nucleotide spacer is embedded into a contiguous, positively charged groove formed by CasC1–6. The residues 1 to 5, 7 to 11, 13 to 17, 19 to 23 and 25 to 29 are continuously stacked. Intriguingly, these stacks are interrupted by five distinct breaks at steps 5–6, 11–12, 17–18, 23–24 and 29–30. The 6th, 12th, 18th, 24th and 30th residues flip out of the stacked helix, generating five regularly spaced kinks across the spacer, thus creating a discontinuous stacking arrangement of five similar fragments, each containing five stacked bases and one base splayed out (Fig. 3a, b). Interestingly, the kink positions precisely colocalize with those regions of the five β-hairpins projecting from CasC2–6 that cross the narrow gap at the break steps. Those two features are stabilized by the side-chains of $F200^{CasC}$ and $V203^{CasC}$ stacking with the respective two bases at each break site, and by hydrogen bonds formed between the β-hairpin and the crRNA phosphate backbone (Fig. 3b).

The 5-nucleotide stacked fragments between the kinks are positioned on the concave surface of the outer layer and are bound to their respective CasC motifs (Fig. 3a and Extended Data Fig. 2). Importantly, the 5-nucleotide segments are organized into an A-form helix, with the bases extending outwards, which is crucial for proper base pairing with target DNA. In contrast, the splayed out bases at the kink sites all face the distal domain of CasC, preventing any base pairing with target DNA, and most likely resulting in a non-continuous crRNA–DNA target duplex. Consistent with our structural finding, five site mismatches at the kink positions have a minimal effect on the target binding affinity, whereas mismatches positioned immediately upstream or downstream of the kink sites reduce the binding affinity by 20–30-fold (Extended Data Fig. 5e, f).

Residues 1–5 and 7–8, located within the first and second 5-nucleotide stacked segments of the spacer, are essential for target-DNA recognition[20–22]. In our structure, nucleotides 1–5, which are present on the surface of CasC6, are continuously stacked and extend outwards (Fig. 3c and Extended Data Fig. 6a). Residue 6 flips away from stack residues 1–5 by 180°, and therefore cannot pair with the DNA target, suggesting that it is not crucial for target recognition and explaining why this residue is not part of the seed sequence. Following this kink at position 6, the base edges of residues 7–8 face in a similar direction as residues 1–5. Residues 7–8 continuously stack with residues 9–11, thus pairing between residues 7–8 and the target DNA should promote the formation of the crRNA–DNA target duplex, thereby facilitating the propagation to the 3′ terminus of crRNA. Compared with the other 5-nucleotide stacked segments, partially obstructed by the CasB dimer, the first two segments containing the seed region are more accessible for the target strand. Moreover, the binding of proto-spacer adjacent motif (PAM)[23,24] to CasA allows for target recognition to start from the 5′-terminus of the spacer (Extended Data Fig. 6b–g). In conclusion, the fact that the seed bases face outwards towards solvent, together with the specific position near the tail end of the Cascade complex, allow for easy pairing with the DNA target, reminiscent of a short interfering RNA seed segment in Argonaute complexes[25].

Structural comparison of the six CasC subunits revealed that, unlike the long β-hairpin arms of other CasC subunits, the corresponding region of CasC1 forms an α-helix, rotates by ~90° (Fig. 3d), and extends to the cleft of CasE (Fig. 2c and Extended Data Fig. 3d). $F200^{CasC2–6}$ and $V203^{CasC2–6}$ stack with the nucleotides at the break site, whereas $F200^{CasC1}$ and $V203^{CasC1}$ form hydrophobic interactions with CasE, thereby promoting Cascade complex formation. As observed in previous cryo-electron microscopy structures[1,10], CasC6 adopts a different conformation from other CasC proteins. The rotation of the distal domain enables CasC6 to interact with CasD forming the 5′ repeat segment binding pocket, and the particular conformation of the β5–β6 hairpin arm of CasD prevents CasC6 from taking on the same conformation as the other CasC subunits (Fig. 3e and Extended Data Fig. 5b).

The inner layer of Cascade is comprised of CasA and CasB1–2 subunits. As mentioned above, one loop of CasA inserts into the concave surface of CasD, while the C-terminal domain of CasA contacts CasB2 and CasC5–6 (Fig. 4 and Extended Data Fig. 7a, b). The elongated dimer of CasB1–2, located along the inner surface of the crRNA–CasC spine, is not involved in crRNA binding. Instead, it interacts with CasA, CasC and CasD, thus facilitating the Cascade assembly. Due to the curved shape of each CasB subunit, the asymmetric CasB–CasB interaction turns the CasB dimer into a curved structure as well, which is important for the interaction of CasB with three continuous CasC subunits (CasC1–3 with CasB1 and CasC3–5 with CasB2, Figs 1c, 4 and Extended Data Fig. 7c–f). Binding of the inner layer to the CasC helix is mediated by five unique binding spots, each of which is formed by a negatively charged

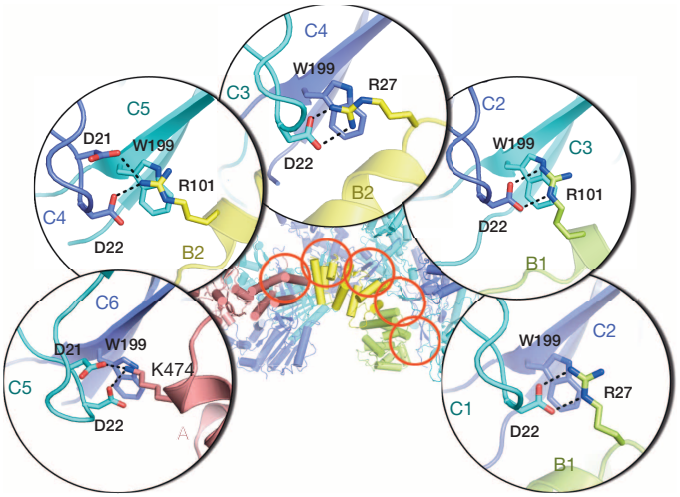


Figure 4 | Assembly of CasA and CasB to the six-CasC helix mediated by the unique Asp-Arg/Lys-Trp triad. The ribbon representation of the inner layer and the CasC helix is shown in the background. Each triad alignment is highlighted by red circle and the structural details are presented accordingly.

residue $D22^{CasC}$ that bridges to a positively charged residue (R27 or R101 of CasB1 or CasB2, as well as K474 of CasA), and is stabilized by stacking with the aromatic side chain of W199 of the following CasC molecule. Thus, the Asp-Arg/Lys-Trp triad unites two neighbouring CasC with either CasA or CasB (Fig. 4 and Extended Data Fig. 7).

Upon viral invasion, the Cascade complex binds double-stranded DNA exhibiting sequence complementary to the crRNA spacer sequence, and then recruits Cas3, a nuclease-helicase that cleaves the bound target DNA[24,26–30]. To understand the molecular basis for target recognition, we fitted each subunit structure into the cryo-electron microscopy map[10] and compared the structures with target bound and unbound. Overall, the outer layer displays negligible intrinsic motion, except that CasE rotates around the CasC1 binding site by ~15° in an anti-clockwise direction, moving the crRNA stem-loop upward as a result (Extended Data Fig. 8a–d). In contrast, the inner layer undergoes a large conformational change upon target binding. While CasB1 and CasB2 move along the CasC helix for ~10–12 Å, CasA rotates by ~25°, resulting in new interfaces of CasA–CasC and CasB–CasC (Extended Data Fig. 8c, e). Due to these conformational changes, the binding channel widens at the seed region, allowing the crRNA–target DNA duplex to fit in, and the motion of the CasB dimer stabilizing the target (Extended Data Fig. 8f, g).

In conclusion, our structural study has revealed that the crRNA plays an essential role not only in target recognition, but also in cascade complex assembly. Furthermore, the elongated β-hairpin loop from Cas protein represents a multifunctional element that serves a critical role in spacer RNA stabilization, CasC helix formation, and CasA–CasD as well as CasC–CasA and CasC–CasB interactions. Our structure also reveals a unique RNA-binding motif, in which a specific crRNA binding mode is used several times to assemble a multi-component complex.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 June; accepted 5 August 2014.
Published online 12 August 2014.

1. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482,** 331–338 (2012).
2. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
3. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).
4. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320,** 1047–1050 (2008).
5. Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–1575 (2002).
6. 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).
7. Westra, E. R. *et al.* The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu. Rev. Genet.* **46,** 311–339 (2012).
8. van der Oost, J., Westra, E. R., Jackson, R. N. & Wiedenheft, B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature Rev. Microbiol.* **12,** 479–492 (2014).
9. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. The basic building blocks and evolution of CRISPR-cas systems. *Biochem. Soc. Trans.* **41,** 1392–1400 (2013).
10. Wiedenheft, B. *et al.* Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477,** 486–489 (2011).
11. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Struct. Mol. Biol.* **18,** 529–536 (2011).
12. Reeks, J., Naismith, J. H. & White, M. F. CRISPR interference: a structural perspective. *Biochem. J.* **453,** 155–166 (2013).
13. Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321,** 960–964 (2008).
14. Sashital, D. G., Jinek, M. & Doudna, J. A. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nature Struct. Mol. Biol.* **18,** 680–687 (2011).
15. 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).
16. 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).
17. Nam, K. H. *et al.* Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I–C/Dvulg CRISPR-Cas system. *Structure* **20,** 1574–1584 (2012).
18. Lintner, N. G. *et al.* Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem.* **286,** 21643–21656 (2011).
19. Hrle, A. *et al.* Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3. *RNA Biol.* **10,** 1670–1678 (2013).
20. 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).
21. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA* **108,** 10098–10103 (2011).
22. Künne, T., Swarts, D. C. & Brouns, S. J. Planting the seed: target recognition of short guide RNAs. *Trends Microbiol.* **22,** 74–83 (2014).
23. Sashital, D. G., Wiedenheft, B. & Doudna, J. A. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell* **46,** 606–615 (2012).
24. Hochstrasser, M. L. *et al.* CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl Acad. Sci. USA* **111,** 6618–6623 (2014).
25. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456,** 209–213 (2008).
26. Westra, E. R. *et al.* CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* **46,** 595–605 (2012).
27. Mulepati, S. & Bailey, S. *In vitro* reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J. Biol. Chem.* **288,** 22184–22192 (2013).
28. Sinkunas, T. *et al. In vitro* reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.* **32,** 385–394 (2013).
29. Jackson, R. N., Lavin, M., Carter, J. & Wiedenheft, B. Fitting CRISPR-associated Cas3 into the helicase family tree. *Curr. Opin. Struct. Biol.* **24,** 106–114 (2014).
30. Sinkunas, T. *et al.* Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30,** 1335–1342 (2011).

Acknowledgements We thank the staff at beamline BL-17U at Shanghai Synchrotron Radiation Facility (SSRF), and beamlines BL-1A, BL-5A and BL-17A at Photon Factory. The research was funded by Chinese Ministry of Science and Technology (2014CB910102 and 2011CBA01105), the Natural Science Foundation of China (31222014 and 31170705), and the Strategic Priority Research program of the Chinese Academy of Sciences (XDB08010203) to Y.W., and was supported by the Research Startup Fund from South University of Science and Technology of China and Shenzhen Government to Z.W. We thank D. Patel for assistance with manuscript editing, H. Wang, H. Li and F. Sun for experimental assistance, and T. Juelich for critical reading and editing of our manuscript.

Author Contributions H.Z., G.S. and M.W. expressed and purified and grew crystals of the Cascade complex. H.Z. and Y.W. collected x-ray diffraction data, J.W. made all constructs and did biochemical assays and Z.W., Y.W., G.B., H.Z. and W.G. solved the Cascade complex structure, Y.W. and Z.W. wrote the paper. All studies were undertaken under the supervision of Y.W.

Author Information Coordinates and structure factors have been deposited with Protein Data Bank accession code 4U7U. 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 the paper. Correspondence and requests for materials should be addressed to Y.W. (ylwang@ibp.ac.cn) or Z.W. (wei.zy@sustc.edu.cn).

METHODS

Cloning, expression and purification of the Cascade complex. The five Cas genes, *CasA*, *CasB*, *CasC*, *CasD* and *CasE*, were amplified from *E. coli* K12 genomic DNA by polymerase chain reaction (PCR) and subcloned into different expression vectors. *CasA* was inserted into pET22b (Novagen, Amp^R). *CasB* was inserted into pRSFDuet-1 (Novagen, Kan^R) containing an N-terminal His_6-tag. *CasC*, *CasD* and *CasE* were inserted into a modified pCDFDuet-1 (Novagen, Str^R). Template CRISPR was chemically synthesized from Sangon Biotech (sequences provided below) and subcloned into pACYCDuet-1 (Novagen, Chl^R). Generation of the CasD and CasC point mutants was performed by site-directed mutagenesis. All constructs were confirmed by sequencing. All four vectors were then co-transformed into the *E. coli* BL21 (DE3). Overexpression of Cascade was induced with 0.2 mM isopropyl-β-D-thiogalactoside (IPTG) at an $OD_{600\,nm}$ of 0.4. After growing for 20 h at 16 °C, cells were harvested by centrifugation, homogenized in 1× PBS buffer A (pH 7.3) containing 5% glycerol. After sonication and centrifugation, supernatants were incubated with Ni Sepharose resin (GE Healthcare) for 2 h and eluted with an increasing imidazole gradient in buffer A. Then, the recombinant complex was further purified by Heparin HP column (GE Healthcare) and MonoQ 5/50 (GE Healthcare) with buffer B (20 mM HEPES, pH 7.5, 100 mM NaCl) and buffer C (20 mM HEPES, pH 7.5, 1 M NaCl). For crystallization trials, the protein complex was purified by gel filtration chromatography (Superdex 200 10/300 GL, GE Healthcare) in buffer D (20 mM HEPES, pH 7.5, 10 mM DTT, 1 mM EDTA). The peak fractions were collected and concentrated to about 4 mg ml^{-1} for crystallization. The entire purification procedure was performed at 4 °C and purity of the Cascade complex was verified by SDS–PAGE.

Crystallization and data collection of Cascade complex. Crystals of the Cascade complex were grown at 16 °C using the hanging drop vapour diffusion method with a reservoir solution containing 100 mM Tris-HCl, pH 8.5, 200 mM NaAc, 100 mM Glycine, 50 mM LiAc and 11–13% (w/v) polyethylene glycol (PEG) 4,000. Crystals appeared after 4 days and grew to full size within 2 weeks. For data collection, crystals were transferred into mother liquor supplemented with 20% (v/v) glycerol and flash-cooled in liquid nitrogen. Diffraction data were measured and collected at 100 K with wavelength of 1.1 Å at beamline BL1A, BL5A and BL17A of Photon Factory (Japan, KEK) or at beamline BL17U of Shanghai Synchrotron Radiation Facility (SSRF). All data sets were processed using HKL2000 or iMosflm[31,32]. The crystals were identified as belonging to space group *P*1, with two Cascade complexes per asymmetric unit.

Structure determination and refinement of the Cascade complex. By using the 8.8-Å electron microscopy map of the target-unbound Cascade complex as a search model, we found an optimal solution for molecular replacement in PHASER[33,34]. To improve the phase the non-crystallographic symmetries (NCS) for two Cascade complexes and for five CasC molecules were identified and applied in density modification process using RESOLVE[35]. The phase was dramatically improved with the powerful NCS average and then was gradually extended to 3.5 Å (Extended Data Fig. 9). The structural model was manually built and refined in Phenix[36] against the 3.05 Å data set. COOT[37] was used for model building and adjusting. The final model was validated using MolProbity[38] and Procheck[39]. The refinement statistics are listed in Extended Data Table 1. All structure figures were prepared using PyMOL (http://www.pymol.org/). Individual Cas proteins are colour-coded. The repeat segments of crRNA are in orange and the spacer is in red.

Cascade mutagenesis and gel filtration assay. Two types of Cascade mutants, CasCΔ and CasDΔ, were produced. In CasCΔ, residues Phe 200, Asp 204, Asp 205 and Leu 206 of CasC were mutated to Ala. In CasDΔ, residues 75–104 of CasD were replaced by a $(GGS)_4$ linker. CasCΔ and CasDΔ were cloned and expressed using the same strategy as used for wild-type Cascade. Both mutants were purified using Ni-affinity resin and analysed by gel filtration assay (Superdex200 10/300 GL, GE Healthcare). The assays were performed in a buffer containing 20 mM HEPES pH 7.5 and 100 mM NaCl. The relevant fractions were then analysed by SDS–PAGE and visualized by Coomassie blue staining.

pACYC-target invasion assay. For our invasion assay, we used a CRISPR system containing a 32-nucleotide spacer: 5′-GGCTCCCTGTCGGTTGTAATTGATAATGTTGA-3′. The complementary sequence of the spacer (5′-TCAACATTATCAATTACAACCGACAGGGAGCC-3′) followed by a PAM sequence (ATG) was constructed into pACYC-1 between the restriction sites NdeI and XhoI, yielding plasmids that contain an artificial protospacer target (pACYC-target). *E. coli* BL21 (DE3) cells that contain CRISPR-encoding plasmids with either the wild-type or mutant *Cascade* genes together with Cas3 were inoculated in LB containing kanamycin (50 μg ml^{-1}), ampicillin (100 μg ml^{-1}) and streptomycin (50 μg ml^{-1}), and grown to an $OD_{600\,nm}$ of 0.3. Expression of both Cas genes and CRISPR was induced for 45 min with 0.5 mM IPTG. Cells were then collected at 4 °C and made competent according to standard protocols. Transformation was performed by adding either 60 ng pACYC-target or 60 ng pACYC.

Next, cells were grown in LB for 60 min at 37 °C before plating on LB-agar plates containing 0.5 mM IPTG, ampicillin (100 μg ml^{-1}), kanamycin (50 μg ml^{-1}), streptomycin (50 μg ml^{-1}) and chloramphenicol (34 μg ml^{-1}) . Plates were incubated for 12 h at 37 °C before observation.

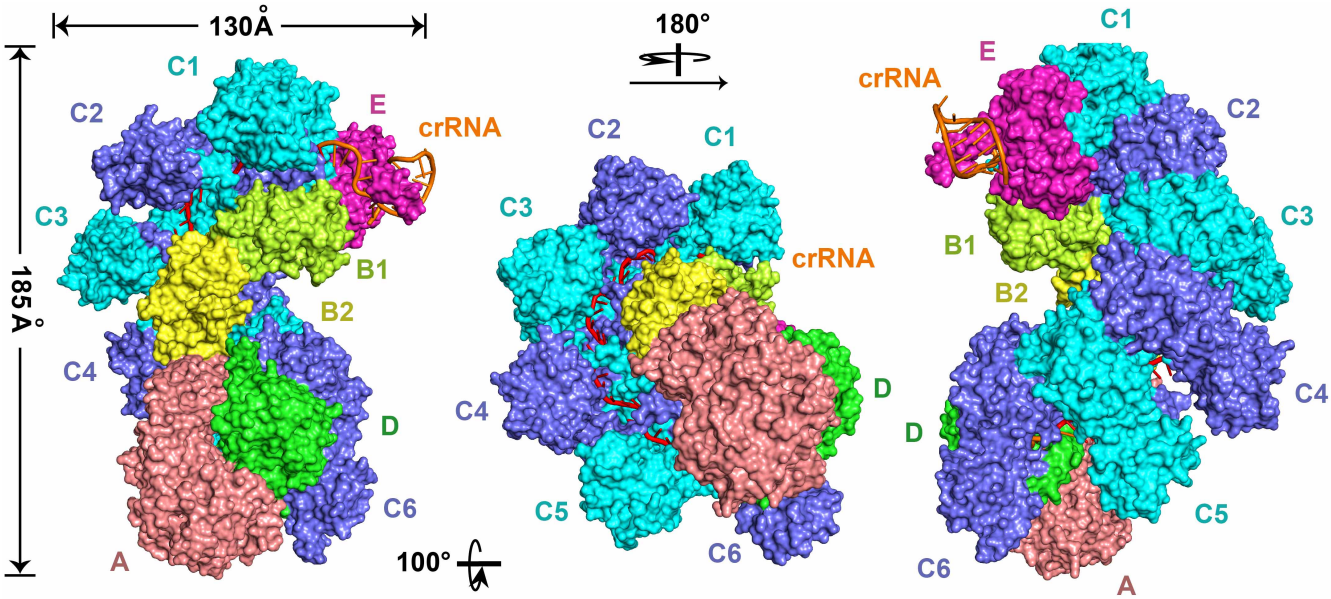
The pACYC control without target sequence did not trigger CRISPR/Cas interference, and was therefore retained, with no inhibition of growth observed. In contrast, the plasmid pACYC-target containing target sequence was recognized and degraded by CRISPR/Cas system, growth inhibition observed on selection medium.

Plasmid loss assay. *E. coli* BL21 (DE3) strains expressing either wild-type Cascade (WT) or Cascade complex in which the hairpin-arm of CasD was replaced with a $(GGS)_4$ linker (DΔ), were transformed with pACYC-target plasmid. Cultures were inoculated in LB containing kanamycin (50 μg ml^{-1}), ampicillin (100 μg ml^{-1}), streptomycin (50 μg ml^{-1}) and chloramphenicol (34 μg ml^{-1}), and grown to an $OD_{600\,nm}$ of 0.6. Expression of *cas* genes and CRISPR was induced for 5 h with 0.5 mM IPTG. Cells were serially diluted and plated on LB-agar non-selective plates (containing 50 μg ml^{-1} kanamycin, 100 μg ml^{-1} ampicillin and 50 μg ml^{-1} streptomycin), or selective plates containing additional 34 μg ml^{-1} chloramphenicol. All plates were incubated for 12 h at 37 °C before observation. If the plasmid can trigger CRISPR/Cas interference, it will be recognized and degraded. As a consequence, growth will be inhibited on selection medium. Alternatively, the plasmid will remain intact, and colonies will appear on selection medium.

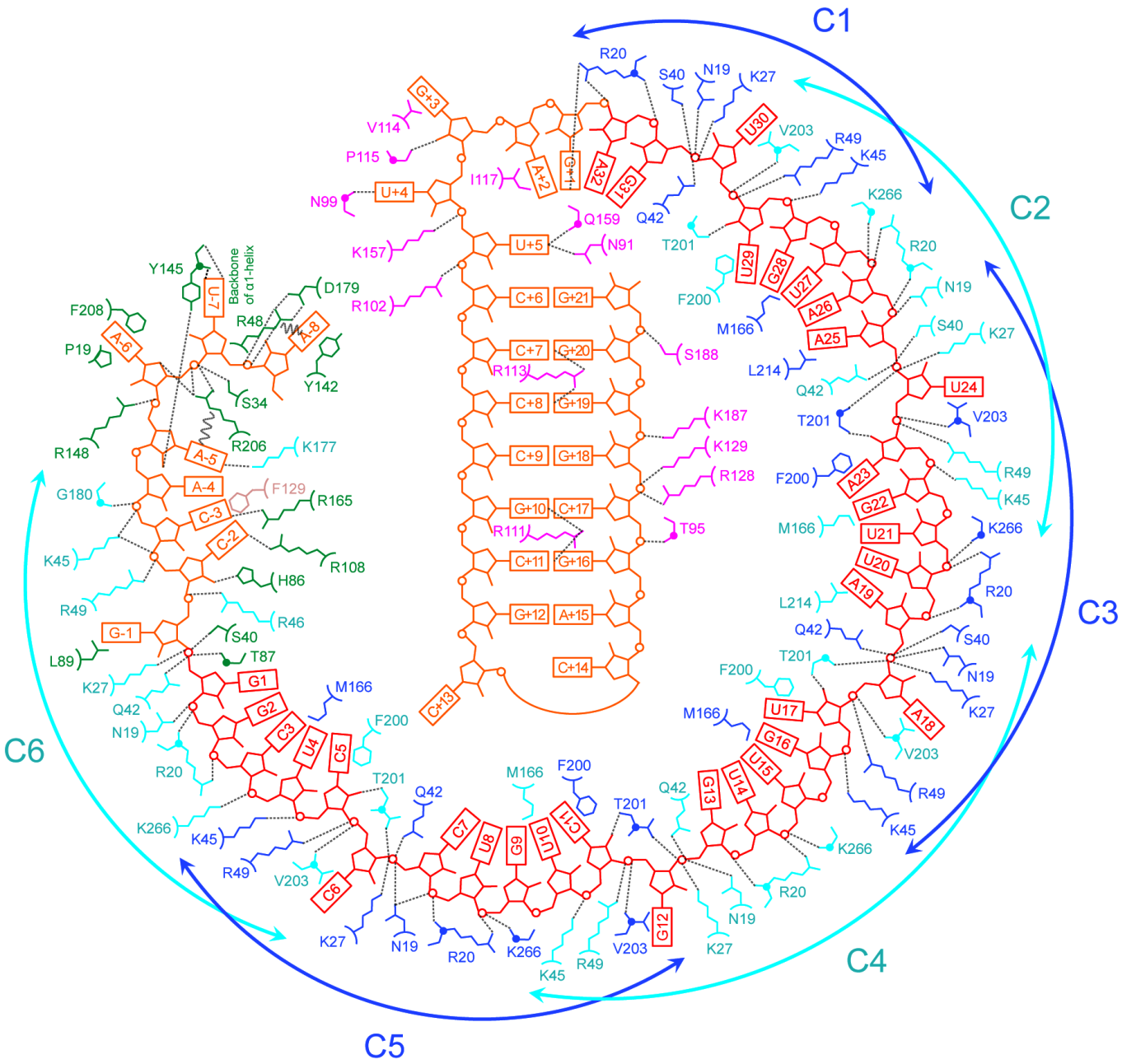
Isothermal titration calorimetry (ITC). DNA oligonucleotides for ITC were dissolved in buffer containing 20 mM HEPES, pH 7.5, and 100 mM NaCl. The Cascade complex for ITC measurements was purified by Superdex 200 with HEPES buffer (20 mM, pH 7.5) containing 100 mM NaCl. The purified Cascade complex (1 μM) was loaded into the cell, and the ssDNA solution (10 μM) in the syringe. Experiments were carried out at 20 °C using a MicroCal ITC200 calorimeter (GE Healthcare). The titration data were processed and fitted using Origin 7.0 software.

Template CRISPR sequence (from 5′ to 3′). GAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGAGTTCCCCGCGCCAGCGGGGATAAACCG.

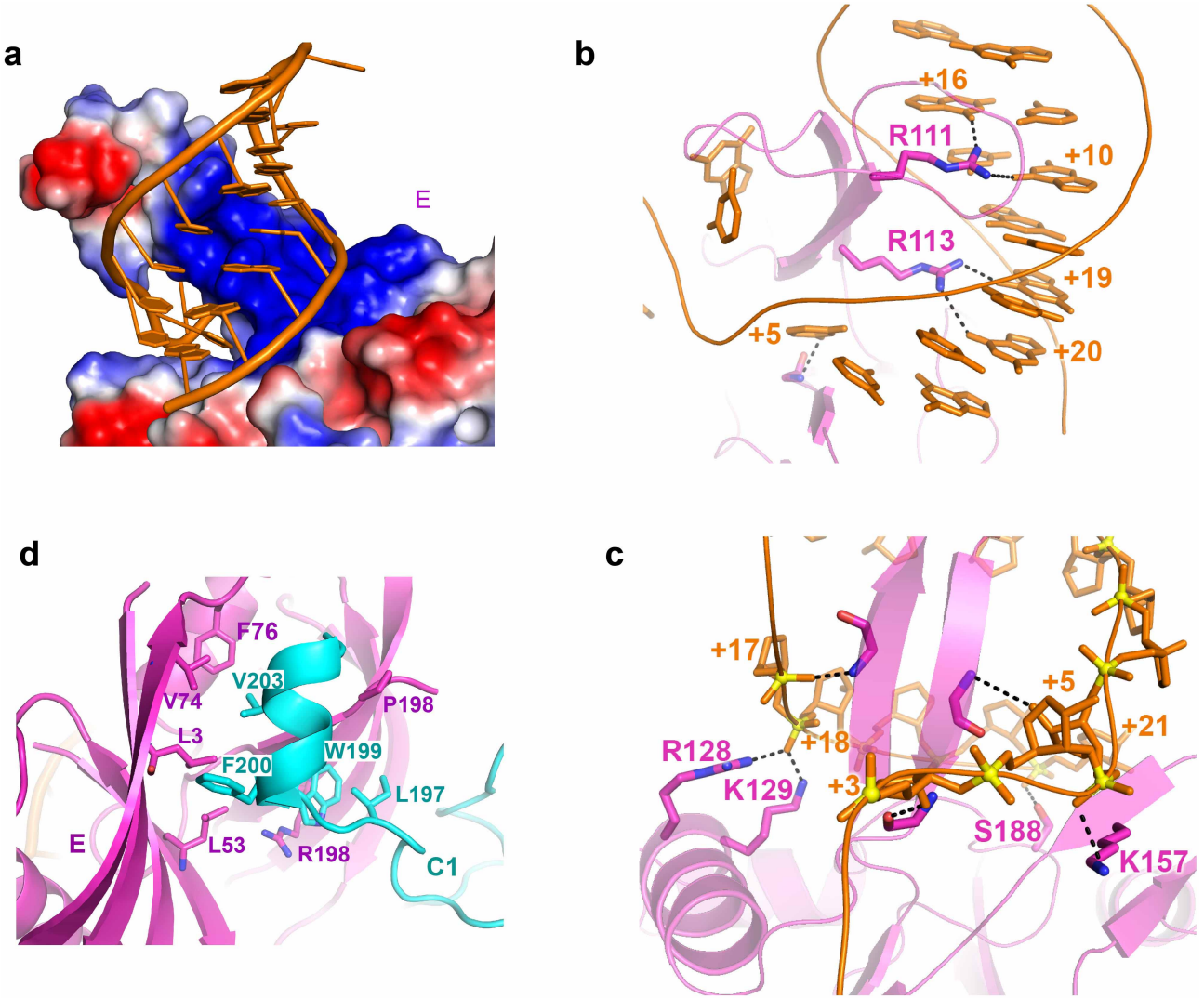
31. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D* **67,** 271–281 (2011).
32. Otwinowski, Z. & Minor, W. Processing of X-Ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
33. Xiong, Y. From electron microscopy to X-ray crystallography: molecular-replacement case studies. *Acta Crystallogr. D* **64,** 76–82 (2008).
34. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
35. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56,** 965–972 (2000).
36. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
38. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
39. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283–291 (1993).



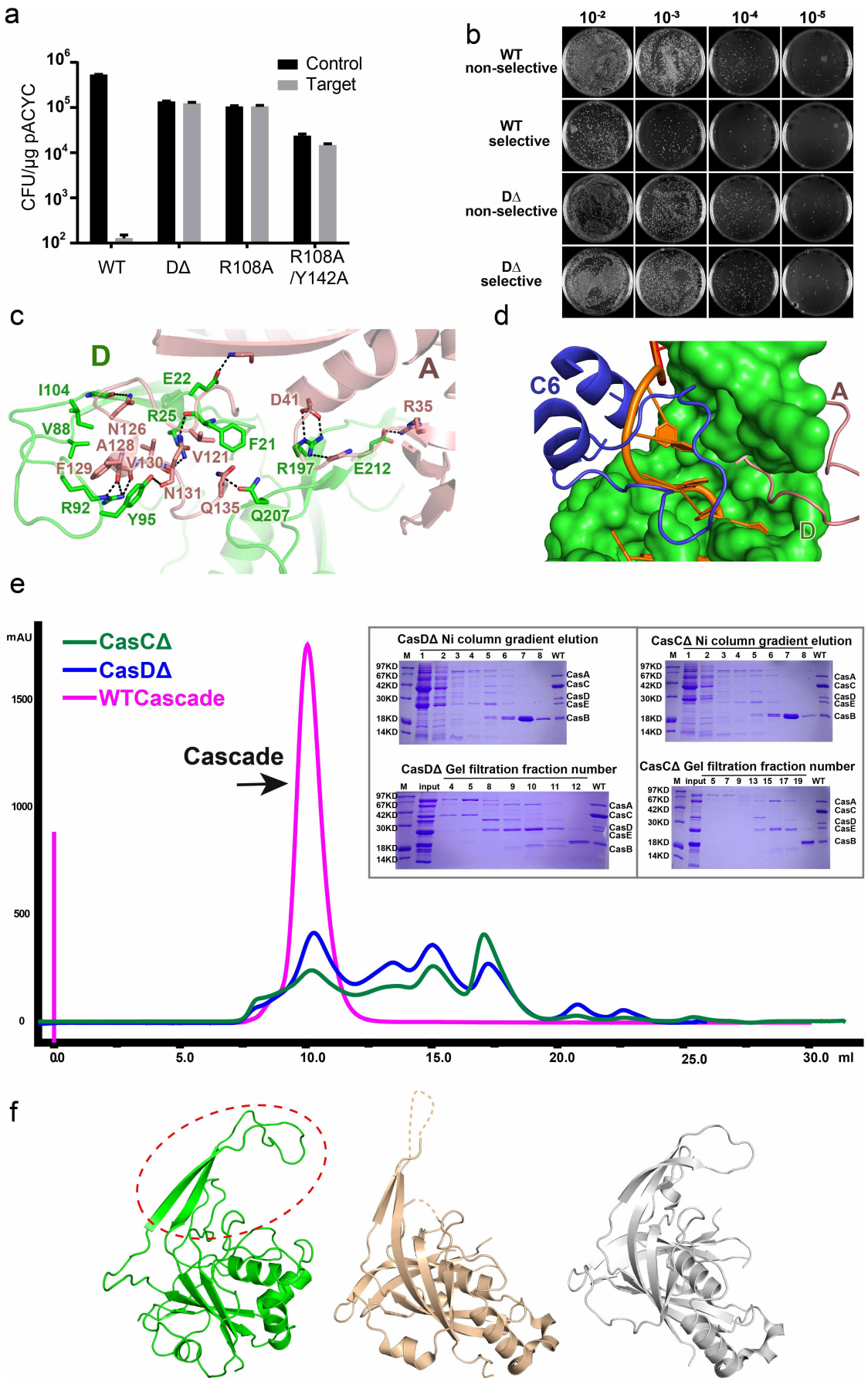
Extended Data Figure 1 | Overall surface view of the Cascade complex with the same orientations as those used for Fig. 1b.



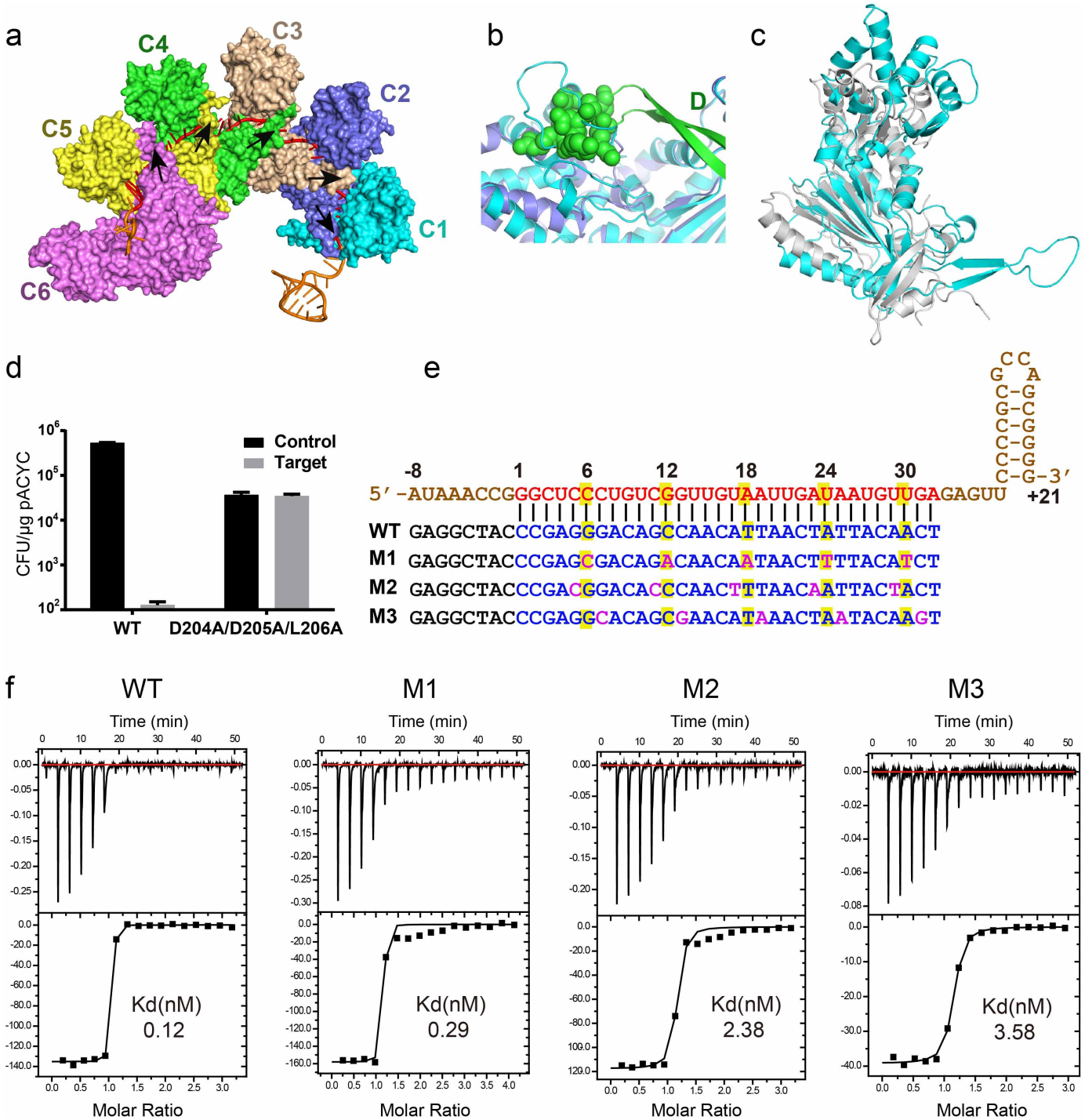
Extended Data Figure 2 | Schematic summary of Cas-protein–crRNA interactions. Hydrogen bonds and salt bridges are indicated by dashed lines. Cation-π interactions are indicated by wavy lines.



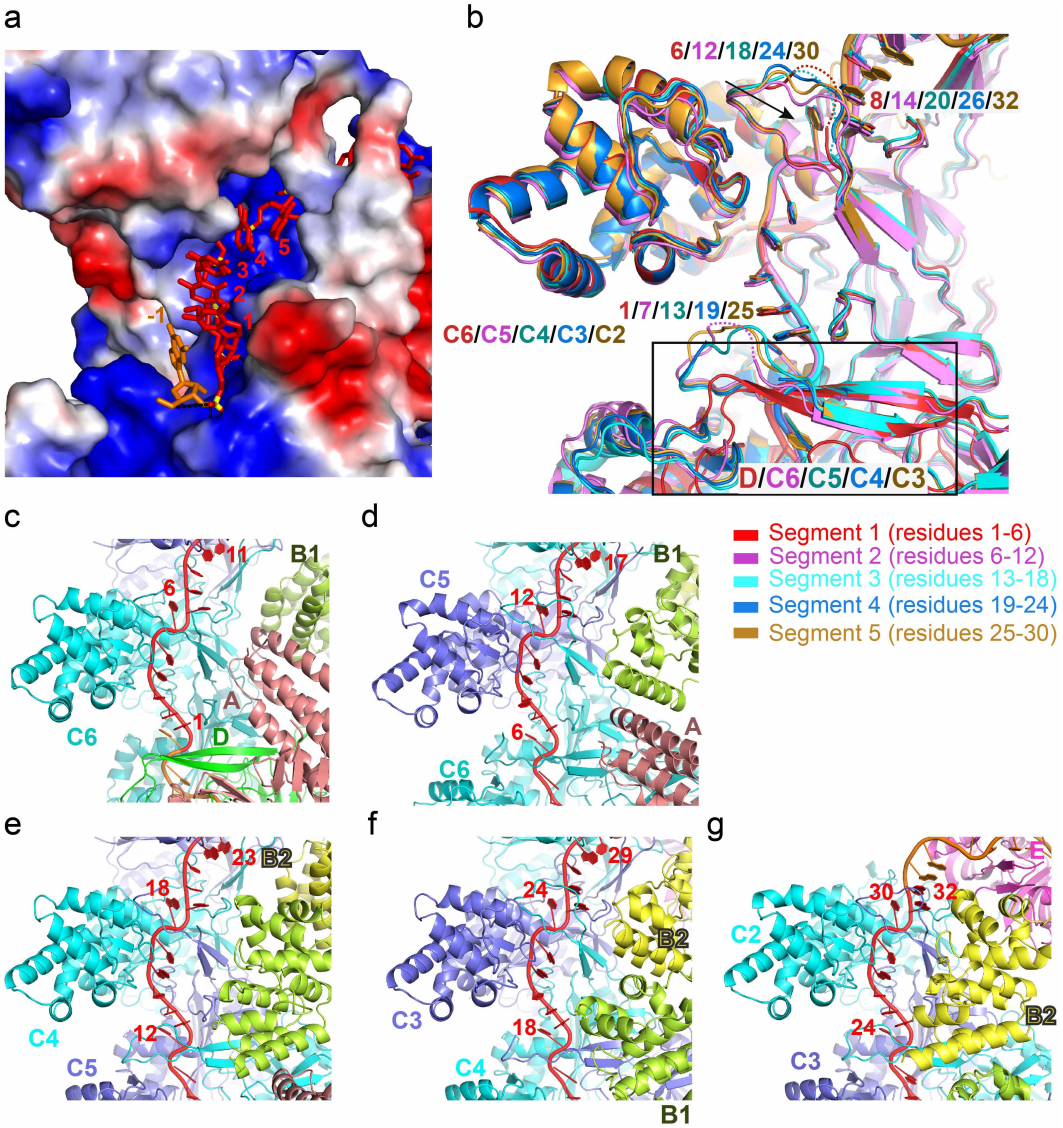
Extended Data Figure 3 | Structure of CasE subunit in the Cascade complex. **a**, CasE is shown as a surface representation, and is labelled according to electrostatic potential (red, negative charge; blue, positive charge), and RNA is shown in ribbon representation (orange). **b**, **c**, Magnified view of the sequence-specific interaction between CasE and the major groove of 3′-end repeat (**b**), and the interaction between CasE and 3′-end repeat backbone (**c**). **d**, Expanded view of the interaction between CasE and CasC1.



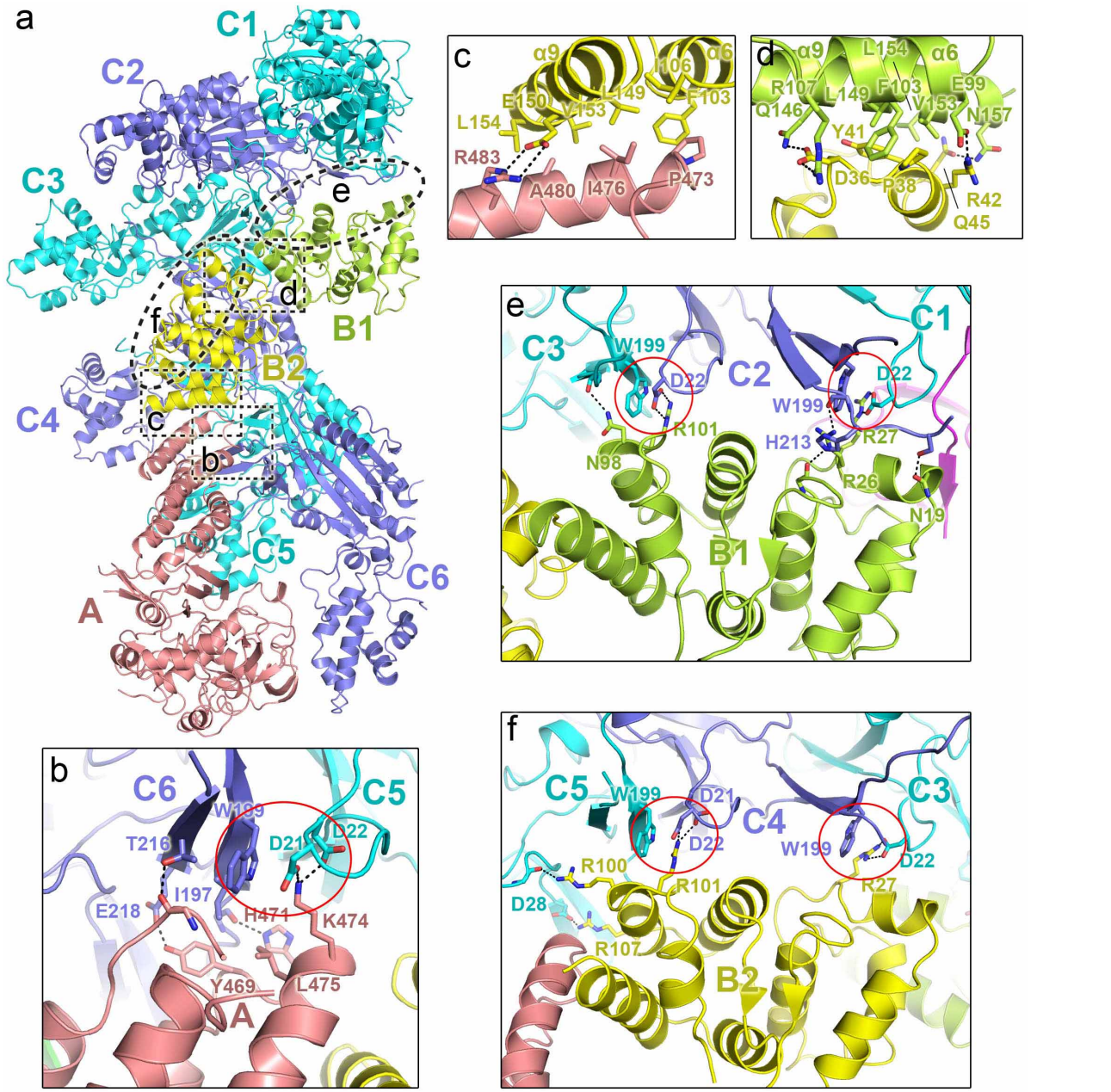
Extended Data Figure 4 | The β-hairpin arm of CasD is critical for Cascade function. a, pACYC-target invasion assay. Competent *E. coli* BL21 (DE3) cells expressing either wild type Cascade (WT) or a mutant form in which the hairpin-arm was replaced with a $(GGS)_4$ linker CasD mutant (DΔ), or R108A, R108A/Y142A CasD mutants, were transformed with either pACYC plasmid (Control) or pACYC-target plasmids (Target). Colony-forming units per microgram pACYC (CFU per μg) are depicted for each of the strains. The statistics was based on 10 replicates. Error bars represent the standard error of the mean (s.e.m.). Upon transformation with pACYC plasmid control, *E. coli* cells expressing WT Cascade exhibited much higher transformation efficiencies than cells transformed with pACYC-target plasmid. However, the transformation efficiencies were high in *E. coli* cells expressing mutant CasD and transformed with either pACYC control or pACYC-target. These results show that the hairpin-arm of CasD is essential for proper function of the Cascade complex. **b**, Plasmid loss assay. *E. coli* BL21 (DE3) cells expressing wild-type or casDΔ(75–104)-replaced Cascade complex and pACYC-target, were serially diluted and grown on either non-selective or selective plates upon induction with 0.5 mM IPTG. Growth of BL21 cells expressing WT Cascade was seriously inhibited on selection medium compared with growth on non-selection medium. However, BL21 cells with mutant CasD displayed similar growth on selection and non-selection medium, indicating that the target sequence was retained upon introduction of the CasD mutant. **c**, Interface between CasD (green) and CasA (salmon). **d**, Interface between CasD (green) and CasC6 (blue). **e**, Pull-down and gel-filtration analysis of the Cascade assembly. WT Cascade complex was formed effectively in the Ni-NTA column and was eluted out with elution buffer containing 100 mM imidazole. In contrast, very limited amounts of the co-expressed Cascade complex with the DΔ mutant or CΔ mutant were obtained in the 100 mM imidazole elution, with the majority of the Cas proteins were in the flow-through. The fractions eluted with 100 mM imidazole were collected and applied for the analytical gel filtration assay. Compared with the WT Cascade, both DΔ mutant and CΔ mutant Cascade did not form a stable complex in gel filtration, further confirming the critical role of the β-hairpin arm of CasD and CasC for the assembly of the Cascade complex. The CasC mutant (CΔ) contains F200A mutation as well as replacement mutation, where residues D204 to L206 were replaced by Ala. **f**, Structural comparison among *E. coli* CasD (green) and Cas5d in the RNA-free state of *Streptococcus pyogenes* (wheat, PDB id: 3VZH) and *Bacillus halodurans* (grey, PDB id: 4F3M). The β-hairpin arm of CasD is highlighted by a red dashed circle.



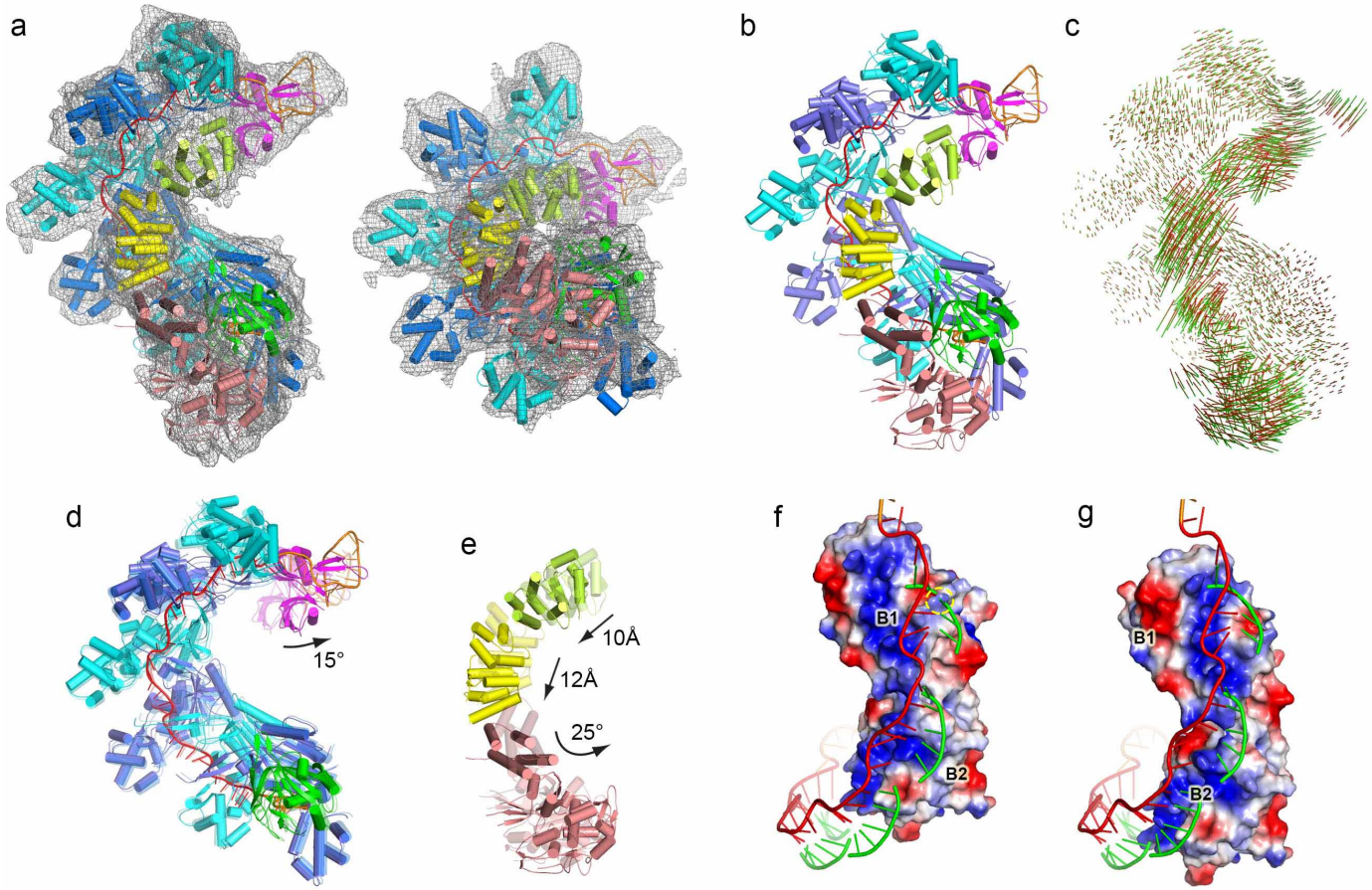
Extended Data Figure 5 | The unique multi-kinked conformation within the crRNA spacer region. **a**, The six CasC subunits form a right-handed helix, with a groove for crRNA binding. Except for CasC1, the β-hairpin (highlighted by a black arrow) of each CasC molecule extends up into the groove of its preceding CasC, and thus results in five kinks observed for the crRNA spacer. **b**, Superposition of CasC5 (cyan) and CasC6 (blue). The top of long β-hairpin in CasD is shown as green spheres, depicting the clash between the distal domain of CasC6 and the long β-hairpin top given it adopts the same conformation as CasC5. **c**, CasC (cyan) shares a similar overall fold with *Sulfolobus solfataricus* Cas7 (grey, PDB id 3PS0). **d**, The pACYC-target invasion assay showing that mutating three interacting residues ($D204^{CasC}$, $D205^{CasC}$, $L206^{CasC}$) at the β-hairpin arm largely impairs Cascade activity. **e**, Four designed DNA targets with complementary sequence (WT) to the spacer segment of crRNA or with non-complementary mutations at kinked sites (M1), at immediately upstream nucleotides (M2), and at immediately downstream nucleotides (M3). The five kinked sites are highlighted as yellow background, with the mutated sequences in purple. **f**, ITC-based analysis of the Cascade complex and four DNA targets interactions. ITC analysis showed that the mutations of the flipping of these residues had no obvious effect on target recognition. However, mutations in either the first or the last nucleotides in the 5-nucleotide stacked region largely affected target binding.



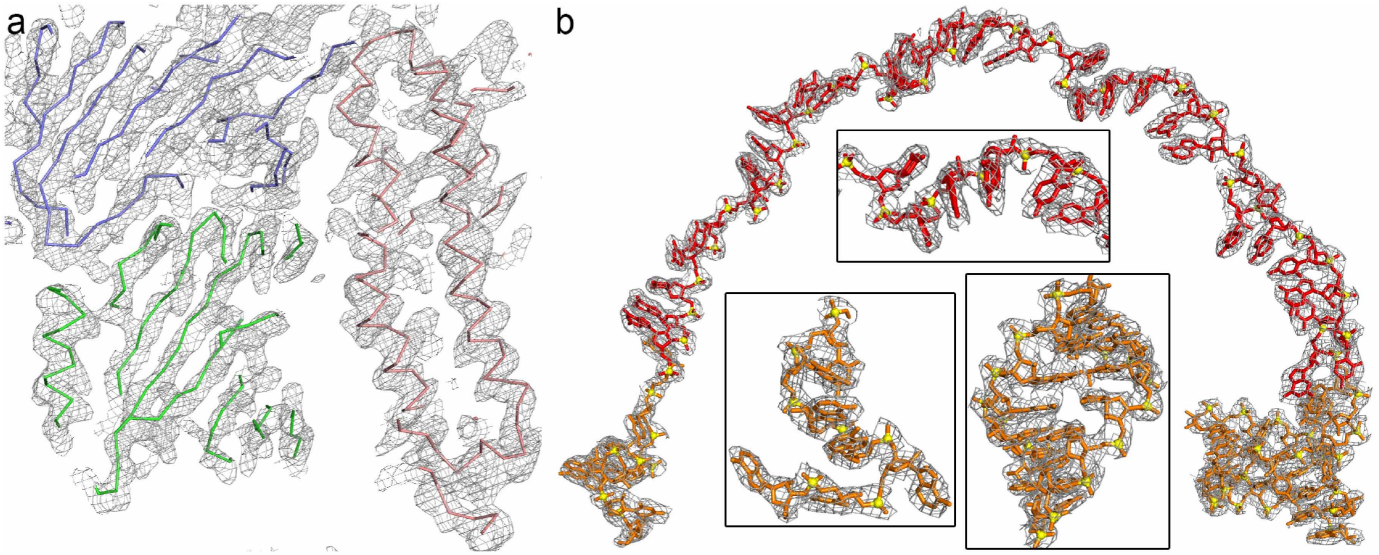
Extended Data Figure 6 | Structural analysis of the crRNA spacer segments. **a**, The first 5-nucleotide segment in the spacer region is positioned in the positively charged groove of CasC. **b**, Overlap of the five 5-nucleotide segments within the crRNA space region and their interacting Cas proteins. The 5-nucleotide segments are interrupted by single flipped-out bases, which are indicated by a black arrow, thereby resulting in kinks at these positions. The conformations of the five segments together with the kink sites are essentially identical. Like those of CasC2–6, the β-hairpin arm of CasD is involved in maintaining the kink at G(−1). Interestingly, although the overall foldings are distinct between CasD and CasC, their β-hairpin arms share similar conformations, which are highlighted by a black box. **c**–**g**, Structural comparison of the five 5-nucleotide stacked segments with Cas proteins shown in the same orientation. The Cascade complex is colour-coded by chains as shown in Fig. 1.



Extended Data Figure 7 | Inter-subunit interactions required for Cascade assembly. **a**, Ribbon representation of the organization of CasA and CasB dimers and their associations with the CasC helix. The inter-subunit interfaces are highlighted with dashed-line boxes. **b**–**f**, The structural details of the CasA–CasC (**b**), CasA–CasB2 (**c**), CasB1–CasB2 (**d**), CasB1–CasC (**e**) and CasB2–CasC (**f**) interactions. The five triads, which connect CasA and the CasB dimer to the CasC helix, are highlighted with red circles. Hydrogen bonds and salt bridges are indicated by dashed lines.



Extended Data Figure 8 | Structural comparison between target-bound and unbound Cascade. **a**, The building of the structural model of the target-bound Cascade. The subunits were individually docked into the 9Å-electron microscopy map. The fitted model is shown in two views with the electron microscopy density overlapping (contoured at 2σ). **b**, Ribbon representation of the target-bound Cascade. **c**, Subunit motions. The movement was represented by red-green lines, which were drawn by connecting each Cα atom (green) in the target-bound Cascade to corresponding Cα atom (red) in the unbound Cascade after the two Cascade structures were superimposed. Thus, the lengths of the lines correlate with the motion scale. The motions of the outer and inner layers are shown by superimposing the target-bound with unbound Cascade structure in **d** and **e**, respectively. The target-free structures are rendered semi-transparent. The translational and rotational motions are indicated by arrows showing the motional direction with translation distances and rotation angles labelled, respectively. **f**, **g**, Comparison of electrostatic surface potentials for the CasB dimer in the target-free (**f**) and target-bound Cascade (**g**) with cartoon representations of the crRNA spacer and its paring target (green). A highly positively charged groove of the CasB dimer fits well with the negatively charged target backbone in **g**, but not in **f**. Also, the paring target partially clashes with the CasB dimer in **f**, with the clashing site highlighted by a yellow circle. The structural analysis suggests that the motion of the CasB dimer identified in the target-bound Cascade complex is required for proper target binding.



Extended Data Figure 9 | Electron density maps showing the model quality. **a**, The 3.5 Å-electron density map contoured at 1.5σ generated from the phase after density modification treatment with the final structural model superimposed. Cα backbones are shown in with the same colour-coding as used in Fig. 1. **b**, The $2F_o - F_c$ map contoured at 1σ of the whole crRNA is shown with the refined RNA model superimposed. Zoom-in views of the 5′-end, 3′-end and the 5-nucleotide stacked segment are shown in the lower left, lower right and upper insets, respectively.

Extended Data Table 1 | Statistics of data collection and refinement

	Cascade
Data collection	
Space group	P1
Cell dimensions	
a, b, c (Å)	111.41, 118.14, 225.87
A, β, γ (˚)	92.24, 93.55, 106.06
Wavelength (Å)	1.1
Resolution (Å)	50-3.05 (3.16-3.05)
R_{merge} (%)	9.1 (40.5)
I/sigma	11.7 (2.1)
Completeness (%)	98.8 (98.3)
Redundancy	3.05 (2.9)
Refinement	
Resolution (Å)	45-3.05
No. reflections	195,250
R_{work} / R_{free} (%)	16.4/20.7
No. atoms	
Protein	53257
RNA	2596
Average B (Å^2)	
Protein	43.4
RNA	45.6
R.m.s.deviations	
Bonds (Å)	0.006
Angle (˚)	1.2
Ramachandran statistics (%)	
Most favorable	89.8
Additionally allowed	9.8
Generously allowed	0.3
Disallowed	0.1

Values in parentheses are for the highest resolution shell.