

Research Article

# Mechanistic insights into Lhr helicase function in DNA repair

Ryan J. Buckley<sup>1,\*</sup>, Kevin Kramm<sup>2,\*</sup>,  Christopher D. O. Cooper<sup>3</sup>, Dina Grohmann<sup>2</sup> and  Edward L. Bolt<sup>1</sup>

<sup>1</sup>School of Life Sciences, University of Nottingham, Nottingham, U.K.; <sup>2</sup>Institute of Microbiology and Archaea Centre, University of Regensburg, 93053 Regensburg, Germany; <sup>3</sup>Department of Biological and Geographical Sciences, School of Applied Sciences, University of Huddersfield, Huddersfield, U.K.

**Correspondence:** Edward L. Bolt (ed.bolt@nottingham.ac.uk) or Dina Grohmann (dina.grohmann@ur.de)



The DNA helicase *Large helicase-related* (Lhr) is present throughout archaea, including in the Asgard and Nanoarchaea, and has homologues in bacteria and eukaryotes. It is thought to function in DNA repair but in a context that is not known. Our data show that archaeal Lhr preferentially targets DNA replication fork structures. In a genetic assay, expression of archaeal Lhr gave a phenotype identical to the replication-coupled DNA repair enzymes Hel308 and RecQ. Purified archaeal Lhr preferentially unwound model forked DNA substrates compared with DNA duplexes, flaps and Holliday junctions, and unwound them with directionality. Single-molecule FRET measurements showed that binding of Lhr to a DNA fork causes ATP-independent distortion and base-pair melting at, or close to, the fork branchpoint. ATP-dependent directional translocation of Lhr resulted in fork DNA unwinding through the ‘parental’ DNA strands. Interaction of Lhr with replication forks *in vivo* and *in vitro* suggests that it contributes to DNA repair at stalled or broken DNA replication.

## Introduction

Lhr (*Large helicase-related*) protein is an ATP-dependent DNA translocase and helicase that forms a distinct group within Superfamily 2 helicases [1,2]. Lhr was discovered and named in bacteria [2], in which it is present in eight of ~30 phyla [2,3]. It is widespread in archaea [4], and the archaeal Lhr is a sequence homologue of the DDX-family of uncharacterized putative helicases found in eukaryotes including in humans [5–7]. Archaeal and bacterial Lhr proteins show high amino acid sequence identity (typically ~30%) between their N-terminal 800–900 amino acids, which is referred to as the ‘Lhr-Core’, that comprises their helicase domains [8]. Bacterial Lhr is extended to 1300–1500 amino acids by a region of unknown function that lacks obvious sequence homologues. Biochemical analysis of the Lhr-Core from the bacteria *Mycobacterium smegmatis* and *Pseudomonas putida* identified ATP-dependent ssDNA translocation with 3′ to 5′ directionality [9,10]. A crystal structure of bacterial Lhr-Core highlights significant similarities with the archaeal DNA repair helicase Hel308 [9,11], most notably in the orientation and interaction of its winged helix domain (WHD) with RecA-like domains typical of Ski2-like helicases [12,13].

Lhr-Core is conserved in many archaea and bacteria, in a genomic context adjacent to a manganese-dependent phosphodiesterase (MPE), an enzyme with active site architecture resembling Mre11 [8]. In other bacteria, full-length Lhr frequently occurs adjacent to the gene encoding RNaseT, which has roles in DNA repair and RNA maturation [14,15]. Deletion of the Lhr-Core gene (Saci\_1500) in the archaeon *Sulfolobus acidocaldarius* resulted in a mild, ~4-fold, sensitivity to UV irradiation in comparison with wild type cells [16]. In contrast, genetic analysis of Lhr in *E. coli* revealed a phenotype in cells treated with the replication inhibitor AZT — deletion of gene *lhr* was synergistic with deletion of the gene encoding the replication-recombination-repair protein RadA [17]. These observations, and reported 4-fold up-regulation in transcription of *lhr* in *M. tuberculosis* in response to mitomycin C [18], suggest that Lhr may be part of a prokaryotic replication-coupled

\*These authors contributed equally to this work.

Received: 12 May 2020  
 Revised: 23 July 2020  
 Accepted: 23 July 2020

Accepted Manuscript online:  
 24 July 2020  
 Version of Record published:  
 19 August 2020

DNA repair pathway. In this work we investigated the properties of Lhr protein from archaea, a homologue of the eukaryotic DDX proteins. We provide evidence that archaeal Lhr interacts with stalled DNA replication, and that the purified Lhr protein has a preference for targeting forked DNA, remodelling it at the fork branch-point prior to its dissociation.

## Materials and methods

### Molecular cloning of archaeal Lhr

The *lhr* gene (open reading frame mt\_1802) from the euryarchaeon *Methanothermobacter thermautotrophicus* (Mth) was first cloned into pBluescript using *SalI* and *XbaI* restriction endonuclease sites (pEB307) after PCR amplification from Mth genomic DNA (a kind gift from Prof. James Chong, University of York). The Mth *lhr* gene contains an internal *NdeI* restriction site that was altered by silent mutation using QuikChange II site-directed mutagenesis (Agilent). This allowed sub-cloning through a second PCR amplification into pET22b and pT7-7 using *NdeI* and *EcoRI* restriction sites (respectively, pEB352 and pEB353). DNA sequences of these constructs were verified to confirm that plasmids were suitable for protein expression and genetic analysis in *E. coli*.

### Genetic analysis of archaeal Lhr

The basis and details for the genetic assay using *E. coli* strain *dnaE486 ΔrecQ* (Figure 1) are detailed in reference [19]. *E. coli* cells were transformed with empty plasmid vector pT7-7, or with pT7-7 constitutively expressing either bacterial RecQ as a control [20], verified helicases from *M. thermautotrophicus* — Hel308 [19], Cas3 [21] and Hef [22] — or putative archaeal helicases, also from *M. thermautotrophicus* — mt1347 and mt0203. Transformed cells were grown in a shaking water bath at 30°C from colonies inoculated in LB broth containing ampicillin (50 µg/ml), until OD<sub>600</sub> of 0.5. Then 100 µl of culture was spread onto a sector of each agar ampicillin plate for incubation at 30°C, 37°C or 42°C.

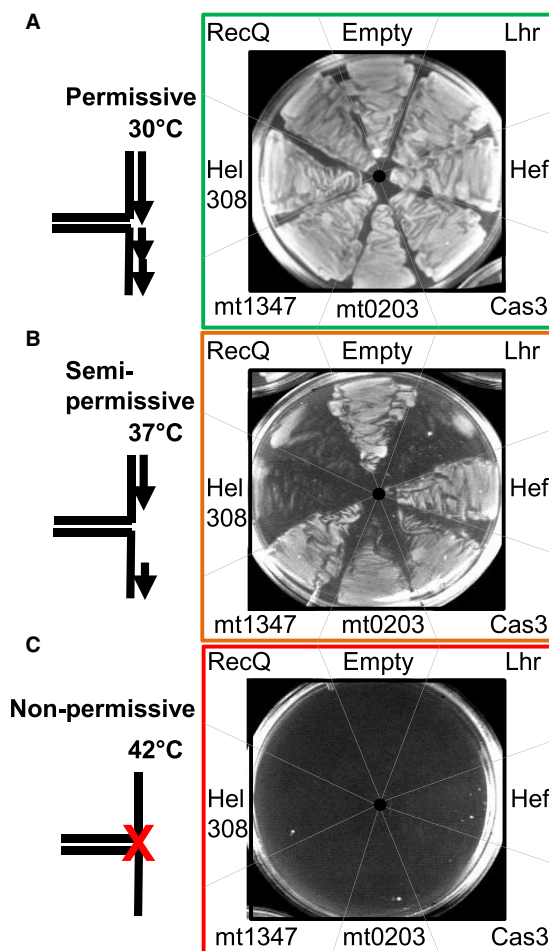
### Purification of archaeal Lhr protein

Plasmid pEB352 was transformed into *E. coli* strain BL21 Codon+ (Agilent) for overexpression of Lhr protein from the archaeon *Methanothermobacter thermautotrophicus*. An overnight culture of this (20 ml) was added to 2 L of LB-ampicillin (50 µg/ml) and chloramphenicol (10 µg/ml) and grown at 30°C with shaking in baffled flasks. At an O.D<sub>600</sub> of 0.5, Lhr expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG, 0.8 mM) and growth was continued for a further two hours. Harvested cells were resuspended in buffer C (20 mM Tris.HCl pH 8.0, 10% glycerol, 100 mM potassium chloride and 2 mM DTT) for –80°C storage.

To purify Lhr protein the biomass was thawed on ice, sonicated and clarified by centrifugation. Soluble protein supernatant was loaded in buffer C into a 5 ml HiTrap Heparin column, and Lhr was eluted in a linear gradient of 0.1–1.5 M potassium chloride in buffer C at ~0.7–0.9 M. Peak Lhr fractions were pooled and loaded directly onto a 16/60 sephacryl S200 column in buffer C, and peak fractions were pooled and dialyzed overnight in buffer C. Dialyzed Lhr was loaded onto a 1 ml HiTrap Q sepharose column and was eluted in a linear gradient of 0.1–1.5 M potassium chloride in buffer C, at ~0.6–0.8 M potassium chloride. Peak Lhr fractions were pooled and dialyzed into buffer C containing 35% glycerol, and stored as aliquots after flash-freezing for storage at –80°C.

### Preparation of DNA substrates for helicase and DNA binding assays

Nucleotide sequences used to generate all substrates are given in Supplementary Table S1. One DNA strand (900 ng in a 20 µl reaction volume) for each substrate was 5'-end labelled with <sup>32</sup>P using T4 polynucleotide kinase and γ<sup>32</sup>-P-ATP. The radio-labelled DNA strand was separated from unincorporated γ<sup>32</sup>-P-ATP using a BioSpin 6 column and the resulting labelled DNA was mixed with 900 ng of each appropriate unlabelled strand in 1× SSC buffer (150 mM sodium chloride, 15 mM sodium citrate at pH 7.0), heated to 95°C for 5 min and allowed to anneal by cooling overnight to room temperature. Resulting DNA was mixed with gel loading dye and loaded onto a 10% TBE gel for electrophoresis at 150 volts for 2 h. The gel was then exposed to autoradiography film and the developed film revealed the positions of the desired substrates for excision from the gel. DNA was eluted from excised gel slices by soaking overnight at 4°C in 20 mM Tris.HCl pH 7.5 containing 20 mM sodium chloride. DNA in buffer recovered from gel debris was quantified by scintillation counting using as standards the scintillation counts of samples taken throughout the procedure that were of known



**Figure 1. Archaeal Lhr interacts with stalling replication forks in *E. coli* *dnaE486*  $\Delta$ *recQ* cells.**

Panels are colour-coded to illustrate temperatures at which replication is unhindered (permissive, 30°C), destabilized (semi-permissive, 37°C) or stopped (non-permissive, 42°C). For each temperature cells were spread onto ampicillin agar after expressing the protein indicated from a plasmid. (A) At 30°C cells replicate normally resulting in fully viable growth in each sector. (B) At 37°C replication is destabilized by the *dnaE486* allele [20]. Hel308 and RecQ, gave inviability phenotypes as expected [19,20,31], and Lhr gave the same phenotype. (C) At 42°C the *dnaE486* allele makes cells inviable — this is used as a control that *dnaE486* suppressor mutations have not arisen.

DNA mass (ng). This established the final yield of substrate DNA in ng that was converted to a final concentration of DNA (nM) for use in assays.

## Helicase assays and EMSAs

See Supplementary Table S1 for substrates. Helicase reactions were in buffer HB (20 mM Tris.HCl pH 7.5, 2 mM DTT, 100  $\mu$ g/ml BSA and 7% glycerol) supplemented with 2 mM ATP (at pH 7.5) and 1 mM magnesium chloride. Helicase assays were at 45°C for either 20 min or in reactions over a time course as shown. Reactions were stopped by addition of de-proteinising buffer (1 $\times$  is 0.625% SDS, 50 mM EDTA and 2.5 mg/ml proteinase K) and gel loading dye was added prior to electrophoresis at 150 volts for 1 h through a 10% acrylamide TBE gel. Assay products were imaged on a storm<sup>TM</sup> scanner (Amersham) from phosphorimaging screens, after drying the gels under a vacuum on a flatbed gel dryer. Assay products were quantified from TIF files of gel images using the GelEval software. For EMSAs, Lhr (100 nM) was mixed with DNA (10 nM) in buffer HB at room temperature with reactions loaded directly onto a 5% acrylamide TBE gel and were imaged using the ChemiDoc MP imaging system (Bio-Rad).

## Assays using fluorescent DNA fork-2 and confocal single-molecule FRET measurements

Fluorescent fork-2 DNA was formed from the four fork-1 oligonucleotides (Supplementary Table S1) mixed in equimolar concentration (10  $\mu$ M) in annealing buffer (10 mM Tris.HCl pH 7.8, 50 mM NaCl, 1 mM EDTA), heated to 95°C for 3 min and cooled to room temperature (23°C) over 1.5 h. DNA was stored at –20°C. For EMSAs, Lhr (100 nM) was mixed with DNA (10 nM) in buffer HB at room temperature with addition of ATP and magnesium chloride (1 : 2 mM) as indicated in Figure 3, and reactions loaded directly onto a 5% acrylamide TBE gel. Gels were imaged using the ChemiDoc MP imaging system (Bio-Rad).

Prior to FRET measurements, the sample chambers (Cellview slide, Greiner Bio-One) were passivated with 2 mg/ml BSA in 10 mM Tris.HCl pH 8.0 for 10 min and washed once with Millipore water. For formation of complexes, 1 nM DNA, 1  $\mu$ M LHR, 1 mM MgCl<sub>2</sub> and 2 mM ATP were mixed in H78 buffer (20 mM NaHEPES pH 7.8, 10% (v/v) glycerol, 100 mM potassium acetate, 1 mM EDTA, 2 mM DTT) and incubated for up to 20 min at room temperature or 45°C. Afterwards, samples were diluted by a factor of 10 in H78 buffer and added to the sample chamber.

Single-molecule fluorescence of diffusing complexes was detected with a MicroTime 200 confocal microscope (PicoQuant) equipped with pulsed laser diodes (532 nm: LDH-P-FA-530B; 636 nm: LDH-D-C-640; PicoQuant/cleanup filter: zet635; Chroma). The fluorophores were excited at 20  $\mu$ W using pulsed interleaved excitation (40 MHz). Emitted fluorescence was collected using a 1.2 NA,  $\times$ 60 microscope objective (UplanSApo  $\times$ 60/1.20W; Olympus) and focused through a 50  $\mu$ m confocal pinhole. A dichroic mirror (T635lpxr; Chroma) was used to separate donor and acceptor fluorescence. Additional bandpass filters (donor: ff01-582/64; Chroma; acceptor: H690/70; Chroma) completed spectral separation of the sample fluorescence. Each filtered photon stream was detected by an individual APD (SPCM-AQRH-14-TR, Excelitas Technologies) and recorded by a time-correlated single photon counting (TCSPC) capable HydraHarp 400 (PicoQuant).

## FRET data analysis

Data analysis of confocal FRET measurements was performed with the software package PAM [23]. Photon bursts of diffusing molecules were selected based on an all-photon burst search (APBS, parameters: L = 100, M = 10, and T = 500  $\mu$ s) and an additional dual-channel burst search (DCBS, parameters: L = 100, M<sub>GG+GR</sub> = 20, M<sub>RR</sub> = 20, and T = 500  $\mu$ s).

For an APBS, the FRET efficiency of each burst (calculated as proximity ratio  $E_{PR}$ ) and the raw stoichiometry factor  $S_{raw}$  was calculated as:

$$E_{PR} = \frac{N_{DA}}{N_{DD} + N_{DA}} \quad (1)$$

$$S_{raw} = \frac{N_{DD} + N_{DA}}{N_{DD} + N_{DA} + N_{AA}} \quad (2)$$

where  $N_{DD}$ ,  $N_{DA}$  and  $N_{AA}$  are the number of detected photons. Indices refer to donor donor emission upon donor excitation (DD), acceptor emission upon donor excitation (DA) and acceptor emission upon acceptor excitation (AA). These were used to calculate the donor leakage and direct excitation correction factors. For DCBS, the FRET efficiency  $E$  and the stoichiometry factor  $S$  of each burst were calculated as:

$$E = \frac{N_{DA} - (c_{leak} \cdot N_{DD} + c_{dir} \cdot N_{AA})}{\gamma \cdot N_{DD} + N_{DA} - (c_{leak} \cdot N_{DD} + c_{dir} \cdot N_{AA})} \quad (3)$$

$$S = \frac{\gamma \cdot N_{DD} + N_{DA} - (c_{leak} \cdot N_{DD} + c_{dir} \cdot N_{AA})}{\gamma \cdot N_{DD} + N_{DA} + \beta \cdot N_{AA} - (c_{leak} \cdot N_{DD} + c_{dir} \cdot N_{AA})} \quad (4)$$

where  $c_{leak}$  is the correction factor for donor leakage,  $c_{dir}$  is the correction factor for direct excitation of the acceptor,  $\gamma$  and  $\beta$  are the detection and excitation correction factors. Burst data were corrected for donor leakage and direct excitation of the acceptor (determined from APBS according to [24], as well as  $\gamma$  and  $\beta$  (determined from DCBS ES-histograms using an internal fit on multiple E/S separated FRET populations). The data were binned (bin size = 0.025), plotted as E histogram and fitted with a single (DNA) or multiple Gaussian fits using the Origin software.

The inter-fluorophore distance  $r$  was calculated from corrected  $E$  values according to:

$$r = R_0 \sqrt[6]{\frac{1 - E}{E}} \quad (5)$$

using the following Förster radius:  $R_0 = 5.9$  nm of the ATTO 532-ATTO 647N dye pair.

## Analysis of Lhr and DDX52 structures

Protein sequence homology was assessed using BLASTP [25] against sequences with a Protein DataBank [26] record, using the *Methanothermobacter thermautotrophicus*  $\Delta$ H open reading frame Mth1802 (UniProt: O27830) and human DDX52 (UniProt: Q9Y2R4) helicase protein sequences as search queries. Protein fold, secondary structure and structural homology searches were performed with Phyre2 [27] under Intensive mode. Predicted structure models were analyzed, superimposed and RMSD calculated with DALI [28], superimposing against the *M. smegmatis* Lhr [9] (PDB: 5V9X) helicase structure. Protein secondary structure was predicted in PSIPRED [29]. Structural models rendered in PyMOL were superimposed using the  $C_\alpha$  chain.

## Results

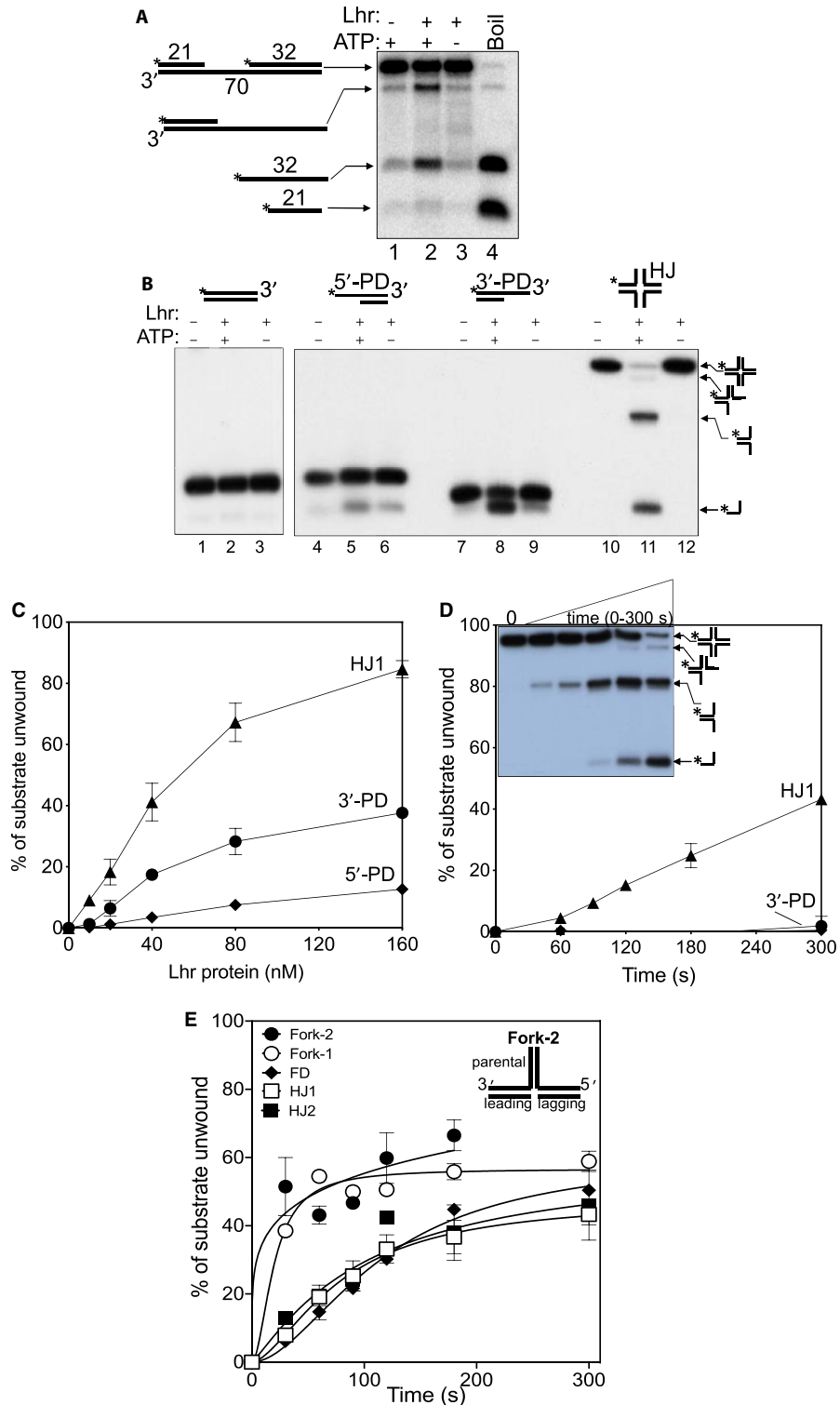
### Genetic analysis of archaeal Lhr indicates interaction with stalled DNA replication

Lhr is distributed throughout the archaeal domain, including in all classes of the Asgardarchaeota that is most closely related to eukaryotes, and in the extremely reduced genomes of Nanoarchaeota — details are presented as Supplementary Data in Supplementary Table S2. We utilized Lhr from the euryarchaeal species *Methanothermobacter thermautotrophicus* (Mth), and first analyzed this Lhr using genetics. Two previous studies in archaea had deleted the *lhr* gene — in *Haloferax volcanii* this gave no discernible phenotype in response to UV or  $\gamma$  irradiation [30], and in *Sulfolobus acidocaldarius* there was very modest (4-fold) UV sensitivity [16]. Here, we observed a robust phenotype from Mth Lhr in a genetic assay that detects interaction with stalled DNA replication [20]. This assay uses *E. coli* cells with a conditional mutation in the gene encoding DNA polymerase III (*dnaE*), the replicative polymerase. This particular mutation, *dnaE486*, causes structural instability of DNA polymerase III at 37°C that triggers stalling of DNA replication, mimicking DNA damage. Cells survive this by activating replication-coupled DNA repair, therefore 37°C is called a ‘semi-permissive’ temperature. However, interference with de-stabilised replication at 37°C by heterologously expressed protein causes low cell viability because native replication-coupled repair is impeded. This assay had previously identified DNA repair phenotypes for archaeal Hel308 and RecQ [19,20,31], and was re-visited to assess other putative archaeal helicases including Lhr (Figure 1). As expected from previous findings [19,20], expression of bacterial RecQ or Hel308 in these cells at permissive temperature (30°C) had no effect on viability (Figure 1A), indicating that these proteins are not toxic when expressed in *E. coli* cells replicating normally, but both caused inviability at 37°C indicating interaction with unstable replication (Figure 1B). Expression of Lhr also caused cell inviability at 37°C, and the normal viability of cells at 30°C confirmed that Lhr protein does not confer toxicity to normal replication. Expression of other known or putative archaeal helicases had no observable effect on cell viability at 37°C (Figure 1A,B). All cells were inviable at 42°C (Figure 1C), a temperature at which the replisome cannot function because of the *dnaE486* mutation — this ensures that suppressor mutations have not arisen to give false positive results at 37°C. This genetic analysis suggests that Lhr, like archaeal Hel308 and bacterial RecQ, interacts with de-stabilised replication forks. This information was taken forward for biochemical analysis of the Mth Lhr protein.

### Archaeal Lhr protein preferentially targets fork-DNA for DNA translocation

The bacterial ‘core’ Lhr (Lhr-Core), which lacks a 700 amino acid C-terminal region present in the bacterial but not archaeal Lhr enzymes, is a ssDNA-stimulated ATPase that translocates ssDNA with 3′ to 5′ directionality [10]. Purified full-length archaeal Lhr (Supplementary Figure S1) was challenged with a gapped DNA duplex substrate to determine if it had similar properties (Figure 2A). In this assay, loading of Lhr onto ssDNA revealed 3′ to 5′ translocation directionality by displacement of the 32 nt strand in preference to the 21 nt strand (Figure 2B lanes 2 and 3). DNA unwinding of the gapped duplex by Lhr *in vitro* was most effective at 2 mM ATP and 1 mM magnesium chloride (Supplementary Figure S2), conditions that were used for





**Figure 2. Lhr is most effective at unwinding branched DNA molecules.**

Part 1 of 2

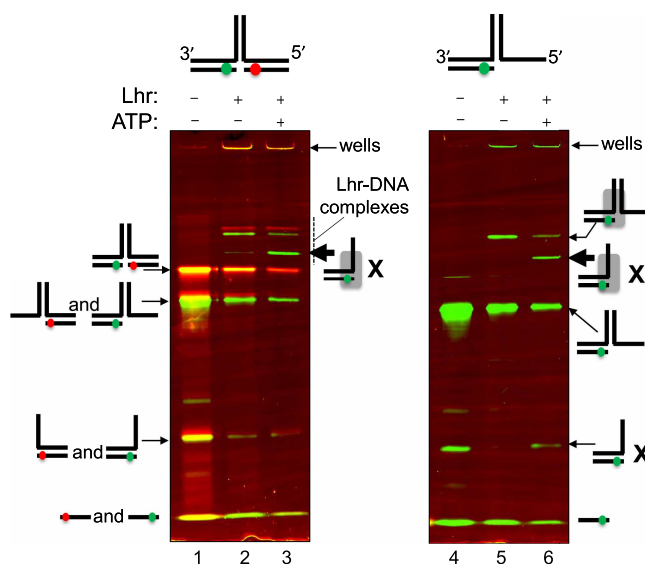
All parts show results of Lhr helicase reactions observed in TBE 10% acrylamide gels. Asterisks indicate 5'-<sup>32</sup>P end labelling of a DNA strand and DNA was used at 10 nM unless stated. (A) Lhr (100 nM) gave ATP-dependent displacement of the 32 nt strand from the gapped duplex (1 nM) indicating 3' to 5' directionality. (B) Lhr (100 nM) did not significantly unwind fully base paired DNA duplex or a partial duplex with a 5' ssDNA tail (5'-PD, lanes 4–6), but unwound a partial duplex DNA with a 3'-ssDNA-tail (3'-PD, lanes 7–9). A Holliday junction (HJ) was unwound more effectively in this assay to generate three-strand,

**Figure 2. Lhr is most effective at unwinding branched DNA molecules.**

Part 2 of 2

two strand and ssDNA products as indicated at the side of the gel panel. The apparent proficiency of Lhr in unwinding the Holliday junction compared with partial duplex DNA Holliday junction was confirmed in part (C), in which Lhr was added to DNA at 10, 20, 80 and 160 nM as indicated. Reactions were repeated three times — the range of standard error is shown. (D) Holliday junction DNA (HJ) was unwound by Lhr (40 nM) at least 10-fold more effectively than the 3'-tailed partial duplex (3'-PD) over the course of time (0–300 s). The inset gel summarizes that Lhr unwound Holliday junctions into products that were further unwound, indicating that Lhr is not specific for targeting Holliday junctions. Reactions were carried out twice, and bars show the standard error. (E) Lhr (40 nM) unwound fork-1 and fork-2 DNA most effectively over time, compared with Holliday junctions. A cartoon of the fork-2 structure is shown for reference to the fork parental, leading and lagging strands. For comparison, the graph also shows a flayed duplex (FD) DNA, comprising the fork parental duplex but neither lagging nor leading strands. Details of the substrates are given in the Supplementary data along with representative gels. Reactions were done three times and bars show standard error from mean.

subsequent assays. We next assessed unwinding of different model synthetic DNA substrates to establish if Lhr had a substrate preference that could be used to gain insight into its DNA unwinding mechanism. In agreement with a requirement for ssDNA to trigger DNA translocation, Lhr did not unwind DNA in a fully base-paired DNA duplex (Figure 2B lanes 1–3). It was weakly active at unwinding a partial duplex with 25 nt of 5' tailed ssDNA (5'-PD, lanes 4–6) but substantially unwound a partial duplex with a 3' ssDNA tail (3'-PD, lanes 7–9). This is in agreement with the 3' to 5' directionality observed when unwinding the gapped duplex (Figure 2A), but some dissociation of the 5' tailed substrate suggested that Lhr may more generally distort DNA base-pairing, leading to low levels DNA strand dissociation, when bound to DNA – further investigation



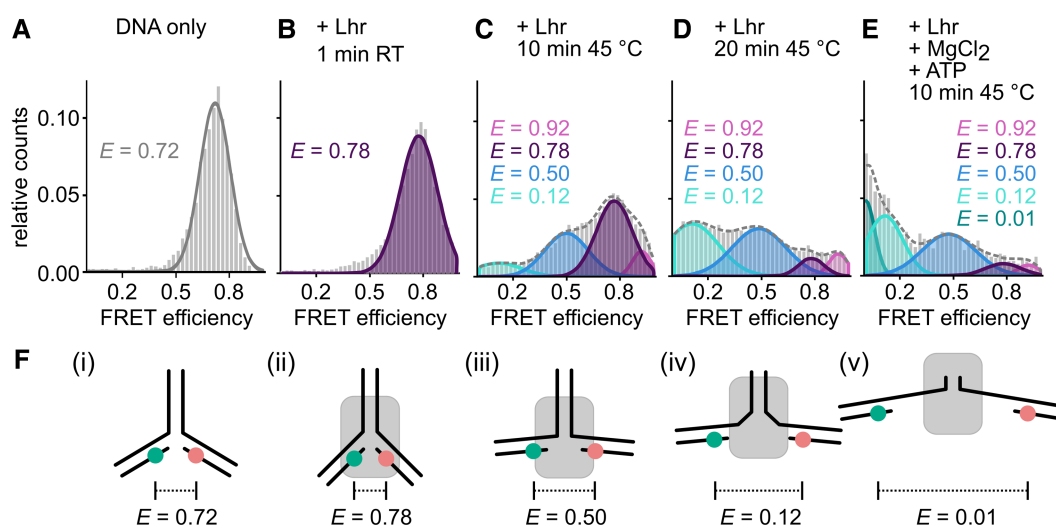
**Figure 3. Evidence for directionality of fork dissociation by Lhr.**

Both panels are from the same native acrylamide EMSA gel, divided to be able to annotate each part with substrate and product DNA. Fork-2 was labelled with ATTO 532 (green) and ATTO 647N (red) fluorophores at the indicated positions in the cartoon representations shown above the gels, a full fork corresponding to lanes 1–3, and a partial fork lacking a lagging strand corresponding to lanes 4–6. The fluorescence signal of the fluorophores was detected using a fluorescence scanner. Green and red bands in the gel correspond to fork-2 and fork-2-Lhr bound reaction products that contain one or both of the labelled DNA strands. Lanes 1 and 4 show bands corresponding to each full substrate as naked DNA, and its component intermediate DNA molecules; each form as shown to the side of the panel. Addition of Lhr (100 nM) and ATP-Mg<sup>2+</sup> to reactions is indicated above each panel. Free substrate (lanes 1 and 4) was bound by Lhr (lanes 2 and 5), and indicated by the label for Lhr-DNA complexes and a grey rectangle denoting DNA-bound protein. Addition of ATP triggered fork dissociation into the products indicated with the letter X. Helicase products could remain bound to Lhr protein, as indicated by the grey rectangles representing Lhr. ATP-dependent formation of Lhr-bound two-strand DNA product is highlighted in lanes 3 and 6.

of this is presented later (Figure 4). Lhr also unwound a partial duplex comprising an RNA–DNA hybrid with a 3′ single stranded ‘tail’ as well as the corresponding tailed DNA duplex (Supplementary Figure S3). These data indicate that Lhr requires single-stranded DNA (ssDNA) to trigger directional translocation/helicase activity.

Unwinding of the 3′ tailed partial duplex (3′-PD) was quite modest — maximally 30% of substrate was unwound when Lhr was used at 10-fold molar excess over DNA (Figure 2C). Lhr unwound an equivalent branched substrate, a Holliday junction (indicated as HJ in the figures), 3-fold to 10-fold more effectively than tailed duplexes measured in, respectively, endpoint (Figure 2C) and time course assays (Figure 2D). This Holliday junction (HJ1) was generated by annealing of the same DNA strand, and its complements, that was used to generate the 3′ tailed partial duplex to ensure DNA sequences were consistent, as detailed in Supplementary Table S1. Lhr generated two major products from unwinding of HJ1 — these products were identifiable as labelled in Figure 2D by comparing them with the single forked product generated by the Holliday junction specific helicase RuvAB (Supplementary Figure S4A), and with ssDNA product of Lhr unwinding the 3′ tailed duplex (Figure 2B lanes 8 and 11, and Supplementary Figure S4B). The structural specificity of RuvAB for unwinding Holliday junctions to only a fork without further unwinding of the fork into ssDNA [32,33], therefore contrasted with Lhr, suggesting that Lhr may be able to target forked DNA for unwinding.

To narrow down the substrate preferences of Lhr *in vitro* we compared unwinding of forked DNA with Holliday junction DNA as a function of time (Figure 2E). Two different Holliday junctions were compared with equivalent forked DNA that comprised a fully base-paired ‘parental’ DNA duplex and leading and lagging strand duplexes of the same DNA sequences as Holliday junctions (Figure 2E and Supplementary Table S1). These assays, using 20 nM of DNA and 40 nM of Lhr protein, indicated modest preference for forked DNA compared with Holliday junctions (Figure 2E). Multiple products from Lhr unwinding Holliday junctions were again apparent (Supplementary Figure S5).



**Figure 4. Single-molecule FRET analysis of conformational changes induced by Lhr on fork-2 DNA.**

Measurements were performed on freely diffusing DNA/protein complexes to monitor Lhr-induced conformational changes on forked DNA substrate. (A) ATTO532 (donor) and ATTO647N (acceptor) labelled fork-2 in the absence of protein, (B) after addition of Lhr (1  $\mu$ M) at room temperature (RT), then (C and D) after 10 min and 20 min of incubation at 45°C. (E) After 10 min at 45°C with addition of 1 mM MgCl<sub>2</sub> and 2 mM ATP. FRET populations were fitted with multiple Gaussian distributions. The mean FRET efficiency *E* of the fitted peaks are shown. The dashed line is the cumulative fit curve. Each measurement was carried out at least three times — see Supplementary Figure S6). (F) Putative model for the mechanism of Lhr-dependent fork DNA unwinding. Conformations are based on inter-fluorophore distances derived from the measured FRET efficiencies that are presented as a data table in Supplementary Figure S6: i. relaxed conformation of fork DNA labelled with donor (green, leading strand) and acceptor (red, lagging strand); ii. compacted fork bound by Lhr (grey); iii. stretched conformation after heat activation of LHR; iv. partially melted fork DNA after ATP-Mg<sup>2+</sup> addition; v. mostly unwound fork still bound to Lhr.



The preference of Lhr for forked DNA that we observed *in vitro* is consistent with Lhr targeting replication forks in genetic assays (Figure 1). But it raised the question, how does Lhr most effectively unwind fully base-paired forks, when it requires access to ssDNA for translocation leading to DNA unwinding? We reasoned targeting of a fork branch-point by Lhr may disrupt base pairing allowing ssDNA loading and translocation, which we investigated using single-molecule Förster resonance energy transfer (smFRET) measurements.

## smFRET measurements reveal ATP-independent remodeling of a DNA fork by Lhr, and ATP-dependent dissociation of the fork-lagging strand

Lhr unwound model DNA forks most effectively in ensemble reactions *in vitro* (Figure 2E), therefore the same fork-2 substrate was used for smFRET analysis. Here, a donor-acceptor dye-pair was positioned in the fork lagging strand (ATTO 647N) and leading strand (ATTO 532) (Supplementary Table S1). We began by assessing Lhr binding and unwinding of this fork-2 in EMSAs, exploiting the dual ATTO labelling that allows for greater differentiation of reaction products than the single <sup>32</sup>P end-radiolabel (Figure 3 lanes 1–3). The reactions were not de-proteinised and consequently LHR in complex with either the complete fork substrate or unwinding intermediates was detected. In reactions lacking ATP, Lhr-fork DNA complexes were observed (Figure 3 lane 2). With ATP, Lhr helicase products primarily result from unwinding of the fork ‘parental’ DNA not fork lagging or leading strands, visible as a single product. The resulting green fluorescing DNA-LHR complex is consistent with the two-strand molecule indicated that would be generated by 3′ to 5′ directionality of Lhr (product X in lane 3). To verify this, we repeated the Lhr binding and unwinding reactions using a partial fork-2 that lacked the red fluorescing lagging strand (Figure 3 lanes 4–6). As expected in the absence of ATP, Lhr bound to the partial fork-2 resulting in a single complex representing Lhr-DNA binding (lane 5). Addition of ATP gave the same two-strand DNA product both bound with Lhr and not bound (both marked X in Figure 3 lanes 4–6), also consistent with the green ATTO labelled partial fork-2 being unwound 3′ to 5′ through the ‘parental’ duplex.

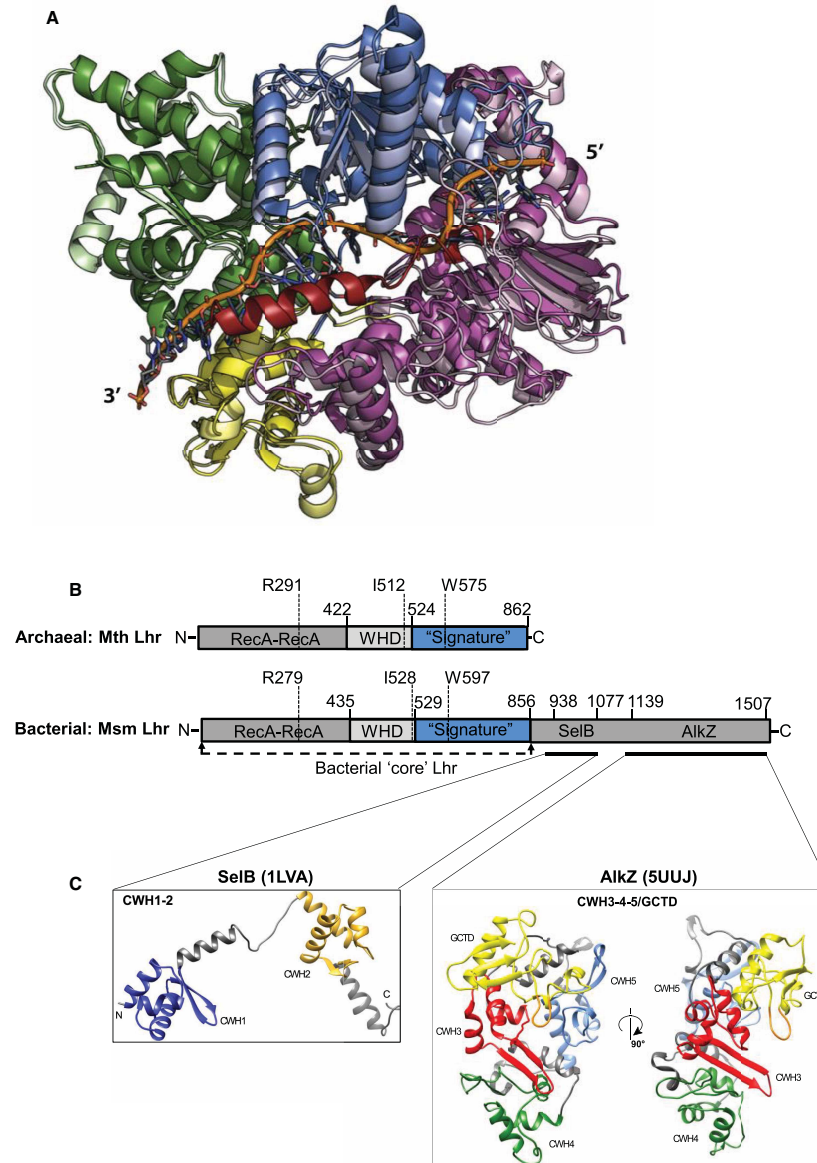
Having gained some qualitative insight into unwinding of the ATTO labelled fork-2 by Lhr we next assessed the effect of Lhr on DNA conformation within the fork at the single-molecule level (Figure 4), by determining the efficiency of energy transfer from donor to acceptor. Higher FRET efficiency (E) values denote shorter inter-dye distances giving a readout of fork conformation at the branchpoint. In the absence of Lhr, the fork DNA gave a single population (E = 0.72) (Figure 4A,Fi) representing a relaxed state with angles of ~130° between the lagging and leading strand DNA. Addition of Lhr at room temperature in buffer without ATP-Mg<sup>2+</sup> shifted the signal to E = 0.78, representing a shortening of the inter-dye distance due to fork compaction or DNA rotation induced by Lhr (Figures 4B and 3F ii). Activating Lhr at 45°C (but without ATP-Mg<sup>2+</sup>) resulted in significant additional FRET populations corresponding to fork DNA undergoing changes into both stretched (E = 0.50) and further compacted (E = 0.92) conformations (Figure 4C,F iii). In these conditions, we also observed decreased signal intensities for compacted forks (E = 0.92 and E = 0.78) that corresponded with an increase in the low FRET efficiency population (E = 0.12), representing a highly stretched or partially unwound fork DNA conformation (Figure 4D,F iv) — the increased inter-dye distance indicated disruption of multiple base pairs close to the fork branch-point. The data indicate that fork DNA binding by Lhr in the absence of ATP causes multiple changes in fork conformation, including partial melting of DNA close to the fork branch-point. Addition of ATP-Mg<sup>2+</sup> resulted in disappearance of the stretched fork signal (E = 0.5, Figure 4E) and appearance of a population with E ~ 0 that results from the fork being mostly unwound, fully separating the FRET dye pair (Figure 4E,F part v).

## Discussion

Lhr protein is highly conserved throughout archaea and has sequence homology with DDX damage repair proteins found in humans and other eukaryotes [7]. Lhr proteins form two sub-groups, Lhr and Lhr-Core, the latter including the archaeal proteins of 800–900 amino acids arranged into RecA-like and accessory domains required for helicase activity. Bacterial *lhr* and bacterial/archaeal *lhr-core* are often located in a conserved genome context with at least one gene encoding a nuclease enzyme; *lhr* with *rnt* that encodes a 3′ to 5′ exonuclease implicated in DNA repair [14,15], and *lhr-core* with MPE, a manganese dependent exonuclease [10]. Our observation of a replication phenotype from expression of archaeal Lhr (Figure 1) is consistent with a role in replication-coupled DNA repair suggested from genetic analyses of Lhr from *E. coli* and *M. tuberculosis*

[17,18]. It is also consistent with our data from *in vitro* helicase assays (Figures 2 and 3) and smFRET (Figure 4) showing that purified Lhr protein targets and unwinds DNA forks.

The 3' to 5' directional DNA translocation of archaeal Lhr is the same as bacterial Lhr [10], and in addition we observe a strong preference for unwinding of DNA within three- or four-stranded forked and Holliday



**Figure 5. Novel features of Lhr protein structures.**

(A) Structural model of the archaeal Lhr used in this work (Mth1802) superimposed onto *M. smegmatis* Lhr (PDB: 5V9X). Lhr is orientated and coloured according to the original description [9] (green, RecA domain 1; blue, RecA domain 2; yellow, winged helix; pink, domain of unknown function), with the Mth\_1802 model in lighter shades. The ssDNA in the Lhr structure is shaded orange, and the *ab initio* modelled Mth1802 C-terminal 30 residues referred to in the text is shaded red. (B) Cartoon summary of the domain organization of Lhr proteins from archaea and bacteria. Labelled are the tandem RecA-like domains, winged helix domains (WHD) and a 'signature' domain of Lhr proteins that is of unknown functions. Amino acid positions are indicated, including invariant amino acids that are required for helicase activity of the bacterial Lhr [9]. Also highlighted is the 'core' helicase of the bacterial Lhr protein, and the C-terminal region of bacterial Lhr that is absent in archaea. (C) Summarizes two parts of the bacterial C-terminal Lhr region that match with structural folds of AlkZ and SelB proteins: CWH, C-terminal winged helix-turn-helix motif; GCTD, glycosylase C-terminal domain.

junction molecules, compared with ssDNA tailed-duplexes. Mycobacterial Lhr-Core was most active on RNA–DNA hybrids that have a 3′ ssDNA tail, although 3- or 4-strand forks or Holliday junctions were not tested [1]. Lhr does not seem to be a bona fide Holliday junction ‘branch migration’ helicase because it unwound model forked DNA better than model Holliday junctions, and because the products formed by Lhr unwinding Holliday junctions differed from the RuvAB branch migration complex. In addition, previous genetic studies on bacterial Lhr showed no strong phenotypes for Lhr associated with RuvABC or RecG-promoted recombination-repair, either epistatic or synergistic.

Our data showed more efficient unwinding of DNA forks by Lhr compared with unwinding of DNA from 3′ ssDNA tail provided to load Lhr for 3′ to 5′ translocation. This was despite the forked substrates being fully base-paired. Using single-molecule FRET we observed substantial melting and remodeling of the fork-2 substrate that would yield the ssDNA needed to trigger the ATP-dependent DNA translocation, thus unwinding the fork. The crystal structure of a mycobacterial Lhr-Core helicase bound to ssDNA most closely resembles the DNA repair helicase Hel308 [9,11,34], another Ski2-like helicase which has the same genetic phenotype as Lhr reported in this work and in previous studies [19,31]. The Lhr crystal structure represents the active translocation stage of Lhr, and the archaeal Lhr used in this work superimposes well when structurally modelled against it (RMSD 0.8 Å). Including a region of the core bacterial and archaeal Lhr proteins, approximately amino acids 520–860, that is of unknown function that has been referred to as a ‘signature’ domain for Lhr proteins ([9] and Figure 5A). In addition, PHYRE2 *ab initio* modelling and PSIPRED searches of archaeal Lhr both predicted additional alpha helical content that was not resolved in the mycobacterial structure, including a 30-residue alpha helical extension intriguingly positioned relative to RecA-like domains and the translocating DNA strand (Figure 5A). We speculate that this may be significant for additional Lhr–DNA interactions, including with forked DNA, although it has not been possible to model a forked DNA structure onto these structures. Lhr is widespread across archaeal phyla (Supplementary Table S2) and can be easily identified in 30 bacterial phyla (Supplementary Excel File), although bacterial Lhr is distinguished from archaeal Lhr by the addition of a C-terminal 500–600 amino acids of unknown function that lacks obvious sequence homology to other proteins (Figure 5B). Structural homology searches and modelling using bacterial Lhr C-terminal residues against the PHYRE2 and DALI servers identified a region strongly matching protein folds in the DNA glycosylase enzyme AlkZ that contributes to replication-coupled DNA repair [35], and a smaller region matching tandem winged helix domains of the elongation factor SelB [36] (1.3 Å and 6.9 Å RMSD, respectively). We also noted interesting structural similarities between Lhr proteins and the human putative helicase DDX52, data that is presented in supplementary results (Supplementary Figure S7).

We conclude that our analyses indicate that archaeal Lhr proteins most likely target DNA arising at compromised replication forks, which may include RNA–DNA hybrids present as lagging strand Okazaki fragments. We propose that remodeling of fork DNA after binding by Lhr generates ssDNA for ATP-dependent DNA translocation to unwind the fork as part of DNA repair.

### Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

### Funding

This work was supported by the Wellcome Trust (RCDF grant 023219 to ELB) and the Biotechnology and Biological Sciences Research Council (grant BB/M020541/1 to ELB).

### Author Contributions

E.L.B. and D.G. designed the project; E.L.B., D.G., K.K. and C.D.O.C. wrote the manuscript; E.L.B., K.K., R.J.B. and C.D.O.C. carried out experiments.

### Abbreviations

AA, acceptor excitation; APBS, all-photon burst search; DCBS, dual-channel burst search; HJ, Holliday junction; Lhr, large helicase-related; MPE, manganese-dependent phosphodiesterase; WHD, winged helix domain.

### References

- 1 Ordonez, H. and Shuman, S. (2013) *Mycobacterium smegmatis* Lhr is a DNA-dependent ATPase and a 3′-to-5′ DNA translocase and helicase that prefers to unwind 3′-tailed RNA:DNA hybrids. *J. Biol. Chem.* **288**, 14125–14134 <https://doi.org/10.1074/jbc.M113.466854>

- 2 Reuven, N.B., Koonin, E.V., Rudd, K.E. and Deutscher, M.P. (1995) The gene for the longest known *Escherichia coli* protein is a member of helicase superfamily II. *J. Bacteriol.* **177**, 5393–5400 <https://doi.org/10.1128/JB.177.19.5393-5400.1995>
- 3 Ambur, O.H., Davidsen, T., Frye, S.A., Balasingham, S.V., Lagesen, K., Rognes, T. et al. (2009) Genome dynamics in major bacterial pathogens. *FEMS Microbiol. Rev.* **33**, 453–470 <https://doi.org/10.1111/j.1574-6976.2009.00173.x>
- 4 Chamieh, H., Ibrahim, H. and Kozah, J. (2016) Genome-wide identification of SF1 and SF2 helicases from archaea. *Gene* **576**, 214–228 <https://doi.org/10.1016/j.gene.2015.10.007>
- 5 Schutz, P., Karlberg, T., van den Berg, S., Collins, R., Lehtio, L., Högbo, M. et al. (2010) Comparative structural analysis of human DEAD-box RNA helicases. *PLoS ONE* **5**, e12791 <https://doi.org/10.1371/journal.pone.0012791>
- 6 Xia, J., Chiu, L.Y., Nehring, R.B., Bravo Nunez, M.A., Mei, Q., Perez, M. et al. (2019) Bacteria-to-human protein networks reveal origins of endogenous DNA damage. *Cell* **176**, 127–143 e124 <https://doi.org/10.1016/j.cell.2018.12.008>
- 7 Abdelhaleem, M., Maltais, L. and Wain, H. (2003) The human DDX and DHX gene families of putative RNA helicases. *Genomics* **81**, 618–622 [https://doi.org/10.1016/S0888-7543\(03\)00049-1](https://doi.org/10.1016/S0888-7543(03)00049-1)
- 8 Ejaz, A., Goldgur, Y. and Shuman, S. (2019) Activity and structure of *Pseudomonas putida* MPE, a manganese-dependent single-strand DNA endonuclease encoded in a nucleic acid repair gene cluster. *J. Biol. Chem.* **294**, 7931–7941 <https://doi.org/10.1074/jbc.RA119.008049>
- 9 Ejaz, A., Ordonez, H., Jacewicz, A., Ferrao, R. and Shuman, S. (2018) Structure of mycobacterial 3'-to-5' RNA:DNA helicase Lhr bound to a ssDNA tracking strand highlights distinctive features of a novel family of bacterial helicases. *Nucleic Acids Res.* **46**, 442–455 <https://doi.org/10.1093/nar/gkx1163>
- 10 Ejaz, A. and Shuman, S. (2018) Characterization of Lhr-Core DNA helicase and manganese-dependent DNA nuclease components of a bacterial gene cluster encoding nucleic acid repair enzymes. *J. Biol. Chem.* **293**, 17491–17504 <https://doi.org/10.1074/jbc.RA118.005296>
- 11 Buttner, K., Nehring, S. and Hopfner, K.P. (2007) Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* **14**, 647–652 <https://doi.org/10.1038/nsmb1246>
- 12 Northall, S.J., Buckley, R., Jones, N., Penedo, J.C., Soultanas, P. and Bolt, E.L. (2017) DNA binding and unwinding by Hel308 helicase requires dual functions of a winged helix domain. *DNA Repair (Amst)* **57**, 125–132 <https://doi.org/10.1016/j.dnarep.2017.07.005>
- 13 Johnson, S.J. and Jackson, R.N. (2013) Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol.* **10**, 33–43 <https://doi.org/10.4161/rna.22101>
- 14 Viswanathan, M., Lanjuin, A. and Lovett, S.T. (1999) Identification of RNase T as a high-copy suppressor of the UV sensitivity associated with single-strand DNA exonuclease deficiency in *Escherichia coli*. *Genetics* **151**, 929–934 PMID: 10049912
- 15 Hsiao, Y.Y., Fang, W.H., Lee, C.C., Chen, Y.P. and Yuan, H.S. (2014) Structural insights into DNA repair by RNase T—an exonuclease processing 3' end of structured DNA in repair pathways. *PLoS Biol.* **12**, e1001803 <https://doi.org/10.1371/journal.pbio.1001803>
- 16 van Wolferen, M., Ma, X. and Albers, S.V. (2015) DNA processing proteins involved in the UV-induced stress response of sulfobacterales. *J. Bacteriol.* **197**, 2941–2951 <https://doi.org/10.1128/JB.00344-15>
- 17 Cooper, D.L., Boyle, D.C. and Lovett, S.T. (2015) Genetic analysis of *Escherichia coli* RadA: functional motifs and genetic interactions. *Mol. Microbiol.* **95**, 769–779 <https://doi.org/10.1111/mmi.12899>
- 18 Rand, L., Hinds, J., Springer, B., Sander, P., Buxton, R.S. and Davis, E.O. (2003) The majority of inducible DNA repair genes in *Mycobacterium tuberculosis* are induced independently of RecA. *Mol. Microbiol.* **50**, 1031–1042 <https://doi.org/10.1046/j.1365-2958.2003.03765.x>
- 19 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 <https://doi.org/10.1093/nar/gki685>
- 20 Hishida, T., Han, Y.W., Shibata, T., Kubota, Y., Ishino, Y., Iwasaki, H. et al. (2004) Role of the *Escherichia coli* RecQ DNA helicase in SOS signaling and genome stabilization at stalled replication forks. *Genes Dev.* **18**, 1886–1897 <https://doi.org/10.1101/gad.1223804>
- 21 Howard, J.A., Delmas, S., Ivancic-Bace, I. and Bolt, E.L. (2011) Helicase dissociation and annealing of RNA-DNA hybrids by *Escherichia coli* Cas3 protein. *Biochem. J.* **439**, 85–95 <https://doi.org/10.1042/BJ20110901>
- 22 Komori, K., Hidaka, M., Horiuchi, T., Fujikane, R., Shinagawa, H. and Ishino, Y. (2004) Cooperation of the N-terminal helicase and C-terminal endonuclease activities of archaeal Hef protein in processing stalled replication fork. *J. Biol. Chem.* **279**, 53175–53185 <https://doi.org/10.1074/jbc.M409243200>
- 23 Schrimpf, W., Barth, A., Hendrix, J. and Lamb, D.C. (2018) PAM: a framework for integrated analysis of imaging, single-molecule, and ensemble fluorescence data. *Biophys. J.* **114**, 1518–1528 <https://doi.org/10.1016/j.bpj.2018.02.035>
- 24 Hellenkamp, B., Schmid, S., Doroshenko, O., Opanasyuk, O., Kuhnemuth, R., Rezaei Adariani, S. et al. (2018) Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nat. Methods* **15**, 669–676 <https://doi.org/10.1038/s41592-018-0085-0>
- 25 Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- 26 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H. et al. (2000) The protein data bank. *Nucleic Acids Res.* **28**, 235–242 <https://doi.org/10.1093/nar/28.1.235>
- 27 Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 <https://doi.org/10.1038/nprot.2015.053>
- 28 Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 <https://doi.org/10.1093/nar/gkq366>
- 29 Buchan, D.W.A. and Jones, D.T. (2019) The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 <https://doi.org/10.1093/nar/gkz297>
- 30 Allers, T., Ngo, H.P., Mevarech, M. and Lloyd, R.G. (2004) Development of additional selectable markers for the halophilic archaeon *Haloflex volcanii* based on the leuB and trpA genes. *Appl. Environ. Microbiol.* **70**, 943–953 <https://doi.org/10.1128/AEM.70.2.943-953.2004>
- 31 Fujikane, R., Shinagawa, H. and Ishino, Y. (2006) The archaeal Hjm helicase has recQ-like functions, and may be involved in repair of stalled replication fork. *Genes Cells* **11**, 99–110 <https://doi.org/10.1111/j.1365-2443.2006.00925.x>
- 32 Dickman, M.J., Ingleston, S.M., Sedelnikova, S.E., Rafferty, J.B., Lloyd, R.G., Grasby, J.A. et al. (2002) The RuvABC resolvosome. *Eur. J. Biochem.* **269**, 5492–5501 <https://doi.org/10.1046/j.1432-1033.2002.03250.x>
- 33 West, S.C. (1997) Processing of recombination intermediates by the RuvABC proteins. *Annu. Rev. Genet.* **31**, 213–244 <https://doi.org/10.1146/annurev.genet.31.1.213>

- 34 Richards, J.D., Johnson, K.A., Liu, H., McRobbie, A.M., McMahon, S., Oke, M. et al. (2008) Structure of the DNA repair helicase hel308 reveals DNA binding and autoinhibitory domains. *J. Biol. Chem.* **283**, 5118–5126 <https://doi.org/10.1074/jbc.M707548200>
- 35 Mullins, E.A., Warren, G.M., Bradley, N.P. and Eichman, B.F. (2017) Structure of a DNA glycosylase that unhooks interstrand cross-links. *Proc. Natl Acad. Sci. U.S.A.* **114**, 4400–4405 <https://doi.org/10.1073/pnas.1703066114>
- 36 Selmer, M. and Su, X.D. (2002) Crystal structure of an mRNA-binding fragment of *Moorella thermoacetica* elongation factor SelB. *EMBO J.* **21**, 4145–4153 <https://doi.org/10.1093/emboj/cdf408>