

Helicase dissociation and annealing of RNA-DNA hybrids by *Escherichia coli* Cas3 protein

Jamieson A. L. HOWARD*, Stephane DELMAS†, Ivana IVANČIĆ-BAČE‡ and Edward L. BOLT*¹

*School of Biomedical Sciences, The University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, U.K., †School of Biology, The University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, U.K., and ‡Department of Molecular Biology, University of Zagreb, Horvatovac 102a, 10000 Zagreb, Croatia

CRISPR (clustered regularly interspaced short palindromic repeat)/Cas (CRISPR-associated) is a nucleic acid processing system in bacteria and archaea that interacts with mobile genetic elements. CRISPR DNA and RNA sequences are processed by Cas proteins: in *Escherichia coli* K-12, one CRISPR locus links to eight *cas* genes (*cas1*, 2, 3 and *casABCDE*), whose protein products promote protection against phage. In the present paper, we report that purified *E. coli* Cas3 catalyses ATP-independent annealing of RNA with DNA forming R-loops, hybrids of RNA base-paired into duplex DNA. ATP abolishes Cas3 R-loop formation and instead powers Cas3 helicase unwinding of the invading RNA strand of a model R-loop substrate. R-loop formation by Cas3 requires magnesium as a co-factor and is

inactivated by mutagenesis of a conserved amino acid motif. Cells expressing the mutant Cas3 protein are more sensitive to plaque formation by the phage λ vir. A complex of CasABCDE ('Cascade') also promotes R-loop formation and we discuss possible overlapping roles of Cas3 and Cascade in *E. coli*, and the apparently antagonistic roles of Cas3 catalysing RNA-DNA annealing and ATP-dependent helicase unwinding.

Key words: annealing, CRISPR-associated 3 (Cas3), clustered regularly interspaced short palindromic repeat (CRISPR), helicase, R-loop.

INTRODUCTION

CRISPR (clustered regularly interspaced short palindromic repeat [1])/Cas (CRISPR-associated) is a nucleic acid processing system in bacteria and archaea that interacts with mobile genetic elements. CRISPR loci are functionally coupled to Cas proteins and can protect cells against infection by phage or plasmids [2–8]. CRISPR/Cas also contributes to other physiological roles in some organisms: prevention of lysogeny-dependent biofilm formation in *Pseudomonas aeruginosa* [9,10], DNA repair or recombination in *Escherichia coli* [11] and partitioning of newly replicated plasmids during cell division in halophilic archaea [12]. In some organisms, CRISPR/Cas may self-target, interacting with host cell genes [13,14]. The various functions of CRISPR/Cas probably reflect the impressive variability in organization and distribution of CRISPR loci and *cas* genes across bacterial and archaeal domains, prompting their classification into CRISPR/Cas subtypes. This and other aspects of CRISPR/Cas function have been detailed in recent reviews [15–19].

Key events in CRISPR/Cas processing include acquisition of DNA fragments from mobile genetic elements into host cell CRISPR loci as 'spacers' alongside new repeat sequences, and transcription of CRISPR loci to RNA (pre-crRNA) that is processed into shorter RNA (crRNA). Therefore crRNA contains sequence that is complementary to antecedent mobile elements, and can be utilised to target invading nucleic acid. In bacteria, crRNA targets DNA [3,5,6,20], and in the archaeon *Pyrococcus furiosus*, invading RNA is targeted, leading to its destruction [21]. Processing and targeting of crRNA is catalysed by 'core' Cas proteins, working with less widespread Csx (Cas subtype-specific) proteins or archaeal CMR proteins [Cas RAMP (repeat-associated mysterious proteins) module]. Acquisition

of DNA spacer-repeats into CRISPR loci is by unknown mechanisms that may involve Cas1 or Cas2 nucleases [11,22,23]. Endonucleolytic trimming of transcribed pre-crRNA into shorter (50–100 nucleotide) RNA (crRNA) [24,25] is catalysed by subtype-specific Cas proteins; Csy4 in *Ps. aeruginosa* [26], Cas6 in the archaeon *P. furiosus* [27] or CasE (also called Cse3) in *E. coli* [3]. *E. coli* CasE is part of a larger protein-RNA complex known as Cascade, encoded by *casABCDE* [20].

CRISPR loci were first described in *E. coli* [28]. Four such loci have been analysed in this organism [29,30], one of which (CRISPR-1) is associated with eight *cas* genes (*Cas1*, 2, 3 and *casA*, *casB*, *casC*, *casD* and *casE*) (Figure 1A). Transcription of CRISPR-1 and *casA–E* is repressed by H-NS [31] and de-repressed by LeuO [32] or BaeR [33]. Activation of *E. coli* CRISPR-1/Cas improved the survival of cells attacked by lysogenic and lytic cycles of infection by λ phage [3,4,7]. This required expression of a crRNA sequence engineered with an anti- λ spacer (T3). In two studies [4,7] the T3 spacer was recombineered into the CRISPR-1 locus, and increased protection from λ phage required *casA–E*. In another study [3], protection from phage was induced by expression of crRNA, *casA–E* and *cas3* in multi-copy plasmids. *E. coli* chromosomal *cas3* is located upstream of *casA–E* (Figure 1A) and is not controlled by H-NS, although *cas3* transcription in *Thermus thermophilus* is increased 2.5-fold in response to infection by phage ϕ YS40 [34].

Cas3 is a core Cas protein found in most CRISPR/Cas subtypes. Evidence from *in vivo* studies indicates that Cas3 functions downstream of endonucleolytic crRNA processing in *Ps. aeruginosa* [9] and co-operates with the Cascade-crRNA complex in *E. coli* [3]. Amino acid sequence motifs in Cas3 predict ATP-dependent helicase activities for Cas3 and a

Abbreviations used: ATP[S], adenosine 5'-[γ -thio]triphosphate; Cas, CRISPR-associated; Cascade, CasABCDE; CRISPR, clustered regularly interspaced short palindromic repeat; crRNA, CRISPR RNA; dsDNA, double-stranded DNA; DTT, dithiothreitol; IPTG, isopropyl β -D-thiogalactopyranoside; LB, Luria-Bertani; MBP, maltose-binding protein; Mth, *Methanothermobacter thermoautotrophicus*; ssDNA, single-stranded DNA; ssRNA, single-stranded RNA; TBE, Tris/borate/EDTA.

¹ To whom correspondence should be addressed (email ed.bolt@nottingham.ac.uk).

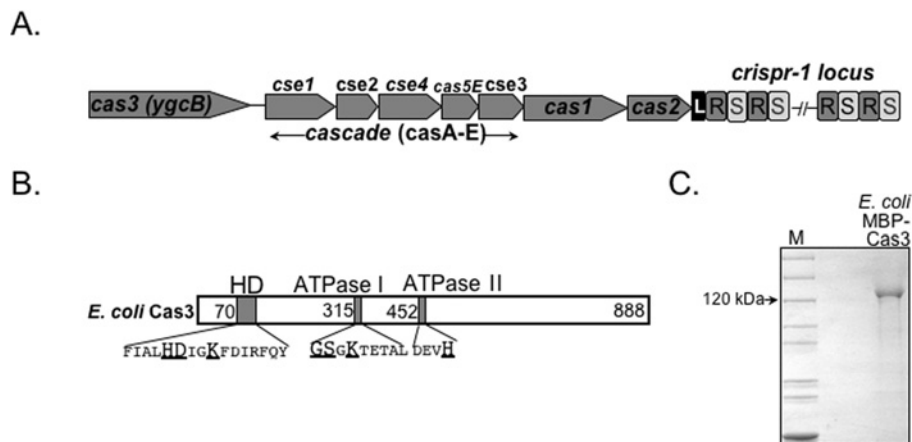


Figure 1 CRISPR/Cas of *E. coli* K-12 MG1655 and its Cas3 protein

(A) Representation of CRISPR-1 in *E. coli* MG1655 and eight *cas* genes. Open reading frames encoding the Cascade (CasA–E) are indicated, downstream of *cas3* encoded by *ygcB*. Leader sequence (L) is highlighted from which crRNA is transcribed, comprising 29 bp identical CRISPR repeats (R) and 32 bp variable spacer sequences (S). There are 12 repeats in CRISPR-1. (B) Cartoon of *E. coli* MG1655 Cas3 showing conserved HD motif, motif I (P-loop/Walker A) and motif II (Walker B) ATPase motifs. Amino acid positions of the first residues of each motif are given: 70, 315 and 452. Highlighted in each motif are residues mutated in the present study. (C) Coomassie Blue stained SDS/PAGE gel showing purified *E. coli* MBP–Cas3, giving an approximate molecular mass of 143 kDa. Molecular mass markers (M) are shown on the left-hand side.

HD-phosphohydrolase domain [35]. As is consistent with containing a helicase domain, purified Cas3 from *Streptococcus thermophilus* catalysed displacement of 20 nucleotide DNA and RNA strands annealed to M13 closed circular ssDNA (single-stranded DNA) [36]. Although the HD domain of *S. thermophilus* Cas3 was required for ssDNA nuclease activity on M13, the same domain in *Sulfolobus solfataricus* Cas3 was active at nicking dsDNA (double-stranded DNA) in preference to ssDNA [37], a nicking activity not tested by Sinkunas et al. [36]. In the present study, we describe the ATP-dependent helicase activity of *E. coli* Cas3 that displaces an RNA strand from an R-loop, a substrate in which RNA is base paired into complementary duplex DNA. We also describe a new biochemical activity for Cas3: ATP-independent annealing of RNA and DNA strands. Cas3 purified from the archaeal species Mth (*Methanothermobacter thermautotrophicus*) showed similar annealing and helicase activities, indicating that these Cas3 functions are evolutionarily conserved. We discuss apparently antagonistic functions of Cas3 that depend on the presence of ATP.

EXPERIMENTAL

Cloning *cas3* from *E. coli* MG1655 and Mth Δ H

E. coli K12 MG1655 *ygcB* (encoding Cas3) was cloned via PCR from genomic DNA into plasmid pET14b using NdeI and EcoRI sites and verified by sequencing generating pEB489. pEB489 was used as a PCR template to clone *ygcB* into pMal-c2 using EcoRI and XbaI sites in-frame with *malE*, creating plasmid pAH1. This was successful at overproducing soluble Cas3 fused at its N-terminus to MBP (maltose-binding protein), creating a protein of approximately 143 kDa as shown in Figure 1(C). Supplementary Table S1 (at <http://www.BiochemJ.org/bj/439/bj4390085add.htm>) lists primer sequences used for cloning *ygcB* into pMal-c2 (*ygcB*-X2 and *ygcB*-Y2) and for Cas3 mutagenesis to generate H74G, D75G, K78L, K320L and G317SS318Y.

Mth ORF1086 (encoding Cas3) was cloned from Mth genomic DNA (a gift from Dr James Chong, Department of Biology, University of York, York, U.K.) via PCR using primers with NdeI and HindIII sites as detailed in Supplementary Table S1. PCR

products were cloned into pT7-7 generating pEB359, and this construct was effective at overexpressing soluble MthCas3. Mth D347A Cas3 was generated by mutagenic PCR on pEB359 using primers listed in Supplementary Table S1.

Purification of *E. coli* and Mth Cas3 proteins

E. coli MBP–Cas3 was purified from pAH1 in DH5 α grown in MU broth containing ampicillin (50 μ g/ml). At $D = 0.5$, cells were induced by IPTG (isopropyl β -D-thiogalactopyranoside; 0.5 mM) for 30 min at 37 °C. Recovered cells were resuspended in amylose column buffer [20 mM Tris/HCl, pH 7.5, 200 mM NaCl, 1 mM EDTA, 1 mM DTT (dithiothreitol) and 0.1 mM PMSF] and frozen at –80 °C. Cas3 was purified by FPLC at room temperature (22 °C) and SDS/PAGE, and Coomassie Blue staining of proteins followed each purification step. Protein concentrations were measured against a BSA standard curve using Bradford's reagent (Bio-Rad Laboratories).

Biomass was lysed by sonication and centrifuged (20000 rev./min, 30 min in an Avanti J25 machine with rotor JA25.50). Soluble material (S1) was passed into a 7 ml amylose (High Flow, New England Biolabs) column equilibrated with amylose column buffer. MBP–Cas3 bound to amylose and was eluted in 2 mM maltose in amylose column buffer. Cas3 was loaded directly on to a 10 ml heparin column equilibrated with buffer C (20 mM Tris/HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 1 mM DTT and 0.1 mM PMSF). Cas3 did not bind to heparin: the flow-through was loaded directly on to a 10 ml DEAE-Sepharose column, eluting in 300–400 mM NaCl. Cas3 pooled fractions were loaded directly on to a 5 ml fibrous cellulose phosphate column pre-equilibrated with buffer C containing 350 mM NaCl. Cas3 did not bind cellulose phosphate and was finally re-loaded on to the amylose column to concentrate the protein. Cas3 fractions were pooled and dialysed against buffer C containing 35 % glycerol for storage at –80 °C. Mutant Cas3 proteins were purified in the same way.

Mth Cas3 was purified from pEB359 transformed into BL21 CodonPlus, inducing Cas3 expression at 37 °C for 2 h with IPTG (0.5 mM). Cells were harvested into buffer C and frozen at –80 °C. Thawed biomass was lysed by sonication and clarified (20000 rev./min, 20 min in an Avanti J25 machine with rotor

JA25.50) to obtain soluble proteins (S1). S1 containing Mth Cas3 was passed through a heparin column in buffer C and did not bind: the flow-through was bound on to a HiTrap Q-Sepharose column and eluted at approximately 600 mM NaCl. Mth Cas3 was pooled and loaded directly on to a 10 ml phenyl-Sepharose column from which Mth Cas3 eluted when the column was washed with water. Mth Cas3 was then dialysed into buffer C for storage.

Cloning, overexpression and purification of *E. coli* CasC, CasD and CasE proteins

E. coli casC (*ygcJ*), *casD* (*ygcI*) and *CasE* (*ygcH*) were cloned by PCR from MG1655 genomic DNA using primers shown in Supplementary Table S1. Primers for *casC* included restriction sites for EcoRI and BamHI for cloning PCR products into pT7-7, giving plasmid pEB488. Primers for *casD* and *casE* included restriction sites for NdeI and XhoI, or NdeI and EcoRI respectively for cloning of PCR products into pET14b giving plasmids pSDC3 and pSDC4 respectively.

CasC was overexpressed from pEB488 in BL21 Codon Plus in flasks containing 400 ml of MU broth with ampicillin (50 µg/ml) and chloramphenicol (20 µg/ml) grown at 37°C. At $D = 0.5$, CasC expression was induced with IPTG (0.5 mM) for 2 h. Cells were harvested, resuspended in <5 ml of buffer CP (20 mM Tris/HCl, pH 8.0, 1 mM EDTA, 1 mM DTT, 0.1 mM PMSF and 100 mM potassium acetate) and frozen at -80°C . Thawed cells were lysed by sonication generating soluble material (S1) that was recovered by centrifugation (20000 rev./min for 30 min in an Avanti J25 machine with rotor JA25.50). CasC was purified by FPLC at room temperature, with protein retrieval monitored by SDS/PAGE and Coomassie Blue staining. S1 was passed through a tandem arrangement of a 5 ml HiTrap Q-Sepharose column linked to a 5 ml HiTrap SP-Sepharose column and this array was developed using a gradient of potassium acetate (0.1–2.0 M), with CasC eluting at 500–750 mM. CasC fractions were dialysed into buffer CP and loaded on to a 1 ml HiTrap heparin column developed with potassium acetate (0.1–1.5 M), CasC eluting at approximately 550–700 mM. CasC fractions were pooled, dialysed into buffer CP, and loaded on to an S200 gel-filtration column equilibrated in buffer CP. Pooled fractions containing CasC were dialysed into buffer CP containing 30% glycerol for storage.

CasD and CasE were purified as hexahistidine-tagged proteins from biomass overexpressed and harvested as described for CasC. Lysis of thawed cells by sonication provided CasE that was present in soluble material (S1), but CasD was insoluble (P1). Hexahistidine-tagged CasE was purified from S1 by using the HisGraviTrap procedure (GE Healthcare) and purified CasE was assessed by SDS/PAGE prior to storage after dialysis into buffer C containing 25% glycerol. Insoluble (P1) CasD was solubilized and refolded on a HisGraviTrap column: samples were washed with buffer C containing 1 M NaCl and the pellet resulting from the washing (P2) was recovered by centrifugation at 18000 rev./min in a JA-25.50 rotor (Beckman Coulter). P2 was then dissolved by gentle stirring at room temperature in 6 M urea and the solution was clarified by centrifugation, giving soluble material that was loaded directly on to a nickel-charged HisGraviTrap column.

Other proteins in the present study

Purified *E. coli* MBP for negative control reactions was bought from New England Biolabs, as were RNase H, exonuclease I and enzymes for molecular cloning (Vent DNA polymerase, T4 DNA ligase, restriction endonucleases, Antarctic phosphatase) and end-

labelling of nucleic acids (T4 polynucleotide kinase). RecG was a gift from Professor Bob Lloyd (School of Biology, Centre for Genetics and Genomics, University of Nottingham, Nottingham, U.K.). T3 and T7 RNA polymerases for *in vitro* transcription of crRNA were bought from Invitrogen. Hel308 protein, used as controls for helicase and ATPase reactions, was purified as described previously [38].

DNA substrates

DNA substrates used in the present study are summarized in Table 1 and detailed in the Supplementary Experimental section and Supplementary Table S2 (at <http://www.BiochemJ.org/bj/439/bj4390085add.htm>). Oligonucleotides were purchased from MWG and were used to generate Cas3 substrates as detailed in [39]. Plasmid DNA used in R-loop assays was generated from *E. coli* strain DH5 α . Any ssDNA in these plasmid preparations was controlled for by digestion with exonuclease I, as described in the Results section. In reactions using linearized plasmids, plasmid was incubated with XbaI for 1 h and then heated to inactivate XbaI. R-loop-forming reactions described in the present study used radiolabelled RNA molecules comprising sequences of *E. coli* CRISPR-1 (crRNA) that were complementary to target DNA. The cloning fragments of *E. coli* CRISPR-1 locus into pBluescript or pT7-7 generated target DNA. Details of how these substrates were generated are listed in the Supplementary Experimental section.

RNA substrates

RNA substrates (crRNA) were generated in the present study by *in vitro* transcription, and their DNA templates are summarized in Table 1 and detailed in Supplementary Table S2. RNA substrates ranged from 250 to 900 nucleotides in length, comparable with the 600-nt crRNA expressed with Cas3 and Cascade during anti-viral defence described previously [3].

R-loop assays

R-loop reactions were in 20 µl volumes for the times stated in the Figures, at 37°C for *E. coli* Cas3 or 44°C for Cas3 from the moderately thermophilic archaeon Mth. Standard reactions contained Cas3 protein, 5 mM magnesium chloride, 5 mM DTT, 7% glycerol, plasmid DNA (300 ng) and radiolabelled RNA (25 ng) at pH 7.5 in Tris/HCl and RNase-free water. Reactions were terminated by the addition of proteinase K (5 µl of 10 mg/ml stock) and EDTA (5 mM). Reactions were loaded directly on to 1.2% agarose gels for electrophoresis in Tris/borate/EDTA buffer (45 mM Tris/borate and 1 mM EDTA) for approximately 3 h at 50 V. Gels were dried on a heated bed with vacuum suction and laid down for phosphorimaging on a Fuji FLA3000 machine.

R-loops were recovered from de-proteinized Cas3 reactions by precipitation at -20°C for at least 2 h in ammonium acetate (1 M) and isopropyl alcohol (50%, v/v), using glycogen (1 µg/µl) as a carrier. The nucleic acid pellet was resuspended in buffer for RNaseH digestion (New England Biolabs) or helicase buffer (5 mM MgCl₂, 5 mM ATP and 10 mM Tris/HCl, pH 7.5) for dissociation by RecG. RNase H and RecG reactions were for 30 min at 37°C.

In vivo phage protection assays

We used *E. coli* strain MG1655, and a modified version of the method described for *E. coli* strain W3110 described previously [3], in which crRNA, Cas3 and Cascade were co-expressed from

Table 1 Summary of DNA and RNA substrates used in the present study

Further details are given in Supplementary Table S2 (at <http://www.BiochemJ.org/bj/439/bj4390085add.htm>).

Substrate	Description
crRNA3	Single-strand RNA (269 nucleotides) comprising 61 nucleotides of <i>E. coli</i> CRISPR-1 and 208 nucleotides of pBluescript
pJLH13	Target DNA plasmid for crRNA3. Is 100% complementary to only 61 bp in the CRISPR part of crRNA3.
pJLH14	Target DNA plasmid for crRNA3. Is 100% complementary with all 269 bp of crRNA3.
crRNA2	Single-strand RNA (1 kb nucleotide) comprising the entire <i>E. coli</i> CRISPR-1 and 110 nucleotides derived from pBluescript transcription.
pJRW2	Target DNA plasmid for crRNA2 and is 100% complementary to crRNA2
R-loop	A model R-loop: duplex DNA containing a 21 nucleotide mismatch into which a 21 nucleotide RNA strand is base paired.
ELB58B	ssDNA that anneals to ϕ X174 virion ssDNA

three plasmids. crRNA was expressed from pWUR478 (a gift from Professor John van der Oost, Laboratory of Microbiology, University of Wageningen, Wageningen, The Netherlands). *cas3* or K78L *cas3* and the cascade operon were cloned in the present study into pRSF-1b (pEB547 and pEB548) or pET-Duet (pEB550) respectively. Cloning procedures for these plasmids are detailed in the Supplementary Experimental section with primer details given in Supplementary Table S1.

For plaque assays, BL21 DE3 cells, which lack CRISPR/*cas*, were transformed with pEB550, pWUR478 and pEB547 or pEB548. Overnight cultures were inoculated in TB broth containing maltose (0.02%), ampicillin (50 μ g/ml), kanamycin (30 μ g/ml) and chloramphenicol (15 μ g/ml). Cultures were grown to $D = 0.5$ and then 500 μ l of each culture was mixed with 3 ml of 0.6% LB (Luria–Bertani) agar before pouring over LB agar plates containing each of the three aforementioned antibiotics and either IPTG (0.6 mM) or no IPTG as indicated. Serial dilutions (10^{-2} to 10^{-7}) of stock phage λ vir were applied as 10 μ l spots to the soft agar top and incubated at 30 °C, as shown in Figure 4(C).

Malachite Green ATPase assays

ATP hydrolysis to phosphate by Cas3 and Cas3 K320L was measured by absorbance at 660 nm in Malachite Green reporter assays and compared with reactions containing no protein. For Cas3, the magnesium chloride concentration was kept constant at 5 mM and ATP (or other NTP) was titrated from 1 to 10 mM. Figure 5(A) shows optimal ATPase activity against time arising from Cas3, in 4 mM ATP:2 mM Mg^{2+} at 37 °C. A phosphate standard curve was used to measure nmol of phosphate produced per nmol of Cas3.

Helicase and nuclease assays

R-loop helicase reactions were performed at 37 °C for *E. coli* Cas3 or 44 °C for Mth Cas3 in a 20 μ l final reaction volume. Standard reactions contained the indicated Cas3 protein, 5 mM magnesium chloride, 5 mM DTT, 7% glycerol and R-loop substrate (5 ng). Reactions contained either 5 mM ATP or 5 mM ATP[S] (adenosine 5'-[γ -thio]triphosphate) and were performed at pH 7.5 in Tris/HCl, and were stopped by the addition of proteinase K and EDTA before being loaded on to 10% acrylamide TBE gels.

The reactions for nuclease digestion by *E. coli* Cas3 contained either 32 P end-labelled ssDNA or dsDNA (5 ng), or M13 or ϕ X174 circular ssDNA (200 ng), as indicated in the Supplementary Figures. The reactions contained $MgCl_2$ from 1–10 mM and were incubated at 37 °C for 10 min to 3 h. For short linear duplex substrates (Table 1), products of exonuclease or endonuclease digestion were detected by deproteinized reactions running on

15% acrylamide TBE gels and phosphorimaging. For nicking reactions, products were mixed with formamide-loading dye, heated at 75 °C for 5 min and loaded on to 15% acrylamide TBE gels containing 4 M urea.

RESULTS

E. coli Cas3 promotes ATP-independent annealing of complementary nucleic acids

Cas3 encoded within CRISPR-1/Cas of *E. coli* K-12 MG1655 (Figure 1A) has amino acid sequence motifs characteristic of HD-superfamily proteins [35] and ATPases of superfamily 2 (SF2) helicases (Figure 1B). *E. coli* Cas3 has been implicated as an effector of CRISPR protection against λ phage by its actions with the Cascade protein complex (encoded by *casA–E*, Figure 1A) [3]. We investigated the biochemical activities of purified *E. coli* Cas3 fused at its N-terminus to *E. coli* MBP (MBP–Cas3, 143 kDa; Figure 1C). Fusion of Cas3 to MBP overcame significant problems encountered when trying to generate stable Cas3 amenable to biochemistry.

ATP-dependent displacement of 20–22 nt oligonucleotides paired to M13 ssDNA has been described previously for Cas3 from *S. thermophilus* [36]. In the present study, we confirm a helicase activity for Cas3 using the *E. coli* enzyme on R-loop substrates, described below (see also Figure 5). In addition, we observed that Cas3 promoted ATP-independent annealing of a 70 nt DNA strand to complementary sequence in ϕ X174 ssDNA (Supplementary Figure S1A at <http://www.BiochemJ.org/bj/439/bj4390085add.htm>). In these reactions, Cas3 could anneal 70% of free radiolabelled oligonucleotide compared with 13% spontaneous annealing in the absence of Cas3 (Supplementary Figure S1B). One interpretation of these observations is that *E. coli* Cas3 has biochemical activities additional to ATP-dependent helicase unwinding. Annealing by Cas3 bore similarities to that reported for other helicases, most notably those of the RecQ family [40–43]. We wanted to test this activity of Cas3 in more detail, using CRISPR RNA and duplex DNA substrates, which may be more relevant to the function of *E. coli* CRISPR/Cas than M13 ssDNA or ϕ X174 ssDNA.

Cas3 annealing activity base pairs RNA into complementary duplex DNA

We investigated whether *E. coli* Cas3 promoted ATP-independent annealing of RNA into complementary sequence within duplex DNA. To monitor these reactions, radiolabelled RNA was generated by *in vitro* transcription of *E. coli* CRISPR-1 fragments cloned into pBluescript, summarized in Table 1. In one set of reactions, Cas3 (50–800 nM) was incubated for 45 min in buffered

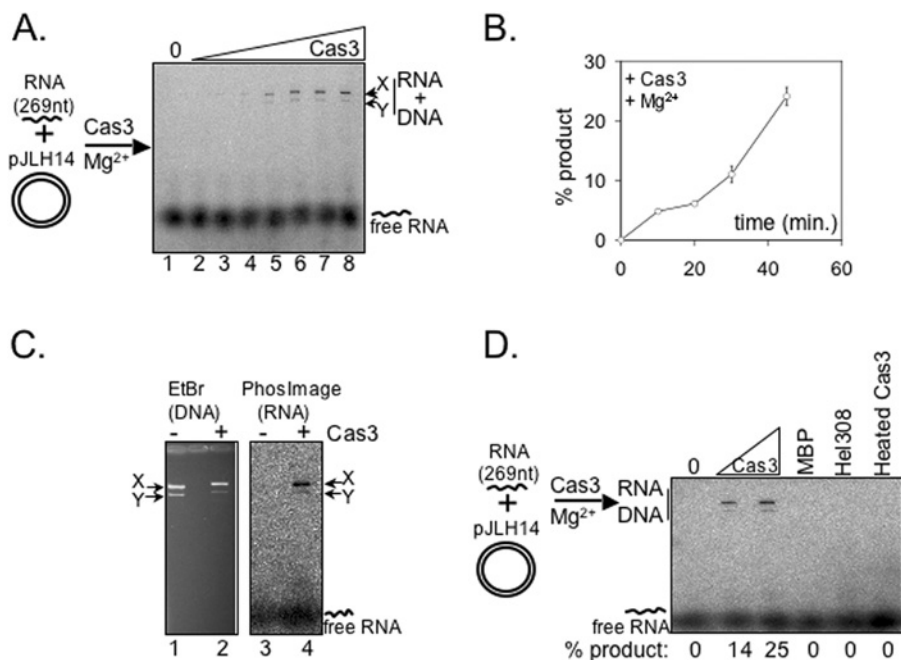


Figure 2 *E. coli* Cas3 promotes ATP-independent association and annealing of RNA with complementary dsDNA

(A) Phosphorimage of Cas3 reactions after agarose gel electrophoresis to detect radiolabelled RNA. Reactions contained purified Cas3 (0, 50, 100, 200, 300, 400, 500 and 800 nM), 5 mM MgCl₂, crRNA3 and uncut plasmid pJLH14. RNA and DNA substrates are summarized in Table 1. Deproteinized reactions gave two RNA products labelled X and Y. (B) Time-course reactions of free crRNA3 converted into product with pJLH14 by Cas3 (200 nM), measured using ImageQuant. The RNA product was expressed as a proportion of total RNA detected in each lane. Reactions were in duplicate and the graph shows means \pm S.E.M. (C) Agarose gels monitoring positions of plasmid DNA and radiolabelled RNA after deproteinizing identical reactions that contained crRNA3, pJLH14 added to Cas3 (300 nM, lanes 2 and 4) or without Cas3 (lanes 1 and 3). Lanes 1 and 2 are stained with ethidium bromide (1 μ g/ml; EtBr), showing slower mobility of plasmid DNA in the presence of Cas3 and crRNA3 (lane 2). Lanes 3 and 4 show identical reactions after phosphorimaging to detect radiolabelled RNA. RNA was detected as the slowest migrating product only in the reaction containing Cas3 (lane 4), in which the RNA product was positioned corresponding exactly to DNA in lane 2. (D) Products of mixing crRNA3 and pJLH14 in magnesium with Cas3 (200, 400 nM) or 'mock' proteins, MBP (400 nM), the superfamily 2 helicase Hel308 (400 nM) or Cas3 (400 nM) which had been pre-heated to 85 °C for 10 min. The amount of RNA product is given below the panel as a percentage of the total RNA present in each reaction.

magnesium chloride (5 mM) with 269-nt RNA (25 ng) that was 100% complementary in sequence to part of the uncut plasmid DNA (pJLH14, 250 ng). Reactions were deproteinized and products were separated by electrophoresis to detect positions of radiolabelled RNA by phosphorimaging. Two products containing RNA migrated much more slowly than free RNA, their formation corresponding to increasing Cas3 (Figure 2A, lanes 4–8). Using a fixed Cas3 concentration converted approximately 20% of free RNA into product in 45 min (Figure 2B, but see also Figure 3). Products were consistent with Cas3 promoting association of RNA with plasmid DNA in a joint molecule that was stable during migration through agarose gels. This observation was supported by comparing positions of ethidium bromide-stained plasmid DNA with radiolabelled RNA in agarose gels of identical de-proteinized reaction products (Figure 2C). Plasmid DNA in reactions lacking Cas3 (Figure 2C, lane 1) or with Cas3 added (Figure 2C, lane 2) migrated as two species, but DNA in the Cas3 reaction was slower migrating than when Cas3 was absent. This slower migrating DNA (Figure 2C, lane 2) corresponded exactly to the position of radiolabelled RNA from an identical reaction imaged to detect RNA (Figure 2C, lane 4). We concluded that Cas3 promoted association of RNA with plasmid DNA.

Control reactions were used to ensure that formation of the observed RNA–DNA products was specific to Cas3. Reactions containing 'mock' proteins, MBP, Hel308 helicase [38] or Cas3 that had been pre-heated to 85 °C showed no detectable RNA–DNA product (Figure 2D). Further support for ATP-independent RNA–DNA products of Cas3 was obtained by substituting *E. coli* Cas3 with Cas3 purified

from the archaeon Mth (Supplementary Figures S2A–S2C at <http://www.BiochemJ.org/bj/439/bj4390085add.htm>). Mth Cas3 also promoted formation of the RNA–DNA product (Supplementary Figure S1C), indicating that this Cas3 activity is evolutionarily conserved.

Cas3 also promoted formation of RNA–DNA products in reactions containing RNA mixed with uncut plasmid DNA molecules in which sequence complementarity was 60 nt (Supplementary Figure S3A, lanes 1–8 at <http://www.BiochemJ.org/bj/439/bj4390085add.htm>) or 1000 nt (Supplementary Figure S3A lanes 9–14). In all of these assays, RNA was generated by *in vitro* transcription to generate sense strand (+) RNA from cloned *E. coli* CRISPR-1 fragments, as described in Table 1. Cas3 also promoted RNA–DNA product formation from RNA sequences that were antisense (–) CRISPR-1 RNA, or from RNA sequences unrelated to CRISPR-1 (Supplementary Figure S3B), indicating that Cas3-promoted ATP-independent annealing of RNA to DNA was independent of nucleic acid sequence or RNA structure specificity.

Cas3 anneals RNA into duplex DNA to form R-loops, and is most efficient when target duplex DNA is linearized

When uncut plasmid was used as a target for annealing complementary RNA by Cas3, we observed approximately 20% of free RNA converted into product. This activity was improved 3–6-fold if plasmid was linearized prior to mixing with RNA and Cas3, measured with fixed Cas3 concentrations (Figure 3A and 3B). A single reaction product was observed from linearized

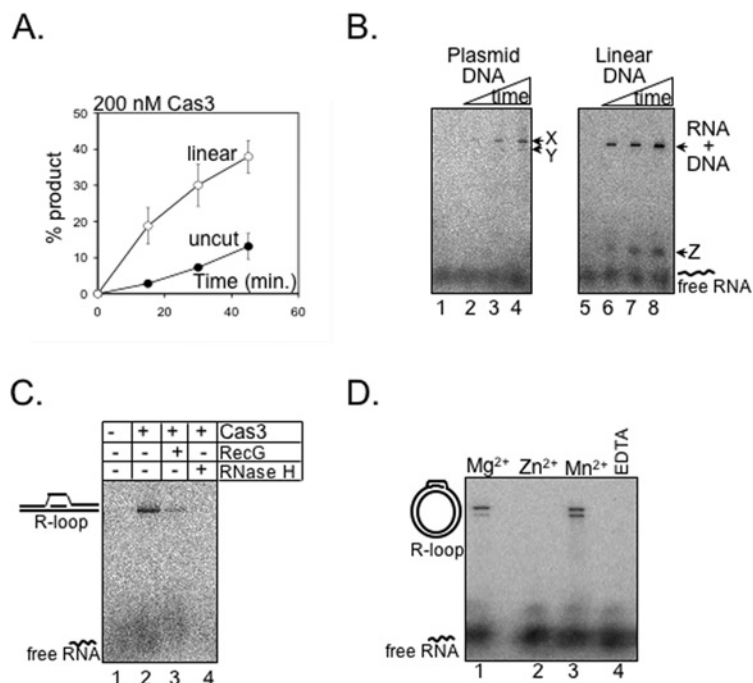


Figure 3 ATP-independent RNA–DNA products of Cas3 are R-loops most efficiently formed with linearized plasmid using magnesium as a co-factor

(A) Graph comparing RNA–DNA product formed by fixed amounts of Cas3 (200 nM) as a function of time in reactions containing radiolabelled crRNA and magnesium (5 mM), and pJLH14 that was either uncut or linearized, as indicated. Reactions were in duplicate, quantified by ImageQuant software calculating RNA product as a proportion of total RNA for each lane. (B) Phosphorimages of one set of reactions from (A) showing products from uncut pJLH14 (lanes 1–4) compared with linearized pJLH14 (lanes 5–8). A new product is present in reactions containing linearized DNA (labelled Z). (C) R-loop formation by Cas3. Nucleic acid products from Cas3 (200 nM) RNA–DNA annealing reactions (crRNA3 + pJLH14) had been isolated prior to incubation with RecG (+ MgCl₂ and ATP) or RNaseH. R-loop was substantially reduced after RecG treatment and RNA is almost entirely absent after RNaseH treatment, consistent with Cas3 promoting formation of stable RNA–DNA hybrids. (D) ATP-independent Cas3 R-loop-forming reactions require magnesium or manganese ions as a co-factor. Reactions contained Cas3 (200 nM), crRNA3 and uncut pJLH14 DNA with either metal ions (5 mM) or 2 mM EDTA, as indicated.

plasmid DNA compared with two from uncut plasmid (Figure 3B, compare lanes 4 and 8). We also observed a new product containing radiolabelled RNA that migrated close to free RNA (Figure 2C, lane 6, labelled Z).

We next asked whether the RNA–DNA product promoted by Cas3 was an R-loop: an RNA strand invaded into duplex DNA by base pairing to complementary sequences displacing a ‘loop’ of ssDNA. R-loops are stable, consistent with the observed migratory properties of Cas3 reaction products. We assessed directly whether RNA–DNA products of Cas3 reactions might be R-loops using RNase H and RecG, two bacterial enzymes that degrade or dissociate R-loops [44,45]. Cas3 reactions set up as in Figure 3(B) were deproteinized and nucleic acid products were recovered by precipitation. The precipitation step was required because we observed RNase H to be inactive when added directly to Cas3 R-loop reactions. RNA–DNA product could be recovered by precipitation (Figure 3C, lane 2), and this product was partially dissociated by RecG (in the presence of ATP), re-generating free radiolabelled RNA (Figure 2D, lane 3). Most significantly, products of Cas3 reactions were substrates for RNaseH, which efficiently degraded RNA within them, consistent with RNA base-paired to complementary DNA (Figure 3C, compare lanes 2 and 4). Taken together, these data are consistent with an activity of Cas3 promoting the formation of hydrogen-bonded base pairs between complementary RNA invaded into duplex DNA as R-loops.

We tested the possibility that Cas3 annealed RNA only to regions of complementary ssDNA that can arise during preparation of plasmid DNA, although this seemed unlikely considering the observed enhancement of product formation by

linearizing plasmids shown in Figure 3(B). However, control reactions were treated with exonuclease I to remove ssDNA in the target plasmid DNA substrate, followed by incubation of this treated DNA with RNA and Cas3. Treatment of plasmid with exonuclease I had no effect on product formation by Cas3 (Supplementary Figure S4A, compare lanes 2 and 3 at <http://www.BiochemJ.org/bj/439/bj4390085add.htm>). We also examined the possibility that the nuclease activity of *E. coli* Cas3 could facilitate spontaneous pairing of RNA to ssDNA. Despite efforts using several substrates and a variety of incubation conditions, we were unable to detect exo- or endo-nuclease activity of *E. coli* Cas3 on plasmid DNA, M13 closed covalent ssDNA or on end-labelled model DNA substrates analysed on neutral or denaturing gels (summarized in Supplementary Figure S4B–S4D). This indicated that Cas3 could promote the ATP-independent annealing of RNA substrates into complementary duplex DNA (R-loop formation) without prior detectable endo- or exo-nuclease processing of DNA.

R-loop formation by *E. coli* Cas3 requires magnesium and a HD amino acid motif

Cas3 R-loop formation required magnesium or manganese as a co-factor, but was not supported by zinc or in EDTA (Figure 3D). To explore this requirement of Cas3, we focussed on its HD motif (Figure 1B), which binds to metal ions in HD superfamily proteins of diverse functions [46–48]. We purified Cas3 proteins with amino acid substitutions in the HD motif (D75G and K78L, Figure 1B) and measured their ability to promote R-

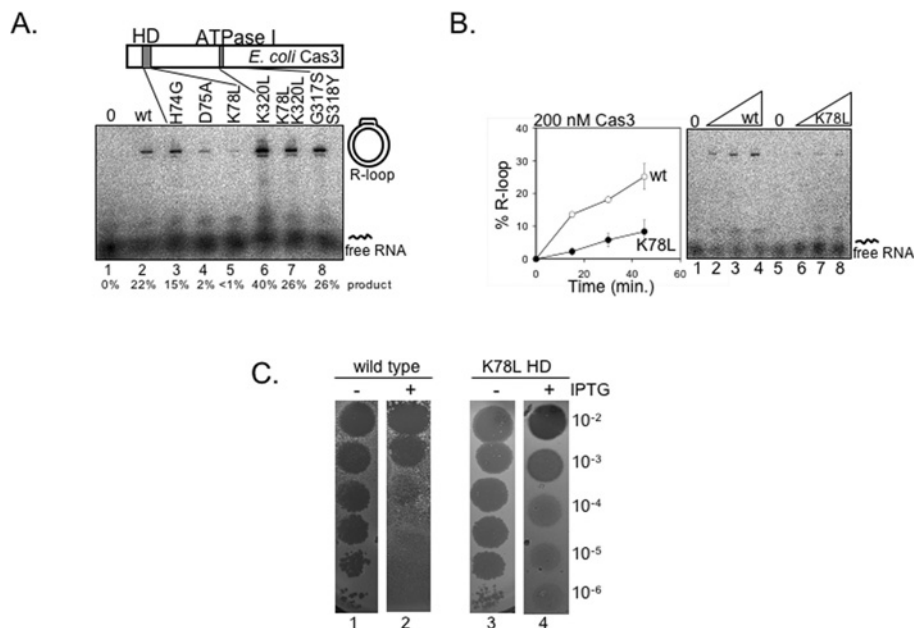


Figure 4 Activity of Cas3 HD motif mutant proteins *in vitro* and *in vivo*

(A) Cas3 with amino acid substitutions in the HD motif (H74G, D75A or K78L, lanes 3–5), or in ATPase motif I (K320L, K78L/K320L and G317SS318Y, lanes 6–8) were assessed for their ability to form R-loops at a fixed concentration (200 nM) compared with wild-type Cas3 (200 nM, lane 2). Reactions contained crRNA3, uncut pJLH14 and magnesium. RNA product was quantified in ImageQuant as a proportion of the total RNA for each lane, using free RNA and product in lane 1 (zero Cas3) as the baseline activity. The amount of free RNA converted into R-loop product is given beneath each lane. (B) Graph and corresponding agarose gel showing R-loop formation as a function of time by using 200 nM wild-type Cas3 compared with 200 nM K78L HD motif mutant Cas3 (gel lanes 4–8) in otherwise identical reactions containing radiolabelled crRNA3 with linearized pJLH14 DNA and 5 mM magnesium chloride. Values are means \pm S.E.M. ($n = 2$). (C) Assays comparing wild-type and K78L Cas3 for their ability to contribute to protection of *E. coli* cells against plaque formation by phage λ .vir. Phage was diluted from stock as shown to the right of the panels and spotted on to a lawn of *E. coli* BL21 DE3 containing plasmids encoding both Cascade (pEB550) and crRNA (pWUR478) plus either pEB547 (wild-type Cas3, lanes 1 and 2) or pEB548 (K78L Cas3, lanes 3 and 4). Expression of Cascade, Cas3 and crRNA from plasmids to effect protection was induced by adding IPTG (0.6 mM) to plates (lanes 2 and 4).

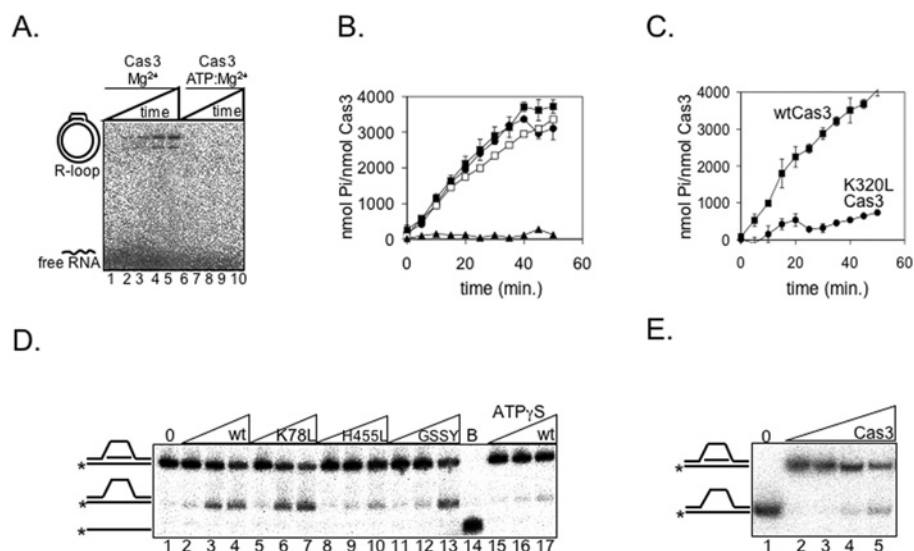


Figure 5 Helicase unwinding of R-loop by *E. coli* Cas3

(A) Phosphorimage of reactions incubating Cas3 (200 nM) with crRNA3 and uncut pJLH14 in magnesium (lanes 1–5) or magnesium and ATP (5:5 mM, lanes 6–10). Samples from a reaction mixture were removed at various time points (5, 15, 25 and 45 min) and deproteinized. (B) ATPase activity of *E. coli* Cas3. Values are means \pm S.E.M. ($n = 2$). Reactions contained 2 mM magnesium and 4 mM ATP, and either no nucleic acid (\square), 1 μ g of ϕ X174 ssDNA (\blacksquare) or 1 μ g of ϕ X174 duplex DNA (\bullet). ATPase activity was calculated using Malachite Green dye as a reporter for liberation of free phosphate from ATP. The reaction lacking Cas3 was a control for zero ATPase activity at each time point (\blacktriangle). (C) Cas3 with an amino acid substitution in ATPase motif I (K320L, \bullet , see Figure 1B) is defective in ATPase activity when compared with wild-type protein (\blacksquare). (D) Products of incubating *E. coli* Cas3 (10, 100 and 300 nM) with R-loop substrate (1 nM), annotated panel left, end-labelled with 32 P on one strand as indicated (*) in buffer containing 5 mM ATP and 5 mM magnesium (lanes 1–13), or ATP[S] (5 mM) instead of ATP (lanes 15–17). Lane 1 contained zero Cas3 and lane 14 shows boiled (B) substrate. Unwinding activity of wild-type (wt) Cas3 (lanes 2–4) was compared with that of Cas3 proteins mutated in the HD motif (K78L, lanes 5–7), or ATPase motif II (WalkerB/DExH box, H455L, lanes 8–10), or ATPase motif I (Walker A/P loop, G317SS318Y, labelled GSSY, lanes 11–13). (E) This panel is equivalent to reactions shown in lanes 2–5 of (C), containing 0, 10, 100 or 300 nM Cas3 protein as indicated, with reaction products compared with a marker molecule (lane 1) of R-loop lacking the invading strand as indicated to the left of the panel.

loop formation. Cas3 D75G and K78L were less effective at forming R-loops than wild-type protein in reactions containing either uncut plasmid DNA (Figure 4A, compare lane 2 with lanes 4 and 5) or linearized plasmid (Figure 4B). This indicated the importance of the HD motif to this Cas3 function, and confirmed that a trace contaminant of *E. coli* Cas3 purification is unlikely to cause the observed R-loop-forming activity. Similar to wild-type Cas3, binding of K78L Cas3 to duplex DNA was detected in EMSAs (electrophoretic mobility-shift assays; Supplementary Figure S4E), indicating that the K78L amino acid substitution had not caused a more general misfolding or structural aberration that might otherwise account for reduced catalysis. Amino acid substitutions to the Cas3 ATPase motif I/Walker A box (K320L and G317S/S318Y, Figure 1B) did not reduce R-loop formation (Figure 4A, lanes 6 and 8). Interestingly, Cas3 protein with combined substitutions in both Lys⁷⁸ and Lys³²⁰ (K78L/K320L) was proficient at forming R-loops (Figure 4A, lane 7), an unexpected observation given reduced levels of R-loop formation catalysed by K78L Cas3.

Mutation of Cas3 HD motif reduces protection of cells against phage λ vir *in vivo*

E. coli cells can be protected against lysis by phage λ vir using inducible plasmid-based co-expression of Cas3, Cascade and crRNA engineered with λ vir sequence [3,20]. We also observed a protective effect using similar plasmid co-expression of wild-type Cas3, Cascade and crRNA (summarized in Figure 4C). In these assays, serial dilutions of phage λ vir suspensions were applied to a lawn of *E. coli* cells to observe plaque formation (Figure 4C, lane 1) that was reduced 100–1000-fold upon IPTG-induced expression of crRNA, Cas3 and Cascade (Figure 4C, lane 2). Expression of Cas3 K78L HD mutated protein instead of wild-type Cas3 in the same assay (Figure 4C, lanes 3 and 4) led to a 100-fold greater sensitivity to plaque formation. This confirmed the importance of the HD-motif to Cas3 function and its contribution to protection against the phage observed in this assay.

E. coli Cas3 helicase activity dissociates R-loops

Addition of ATP to Cas3 R-loop reactions correlated with a total absence of R-loop product, compared with reactions containing magnesium only (Figure 5A, compare lanes 2–5 with lanes 6–10). This could be explained by ATP-dependent helicase unwinding by Cas3 in dismantling R-loops. To investigate this, we first assayed for the ATPase activity of purified *E. coli* Cas3. ATP (but not GTP, CTP or TTP) was hydrolysed by Cas3 in the absence of nucleic acids, and pre-incubation of Cas3 with ϕ X174 ssDNA or ϕ X174 dsDNA had no stimulatory effect on ATPase activity (Figure 5B). Hel308, a ssDNA-stimulated ATPase helicase [38], was used as a positive control for DNA-stimulated ATPase activity. ATP hydrolysis was most effective at a 2:1 molar ratio of ATP/magnesium. Cas3 K320L, mutated in ATPase motif I, had 2–4-fold reduced ATPase activity when compared with wild-type Cas3 at fixed protein concentrations as a function of time (Figure 5C), as expected and in agreement with *S. thermophilus* Cas3 data [36].

To test directly whether ATPase activity of Cas3 could drive helicase unwinding of an R-loop, we generated a model substrate from oligonucleotides summarized in Table 1 and detailed in Supplementary Table S1. Incubation of this R-loop with Cas3, ATP and magnesium gave a product (Figure 5D, lanes 2–4) that was barely detectable when ATP was substituted for poorly

hydrolysable ATP[S] (lanes 15–17), or if Cas3 was disrupted in ATPase motif II (Walker B/DExH box, H455L; Figure 5D, lanes 8–10). Helicase activity of Cas3 was unaffected by mutation in the HD motif (Figure 5D, lanes 5–7), and was moderately reduced by mutation of glycine and serine residues in ATPase motif I (Walker A, 'GSSY'; Figure 5D, lanes 11–13). The product formed by Cas3 in these helicase reactions was DNA lacking the 'invading' RNA, indicated by comparing its migration through gels with a marker substrate also lacking the RNA strand (Figure 5E). Purified archaeal Mth Cas3 also unwound the R-loop substrate in the same way (Supplementary Figure S2D), indicating that this is a conserved function of Cas3 across domains of life, in addition to ATP-independent R-loop formation. The ability of *E. coli* Cas3 to dissociate a model R-loop substrate extends the known helicase activity of Cas3, but raises the question of how Cas3 catalyses apparently antagonistic functions of forming and dissociating R-loops, discussed below.

E. coli Cascade complex also promotes R-loop formation

Expression of Cascade was essential for the protection of *E. coli* cells against phage λ vir, described in Figure 4(C) and previously [3], and a recent study showed that R-loop formation by purified *E. coli* Cascade complex is independent of ATP [20]. In agreement with this, we observed that a mixture of purified CasC, CasD and CasE proteins, but not individual proteins, promoted formation of R-loop products indistinguishable from those described for Cas3 (Figure 6). This indicated that the crRNA substrates we *in vitro* transcribed for the present study are relevant for recognition not just by Cas3, but also for the key CRISPR effector complex Cascade. This is also consistent with the expression of 600-nt crRNA that was used with Cascade and Cas3 to effect *in vivo* protection of cells against λ vir, described previously [3] and also shown in the present study. These partly overlapping functions of Cas3 and Cascade promoting R-loop formation are discussed below.

DISCUSSION

Cas3 R-loop formation and the HD domain

E. coli Cas3 and Cascade (CasABCDE) participate in processing crRNA and complementary duplex DNA sequences during interference stages of CRISPR immunity. Models of Cascade and Cas3 function place crRNA processing by Cascade upstream of Cas3. Helicase activity of *S. thermophilus* Cas3 can unwind duplexes formed from ssRNA (single-stranded RNA) and ssDNA, and Cas3 nuclease activity is proposed to destroy invading DNA [36,49]. Helicase data shown in the present study supports a model for Cas3 helicase processing, but goes further by showing that Cas3 can dissociate an R-loop substrate of ssRNA invaded into a duplex of DNA, and does so by removing the invaded RNA strand. We also report a novel activity of Cas3: ATP-independent annealing of RNA into duplex DNA to form an R-loop. Helicase and R-loop forming activities were also observed using Cas3 from the archaeal species Mth, supporting an evolutionarily conserved role for each Cas3 function. R-loops have been implicated in several areas of nucleic acid metabolism, including replication of mitochondrial genomes [50] and T4 phage [51], recombination [52], and control of DNA topology [53]. In each case a stable RNA–DNA hybrid is required, formed from duplex DNA and ssRNA, as has been proposed for targeting reactions in CRISPR immunity.

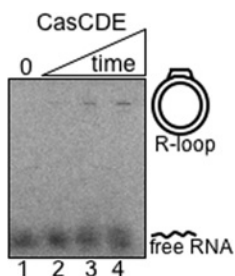


Figure 6 R-loop formation by Cascade

Phosphorimage of time-course reactions incubating purified CasC, CasD and CasE proteins with crRNA3 and pJLH14 in buffer containing 5 mM magnesium. CasC (100 nM), CasD (100 nM) and CasE (100 nM) were pre-incubated together for 10 min at room temperature prior to mixing with RNA and DNA for 10, 30 and 60 min before stopping and deproteinizing reactions for visualizing RNA.

R-loop formation by *E. coli* Cas3 required magnesium as a co-factor and a functional HD motif. These pre-requisites for annealing and R-loop formation by Cas3 indicate a catalytic mechanism rather than a passive effect of Cas3 binding to complementary nucleic acids and promoting their annealing by simply bringing them into proximity. Since HD domains were first described [35] they have been characterized as phosphatases and nucleotidases acting on nucleotide substrates, requiring binding to a metal ion co-factor [47,54–56]. HD motifs of Cas3 proteins support different activities: the *S. thermophilus* Cas3 HD domain was required for robust nuclease degradation on M13 ssDNA substrate, but no nuclease activity was detected on uncut dsDNA plasmid [36]. In contrast, *S. solfataricus* protein Sso2001, which is homologous with the HD domain of fused HD-helicase Cas3, showed nicking activity on dsDNA and dsRNA (double-stranded RNA) but not ssDNA [37]. These differing nuclease activities suggest remarkable variability in substrate specificities of Cas3 HD motifs from different organisms. Amino acid sequence alignments of predicted HD domains of Sso2001 and Cas3 from *S. thermophilus* and *E. coli* do not reveal any obvious differences in conserved residues that are known to be required for catalysis and which might account for these different substrate preferences. It may be that information obtained from atomic structures will illuminate interpretation of substrate preferences that might be controlled by aspects of Cas3 HD domain protein architecture other than individual catalytic residues.

We were unable to detect robust nicking or nuclease activity of *E. coli* Cas3, despite assays in a wide range of buffers and magnesium concentrations analysed on neutral or denaturing gels using substrates similar to those described previously [36,37]. This suggests that the observed R-loop-forming activity of Cas3, although requiring the HD domain, is independent of nuclease activity, at least that which could be detected in the assays in the present study. Our negative outcomes from Cas3 nuclease assays cannot exclude the possibility that *E. coli* Cas3 may have cryptic nuclease activity that is activated *in vivo* by some other factor, or by recognition of a specific DNA/RNA structure. This type of activation has been described for the annealing helicase AH2 [57], and might help to ensure that potentially destructive nuclease activities are not deployed unless strictly targeted. It is unlikely that the N-terminal MBP tag fused to Cas3 impeded nuclease function. Factor Xa digestion of MBP–Cas3 to remove the MBP tag did not reveal a nuclease activity, and in any case R-loop-forming and helicase activities of Cas3 were proficient in the presence of the MBP tag. At this stage there is no understanding of the structure of Cas3 proteins or relative positioning of HD

and helicase domains. However, our observation that the negative effect of HD motif mutants on R-loop formation by Cas3 can be reversed by also mutating the nucleotide-binding ATPase motif I (Walker A) raises an intriguing possibility of communication or other cross-talk between the HD and helicase domains of Cas3. This might indicate that metal binding by the HD domain may be needed for activation of R-loop formation via some kind of structural effect on Cas3 protein. However, we emphasize that this is currently speculative and presentation of detailed atomic structures of Cas3 should lead to major insights into domain functioning.

Roles of Cas3 and Cascade

The importance of the HD motif for Cas3 functioning in protection of *E. coli* cells against plaque formation by λ vir was emphasized by reduced protection of cells expressing K78L Cas3 compared with wild-type Cas3. K78L Cas3 showed 4–6-fold reduced R-loop formation *in vitro*, an observation that is consistent with reduced interference of viral DNA by crRNA *in vivo* and a corresponding increased sensitivity to infection. However, we emphasize that we are not proposing that R-loop formation during CRISPR immunity is exclusive to Cas3, having been demonstrated for Cascade in work appearing during preparation of the present paper [20]. Consistent with this, we also observed R-loops formed by the Cascade CDE complex using the same substrates as those for Cas3. We propose that Cas3 and Cascade overlap in function by their ability to both promote R-loop formation. Cascade catalyses endonucleolytic processing of pre-crRNA to crRNA that is bound stably by the Cascade complex for targeting to complementary DNA sequence. This is likely to control seeding of R-loop formation, but leaves open the question of how Cas3 gains access to an R-loop for its further helicase and other processing. It is clear that Cas3 can recognize R-loops by its ability to unwind them, and it might be that Cascade primes formation of the R-loop that is presented to Cas3 for extension and stabilization. More research will be needed to unravel whether Cascade physically recruits Cas3 or if Cas3 can independently act on an R-loop without association through Cascade. We also note that an RNA–DNA hybrid (R-loop) formed during interference stages of CRISPR immunity would be a provocative substrate for digestion by RNaseH that might need to be protected for effective targeting, a topic that has not yet been explored and may involve a nucleation or protection role for either Cascade or Cas3. Finally, we cannot rule out the possibility that R-loop formation by Cas3 is involved in another aspect of Cas3 function that is not directly part of crRNA–DNA interference stages of CRISPR immunity. This might be hinted at by differing transcriptional control of *cascade* compared with *cas3* in *E. coli*.

Antagonistic roles of Cas3 forming and dissociating similar substrates dependent on ATP status

In line with a helicase activity of Cas3 that removes the RNA strand from an R-loop, we observed that ATP abolished detectable R-loop formation. An important question arising here is how can Cas3 have apparently antagonistic roles in both forming and unwinding R-loops? One possibility is that ATP might be central to Cas3 not only for powering helicase unwinding, but also for some other aspect of Cas3 functioning. Engagement of ATP by Cas3 might switch a change in oligomeric or conformational state of Cas3, in turn controlling biochemical activities. This effect has been reported for Rad52 and RecQ1 proteins, which also promote annealing of complementary nucleic acids in the absence

of ATP, through assembly of higher-order protein oligomers. In the case of RecQ1, these oligomers are dispersed by ATP leading to DNA helicase activity [42]. This type of ATP effect on Cas3, and possibly its interaction with Cascade proteins, might form an interesting avenue for future research.

AUTHOR CONTRIBUTION

Edward Bolt conceived the project and wrote the paper. Experiments were designed and performed by Edward Bolt and Jamieson Howard, and by Ivana Ivančić-Baće and Stephane Delmas for Figures 4 and 5 respectively.

ACKNOWLEDGEMENTS

We thank Alexandra Hughes and Jamie Webster for support, and Peter McGlynn and Mark Dillingham for useful discussion.

FUNDING

This research was funded by a Biotechnology and Biological Sciences Research Council (BBSRC) Ph.D. studentship (to J.L.H.) and by the Croatian Ministry of Science, Education and Sports (grant number 119-1191196-1201 (to I.I.-B.) and an award from Croatian Academy of Sciences and Arts. We are also grateful to The British Council for pump-prime funding the collaboration between E.L.B. and I.I.-B. in 2008.

REFERENCES

- Jansen, R., Embden, J. D., Gastra, W. and Schouls, L. M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712
- Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuys, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V. and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964
- Edgar, R. and Qimron, U. (2010) The *Escherichia coli* CRISPR system protects from lambda lysogenisation, lysogens and prophage induction. *J. Bacteriol.* **192**, 6291–6294
- Garneau, J. E., Dupuis, M. E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A. H. and Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71
- Marraffini, L. A. and Sontheimer, E. J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845
- Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K. A., Djordjevic, M., Wanner, B. L. and Severinov, K. (2010) Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.* **77**, 1367–1379
- Manica, A., Zebec, Z., Teichmann, D. and Schleper, C. (2011) *In vivo* activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Mol. Microbiol.* **80**, 481–491
- Cady, K. C. and O'Toole, G. A. (2011) Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J. Bacteriol.* **193**, 3433–3445
- Zegans, M. E., Wagner, J. C., Cady, K. C., Murphy, D. M., Hammond, J. H. and O'Toole, G. A. (2009) Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. *J. Bacteriol.* **191**, 210–219
- Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A. et al. (2010) A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol. Microbiol.* **79**, 484–502
- Mojica, F. J., Ferrer, C., Juez, G. and Rodriguez-Valera, F. (1995) Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* **17**, 85–93
- Aklujkar, M. and Lovley, D. R. (2010) Interference with histidyl-tRNA synthetase by a CRISPR spacer sequence as a factor in the evolution of *Pelobacter carbinolicus*. *BMC Evol. Biol.* **10**, 230
- Stern, A., Keren, L., Wurtzel, O., Amitai, G. and Sorek, R. (2010) Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* **26**, 335–340
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F. et al. (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477
- Deveau, H., Garneau, J. E. and Moineau, S. (2010) CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* **64**, 475–493
- Karginov, F. V. and Hannon, G. J. (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell* **37**, 7–19
- Vale, P. F. and Little, T. J. (2010) CRISPR-mediated phage resistance and the ghost of coevolution past. *Proc. Biol. Sci.* **277**, 2097–2103
- Terns, M. P. and Terns, R. M. (2011) CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* **14**, 321–327
- Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R. et al. (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536
- Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M. and Terns, M. P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945–956
- Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W. and Doudna, J. A. (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **17**, 904–912
- Beloglazova, N., Brown, G., Zimmerman, M. D., Proudfoot, M., Makarova, K. S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W. et al. (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.* **283**, 20361–20371
- Lillestøl, R. K., Redder, P., Garrett, R. A. and Brugger, K. (2006) A putative viral defence mechanism in archaeal cells. *Archaea* **2**, 59–72
- Tang, T. H., Bachellerie, J. P., Rozhdzestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., Elge, T., Brosius, J. and Huttenhofer, A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7536–7541
- Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. and Doudna, J. A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355–1358
- Carte, J., Wang, R., Li, H., Terns, R. M. and Terns, M. P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433
- Diez-Villasenor, C., Almendros, C., Garcia-Martinez, J. and Mojica, F. J. (2010) Diversity of CRISPR loci in *Escherichia coli*. *Microbiology* **156**, 1351–1361
- Touchon, M. and Rocha, E. P. (2010) The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. *PLoS ONE* **5**, e11126
- Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N. and Wagner, R. (2010) Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* **75**, 1495–1512
- Westra, E. R., Pul, U., Heidrich, N., Jore, M. M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L. et al. (2010) H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* **77**, 1380–1393
- Perez-Rodriguez, R., Haitjema, C., Huang, Q., Nam, K. H., Bernardis, S., Ke, A. and Delisa, M. P. (2011) Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. *Mol. Microbiol.* **79**, 584–599
- Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S. and Shinkai, A. (2009) Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *J. Mol. Biol.* **395**, 270–281
- Aravind, L. and Koonin, E. V. (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem. Sci.* **23**, 469–472
- Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30**, 1335–1342
- Han, D. and Krauss, G. (2009) Characterization of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2. *FEBS Lett.* **583**, 771–776
- Guy, C. P. and Bolt, E. L. (2005) Archaeal Hel308 helicase targets replication forks *in vivo* and *in vitro* and unwinds lagging strands. *Nucleic Acids Res.* **33**, 3678–3690
- Guy, C. P., Haldenby, S., Brindley, A., Walsh, D. A., Briggs, G. S., Warren, M. J., Allers, T. and Bolt, E. L. (2006) Interactions of RadB, a DNA repair protein in archaea, with DNA and ATP. *J. Mol. Biol.* **358**, 46–56
- Cheek, C. F., Wu, L., Garcia, P. L., Janscak, P. and Hickson, I. D. (2005) The Bloom's syndrome helicase promotes the annealing of complementary single-stranded DNA. *Nucleic Acids Res.* **33**, 3932–3941
- Machwe, A., Xiao, L., Groden, J., Matson, S. W. and Orren, D. K. (2005) RecQ family members combine strand pairing and unwinding activities to catalyze strand exchange. *J. Biol. Chem.* **280**, 23397–23407
- Muzzolini, L., Beuron, F., Patwardhan, A., Popuri, V., Cui, S., Niccolini, B., Rappas, M., Freemont, P. S. and Vindigni, A. (2007) Different quaternary structures of human RECQ1 are associated with its dual enzymatic activity. *PLoS Biol.* **5**, e20

- 43 Zhao, X. and Jain, C. (2011) DEAD-box proteins from *Escherichia coli* exhibit multiple ATP-independent activities. *J. Bacteriol.* **193**, 2236–2241
- 44 Nakamura, H., Katayanagi, K., Morikawa, K. and Ikehara, M. (1991) Structural models of ribonuclease H domains in reverse transcriptases from retroviruses. *Nucleic Acids Res.* **19**, 1817–1823
- 45 Vincent, S. D., Mahdi, A. A. and Lloyd, R. G. (1996) The RecG branch migration protein of *Escherichia coli* dissociates R-loops. *J. Mol. Biol.* **264**, 713–721
- 46 Hogg, T., Mechold, U., Malke, H., Cashel, M. and Hilgenfeld, R. (2004) Conformational antagonism between opposing active sites in a bifunctional RelA/SpoT homolog modulates (p)ppGpp metabolism during the stringent response. *Cell* **117**, 57–68
- 47 Yakunin, A. F., Proudfoot, M., Kuznetsova, E., Savchenko, A., Brown, G., Arrowsmith, C. H. and Edwards, A. M. (2004) The HD domain of the *Escherichia coli* tRNA nucleotidyltransferase has 2',3'-cyclic phosphodiesterase, 2'-nucleotidase, and phosphatase activities. *J. Biol. Chem.* **279**, 36819–36827
- 48 Zimmerman, M. D., Proudfoot, M., Yakunin, A. and Minor, W. (2008) Structural insight into the mechanism of substrate specificity and catalytic activity of an HD-domain phosphohydrolase: the 5'-deoxyribonucleotidase YfbR from *Escherichia coli*. *J. Mol. Biol.* **378**, 215–226
- 49 Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copie, V. et al. (2011) Structural and functional characterization of an archaeal CASCADE complex for CRISPR-mediated viral defense. *J. Biol. Chem.* **286**, 21643–21656
- 50 Brown, T. A., Tkachuk, A. N. and Clayton, D. A. (2008) Native R-loops persist throughout the mouse mitochondrial DNA genome. *J. Biol. Chem.* **283**, 36743–36751
- 51 Kreuzer, K. N. (2000) Recombination-dependent DNA replication in phage T4. *Trends Biochem. Sci.* **25**, 165–173
- 52 Kogoma, T. (1997) Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol. Mol. Biol. Rev.* **61**, 212–238
- 53 Drolet, M. (2006) Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol. Microbiol.* **59**, 723–730
- 54 Proudfoot, M., Kuznetsova, E., Brown, G., Rao, N. N., Kitagawa, M., Mori, H., Savchenko, A. and Yakunin, A. F. (2004) General enzymatic screens identify three new nucleotidases in *Escherichia coli*. Biochemical characterization of SurE, YfbR, and YjgG. *J. Biol. Chem.* **279**, 54687–54694
- 55 Seto, D., Bhatnagar, S. K. and Bessman, M. J. (1988) The purification and properties of deoxyguanosine triphosphate triphosphohydrolase from *Escherichia coli*. *J. Biol. Chem.* **263**, 1494–1499
- 56 An, G., Bendiak, D. S., Mamelak, L. A. and Friesen, J. D. (1981) Organization and nucleotide sequence of a new ribosomal operon in *Escherichia coli* containing the genes for ribosomal protein S2 and elongation factor Ts. *Nucleic Acids Res.* **9**, 4163–4172
- 57 Yusufzai, T. and Kadonaga, J. T. (2010) Annealing helicase 2 (AH2), a DNA-rewinding motor with an HNH motif. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20970–20973

Received 20 May 2011/21 June 2011; accepted 23 June 2011

Published as BJ Immediate Publication 23 June 2011, doi:10.1042/BJ20110901

SUPPLEMENTARY ONLINE DATA

Helicase dissociation and annealing of RNA-DNA hybrids by *Escherichia coli* Cas3 protein

Jamieson A. L. HOWARD*, Stephane DELMAS†, Ivana IVANČIĆ-BAČE‡ and Edward L. BOLT*¹

*School of Biomedical Sciences, The University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, U.K., †School of Biology, The University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, U.K., and ‡Molecular Biology, University of Zagreb, Horvatovac 102a, Zagreb, Croatia

EXPERIMENTAL

Cas3 ssDNA–ssDNA annealing assays

Reactions were carried out at 37°C in a 20 µl final reaction volume. Reactions contained Cas3 protein as indicated with 5 mM MgCl₂, 5 mM DTT, 7% glycerol, ϕX174 circular virion DNA (500 ng) and radiolabelled 71-mer ELB58B ssDNA (6 ng, see Table 1 of the main text) at pH 7.5 in 10 mM Tris/HCl. Reactions were stopped and deproteinized, in end-point or time-course assays, by addition of SDS (1%) and proteinase K (1 µg/µl) and loaded directly on to 1% agarose gels which were subjected to electrophoresis at 50 V for 4–5 h. Gels were dried using a heated bed with vacuum suction and laid down for phosphorimaging on a Fuji FLA3000 machine.

Cloning the *E. coli* CRISPR-1 locus to form target substrates for Cas3 R-loop formation

Plasmid pJLH14 DNA was created in two steps. First, a 60 bp fragment of *E. coli* CRISPR-1 (spacer 3 with 19 bp of 5' flanking DNA and 10 bp of 3' flanking repeat) was amplified by PCR using primers listed in Table S1. This was cloned into the pT7-7 AccI site via HinPII sites on the PCR fragment, generating pJLH13. Secondly, the 60 bp CRISPR DNA fragment from pJLH13 was sub-cloned into the polylinker of pBluescript via EcoRI and HindIII sites flanking CRISPR in pJLH13, giving pJLH14. Plasmid pJRW2 was created by PCR and cloning *E. coli* CRISPR-1 chromosomal locus (12 repeats and 12 spacers) into pBluescript via BamHI and XbaI sites using primers listed in Table S1.

Generating RNA substrates for Cas3 by *in vitro* transcription

crRNA2 and crRNA3 were synthesized by *in vitro* transcription. crRNA3 was transcribed from BssHII-linearized pJLH14 template. In this construct a spacer-repeat fragment of CRISPR-1 is 70 bp downstream of the pBluescript T3 promoter. Transcription reactions were carried out for 1 h at 37°C in buffered T3 RNA polymerase, pJLH14 (1 µg), 10 mM DTT, 0.5 mM rATP, rGTP and rUTP, 0.2 mM rCTP and 0.05 MBq of [α -³²P]CTP. crRNA2 was made in a similar manner from linearized pJRW2 DNA template. In each case, after 1 h of transcription, DNA was removed using Turbo DNaseI (Ambion) and reactions were stopped by the addition of EDTA (30 mM). RNA molecules were purified from un-incorporated [α -³²P]CTP using BioSpin 6 columns (Bio-Rad). A 'random' RNA strand was generated from pGB126, which contains an 0.8 kbp fragment of *E. coli* *recG* cloned into pBluescript [39]. Antisense crRNA was generated by *in vitro* transcription from pJLH14 or pJRW2 as described above but using T7 RNA polymerase.

Cloning for *in vivo* phage protection assay

Cloning of *E. coli* MG1655 *cas3* into pRSF-1b required prior manipulation to remove an NcoI site from within the coding sequence of *cas3*. Deletion of the NcoI recognition site by QuikChange[®] mutagenesis was confirmed by restriction digestion and the resulting plasmid was used as a template to amplify *cas3* for cloning into pRSF-1b via NcoI and KpnI sites generating pEB547. The 4.2 kb operon region *casA–E* (Cascade) was amplified by PCR from *E. coli* MG1655 genomic DNA, introducing NcoI and EcoRI sites for cloning into pET-Duet.

¹ To whom correspondence should be addressed (email ed.bolt@nottingham.ac.uk).

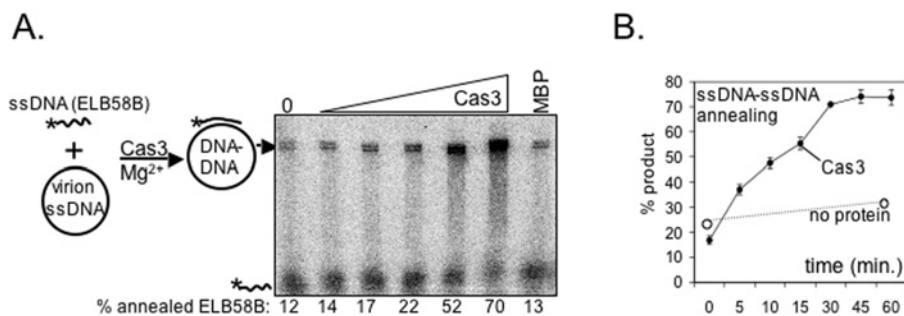


Figure S1 Cas3 promoted annealing of ssDNA to ssDNA and of RNA into duplex DNA

(A) Agarose gel phosphorimaged to show products from mixing increasing concentrations of Cas3 as indicated (0, 50, 100, 200, 400 and 800 nM) with ³²P end-labelled (*) ELB58B ssDNA (6 ng) and circular ssDNA of virion ϕ X174 (200 ng) in buffer containing 5 mM magnesium but lacking ATP. ELB58B is 100 % complementary to sequence close to the XhoI site of ϕ X174. The percentage of product in each reaction is given below the panel. The final lane represents the outcome of a reaction replacing Cas3 with *E. coli* MBP (800 nM) as a control for non-specific annealing promoted by 'mock' protein. (B) Graph of Cas3-promoted annealing of ELB58B (6 ng) to circular ssDNA virion ϕ X174 (200 ng) in reactions containing 5 mM magnesium. Cas3 (400 nM) was incubated to measure annealing as a function of time, presented as means \pm S.E.M. of duplicate reactions.

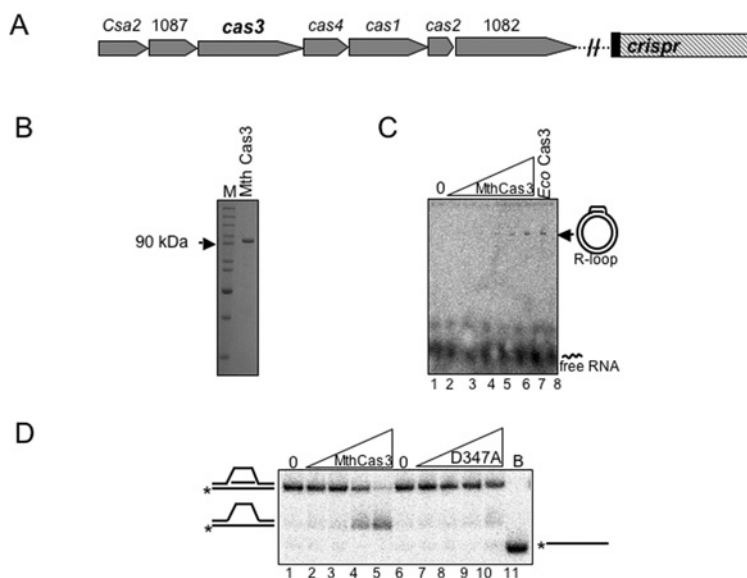


Figure S2 ATP-independent R-loop formation and ATP-dependent helicase activity of archaeal Cas3

(A) Cartoon representation of the overall arrangement of CRISPR/Cas in the moderately thermophilic archaeal species *Mth*. Genes encoding core Cas proteins and sub-type specific Cas proteins (1087 and 1082) are in grey, with Cas3 highlighted in bold. The CRISPR repeat region is represented as a blocked hatched box. (B) Coomassie Blue-stained 10 % acrylamide SDS/PAGE showing purified *Mth* Cas3 protein migrating close to its expected molecular mass of 99 kDa. Molecular mass markers (M) are shown on the left-hand side. (C) Phosphorimages of a dried 1.2 % (w/v) agarose gel monitoring formation of R-loop product by *Mth* Cas3 (lanes 1–7, 0, 50, 100, 200, 300, 400 and 800 nM) in reactions containing plasmid DNA (μ JLH14, 200 ng), radiolabelled crRNA3 (25 ng) in MgCl₂ (5 mM). For comparison, lane 8 shows the product of the same reaction incubating *E. coli* Cas3 (200 nM). (D) Phosphorimage of a gel (10 % TBE) showing products of mixing *Mth* Cas3 proteins (0, 10, 50, 100 and 250 nM) with R-loop substrate (1 nM, see Table 1 of the main text) end-labelled with ³²P on one strand as indicated (*) in buffer containing 5 mM ATP and 5 mM magnesium chloride (lanes 1–13), Unwinding activity of wild-type *Mth* Cas3 (lanes 2–5) was compared with *Mth* Cas3 mutated in the ATPase motif II (Walker B/DExH box D347A, lanes 7–10). Lane 11 (B) indicates the product of boiling the R-loop substrate to liberate the ³²P-labelled DNA strand. See also Figures 5(D) and 5(E) of the main text.

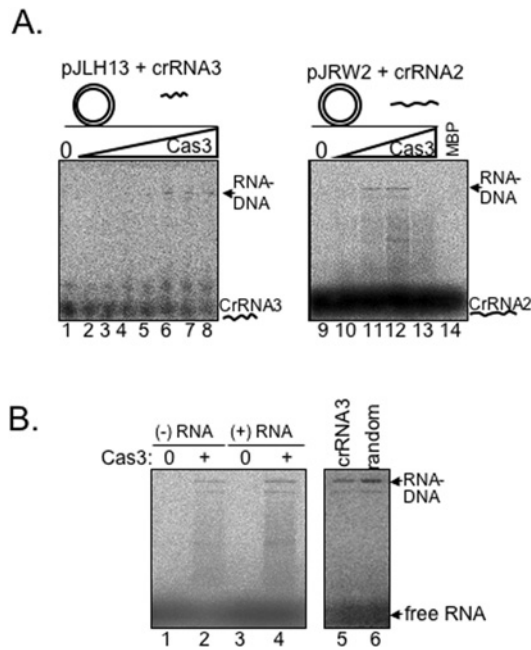


Figure S3 Cas3-promoted R-loop formation is effective on various RNA substrates

(A) Phosphorimages of dried 1.2% (w/v) agarose gels monitoring the position of radiolabelled RNA substrate (25 ng) after mixing with Cas3 (0, 50, 100, 200, 300, 400, 500 and 800 nM) and a target uncut plasmid (200 ng) with complementary sequence in buffer containing 5 mM MgCl₂. crRNA3 and pJLH13 RNA and DNA substrates were used in reactions in lanes 1–8, and have a 60 bp region of complementarity. crRNA2 and pJRW2 RNA and DNA substrates were used in reactions shown in lanes 9–14 and have over 900 bp of 100% complementarity. Lane 14 shows a non-Cas3 protein control reaction, instead containing 800 nM of *E. coli* MBP. (B) Phosphorimages of dried 1.2% (w/v) agarose gels monitoring the position of radiolabelled RNA substrate after mixing with Cas3 (400 nM) and RNA molecules crRNA2 antisense (lane 2), crRNA2 sense (lane 4), crRNA3 (lane 5) or a non-CRISPR RNA (random, lane 6). Target uncut plasmid DNA substrates for each RNA in reactions were: pJRW2 for antisense and sense crRNA2, pJLH14 for crRNA3 and *recG* cloned into pBluescript, for a random RNA sequence not related to *E. coli* CRISPR-1 sequence in crRNA2 or crRNA3. These substrates are described fully in Table S1 and in the Experimental section of the main text. Lanes 1 and 3 contained no Cas3 protein.

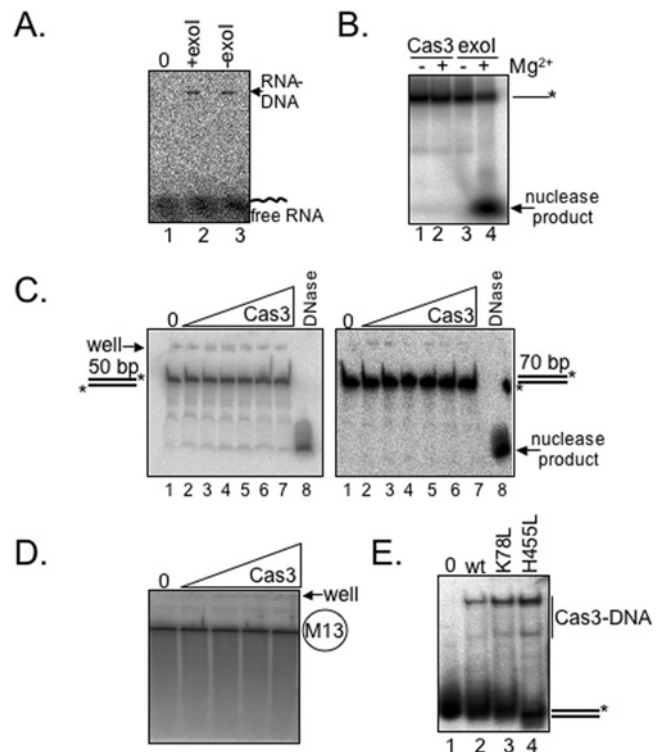


Figure S4 Cas3 has no detectable nuclease activity on a variety of different single- and double-stranded DNA substrates

(A) Phosphorimage of 1.2% (w/v) agarose gels monitoring the position of radiolabelled RNA substrate. Uncut plasmid DNA (pJLH14, 200 ng) pre-incubated with exonuclease I (exol, 10 units) to digest any ssDNA regions showed formation of R-loops when mixed with Cas3 (400 nM) and radiolabelled crRNA3 (25 ng) in MgCl₂ (5 mM) buffer (lane 2), compared with identical reactions containing pJLH14 not pre-treated with exonuclease I (lane 3). Lane 1 shows a reaction without Cas3 protein. (B) Assay for Cas3 exonuclease activity: 15% acrylamide TBE gel of reactions incubating ELB40 ssDNA (Table S2), ³²P end-labelled as indicated (*), with either Cas3 (500 nM, lanes 1 and 2) or exonuclease I (10 units, lanes 3 and 4). Nuclease activity of exonuclease I was apparent in the presence of 5 mM MgCl₂ (lane 4) but not without (lane 3). Cas3 showed no equivalent nuclease activity in this assay (lane 2). (C) Assay for Cas3 nicking activity on duplex DNA molecules (Table S1): products of incubating Cas3 (0–800 nM, lanes 1–7 in each panel) with end-labelled 50 bp (left panel) or 70 bp duplex DNA (right panel) in 5 mM MgCl₂ were denatured and run on 15% acrylamide TBE urea gels to detect nicking of DNA strands. Lane 8 in each panel is a positive control showing the product of DNaseI nicking each substrate. (D) Assay for Cas3 nuclease activity on M13 circular ssDNA (200 ng) when incubated at 0, 50, 100, 200 or 500 nM in buffer containing 5 mM MgCl₂. The panel shows 1.5% agarose gel containing 50 mM NaCl, 50 mM NaOH and 1 mM EDTA. Reactions were stopped after 1 h by addition of proteinase K and 1 mM EDTA. NaOH was added to 10 mM and glycerol to 7% (v/v) prior to electrophoresis for 1 h at 50 V. No Cas3 nuclease activity could be detected. (E) EMSA showing Cas3 binding to duplex DNA (Table 1 of the main text) end-labelled with ³²P as indicated (*) with wild-type, K78L and H455L Cas3 proteins (all at 200 nM). Reactions were at room temperature for 10 min in buffer containing 2 mM MgCl₂. The electrophoresis gel contained 5% (v/v) acrylamide, 6.7 mM Tris/HCl, pH 8.0, 3.3 mM sodium acetate and 2 mM MgCl₂.

Table S1 Primers used to clone *E. coli* CRISPR-1 and *cas* genes, and *Mth cas3*, and their site-directed mutants

Use	Primer	Sequence
Cloning <i>E. coli</i> Cas3 into pMalc2 for MBP-Cas3	ygcb-X2	5'-GGAATTCATGGAACCTTTTAAATATA
	ygcb-Y2	5'-GCTCTAGATATTTGGGATTTGCAGG
HD motif Cas3 mutant, D75G	D75-1	5'-CATTGCTCTTCATGGTATTGAAAGTTTG
	D75-2	5'-CAAACCTTCCAATACCATGAAGAGCAATG
HD motif Cas3 mutant, K78L	K78L-1	5'-CATGATATTGACTGTTTGATATAC
	K78L-2	5'-GTATATCAAACAGTCCAATATCATG
HD motif Cas3 mutant, H74G	HG-1	5'-CATTGCTCTTGGTGATATTGG
	HG-2	5'-CCAATATCACCAGAGCAATG
Walker A, ATPase I mutant	K320L1	5'-CAGGCTCCGGTAAACGAAACAGC
	K320L1b	5'-GCTGTTCCGTTAGACCGGAGCCTG
Walker A, ATPase I mutant	G317SS318Y	5'-ATAGAGGCACCTACAGCTTCGGTAAACGAAAC
	G317SS318Y	5'-GTTCCGTTTACCGAAGCTGTAGGTGCCTCTAT
Walker B, ATPase II mutant	H455L-1	5'-GTGTTTTAATTGTTGATGAACCTTAGCTTACGACACC
	H455L-2	5'-GGTGTCTGAAGCTAGAAGTTCATCAACAATTAACAC
Cloning <i>Mth</i> Cas3	1086-1	5'-GGTGGCTCATATGGAATTAATCG
	1086-2	5'-CAAGAAAAGCTTCATCTCACACTATGA
Walker B, ATPase II mutant of <i>Mth</i> Cas3	D347A-1	5'-TATTTAGTGAGGCTGAAATCGATATAC;
	D347A-2	5'-GTATATCGATTTGAGCTCACTGAAATA
Cloning <i>E. coli</i> Cascade CasCDE complex	YgcJX(CasC)	5'-GGAATTCATGCTAACTTTATCAAT
	YgcJY(casC)	5'-CGGGATCCTCAGCCTCGCCATTATTA
	CasD-1	5'-GGAATTCATATGAGATCTTATTGATCTTGC
	CasD-2	5'-CCGCTCGAGTTACTGAGATACATCCATACC
	CasE-1	5'-GGAATTCATATGATCTCAGTAAAGTC
	CasE-2	5'-GGAATTCACAGTGGAGCCAAAG
Cloning entire <i>E. coli</i> Cascade (<i>casA-casE</i>) operon	CasA1	5'-CATGCCATGGATGAATTTGCTTATTGATAACTGG
	CasE-2	5'-GGAATTCACAGTGGAGCCAAAG
CRISPR-1 cloning into pJLH13, as starting clone for pJLH14 (crRNA3)	Spacer10-A	5'-GGAATTCACAGGGGCGCGCTCGCC
	Spacer10-B	5'-CGGGATCCCGTTGCCAAATAGGGGCGACC
CRISPR-1 cloning into pJRW2	Coli-1	5'-CGGGATCCGAAATAGCTGATCTTTAAT
	Coli-2	5'-GCTCTAGAGCGAAGGCGTCTTGATGG
Removing NcoI site from MG1655 <i>cas3</i> for cloning into pRSF-1b	NcoI-A	5'-GAGTCTGGTTTCCGTGGGTAGAGGCCGTTAC
	NcoI-B	5'-GTAACGGCCTTACCACCGAAACAGGAGCTC
Cloning <i>cas3</i> into pRSF-1b	RSF-A	5'-CATGCCATGGAACCTTTTAAATATA
	RSF-B	5'-GGGGTACCTTATTGGGATTTGCAGGG

Table S2 Details of plasmid and RNA substrates and sequences of oligonucleotides used for model substrates in Cas3 assays

Substrate	Notes/sequences of oligonucleotides
CRISPR RNA3 (crRNA3)	Generated by <i>in vitro</i> transcription from pJLH14
pJLH13	pT7-7 containing 61 bp fragment of <i>E. coli</i> CRISPR-1 locus (32 bp spacer 3 + 19 bp 5' flanking repeat DNA + 10 bp of 3' flanking repeat DNA). Cloning is detailed in the Experimental section.
pJLH14	pBluescript containing the same <i>E. coli</i> CRISPR-1 fragment as in pJLH13, sub-cloned as detailed in the Experimental section.
crRNA2	Generated by <i>in vitro</i> transcription from pJRW2
pJRW2	pBluescript containing entire <i>E. coli</i> CRISPR-1 locus, approximately 1 kb. Cloning is detailed in the Experimental section. This plasmids fully base pairs with crRNA2.
R-loop	DNA strand 1: 5'-GGGTGAACCTGCAGGTGGGCGGCTGCTCATCGTAGGTTAGTTGGTAGAATTCGGCAGCGTC-3' DNA strand 2: 5'-GACGCTGCCGAATCTACCAGTGCCTTGTAGGACA TCTTTGCCACCTGCAGGTTACCC-3' RNA strand: 5'-AAAGAUUCCUAGCAAGGCAC-3'
ELB58B	5'-ACGCGTCACTTTTGGACAGAATCGTTAGTTGATGGCGAAAGTTCGCAAAAGTAAAGATTCTCGAGCTGCAGCAAGGA-3'
ELB40 ssDNA	5'-GGAGCTCCCTAGGCAGGATCGTTGCGGACGATGGCCCTCGAAGAGCTCCAGTTACGGATACGGATCCTGC-3'
Duplex linear DNA	ELB41: 5'-GCAGGATCCGTATCCGTAACCTGGAGCTCTTTCGAAGGCCATCGTCGCGAAGCATCTGCCTAGGGAGCTCC-3'
Duplex linear DNA	RGL16: 5'-ATCGATAGTCTAGACAGCATGCTCTAGCAAGCCAGAATTCGGCAGCGT-3' ELB37: 5'-ACGCTGCCGAATCTGGCTTGTCTAGGACATGCTGTCTAGAGACTATCGAT-3'

Received 20 May 2011/21 June 2011; accepted 23 June 2011

Published as BJ Immediate Publication 23 June 2011, doi:10.1042/BJ20110901