Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation

Yanwu Huo1,5, Ki Hyun Nam1,5,6, Fang Ding1,5, Heejin Lee1,5, Lijie Wu2, Yibei Xiao1, M Daniel Farchione Jr1, Sharleen Zhou3, Kanagalaghatta Rajashankar4, Igor Kurinov4, Rongguang Zhang2 & Ailong Ke1

CRISPR drives prokaryotic adaptation to invasive nucleic acids such as phages and plasmids, using an RNA-mediated interference mechanism. Interference in type I CRISPR-Cas systems requires a targeting Cascade complex and a degradation machine, Cas3, which contains both nuclease and helicase activities. Here we report the crystal structures of Thermotoga maritima Cas3 bound to single-stranded (ss) DNA substrate and show that it is an obligate 3'→5' ssDNase that preferentially accepts substrate directly from the helicase moiety. Conserved residues in the HD-type nuclease coordinate two iron ions for ssDNA cleavage. We demonstrate ATP coordination and conformational flexibility of the SF2-type helicase domain. Cas3 is specifically guided toward Cascade-bound target DNA by a PAM sequence, through physical interactions with both the nontarget substrate strand and the CasA protein. The sequence of recognition events ensures well-controlled DNA targeting and degradation of foreign DNA by Cascade and Cas3.

Clustered regularly interspaced palindromic repeats (CRISPR) drives adaptation to invasive nucleic acids such as phages, conjugative plasmids and transposable elements, using an RNA-mediated interference mechanism that has fundamental similarities to innate and adaptive immune responses1–3. This RNA-based adaptive immunity mechanism has fundamental similarities to innate and adaptive immune responses, and it preferentially cleaves the nontarget-strand DNA ~12 nt into the R-loop region, driven by ATP hydrolysis, subsequently catalyzing a similar degradation action on the target strand13,16. A recent study has suggested that recruitment of Cas3 involves interaction with the CasA component of the Cascade complex24. To fully understand the Cascade-activated mechanism of DNA unwinding and degradation, we determined the crystal structure

Received 2 June; accepted 18 July; published online 17 August 2014; doi:10.1038/nsmb.2875
of the Cas3 protein bound to an ssDNA substrate and biochemically defined its physical interactions with the Cascade complex. We revealed the catalysis mechanism of the HD nuclease with a snapshot of the ssDNA substrate coordinated by two catalytic irons in the active site. We captured the SF2 helicase in open conformation, with and without an ATP molecule bound, thus providing hints about the ATP hydrolysis–driven conformational-switching cycle. We revealed the functions of the Cas3-specific structure features with CRISPR interference assays and biochemical reconstitutions. We showed that Cas3 is specifically guided toward Cascade-bound target DNA in the presence of an optimal protospacer-adjacent motif (PAM) sequence, through physical interactions with the CasA component of Cascade.

Table 1  Data collection and refinement statistics

| Data collection* | Native | ATP | ADP | AMP-PNP | SeMet | Ta6Br12 soak |
|------------------|--------|-----|-----|---------|-------|--------------|
| Space group P21  | P21    | P21 | P21 | P21     | P21   | P21          |
| Cell dimensions | a, b (Å) | 85.79, 218.54, 123.75 | 87.25, 222.81, 125.09 | 87.21, 222.61, 124.85 | 86.92, 222.06, 124.90 | 86.41, 218.85, 124.03 | 86.45, 218.08, 122.90 |
| a, b, c (Å)     | 90.00, 90.00, 105.00 | 90.00, 90.00, 104.10 | 90.00, 90.00, 104.07 | 90.00, 90.00, 104.30 | 90.00, 90.00, 104.45 | 90.00, 90.00, 104.18 |
| Resolution (Å)  | 2.65   | 3.34 | 3.12 | 2.93    | 3.5   | 4.8          |
| Rmerge (Å)       | 58.7   | 64.1 (41.2) | 11.0 (48.8) | 8.2 (39.0) | 12.1 (33.9) | 10.8 (41.2) |
| I / σI          | 18.1 (2.3) | 9.3 (2.6) | 9.3 (3.2) | 9.4 (3.0) | 5.7 (1.5) | 14.7 (3.1) |
| Completeness (%) | 92.1 (86.8) | 99.2 (99.6) | 98.5 (98.3) | 99.1 (99.8) | 94.0 (86.9) | 98.6 (91.7) |
| Redundancy (%)   | 2.2 (1.9) | 2.2 (2.2) | 3.0 (2.9) | 2.6 (2.6) | 2.7 (2.0) | 3.8 (3.5) |
| Refinement       |        |      |      |         |       |              |
| Resolution (Å)   | 2.66   | 3.34 | 3.12 | 2.93    | 3.5   | 4.8          |
| No. reflections  | 115,323 | 66,466 | 80,395 | 97,476 |
| Rwork / Rfree   | 17.3 / 22.6 | 17.2 / 22.7 | 17.0 / 23.6 | 18.2 / 23.4 |
| No. atoms Protein | 28,027 | 27,906 | 27,922 | 28,051 |
| DNA             | 788    | 788  | 788  | 788     |
| Ligand / ion    | 0 / 8  | 124 / 8 | 108 / 8 | 124 / 8 |
| Water           | 215    | 0    | 0    | 0       |
| B factors Protein | 57.9   | 107.2 | 63.4 | 35.0    |
| DNA             | 66.2   | 118.4 | 80.6 | 52.3    |
| Ligand / ion    | 36.9   | 123.2 | 92.2 | 86.7    |
| Water           | 51.0   | 76.6  | 40.8 | 16.0    |
| r.m.s. deviations Bond lengths (Å) | 0.020 | 0.015 | 0.015 | 0.019 |
| Bond angles (°) | 1.41   | 1.56 | 1.51 | 1.49    |

* A single crystal was used for each structure. Values in parentheses are for highest-resolution shell. SeMet, selenomethionine.
and the noncomplementary strand of the dsDNA substrate. The stringent set of recognition events ensures well-controlled DNA targeting and degradation of foreign DNA in the type I CRISPR-Cas system.

RESULTS
Overall structure of ssDNA-bound T. fusca Cas3
We determined the crystal structure of the Cas3 protein in the T. fusca CRISPR-Cas type I-E system at 2.65 Å resolution, with a 12-nt endogenous ssDNA substrate bound (Fig. 1 and Table 1). The structure provides a snapshot of Cas3 in which two enzymatic activities are combined to unwind and degrade its DNA substrate (Fig. 1a,b). The SF2 helicase has a conserved arrangement of two juxtaposed RecA domains followed by Cas3-specific structure features, including a long linker helix and an accessory C-terminal domain (CTD) spanning the top. The HD nuclease domain packs against the first RecA-like domain (RecA1) of the helicase through a large, conserved hydrophobic interface of ~4,200 Å². The key interface residues, including W216, L217 and L260 from HD and W406, R412, L415, F441 and W470 from RecA1, are highly conserved (Fig. 2 and Supplementary Fig. 1). The RecA1 and RecA2 at the helicase core are separated by a cleft, where the conformational changes induced by ATP binding and hydrolysis are expected to take place17,25. Following RecA2, a horizontally packed linker helix spans the entire helicase back to the HD domain. This is followed by a flexible linker projecting toward the CTD, wrapping one side of the DNA-binding platform. The CTD contacts conserved surface loops in each of the RecA-like domains on the opposite side of the platform (Fig. 2b), burying a total surface area of ~2,020 Å², thus leading to the formation of a closed ssDNA-threading channel.

The path to degradation inside Cas3
A ssDNA substrate spanning both enzymatic moieties is captured in the precleavage state in the crystal structure, providing insights about the concerted DNA unwinding and DNA-degradation actions of Cas3. This substrate was endogenous DNA in origin, copurified with the Cas3 protein (Supplementary Fig. 2a), and it was probably processed further during crystallization. Electron densities are visible for 6 nt in the helicase region and 3 nt in the HD nuclease region, and 2 nt can be inferred in between (Fig. 1c–e). The DNA substrate traverses the most-conserved surface regions in Cas3 (Supplementary Fig. 3) and makes multiple sequence-nonspecific contacts along the path (Fig. 1c). The 5’ end of the DNA substrate enters from the RecA2 side of Cas3. A short separation helix is inserted between nts 1 and 2; the conserved R729 and V734 may have a role in threading nt 2 and beyond into the tunnel. The putative separation hairpin (amino acids aa 715–727), which is conserved in many SF2 helicases and wedges into the dsDNA11, is disordered in our structure. The sugar-phosphate backbone of nts 2–4 and 5–7 are contacted by RecA2 and RecA1, respectively, providing two anchoring points to enable the expected ‘inchworm’ movement in the SF2 helicase (Fig. 1d). Contacts from RecA2 include salt-bridge and hydrogen-bond interactions from R622 and Q660 to nt 2; S621 and T659 to nt 3; and R628 and Q664 to nt 4. RecA1–DNA contacts are from T337, M338 and E424 to nt 5; S372, T422 and Q425 to nt 6; and M428 to nt 7. Emerging from the helicase channel, the nucleoside portions of nts 7 and 8 are difficult to trace, owing to a lack of defined contacts, but density reappears toward the active site of the HD nuclease after an ~90° twist at the sugar-phosphate backbone. Here, ssDNA binding is assisted by K411 (nt 10) and W216 (nt 10) as well as active site residues and catalytic metal ions (Fig. 1e). The electron density is poorly resolved beyond the labile phosphate of nt 12 at the HD active site; residual electron density suggests that the leaving nucleotide kinks 90° upward to avoid steric clashes (Fig. 1e). This sharp geometry distortion may form the basis for the HD nuclease to disfavor dsDNA substrates. Overall, the path of ssDNA substrate demonstrates that Cas3 is an obligate 3’–5’ ssDNase preferentially accepting its substrate directly from the helicase moiety.

Metal coordination and catalysis mechanisms inside HD
The HD nuclease has been characterized as a transition-state metal–dependent single-stranded exo- and/or endonuclease15,16,21–23. The metal coordination scheme differs substantially among the published apo crystal structures21–23 (Supplementary Fig. 4). This may be the result of active site perturbation due to harsh purifications and/or the lack of ssDNA substrate21–23, ssDNase activity could be detected from T. fusca Cas3 supplemented with Mg²⁺, Mn²⁺ or Co²⁺, and the cleavage product contained a 5’-P-O-P-3’ and a 3’-OH (Supplementary Fig. 2b–d). The substrate-bound Cas3 structure provides many more mechanistic insights regarding substrate coordination and the catalysis...
mechanism. The labile 3′-P bond spanning the HD active site is trapped in the precleavage state, and the two nonbridging oxygen atoms interact strongly with the two irons (which are probably in the Fe(II) state) in the HD active site (Fig. 3a). We subsequently verified the identity of the unexpected irons with X-ray–absorption fine-structure (XAFS) analysis and an anomalous difference map collected at the absorption edge of iron (Supplementary Fig. 5). The strong affinity suggests that they are constitutive cofactors. The metal ions are coordinated by a cluster of highly conserved active site residues, including the invariable residues H83 and D84 (the HD motif), H37, H115, H149, H150 and the less conserved residue D215. The soft nature of the iron-chelating groups dictates the binding of transition-state metal ions rather than Mg2+ in these two sites. A similar active site configuration is present in a recently deposited apo putative HD nuclease domain structure (PDB 2PQ7; comparison in Supplementary Fig. 4d), thus suggesting that the two-iron architecture is probably conserved in a large percentage of Cas3 proteins. The strategic position of these two irons suggests their involvement in catalysis by coordinating a deprotonated water molecule between the two irons for nuclease attack and by stabilizing the negative charge together with K87 during the transition state26. Other HD nucleases are likely to coordinate transition metal ions at equivalent positions, although the exact configuration has not been fully resolved in the published apo structures, possibly owing to harsh purification conditions, weaker binding affinity and/or the lack of ssDNA substrate.21–23

On the basis of the new structural insight, we subjected HD residues involved in iron coordination (H37, H83, H115 and H149 H150) or ssDNA binding (K23) to alanine-scanning mutagenesis. All resulted in a dramatic loss-of-function phenotype similar to that of the D84A mutant, thus confirming their essential role in the HD nuclease (Fig. 3a). The conserved residue S219 is proximal to the 3′ leaving oxygen of the substrate and may be involved in its protonation. S219A mutation led to a substantial loss of function in the in vivo assay (Fig. 3c).

**Defining the function of HD active site residues in vivo**

We set up an in vivo CRISPR interference assay in *Escherichia coli* to evaluate the function of HD active site residues (Fig. 3b). Induction of *T. fusca* Cascade and Cas3 expression led to efficient CRISPR interference when the target plasmid bore a protospacer adjacent to a strong PAM (5′-WAK, and the strongest PAM, 5′-AAG) was used in subsequent assays. Averages and s.d. from independent experiments (n = 3) are shown here and in d. (d) HD active site mutagenesis with the same assay with shorter induction time (details in Online Methods). The consensus is 5′-WAK, and the strongest PAM, 5′-AAG, was used in subsequent assays. Averages and s.d. from independent experiments (n = 3) are shown here and in d. (d) HD active site mutagenesis with the same assay with shorter Cascade + Cas3 induction time (details in Online Methods). WT, wild type.

**Essential features of type A SF2 helicase of Cas3**

Our crystal structure reveals that 9 of the 12 typical sequences commonly found in SF2 helicases are well conserved in Cas3 motifs (I–VI, Ia–d, Ia–d, Ia and V) and PAM sequences with the same assay with shorter Cascade + Cas3 induction time (details in Online Methods). WT, wild type.
moeity adopts an extended conformation, pointing toward D451 and E452 in motif II (DEAH, Walker B motif). We supplemented Mg\(^{2+}\) in substoichiometric amounts and did not observe it near the \(\beta\) - and \(\gamma\)-phosphates. However, residues responsible for positioning the Mg\(^{2+}\), such as T312 and D451 in motifs I and II as well as E452, the residue responsible for coordinating the attacking water molecule, are well conserved in space, thus suggesting that Cas3 would function as a typical SF2 helicase.

**Functional importance of the Cas3-specific helicase features**

Because these SF2 motifs have been well defined, we focussed the mutagenesis on Cas3-specific structure features. Motif Ic is uniquely found in Cas3, and its position suggests an involvement in coupling the helicase and nuclease activities. Indeed, R410A K411A and R410Y K411F R412A, mutants targeting motif Ic, both led to a strong loss-of-function phenotype similar to that of HD active site mutants (Fig. 4c). Replacing the linker helix (aa 777–815) with a glycine/serine-rich linker (Δlinker helix) was also more detrimental than the helicase-null mutant, thus suggesting an essential function for this structure feature (Fig. 4c). Deletion of the CTD (aa 819–924) reduced CRISPR interference to the helicase-null-mutant level (Fig. 4c).

**Biochemical reconstitution of the Cascade-Cas3 interaction**

A missing link in the type I CRISPR system is the definition of specific interactions that lead to Cas3 recruitment by a target DNA–bound Cascade. To address this, we reconstituted the Cascade-Cas3 interaction, using electrophoretic mobility shift assays (EMSAs) for deeper mechanistic dissection. Unlike the *E. coli* Cascade, the purified *T. fusca* Cascade mostly dissociated into free CasA and a crRNA-containing CasB–E subcomplex in the absence of a proper substrate (Fig. 5a–c); intact Cascade was stable only when bound to a protospacer and PAM-containing dsDNA, a bubbled substrate or a nicked substrate mimicking an R-loop intermediate (Figs. 5d and 6a). When we used a perfectly base-paired dsDNA substrate, CasA was required to recruit CasB–E to the DNA substrate, thus underlining the importance of CasA in initiating the dsDNA unwinding process to form the R loop (Fig. 5d). When dsDNA unwinding was bypassed through the use of a bubbled DNA substrate, CasB–E alone was capable of...
interactions, presumably through base-pairing between its crRNA and the exposed protospacer region in the target-strand DNA (Fig. 6b).

The affinity of Cas3 alone for various DNA substrates was rather weak, and the rate-limiting step appeared to be at the substrate binding-and-exchange step (Fig. 6a, right three lanes). The nicked R-loop mimic supported the most stable Cas3-Cascade interaction, and Cas3 was specifically recruited to Cascade when a series of conditions were met: (i) an intact Cascade with CasA stably bound; (ii) an optimal PAM sequence in the substrate (because PAM substitution in either or both strands of the DNA substrate disrupted binding, presumably because of CasA dissociation (Fig. 6a, lanes 10–18); and (iii) a 3′ overhang of at least 10 nt in the nontarget-strand DNA (because we detected no Cas3 binding with a 5-nt overhang) (Fig. 6a, lanes 1–9)). The 10-nt-overhang requirement is roughly consistent with the 11-nt ssDNA inside our Cas3 crystal structure, thus suggesting that stable Cas3-Cascade association probably requires the nontarget DNA strand to thread through the helicase to reach the HD active site in Cas3 (Fig. 1).

Inclusion of AMP-PNP or use of the D451A helicase mutant strengthened the Cas3-Cascade interaction, whereas ATP incubation led to much weaker interaction, presumably because the elevated helicase activity allowed Cas3 to clear the Cascade (Fig. 6b).

The function of the Cas3-specific structure features became apparent in the Cascade-Cas3 interaction assay (Fig. 6b). Cas3 lacking the CTD (aa 1–818) exhibited weakened affinity for Cascade. Complementing this truncation with CTD in trans restored its Cascade-binding affinity to the wild-type level, consistently with our structure suggesting that the CTD domain regulates HD nuclease activity by functioning as a substrate filter. More importantly, Cascade binding was completely lost when the linker region in Cas3 was replaced by a flexible linker (Fig. 6b), thus pointing to a strong involvement of this region in Cascade interaction. HD nuclease mutant H83A also exhibited weaker affinity for Cascade, consistently with an accessory function of the HD domain in stabilizing the Cascade-Cas3 interaction by tighter binding to the substrate (Fig. 6b). Finally, addition of ATP weakened the Cascade-Cas3 interaction, presumably because it triggered Cas3 to processively unwind and degrade DNA substrate. Indeed, when we used the D451 helicase mutant in the assay with ATP present, Cascade-Cas3 interaction was restored (Fig. 6b).

DISCUSSION

The most widespread type I CRISPR-Cas system can be further classified into six different subtypes (I-A to I-F), each encoding a unique set of subtype-specific genes. Although the composition of the target-searching Cascade complex varies substantially among these subtypes, the effector gene cas3 is universally conserved in all type I systems, thus underlining its functional importance. The analysis of the Cas3 crystal structure in this study reveals the arrangement of the nuclease and helicase activities inside the ~100-kDa Cas3 protein and sheds light into its mechanism of DNA unwinding and degradation. The biochemical reconstitutions further define the set of conditions leading to Cas3 recruitment to the Cascade-bound DNA substrate.

Although the HD nuclease inside Cas3 was correctly defined as a transition-state metal–dependent ssDNase, its exact catalysis mechanism remained elusive. Our Cas3 structure reveals the coordination of two constitutively bound catalytic iron cofactors and their interactions with the ssDNA substrate. Questions remain about why the ssDNA substrate was not cleaved inside the Cas3 crystal. A possibility is that only Fe(II), but not its oxidized Fe(III) form, is capable of supporting the HD nuclease activity and that our T. fusca Cas3 may have been slowly inactivated during purification and crystallization because of Fe(II) oxidation, thus leading to the observation of a trapped ssDNA in its precursor form. Efforts are underway to investigate this possibility and to provide further snapshots of Cas3 in various stages of the enzymatic cycle.

The structure features commonly found in SF2 helicases are well conserved in the Cas3 helicase region. This suggests that Cas3 would function like a typical SF2 helicase to unwind dsDNA. SF1 and SF2 monomeric helicases have been shown to unwind dsDNA with an inchworm mechanism. That is, binding of ATP to RecA induces a rotation and closing movement in RecA to orient important residues for ATP hydrolysis. This structure compression is then relieved after ATP hydrolysis and dissociation. The ATP-induced conformational change, coupled with an alternate tight-loose grip on ssDNA from the two RecA-like domains, leads to the inchworm movement of the helicase on one DNA strand to unwind the dsDNA. The compression of the helicase upon ATP binding was not captured in our Cas3 crystals, presumably owing to crystal-lattice trapping; however,
we see hints of such movement when comparing the four Cas3 molecules in the asymmetric unit of the crystal lattice, because each undergoes a different extent of rigid-body movement in the crystal lattice (Supplementary Fig. 6e). A hinge motion in the ATP-binding cleft results in a closing-in motion in molecule D, such that its HD domain shifts by as much as 3 Å, causing the active site metal ions and ssDNA substrate to move ~1.7 Å (Supplementary Fig. 6f). Such movements give hints about the consequence of conformational changes, induced by ATP binding and hydrolysis, in the Cas3 helicase region and point to a concerted mechanism to couple the unwrapping movement of the helicase with the substrate translocation in the HD nucleic acid site.

When taken together, the Cas3 crystal structure enabled us to resolve some of the major mechanistic questions in the type I CRISPR-Cas system. The two enzymatic activities in Cas3 work in a concerted fashion at several levels. Apart from unwinding and feeding the substrate into the HD nuclease, the helicase moiety also initiates the recruitment of the latter activity through physical interactions with the CasA component of the R-loop–presenting Cascade (Fig. 6b). Subsequently, the HD nuclease strengthens the Cascade-Cas3 interaction with its strong ssDNA binding affinity at the active site (Fig. 6b). Cascade-Cas3 interaction occurs only when a series of conditions are met, thus reflecting an evolutionary pressure to tightly regulate the activation of Cas3-mediated DNA degradation.

In consideration of the initial steps involving Cascade-mediated recruitment and activation of Cas3 at the R-loop region, questions remain about how Cas3 is loaded onto the nontarget strand in the R-loop structure before the cleavage event, because threading a looped ssDNA without an open 3’ end through the caged Cas3 helicase would be topologically challenging. Although it is possible that the HD nuclease makes the first cut by bypassing the helicase, the existing data15,16 are more consistent with a model in which the CTD of Cas3 transiently dissociates, possibly triggered by interaction with Cascade, to allow ssDNA placement into the helicase. Such accessory domain movement has been observed in other helicas, as exemplified in UvrD28. After CTD reassociation, the HD nuclease makes the first cut and converts Cas3 into the processive conformation observed in the crystal structure, and ATP hydrolysis further drives 3’-to-5’ unwinding and degradation. An alternative model to resolve the topological challenge involves the retraction of the long flexible linker sequence (aa 816–832) beyond the HD nuclease to expose the side of the ssDNA-recruitment channel along Cas3. Single-molecule fluorescence resonance energy transfer experiments and cocrystal structures of Cas3 bound to the target-presenting Cascade complex will be required to distinguish these two competing models.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Coordinates and structure factors for Cas3 (4QQW), Cas3–ATP (4QQX), Cas3–AMP-PNP (4QQ2) and Cas3–ADP (4QQY) have been deposited in the Protein Data Bank.

Note: Any Supplementary Information and Source data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by US National Institutes of Health (NIH) grants GM-086766 and GM-102543 to A.K. and Korean Postdoctoral Fellowship NRF-2010-357-C0106 to K.H.N. NE-CAT and MACCHES5 beamlines were supported by NIH grants GM103403 and GM103485, respectively. We thank Y. Chen and K. Perry for technical help and J. van der Oost, J. Grigg, R. Hayes and I. Price for helpful discussions.

AUTHOR CONTRIBUTIONS

Y.H., K.H.N., A.K., K.K. and I.K. collected diffraction data. K.H.N., L.W., R.Z. and A.K. determined the structure. F.D., H.L., A.K., Y.X., M.D.F., Y.H. and S.Z. performed the biochemical analyses. A.K. designed the research and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Terns, M.P. & Tens, R.M. CRISPR-based adaptive immune systems. Curt. Opin. Microbiol. 14, 321–327 (2011).
2. Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. RNA-guided genetic silencing systems in bacteria and archaea. Nature 482, 331–338 (2012).
3. Jore, M.M., Brouns, S.J. & van der Oost, J. RNA in defense: CRISPRs protect prokaryotes against mobile genetic elements. Cold Spring Harb. Perspect. Biol. 4, a003657 (2012).
4. Haft, D.H., Selengut, J., Morgodin, 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).
5. 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 analogues with eukaryotic RNAI, and hypothetical mechanisms of action. Biol. Direct. 1, 7 (2006).
6. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712 (2007).
7. Mărieș, L.A. & Sontheimer, E.J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322, 1843–1845 (2008).
8. Hale, C.R. et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139, 945–956 (2009).
9. Makarova, K.S. et al. Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. 9, 467–477 (2011).
10. Brouns, S.J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960–964 (2008).
11. Jackson, R.N., Lavin, M., Carter, J. & Wiedenheft, B. Fitting CRISPR-associated Cas3 into the helicase family tree. Curt. Opin. Struct. Biol. 24, 106–114 (2014).
12. Jore, M.M. et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. 18, 529–536 (2011).
13. Wiedenheft, B. et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature 477, 486–489 (2011).
14. Westra, E.R. et al. CRISPR immunity relies on the conscription binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol. Cell 46, 595–605 (2012).
15. Sinkunas, T. et al. In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. EMBO J. 32, 385–394 (2013).
16. Mulepali, 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).
17. Singleton, J.M., Dillingham, M.S. & Wigley, D.B. Structure and mechanism of helicases and nucleic acid translocases. Annu. Rev. Biochem. 76, 23–50 (2007).
18. Singleton, J.M., Dillingham, M.S., Gaudier, M., Kowalczykowski, S.C. & Wigley, D.B. Crystal structure of RecBCD exonuclease reveals a machine for processing DNA breaks. Nature 432, 187–193 (2004).
19. Taylor, D.W. et al. Substrate-specific structural rearrangements of human Dicer. Nat. Struct. Mol. Biol. 20, 662–670 (2013).
20. Perry, J.J. et al. WRN exonuclease structure and molecular mechanism imply an editing role in DNA end processing. Nat. Struct. Mol. Biol. 13, 414–422 (2006).
21. Mulepali, S. & Bailey, S. Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). J. Biol. Chem. 286, 31896–31903 (2011).
22. Beloglazova, N. et al. Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. EMBO J. 30, 4616–4627 (2011).
23. 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).
24. Hochsiedler, M.L. et al. Cas3 mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc. Natl. Acad. Sci. USA 111, 6618–6623 (2014).
25. Yang, W. Lessons learned from UvrD helicase: mechanism for directional movement. Annu. Rev. Biophys. 39, 367–385 (2010).
26. Yang, W. Nucleases: diversity of structure, function and mechanism. Q. Rev. Biophys. 44, 1–93 (2011).
27. Böttner, K., Nehring, S. & Hofpner, K.P. Structural basis for DNA duplex separation by a superfamily-2 helicase. Nat. Struct. Mol. Biol. 14, 647–652 (2007).
28. Lee, J.Y. & Yang, W. UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. Cell 127, 1349–1360 (2006).
ONLINE METHODS

Expression, purification, crystallization and diffraction of Cas3 protein. The full-length *T. fusca* cas3 (NCBI Q477P0, gene symbol Tfu_1593) fused to an N-terminal His6 tag was cloned into the pET28b expression vector between Ndel and Xhol restriction sites. *E. coli* BL21 (DE3) star cells containing the sequence-verified Cas3 construct were grown in LB medium at 37 °C to O.D.₆₀₀ of 0.6, protein expression was induced by the addition of 0.5 mM isopropyl β-n-1-thiogalactopyranoside (IPTG), and cells were further cultured at 18 °C for 16 h. The harvested cells were disrupted by sonication in buffer A (20 mM Tris-HCl, pH 7.5, 300 mM NaCl, 10% glycerol, and 2 mM β-mercaptoethanol). The supernatant after centrifugation was applied onto an Ni-NTA column pre-equilibrated in buffer A. After washing with 10–15 volumes of wash buffer in buffer A plus 5 mM imidazole, Cas3 protein was eluted with buffer A plus 250 mM imidazole. After dialysis into low-salt buffer, Cas3 was further purified on aMonoS column, and the peak fractions were concentrated and purified on a Superdex 200 column. The peak fractions were concentrated to 20 mg/ml and flash frozen at −80 °C. A similar purification procedure was followed to obtain selenomethionine-derived proteins. Cas3 crystals were grown with the hanging-drop vapor-diffusion method by mixing of 1 μl of protein solution with 1 μl of mother liquor (100 mM MES, pH 6.0, and 24% PEG 4000) at 18 °C. ATP, AMP-PNP-, and ADP-bound structures were determined from crystals soaked in the well solution supplemented with 2 mM of each compound and 1 mM MgCl₂ overnight. Diffraction data were collected through the mail-in-data-collection service at beamlines in APS NE-CAT and were processed with HKL2000 (ref. 29). Cas3 mutants were constructed with site-directed mutagenesis. Proteins were purified with Ni-NTA and MonoQ columns, concentrated to over 10 mg/ml, and stored at −20 °C in 50% glycerol.

Phasing, model building and structure refinement. The initial experimental phases were determined by the single-wavelength anomalous dispersion (SAD) method with a data set collected at the tantalum L-III absorption edge from a crystal derivatized with Ta₆Br₁₂ clusters. The positions of four cluster sites were identified by Molrep33. The tantalum cluster phases allowed the placement of 95% of all non-H atoms with REFMAC5 (ref. 36) and Phenix.refine37. The nature structural & molecular biology doi:10.1038/nsmb.2875

 procedures in Phenix.autosol31 and Phenix.autobuild32. The phases were further rebuilt into the electron density map with Coot 35. The resulting model was refined with REFMAC5 (ref. 36) and Phenix.autobuild32 by applying the noncrystallography symmetry (NCS) as superatoms. The resulting phases were then improved and extended to 3.0 Å resolution on the built-in multichannel analyzer of X-123SDD was calibrated with known fluorescence emission lines of several metals. The gain of the detector was set to 75%, corresponding to an energy range of 0–16.7 keV. EDS experiments were carried out with incident X-ray energy close to the K absorption edge of iron. The anomalous scattering components, f′ and f″, were derived from the XAFS spectrum with CHOOCH39.

Expression and purification of *T. fusca* Cascade complexes. *T. fusca* cas operon encoding the cas3 to casE genes (Tfu_1592 to Tfu_1588) was PCR amplified from genomic DNA and cloned into a pET expression vector. A His6 tag was inserted via mutagenesis into the N terminus of cas3 for Ni-NTA purification. Coexpression of Tfu_Cas3 with pre-cRNA expression vector led to a stable cRNA-containing CasB–E subcomplex, with CasA weakly associated. Subsequently, CasA was removed from the expression operon, recombinantly expressed separately as a His₆-tagged protein, and purified into homogeneity with Ni-NTA and MonoQ columns. CasB–E subcomplex lacking CasA was purified with Ni-NTA, MonoQ, and Superdex 200 columns (Fig. 5). Even though the Superdex 200 SEC peak fractions contained CasB–E proteins at the correct stoichiometry, the higher-molecular-weight half of the peak contained contaminating nucleic acids, thus resulting in various degrees of aggregation as revealed by the 6% native gel (Fig. 5b). The fractions lacking these contaminating nucleic acids were pooled, concentrated to over 10 mg/ml, and flash frozen in small aliquots at −80 °C for biochemical assays.

In vivo assay for *T. fusca* Cascade and Cas3–mediated target plasmid loss. We modified the assay from a published protocol41 to recombminantly express the *T. fusca* Cascade and *T. fusca* Cas3 for target plasmid destruction in the *E. coli* BL21_AI cell line, which does not express the *E. coli* CRISPR and cas operons. The experimental design and the list of plasmids used are detailed in Figure 3b and Supplementary Table 1, respectively. All Tfu_Cas3 constructs were pET28b (KanR) based. The Tfu_Cascade expression construct was generated by insertion of the Tfu_CasA–E fragment into a pBAD (ApR) vector. The pre-cRNA expression cassette containing four identical CRISPR units under the control of T7 RNA polymerase promoter and terminator was synthesized by GenScript and cloned into the pACYC-Duet-1 (CmR) vector. A 332-bp fragment of the 5′-phase genome containing a matching protospacer and a 3′-bp PAM was cloned into the pCDF-Duet-1 (SmR) vector to serve as the target DNA. All plasmids were sequence verified. The pET28b_Cas3, pBAD_Cascade, pACYC_CRISPR and pCDF_target plasmids were transformed into BL21_AI competent cells and grown on LB plates containing kanamycin (50 μg/ml), ampicillin (100 μg/ml), streptomycin (50 μg/ml) and chloramphenicol (34 μg/ml).

In the PAM sequence–evaluation assays, the four-plasmid–containing BL21_AI cells from a single colony were cultured at 37 °C in nonselective LB medium to O.D.₆₀₀ of 0.3, at which point the expression of *T. fusca* Cascade, Cas3, and pre-cRNA was induced for 12 h by the addition of 0.5% l-arabinose and 2.5 mM IPTG. Each cell culture was then divided into two equal volumes and plated onto Kan + Ap + Cm LB plates (nonselective for pCDF_target) and Kan + Ap + Cm + Sm plates (selective for the target plasmid) in a series of dilutions. The number of colonies on each plate was counted after overnight incubation at 37 °C. The CRISPR interference efficiency was reflected in the ratio of colony-forming units on the nonselective over selective plates. Ten-fold titration series were routinely plated to capture the dynamic range. Each experiment was repeated three times to calculate the s.d.

All Cas3 mutants were evaluated by a slightly different protocol. The target plasmid contains the strongest PAM sequence, 5′-AAG. After induction of Cascade and Cas3 expression, the *E. coli* cells were incubated for 6 h before plating, instead of the 12 h used in the PAM sequence assays. The shorter induction time led to slightly reduced dynamic range in CRISPR interference efficiency. However, the disadvantage is outweighed by a substantially reduced error spread, thus making the mutagenesis data more reliable for quantitative comparison.

In vitro reconstitution of Cas3–Cascade–R-loop interaction. DNA oligos (Supplementary Table 1) were chemically synthesized and gel-purified. The non-target DNA strand was 32P-labeled, annealed with the target strand at 1:2 molar ratio by incubation at 65 °C for 10 min and cooled on bench. The Cascade–R-loop complexes (or mimics) were assembled by incubation of the annealed DNA substrate with 40 nM Cas3–E and 50 nM CasA at 25 °C for 20 min in a buffer containing 20 mM HEPES, pH 7.5, 100 mM NaCl, 10 mM MgCl₂, 10 mM FeCl₂, 2 mM AMP-PNP (or ATP), and 5% glycerol. Wild-type or mutant Cas3 proteins were further carried out on the crystal (Supplementary Fig. 5c) and away from the crystal (Supplementary Fig. 5d) by tuning the incident X-ray energy close to the K absorption edge of iron. The anomalous scattering components, f′ and f″, were derived from the XAFS spectrum with CHOOCH39.

X-ray–absorption fine-structure (XAFS) analysis for metal-ion identity. Energy-dispersive X-ray spectroscopy (EDS) was carried out with a silicon drift detector (model X-123SDM, Amptek) at the NE-CAT 24ID-C beamline. The built-in multichannel analyzer of X-123SDM was calibrated with known fluorescence emission lines of several metals. The gain of the detector was set to 75%, corresponding to an energy range of 0–16.7 keV. EDS experiments were carried out with incident X-ray energy of 12.66 keV, just above the K absorption edge of selenium. First, the EDS spectrum was recorded for 127.85 s with X-rays incident on the crystal in the cryoloop (Supplementary Fig. 5a). As a negative-control experiment, EDS was also recorded for 190.7 s on the cryoloop but away from the crystal (Supplementary Fig. 5b). To verify the presence of iron in the crystal, near-edge X-ray–absorption fine-structure spectroscopy (XAFS)
were then introduced at two different concentrations (20 nM or 80 nM), and the reaction mixture was incubated at 37 °C for an additional 30 min. The reaction mixtures were then electrophoretically separated on a 6% Tris/borate/EDTA (TBE) native PAGE in a cold room. The gel was exposed to a storage phosphor screen, and the radiograph signals were recorded by a Typhoon 9200 machine. All controls were done in the same buffer condition by following the same incubation procedure.

Nuclease assay. Nucleic acid cleavage reactions were performed at 37 °C for 60 min in a buffer containing 10 mM Tris·HCl, pH 8.0, 60 mM KCl, 10 mM MgCl₂, and 1 mM dithiothreitol (DTT). 0.1 µM 5′-Cy5-labeled ssDNA was incubated with 1 µM of wild-type or mutant Cas3 proteins. Reactions were initiated by the addition of protein and stopped by the addition of 3× stop buffer (67.5 mM EDTA, 27% (v/v) glycerol, and 0.3% (w/v) SDS). The reaction mixture was separated by electrophoresis on 10% (w/v) 8 M urea polyacrylamide gel (PAGE), and the Cy5 fluorescence signal was recorded with a Typhoon 2900. The metal-dependency experiment was carried out in essentially the same conditions, except the Mg²⁺ was substituted with 10 mM other metal ions.

29. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).

30. Sheldrick, G.M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr. D Biol. Crystallogr. 66, 479–485 (2010).

31. Terwilliger, T.C. et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallogr. D Biol. Crystallogr. 65, 582–601 (2009).

32. Terwilliger, T.C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr. D Biol. Crystallogr. 64, 61–69 (2008).

33. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. 66, 22–25 (2010).

34. Cowtan, K.D. & Main, P. Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. Acta Crystallogr. D Biol. Crystallogr. 49, 148–157 (1993).

35. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132 (2004).

36. Winn, M.D., Murshudov, G.N. & Papiž, M.Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol. 374, 300–321 (2003).

37. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221 (2010).

38. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21 (2010).

39. Evans, G. & Pettifer, R.F. CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. J. Appl. Crystallogr. 34, 82–86 (2001).