HAL
open science

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

Jamieson Leyland Howard, Stephane Delmas, Ivana Ivancic-Bace, Edward L Bolt

To cite this version:

Jamieson Leyland Howard, Stephane Delmas, Ivana Ivancic-Bace, Edward L Bolt. Helicase dissociation and annealing of RNA-DNA hybrids by Escherichia coli Cas3 protein.. Biochemical Journal, 2011, 439 (1), pp.85-95. 10.1042/BJ20110901 . hal-00628671

HAL Id: hal-00628671

https://hal.science/hal-00628671

Submitted on 4 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Biochemical Journal Immediate Publication. Published on 23 Jun 2011 as manuscript BJ20110901

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

Jamieson A. L. Howard, Stephane Delmas[*], Ivana Ivančić-Baće[#] & Edward L. Bolt

School of Biomedical Sciences, The University of Nottingham, Queen's Medical Centre, Nottingham NG72UH. U.K.

[*]School of Biology, The University of Nottingham, Queen's Medical Centre, Nottingham NG72UH. U.K.

[#]Faculty of Sciences, The University of Zagreb, Croatia.

Correspondence to: Edward L. Bolt
Ed.bolt@nottingham.ac.uk
+44 (0)115 8230194

Keywords: helicase, CRISPR, Cas3, R-loop, annealing

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Abstract

CRISPR/Cas is a nucleic acid processing system in bacteria and archaea that interacts with mobile genetic elements. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) DNA and RNA sequences are processed by Cas (CRISPR-associated) 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. 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 abolished Cas3 R-loop formation and instead powered Cas3 helicase unwinding of the invading RNA strand of a model R-loop substrate. R-loop formation by Cas3 required magnesium as a co-factor and was inactivated by mutagenesis of a conserved amino acid motif. Cells expressing mutant Cas3 protein were more sensitive to plaque formation by phage λvir. A complex of CasABCDE ("Cascade") also promoted R-loop formation and we discuss possible overlapping roles of Cas3 and Cascade in *E. coli*, and apparently antagonistic roles of Cas3 catalysing RNA-DNA annealing and ATP-dependent helicase unwinding.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Introduction

CRISPR/Cas is a nucleic acid processing system in bacteria and archaea that interacts with mobile genetic elements. CRISPR loci (Clustered Regularly Interspaced Short Palindromic Repeats [1]) are functionally coupled to Cas (CRISPR associated) 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]. Varying functions of CRISPR/Cas probably reflect the impressive variability in organisation and distribution of CRISPR loci and *cas* genes across bacterial and archaeal domains, prompting their classification into CRISPR/Cas sub-types. 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 sequence, 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 sub-type-specific) proteins or archaeal CMR proteins (Cas RAMP module; RAMP means Repeat Associated Mysterious Proteins). 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 nt) RNA (crRNA) [24, 25] is catalysed by sub-type specific Cas proteins, Csy4 in *P. aeruginosa* [26], Cas6 in the archaeon *Pyrococcus furiosus* [27], or CasE (also called Cse3) in *Escherichia coli* [3]. *E. coli* CasE is part of a larger protein-RNA complex called Cascade, encoded by *casA-E* [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*) (Fig. 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 survival of cells attacked by lysogenic and lytic cycles of infection by λ phage [3, 4, 7]. This required expression of crRNA sequence engineered with an anti-λ spacer (T3). In two studies [4, 7] 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 ϕYS40 [34].

Cas3 is a core Cas protein found in most CRISPR/Cas sub-types. Evidence from *in vivo* studies indicates that Cas3 functions downstream of endonucleolytic crRNA processing in *P. aeruginosa* [9] and co-operates with the Cascade-crRNA complex in *E. coli* [3]. Amino acid sequence motifs in Cas3 had predicted ATP-dependent helicase activities for Cas3 and a HD-phosphohydrolase domain [35]. Consistent with a helicase domain, purified Cas3 from *Streptococcus thermophilus* catalysed displacement of 20 nucleotide DNA and RNA strands annealed to M13 closed circular single-stranded DNA (ssDNA) [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 double stranded DNA (dsDNA) in preference to ssDNA [37], a nicking activity not tested in reference [36]. We describe 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 *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.

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Experimental: Materials and Methods

Cloning *cas3* from *E. coli* MG1655 and *Methanothermobacter thermautotrophicus* ΔH.

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 Xba I sites in frame with *malE* creating plasmid pAH1. This was successful at over-producing soluble Cas3 fused at its N-terminus to maltose binding protein (MBP-Cas3), a protein of approximately 140 kDa shown in Figure 1C. Table 2 lists primer sequences used for cloning *ygcB* into pMal-c2 (ygcB-X2 and ygcB-Y2) and for its mutagenesis to generate Cas3 proteins H74G, D75G, K78L, K320L and G317SS318Y.

Methanothermobacter thermautotrophicus (Mth) ORF1086 (encoding Cas3) was cloned from Mth genomic DNA (a gift from Dr. James Chong, University of York) *via* PCR using primers with NdeI and HinDIII sites detailed in Table 2. PCR products were cloned into pT7-7 generating pEB359 and this construct was effective at over-expressing soluble MthCas3. Mth D347A Cas3 was generated by mutagenic PCR on pEB359 using primers listed in Table 2.

Purification of *E. coli* and Mth Cas3 proteins

E. coli MBP-Cas3 was purified from pAH1 in DH5α grown in MU broth containing ampicillin (50 μg/mL). At O.D. 0.5 cells were induced by IPTG (0.5 mM) for 30 minutes at 37°C. Recovered cells were resuspended in amylose column buffer (20 mM Tris.HCl pH7.5, 200 mM NaCl, 1 mM EDTA, 1 mM DTT and 0.1 mM phenylmethylsufonylflouride PMSF) and frozen at -80°C. Cas3 was purified by FPLC at room temperature and SDS-PAGE and Coomassie blue staining of proteins followed each purification step. Protein concentrations were measured against a standard curve of bovine serum albumin using Bradford's reagent (BioRad).

Biomass was lysed by sonication and centrifuged (20 k rpm, 30 minutes in an Avanti J25 machine with rotor JA25.50). Soluble material (S1) was passed into a 7 mL amylose (High Flow, NEB) column equilibrated with amylose column buffer. MBP-Cas3 bound to amylose and was eluted in 2 mM maltose added to 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 onto a 10 mL DEAE-sepharose column, eluting in 300-400 mM NaCl. Cas3 pooled fractions were loaded directly into 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 onto 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 hours with IPTG (0.5 mM). Cells were harvested into buffer C and frozen at -80°C. Thawed biomass was lysed by sonication and clarified (20k rpm, 20 minutes) 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 into a HiTrap Q-sepharose column and eluted at about 600 mM NaCl. Mth Cas3 was pooled and loaded directly onto 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, over-expression 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 in Table 2. Primers for *casC* included restriction sites for EcoRI and BamHI for cloning PCR product 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.

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

CasC was over-expressed 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 O.D. 0.5 *casC* expression was induced with IPTG (0.5 mM) for two hours. Cells were harvested, resuspended in <5 mL of buffer CP (20 mm Tris.HCl pH 8.0, 1 mM EDTA, 1 mM DTT and 0.1 mM PMSF, 100 mM potassium acetate) and frozen at -80°C. Thawed cells were lysed by sonication generating soluble material (S1) that was recovered by centrifugation (20 k rpm, 30 minutes in an Avanti J25 machine with rotor JA25.50). CasC was purified by FPLC at room temperature, monitoring by SDS-PAGE and Coomassie staining. S1 was passed through a tandem arrangement of 5 mL HiTrap Q-sepharose linked to a 5 mL HiTrap SP-sepharose and this array was developed using a gradient of potassium acetate (0.1-2.0 M), CasC eluting at 500-750 mM. CasC fractions were dialysed into buffer CP and loaded onto 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 onto 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 over-expressed and harvested as described for CasC. Lysis of thawed cells by sonication gave 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 solubilised and refolded on a HisGraviTrap column: samples were washed with buffer C containing 1 M NaCl and the pellet resulting from washing (P2) was recovered by centrifugation. P2 was then dissolved by gentle stirring at room temperature in 6 M urea and the solution clarified by centrifugation giving soluble material that was loaded directly onto a nickel-charged HisGraviTrap column procedure.

Other proteins in this work

Purified *E.coli* maltose binding protein 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 Prof. Bob Lloyd, FRS (Univ. of Nottingham). 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 had been purified as described in reference [38].

DNA substrates

DNA substrates used in this work are summarised in Table 1 and detailed in supplementary methods Table ST1. Oligonucleotides were purchased from MWG and were used to generate Cas3 substrates as detailed in [39]. Plasmid DNA used in R-loop assays were generated from *E. coli* strain DH5α. Any ssDNA in these plasmid preparations was controlled for by digestion with exonuclease I, as described in the results. In reactions using linearised plasmids, plasmid was incubated with XbaI for one hour and then heated to inactivate XbaI. R-loop forming reactions described in this work used radiolabelled RNA molecules comprising sequences of *E. coli* CRISPR-1 (crRNA) that were complementary to target DNA. Cloning fragments of E. coli CRISPR-1 locus into pBluescript or pT7-7 generated target DNA. Details of how these substrates were generated are in supplementary methods.

RNA substrates

RNA substrates (crRNA) generated in this work by *in vitro* transcription, and their DNA templates, are summarised in table 1 and detailed in supplementary table ST1. RNA substrates ranged from 250-900 nucleotides in length, comparable to the 600-nucleotide crRNA expressed with Cas3 and Cascade during anti-viral defence in reference [3].

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

R-loop assays

Reactions were in 20 μL volumes for times stated in figures at 37°C for *E. coli* Cas3 or 44°C for Cas3 from the moderately thermophilic archaeon *Methanothermobacter thermautotrophicus* (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 addition of proteinase K (5 μL of 10 mg/mL stock) and EDTA (5 mM). Reactions were loaded directly onto 1.2% agarose gels for electrophoresis in Tris-Borate-EDTA buffer for about 3 hours at 50 volts. Gels were dried by on a heated bed with vacuum suction and laid down for phosphorimaging on a Fuji FLA3000 machine.

R-loops were recovered from de-proteinised Cas3 reactions by precipitation at -20°C for at least 2 hours in ammonium acetate (1 M) and isopropyl alcohol (50% v/v), using glycogen (1 μg/ μL) as a carrier. Nucleic acid pellet was resuspended in buffer for RNaseH digestion (New England Biolabs) or helicase buffer (5 mM $MgCl_2$, 5 mM ATP, 10 mM Tris.HCl pH7.5) for dissociation by RecG. RNase H and RecG reactions were for 30 minutes 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 [3], in which crRNA, Cas3 and Cascade (CasABCDE) were co-expressed from three plasmids. crRNA was expressed from pWUR478 (a kind gift from Prof. John van der Oost, University of Wageningen). *cas3* or K78L *cas3*, and the cascade operon were cloned in this work into pRSF-1b (pEB547 and pEB548) or pET-Duet (pEB550) respectively. Cloning procedures for these plasmids are detailed in supplementary methods with primer details given in supplementary table 2.

For plaque assays, BL21 DE3, which lacks CRISPR/cas, was transformed with pEB550, pWUR478 and pEB547 or pEB548. Overnight cultures were inoculated in TB broth containing maltose (0.02%, w/v), ampicillin (50 μg/mL), kanamycin (30 μg/mL) and chloramphenicol (15 μg/mL). Cultures were grown to O.D. 0.5 and then 500 μL of each culture was mixed with 3 mL of 0.6% LB agar before pouring over LB agar plates containing each of the three aforementioned antibiotics and either IPTG (0.6 mM) or no IPTG as indicted. Serial dilutions (10^{-2} to 10^{-7}) of stock phage λvir were applied as 10 μL spots to the soft agar top and incubated at 30°C, as shown in figure 4C.

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 compared to reactions containing no protein. For Cas3, the magnesium chloride concentration was kept constant at 5 mM and ATP (or other NTP) titrated from 1 to 10 mM. Figure 5A 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 phosphate produced per nmol Cas3.

Helicase and nuclease assays

R-loop helicase reactions were at 37°C for *E. coli* Cas3 or 44°C for Mth Cas3, in a 20 μL final reaction volume. Standard reactions contained Cas3 protein as indicated, 5 mM magnesium chloride, 5 mM DTT, 7% glycerol, and R-loop substrate (5 ng). Reactions contained either 5 mM ATP or 5 mM ATPγS and were at pH 7.5 in Tris.HCl, and were stopped by addition of proteinase K and EDTA and loaded onto 10% acrylamide TBE gels.

Reactions for nuclease digestion by *E. coli* Cas3 contained either ^{32}P end-labelled ssDNA or dsDNA (5 ng), or M13 or ϕX174 ss circular DNA (200 ng), as indicated in the supplementary figures. Reactions contained magnesium chloride from 1-10 mM and were incubated at 37°C for 10

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

minutes to 3 hours. For short linear duplex substrates (table 1) products of exonuclease or endonuclease digestion were detected by deproteinised 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 minutes and loaded onto 15% acrylamide TBE gels containing 4 M urea.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Results

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

Cas3 encoded within CRISPR-1/Cas of *E. coli* K-12 MG1655 (Fig. 1A) has amino acid sequence motifs characteristic of HD-superfamily proteins [35] and ATPases of superfamily 2 (SF2) helicases (Fig. 1B). *E. coli* Cas3 has been implicated as an effector of CRISPR protection against λphage by its actions with the Cascade protein complex (encoded by *casA-E*, Fig. 1A) [3]. We investigated biochemical activities of purified *E. coli* Cas3 fused at its N-terminus to *E. coli* maltose binding protein (MBP) (MBP-Cas3, 140 kDa, Fig. 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 nucleotide (nt) oligonucleotides paired to M13 single strand DNA (ssDNA) has been described for Cas3 from *Streptococcus thermophilus* [36]. We confirm a helicase activity for Cas3 using the *E. coli* enzyme on R-loop substrates, described in a later section (Fig. 5). In addition, we observed Cas3 promoted ATP-independent annealing of a 70nt DNA strand to complementary sequence in ϕX174 ssDNA (Fig. S1A). In these reactions Cas3 could anneal 70% of free radiolabelled oligonucleotide compared to 13% spontaneous annealing in the absence of Cas3 (Fig. 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 wished to test this activity of Cas3 in more detail, using CRISPR RNA and duplex DNA substrates, which may be more relevant to functioning of *E. coli* CRISPR/Cas than M13 ssDNA or ϕX174 ssDNA.

Cas3 annealing activity base pairs RNA into complementary duplex DNA

We investigated if *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, summarised in table 1. In one set of reactions Cas3 (50-800 nM) was incubated for 45 minutes in buffered magnesium chloride (5 mM) with 269-nucleotide RNA (25 ng) that was 100% complementary in sequence to part of uncut plasmid DNA (pJLH14, 250 ng). Reactions were de-proteinised and products 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 (Fig. 2A, lanes 4-8). Using fixed Cas3 concentrations converted about 20% of free RNA into product in 45 minutes (Fig. 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-proteinised reaction products (Fig. 2C). Plasmid DNA in reactions lacking Cas3 (lane 1) or with Cas3 added (lane 2) migrated as two species, but DNA in the Cas3 reaction was slower migrating than when Cas3 was absent. This slower migrating DNA (lane 2) corresponded exactly to the position of radiolabelled RNA from an identical reaction imaged to detect RNA (lane 4). We concluded that Cas3 promoted association of RNA with plasmid DNA.

Control reactions were used to ensure formation of observed RNA-DNA products was specific to Cas3. Reactions containing “mock” proteins, maltose binding protein (MBP), Hel308 helicase [38], or Cas3 that had been pre-heated to 85°C, showed no detectable RNA-DNA product (Fig. 2D). Further support for ATP-independent RNA-DNA products of Cas3 was obtained by substituting *E. coli* Cas3 with Cas3 purified from the archaeon *Methanothermobacter thermautotrophicus* (Mth) (Fig. S2A-C). MthCas3 also promoted formation of the RNA-DNA product (Fig. S1C), indicating that this Cas3 activity is conserved evolutionarily.

Cas3 also promoted formation of RNA-DNA products in reactions containing RNA mixed with uncut plasmid DNA molecules in which sequence complementarity was 60 nucleotides (Fig. S3A, lanes 1-8) or 1000 nucleotides (Fig. S3A lanes 9-14). In all of these assays RNA had been 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

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

sequences that were antisense (-) CRISPR-1 RNA, or from RNA sequences unrelated to CRISPR-1 (Fig. 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 linearised

When uncut plasmid was used as a target for annealing complementary RNA by Cas3 we observed about 20% of free RNA converted into product. This activity was improved 3-6 fold if plasmid was linearised prior to mixing with RNA and Cas3, measured at fixed Cas3 concentrations (Fig. 3A and 3B). A single reaction product was observed from linearised plasmid DNA compared to two from uncut plasmid (Fig. 3B, compare lanes 4 and 8). We also observed a new product containing radiolabelled RNA that migrated close to free RNA (Fig. 2C lane 6, labelled Z).

We next asked if 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 observed migratory properties of Cas3 reaction products. We assessed directly if 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 3B were de-proteinised and nucleic acid products 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 (Fig. 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 (compare lanes 2 and 4). 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 single stranded DNA that can arise during preparation of plasmid DNA, although this seemed unlikely considering the observed enhancement of product formation by linearising plasmids shown in Figure 3B. However, control reactions were treated with exonuclease I to remove ssDNA in 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 (Fig. S4A, compare lanes 2 and 3). We also examined the possibility that a 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 single stranded DNA or on end labelled model DNA substrates analysed on neutral or denaturing gels (summarised in Fig. S4B-D). 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 exonuclease 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 (Fig. 3D). To explore this requirement of Cas3 we focussed on its HD-motif (Fig. 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, Fig. 1B) and measured their ability to promote R-loop formation. Cas3 D75G and K78L were less effective at forming R-loops than wild type protein in reactions containing either uncut plasmid DNA (Fig. 4A, compare lane 2 with lanes 4 and 5) or linearised plasmid (Fig. 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. Like wild type Cas3, binding of K78L Cas3 to duplex DNA was detected in mobility shift assays (EMSAs, Fig. 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

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

ATPase motif I/Walker A box (K320L and G317S-S318Y, Fig. 1B) did not reduce R-loop formation (Fig. 4A lanes 6 and 8). Interestingly, Cas3 protein with combined substitutions in both K78 and K320 (K78L-K320L) was proficient at forming R-loops (Fig. 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 λvir *in vivo*.

E. coli cells can be protected against lysis by phage λvir using inducible plasmid-based co-expression of Cas3, Cascade (CasA-E) and CRISPR RNA (crRNA) engineered with λvir sequence [3, 20]. We also observed a protective effect using similar plasmid co-expression of wild type Cas3, Cascade and crRNA (summarised in Fig. 4C). In these assays serial dilutions of phage λvir suspensions were applied to a lawn of *E. coli* cells to observe plaque formation (lane 1) that was reduced 100-1000 fold on IPTG induced expression of crRNA, Cas3 and Cascade (lane 2). Expression of Cas3 K78L HD mutated protein instead of wild type Cas3 in the same assay (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 phage observed in this assay.

***E. coli* Cas3 helicase activity dissociates R-loops.**

Addition of ATP to Cas3 R-loop reactions correlated to a total absence of R-loop product, compared to reactions containing magnesium only (Fig. 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 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 single-stranded DNA (ϕX174 ssDNA) or double-stranded DNA (ϕX174 dsDNA) had no stimulatory effect on ATPase activity (Fig. 5B). Hel308, a ssDNA stimulated ATPase-helicase [38], was used as a positive control for DNA stimulated ATPase activity. ATP hydrolysis was most effective in a 2:1 molar ratio of ATP:magnesium. Cas3 K320L, mutated in ATPase motif I, had 2-4 fold reduced ATPase activity when compared to wild type Cas3 at fixed protein concentrations as a function of time (Fig. 5C), as expected and in agreement with *S. thermophilus* Cas3 data [36].

To test directly if ATPase activity of Cas3 could drive helicase unwinding of an R-loop we generated a model substrate from oligonucleotides summarised in table 1 and detailed in supplementary table 2. Incubation of this R-loop with Cas3, ATP and magnesium gave a product (Fig. 5D lanes 2-4) that was barely detectable when ATP was substituted for poorly hydrolysable ATPγS (lanes 15-17), or if Cas3 was disrupted in ATPase motif II (Walker B/DExH box, H455L, lanes 8-10). Helicase activity of Cas3 was unaffected by mutation in the HD motif (Fig. 5D lanes 5-7), and was moderately reduced by mutation of glycine and serine residues in ATPase motif I (Walker A, "GSSY", 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 (Fig. 5E). Purified archaeal MthCas3 also unwound the R-loop substrate in the same way (Fig. 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 λvir described in Figure 4C and in reference [3], and a recent study showed R-loop formation by purified *E. coli* Cascade complex independently 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 (Fig. 6). This indicated that crRNA substrates we *in vitro* transcribed for this work are relevant for recognition not

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

just by Cas3 but also for the key CRISPR effector complex Cascade. This is also consistent with expression of 600-nucleotide crRNA that was used with Cascade and Cas3 to effect *in vivo* protection of cells against λvir, originally described in reference [3], and also shown in this work. These partly overlapping functions of Cas3 and Cascade promoting R-loop formation are discussed in the next section.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

DISCUSSION

Cas3 R-loop formation and the HD domain

E. coli Cas3 and Cascade (CasABCDE) participate in processing CRISPR RNA (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 *Streptococcus thermophilus* Cas3 can unwind duplexes formed from ssRNA and ssDNA, and Cas3 nuclease activity is proposed to destroy invading DNA [36, 49]. Helicase data presented in this work 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 *Methanothermobacter thermautotrophicus*, 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.

R-loop formation by *E. coli* Cas3 required magnesium 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 characterised 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: *Streptococcus 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, *Sulfolobus solfataricus* protein Sso2001, which is homologous to the HD domain of fused HD-helicase Cas3, showed nicking activity on dsDNA and dsRNA 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 in references [36, 37]. This suggests that 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 this work. 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 maltose binding protein (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 emphasise that this is currently speculative and presentation of detailed atomic structures of Cas3 should lead to major insights into domain functioning.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Roles of Cas3 and Cascade

The importance of the HD motif for Cas3 functioning in protection of *E. coli* cells against plaque formation by λvir was emphasised by reduced protection of cells expressing K78L Cas3 compared to 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 emphasise 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 this manuscript [20]. Consistent with this, we also observed R-loops formed by 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 recognise R-loops by its ability to unwind them, and it might be that Cascade primes formation of R-loop that is presented to Cas3 for extension and stabilisation. More research will be needed to unravel if 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 to *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 of future research.

Funding

Funding for this research was through a BBSRC PhD studentship to JLH and grant 119-1191196-1201 to II-B from the Croatian Ministry of Science, Education and Sports 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 ELB and II-B in 2008.

Acknowledgements

We are grateful to Alexandra Hughes and Jamie Webster for support and to Peter McGlynn and Mark Dillingham for useful discussion.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.

© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

References

1 Jansen, R., Embden, J. D., Gaastra, W. and Schouls, L. M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 43, 1565-1575

2 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

3 Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, 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

4 Edgar, R. and Qimron, U. (2010) The Escherichia coli CRISPR system protects from lambda lysogenisation, lysogens and prophage induction. J. bact. 192, 6291-6294

5 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

6 Marraffini, L. A. and Sontheimer, E. J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322, 1843-1845

7 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

8 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

9 Cady, K. C. and O'Toole, G. A. (2011) Non-identity Targeting of Yersinia-Subtype CRISPR-Prophage Interaction Requires the Csy and Cas3 Proteins. J Bacteriol

10 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

11 Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., Phanse, S., Joachimiak, A., Koonin, E. V., Savchenko, A., Emili, A., Greenblatt, J., Edwards, A. M. and Yakunin, A. F. (2010) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol Microbiol 79, 484-502

12 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

13 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

14 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

15 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., van der Oost, J. and Koonin, E. V. (2011) Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol

16 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

17 Karginov, F. V. and Hannon, G. J. (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol Cell 37, 7-19

18 Vale, P. F. and Little, T. J. (2010) CRISPR-mediated phage resistance and the ghost of coevolution past. Proc Biol Sci 277, 2097-2103

19 Terns, M. P. and Terns, R. M. (2011) CRISPR-based adaptive immune systems. Curr Opin Microbiol In Press

20 Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J., van der Oost, J. and Brouns, S. J. (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

21 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

22 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

23 Beloglazova, N., Brown, G., Zimmerman, M. D., Proudfoot, M., Makarova, K. S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., Koonin, E. V., Edwards, A. M., Savchenko, A. and Yakunin, A. F. (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J Biol Chem 283, 20361-20371

24 Lillestol, R. K., Redder, P., Garrett, R. A. and Brugger, K. (2006) A putative viral defence mechanism in archaeal cells. Archaea 2, 59-72

25 Tang, T. H., Bachellerie, J. P., Rozhdestvensky, 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

26 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

27 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

28 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

29 Diez-Villasenor, C., Almendros, C., Garcia-Martinez, J. and Mojica, F. J. (2010) Diversity of CRISPR loci in Escherichia coli. Microbiology 156, 1351-1361

30 Touchon, M. and Rocha, E. P. (2010) The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella. PLoS One 5, e11126

31 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

32 Westra, E. R., Pul, U., Heidrich, N., Jore, M. M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L., Mastop, M., Wagner, E. G., Schnetz, K., Van Der Oost, J., Wagner, R. and Brouns, S. J. (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

33 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

34 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

35 Aravind, L. and Koonin, E. V. (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem Sci 23, 469-472

36 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

37 Han, D. and Krauss, G. (2009) Characterization of the endonuclease SSO2001 from Sulfolobus solfataricus P2. FEBS Lett 583, 771-776

38 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

39 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

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

40 Cheok, 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

41 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

42 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 [corrected]. 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., Young, M. J., White, M. F. and Lawrence, C. M. (2011) Structural and functional characterization of an archaeal CASCADE complex for CRISPR-mediated viral defense. J Biol Chem

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. TIBS 25

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 YjjG. 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. Nucl. 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

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Figure Legends

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 (ORFs) 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 first residues of each motif are given: 70, 315 and 452. Highlighted in each motif are residues mutagenised in this work. **(C).** Coomassie stained SDS-PAGE gel showing purified *E. coli* Cas3 as an N-terminal fusion to *E. coli* maltose binding protein (MBP-Cas3), giving an approximate molecular weight of 143 kDa.

Figure 2. *E. coli* Cas3 promotes ATP-independent association and annealing of RNA with complementary double stranded DNA. (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 magnesium chloride, crRNA3 and uncut plasmid pJLH14. RNA and DNA substrates are summarised in table 1. Deproteinised reactions gave two RNA products labelled X and Y. **(B).** Time course reactions of free crRNA3 converted to product with pJLH14 by Cas3 (200 nM), measured using ImageQuant. RNA product was expressed as a proportion of total RNA detected in each lane. Reactions were in duplicate and the graph shows standard error bars. **(C).** Agarose gels monitoring positions of plasmid DNA and radiolabelled RNA after deproteinising identical reactions that contained crRNA3, pJLH14 added to Cas3 (300 nM, lanes 2 and 4) or to no Cas3 (lanes 1 and 3). Lanes 1 and 2 are stained with ethidium bromide (1 μg /mL) 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 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) that had been pre-heated to 85°C for 10 minutes. Amount of RNA product is given below the panel as a percentage of total RNA present in each reaction.

Figure 3. ATP-independent RNA-DNA products of Cas3 are R-loops most efficiently formed with linearised plasmid using magnesium 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 linearised, 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 to linearised pJLH14 (lanes 5-8). A new product is present in reactions containing linearised 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 (+$MgCl_2$ 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.

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, K78LK320L and G317SS318Y, lanes 6-8) were assessed for their ability to form R-loops at a fixed concentration (200 nM) compared to wild type Cas3 (200 nM, lane 2). Reactions contained crRNA3, uncut pJLH14 and magnesium. RNA product was quantified in ImageQuant as a proportion of total RNA for each lane, using free RNA and product in lane 1 (zero Cas3) as baseline activity. Amount of free RNA converted to R-loop product is given beneath each lane. **(B).**

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Graph and corresponding agarose gel showing R-loop formation as a function of time by using 200 nM wild type Cas3 compared to 200 nM K78L HD motif mutant Cas3 (gel lanes 4-8) in otherwise identical reactions containing radiolabelled crRNA3 with linearised pJLH14 DNA and 5 mM magnesium chloride. Reactions were in duplicate to generate a graph with standard error bars. **(C).** Assays comparing wild type and K78L Cas3 for their ability to contribute to protection of *E. coli* cells against plaque formation by phage λvir. Phage was diluted from stock as shown to the right of the panels and spotted onto 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 6-9). Samples from a reaction mixture were removed at various time points (5, 15, 25 and 45 min.) and deproteinised. **(B).** ATPase activity of *E. coli* Cas3 as mean averages of reactions in duplicate with standard error bars. Reactions contained 2mM magnesium and 4 mM ATP, and either no nucleic acid (white squares), 1 μg ϕX174 ssDNA (filled squares) or 1 μg of ϕX174 duplex DNA (filled circles). ATPase activity was calculated using malachite green dye as a reporter for liberation of free phosphate from ATP. Reaction lacking Cas3 was a control for zero ATPase activity at each time point (filled triangles). **(C).** Cas3 with an amino acid substitution in ATPase motif I (K320L, filled circles, see Fig. 1B) is defective in ATPase activity when compared to wild-type protein (filled squares). **(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 ATPγS (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 to that of Cas3 proteins mutated in the HD motif (K78L, lanes 5-7), or ATPase motif II (WalkerB/DExH box, H455L), or ATPase motif I (Walker A/Ploop, G317SS318Y, labelled GSSY, lanes 11-13). **(E).** This panel is equivalent to reactions shown in lanes 2-5 of part C, containing zero, 10, 100 or 300 nM Cas3 protein as indicated, with reaction products compared to a marker molecule (lane 1) of R-loop lacking the invading strand as indicated to the left of the panel.

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 minutes at room temperature prior to mixing with RNA and DNA for 10, 30 and 60 minutes before stopping and deproteinising reactions for visualising RNA.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Table 1: Summary of DNA and RNA substrates used in this study. Further details are given in supplementary table ST1.

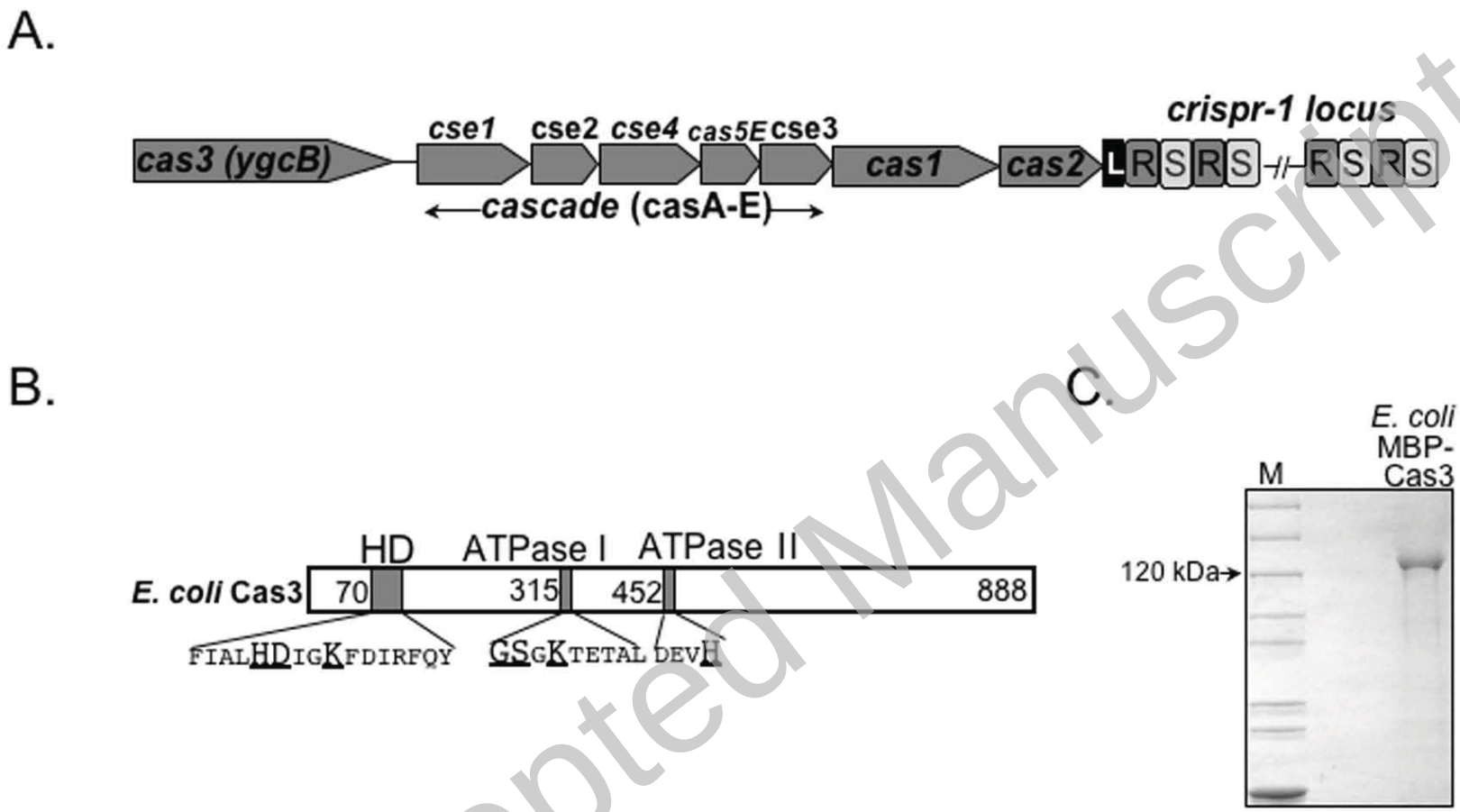
Substrate	Description
CRISPR RNA3 (crRNA3)	269nt single strand RNA comprising 61 nt of *E. coli* CRISPR-1 and 208 nucleotides
pJLH13	Target DNA plasmid for crRNA3. Is 100% complementary to only 61bp sequence in CRISPR part of crRNA3.
pJLH14	Target DNA plasmid for crRNA3. Is 100% complementary with all 269 base pairs of crRNA3.
CRISPR RNA2 (crRNA2)	1 kb nt single strand RNA comprising entire *E. coli* 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 21nt mismatch into which 21nt RNA strand is base paired.
ELB58B	ssDNA that anneals to ϕX174 virion ssDNA

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

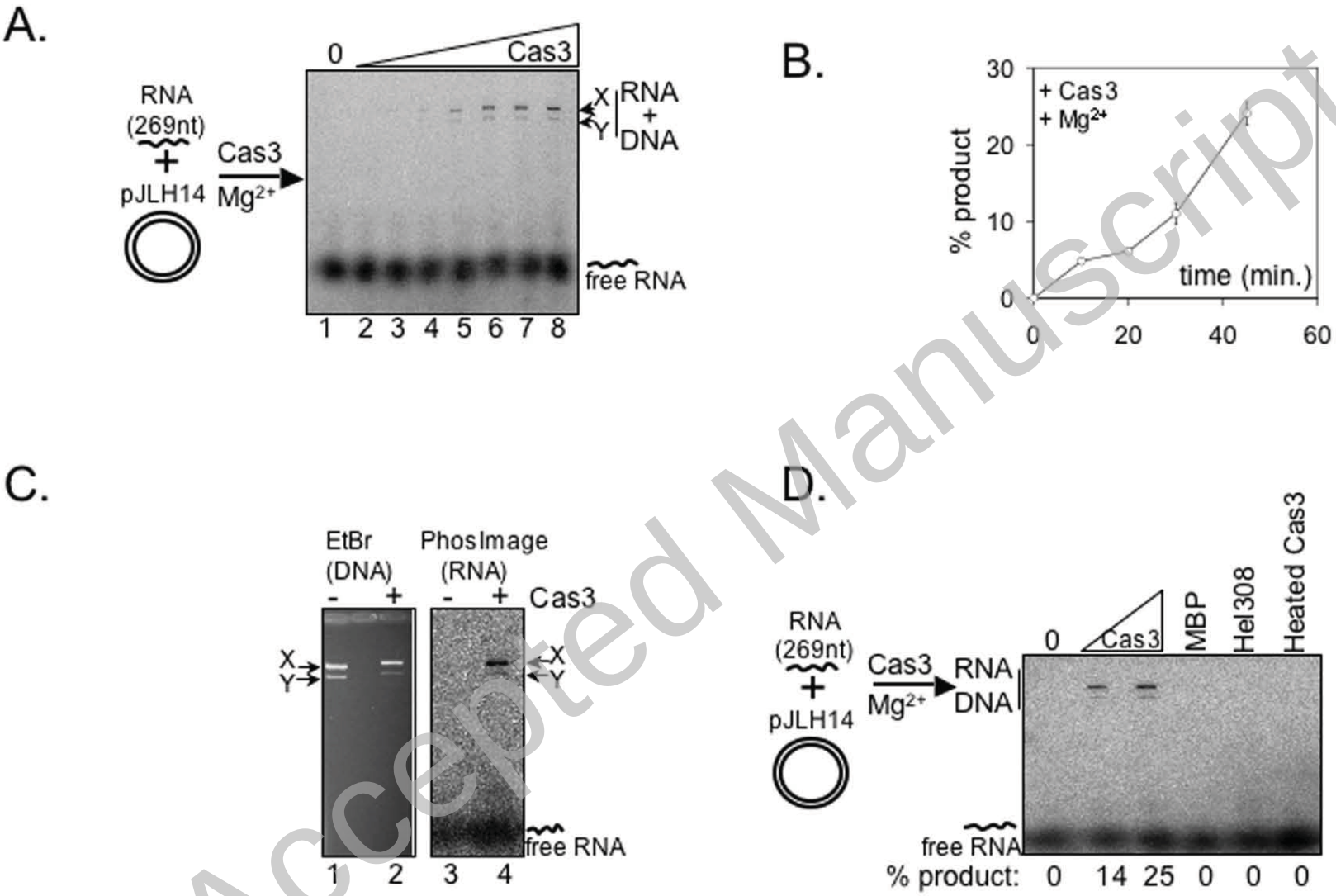


Howard *et al* Fig. 1

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

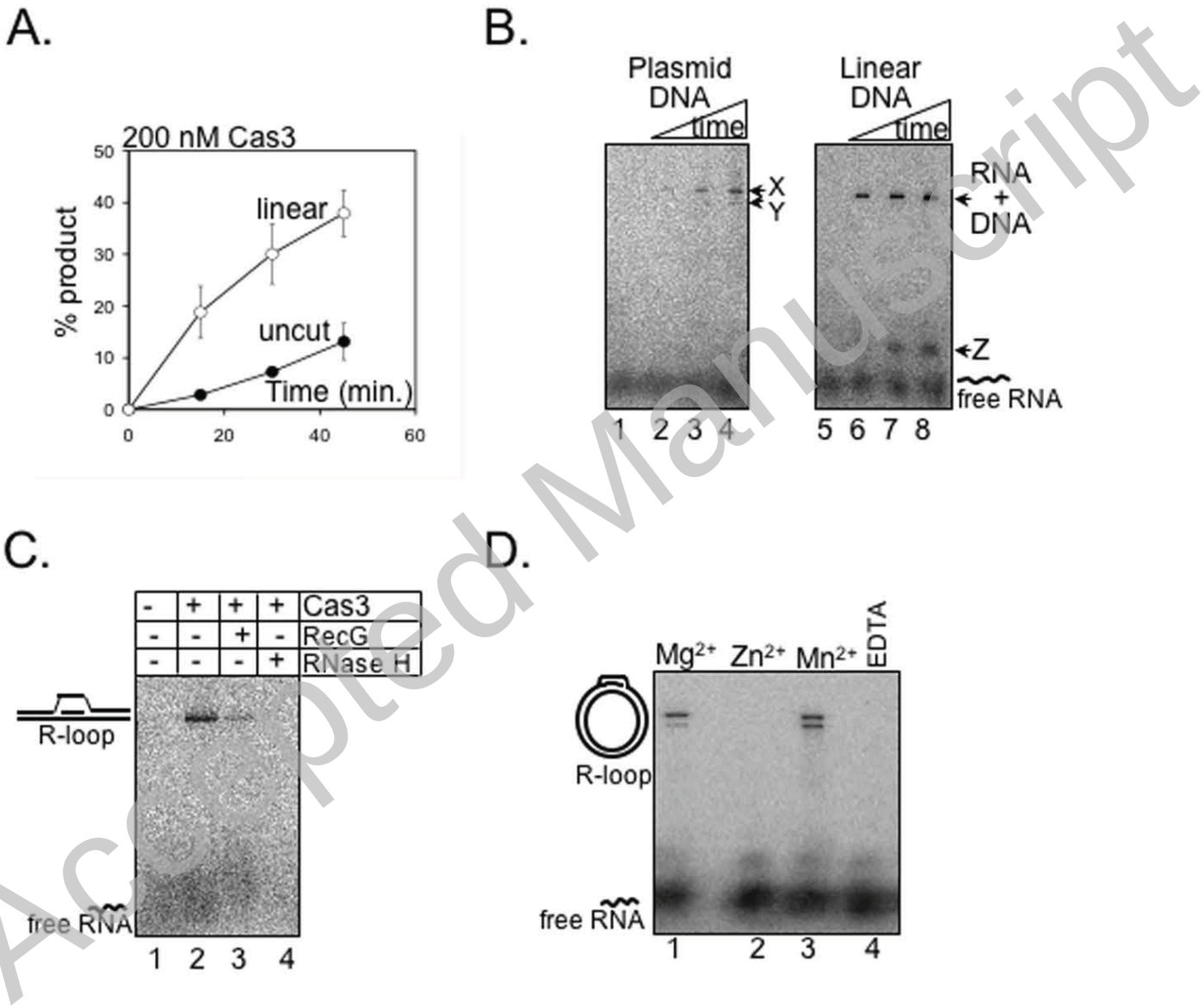


Howard *et al* Fig. 2

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

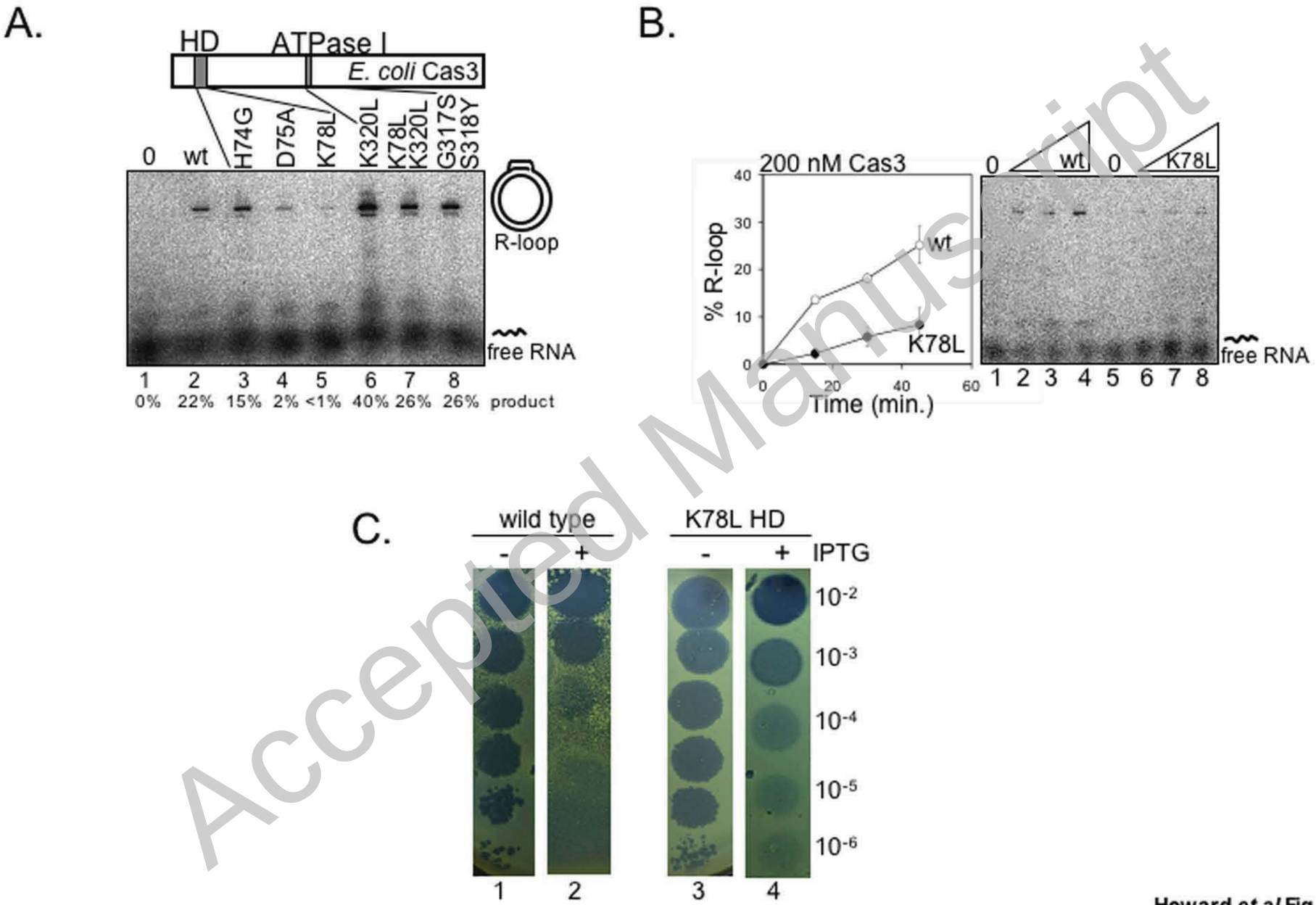


Howard *et al* Fig. 3

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

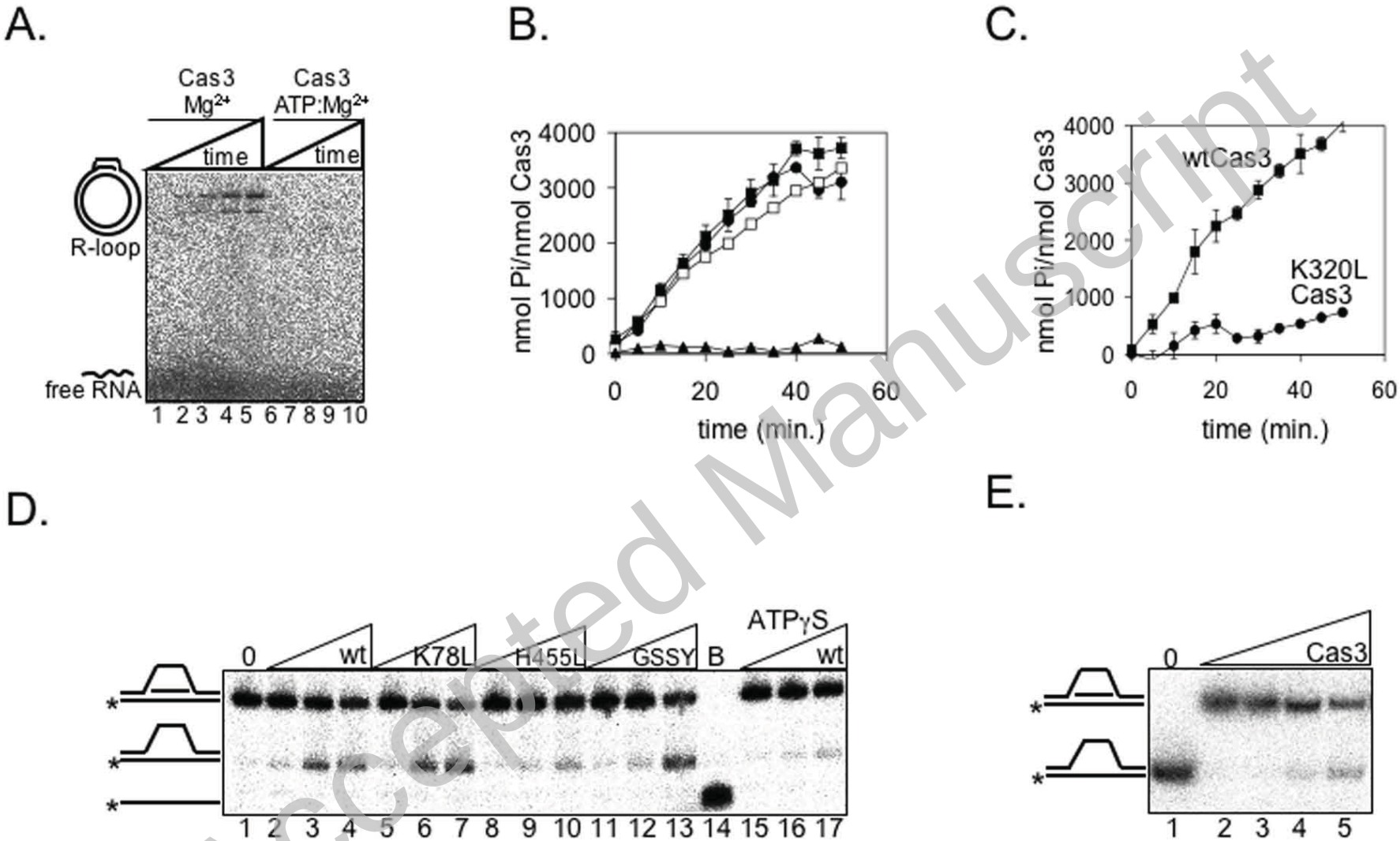


Howard *et al* Fig. 4

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901

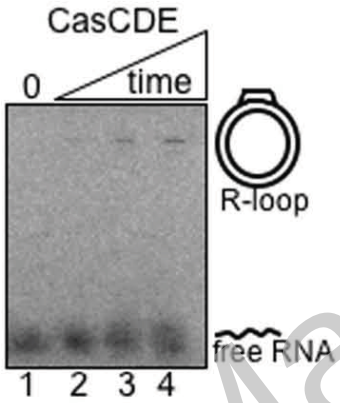


Howard *et al* Fig. 5

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110901



Howard *et al* Fig. 6

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited