442–455 *Nucleic Acids Research, 2018, Vol. 46, No. 1* *Published online 20 November 2017*
doi: 10.1093/nar/gkx1163

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

Anam Ejaz†, Heather Ordonez†, Agata Jacewicz, Ryan Ferrao and Stewart Shuman*

Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

Received October 16, 2017; Revised November 02, 2017; Editorial Decision November 03, 2017; Accepted November 06, 2017

ABSTRACT

Mycobacterial Lhr is a DNA damage-inducible superfamily 2 helicase that uses adenosine triphosphate (ATP) hydrolysis to drive unidirectional 3′-to-5′ translocation along single-stranded DNA (ssDNA) and to unwind RNA:DNA duplexes *en route*. ATPase, translocase and helicase activities are encompassed within the N-terminal 856-amino acid segment. The crystal structure of Lhr-(1–856) in complex with AMPPNP•Mg^{2+} and ssDNA defines a new helicase family. The enzyme comprises two N-terminal RecA-like modules, a winged helix (WH) domain and a unique C-terminal domain. The 3′ ssDNA end binds in a crescent-shaped groove at the interface between the first RecA domain and the WH domain and tracks 5′ into a groove between the second RecA and C domains. A kissing interaction between the second RecA and C domains forms an aperture that demarcates a putative junction between the loading strand tail and the duplex, with the first duplex nucleoside bookended by stacking on Trp597. Intercalation of Ile528 between nucleosides of the loading strand creates another bookend. Coupling of ATP hydrolysis to RNA:DNA unwinding is dependent on Trp597 and Ile528, and on Thr145 and Arg279 that contact phosphates of the loading strand. The structural and functional data suggest a ratchet mechanism of translocation and unwinding coupled to ATP-driven domain movements.

INTRODUCTION

Helicases are nucleic acid-dependent NTPases that play important roles in catalyzing or regulating DNA and RNA transactions (1). Helicases use the energy of nucleoside triphosphate (NTP) hydrolysis to effect mechanical changes in the secondary structure of nucleic acids or to remodel the structures of protein–nucleic acid complexes. Differences in the inventory of DNA helicases between taxa offer clues to the evolution and diversification of replication/repair strategies and, where the inventories diverge in animal hosts versus infectious pathogens, they can suggest anti-infective drug targets. With this in mind, we are focused on the roster of helicases in *Mycobacteria*, a genus of the phylum *Actinobacteria* that includes the human pathogen *Mycobacterium tuberculosis* and its avirulent relative *Mycobacterium smegmatis*. To date, our group has characterized the biochemical activities and/or *in vivo* functions of seven mycobacterial DNA helicases, including AdnAB (in DNA break repair via homologous recombination) (2–4), RecBCD (in DNA break repair via single strand annealing) (5), UvrD1 (in clastogen resistance) (6,7), UvrD2 (essential for viability) (8), SftH (9), RqlH (10) and Lhr (11). Other mycobacterial DNA helicases that have been characterized include: UvrB (12), DinG (13), RecG (14,15), DnaB (16) and XPB (17).

Mycobacterium smegmatis Lhr is the founding member of a proposed new clade of superfamily 2 (SF2) helicases, by virtue of its distinctive biochemical specificities and its signature domain organization (11). Lhr is a 1507-amino acid (aa) monomeric nucleic acid-dependent adenosine triphosphatase (ATPase)/dATPase that uses ATP hydrolysis to drive unidirectional 3′-to-5′ translocation along single-stranded DNA (ssDNA) and to unwind duplexes *en route*. Lhr is better at unwinding an RNA:DNA hybrid than it is at displacing a DNA:DNA duplex. Lhr is upregulated in mycobacteria response to DNA damage by ultraviolet irradiation or mitomycin C (18,19), but its precise role in repair is unknown. Lhr homologs are found in bacteria from eight phyla, being especially prevalent in *Actinobacteria* and *Proteobacteria*. [A list of bacterial genera that encode a full-length homolog of Lhr, spanning 54 genera of *Actinobacteria* (including *M. tuberculosis*), 26 genera of *Proteobacteria*

*To whom correspondence should be addressed. Tel: +1 212 639 7145; Email: s-shuman@ski.mskcc.org
†These authors contributed equally to the paper as first authors.

(including *Escherichia coli*, the organism from which an Lhr gene was first sequenced (20)) and various genera of *Firmicutes*, *Acidobacteria*, *Planctomycetes*, *Verrucomicrobia*, *Nitrospirae* and *Spirochaetes* is published in ref. (11), Supplementary Table S1.]

The ATPase, DNA translocase and RNA:DNA helicase activities of mycobacterial Lhr are encompassed within the N-terminal 856-aa segment of the 1507-aa polypeptide (11). This autonomous helicase consists of a classic SF2 (DExH-box) motor domain embraced within the N-terminal half of the protein. However, the Lhr-(1–562) fragment containing the conserved motor domain does not suffice for ATP hydrolysis, signifying that the module from aa 563–856 contributes structural elements necessary for the DNA-dependent ATPase. This module, of unknown structure or function, is fused in *cis* to an upstream helicase-like NTPase domain in many biochemically uncharacterized proteins in the NCBI database. Indeed, the abbreviated Lhr-(1–856) helicase enzyme *per se* exemplifies a large family of homologous polypeptides, all of similar size (∼800–900 aa), present in the proteomes of scores of diverse bacterial and archaeal taxa. The C-terminal module of Lhr-(1–856) has no instructive motifs or similarity to any proteins outside the clade of Lhr homologs.

The substrate specificity and domain organization of Lhr-(1–856) are sufficiently distinctive to warrant deeper investigations of Lhr structure and mechanism. To this end, we determined the crystal structure of Lhr-(1–856) in complex with AMPPNP•Mg^{2+} and a 'loading/tracking' 3′ DNA single strand. The protein consists of two N-terminal RecA-like domains that bind the AMPPNP•Mg^{2+}, a central winged helix (WH) domain and a C-terminal domain of novel structure. All four domains make contributions to Lhr's unique DNA interface, which we interrogate functionally via structure-guided mutagenesis.

MATERIALS AND METHODS

Native Lhr-(1–856) purification

Plasmid pET16b-Lhr-(1–856) (11) was transformed into *E. coli* BL21(DE3) cells. Cultures (4 × 1-liter) amplified from a single ampicillin-resistant transformant were grown at 37°C in TB broth containing 100 μg/ml ampicillin until the A_{600} reached 0.8. The cultures were chilled on ice for 1 h, then adjusted to 2% (v/v) ethanol and 0.3 mM isopropyl-β-D-thiogalactopyranoside and incubated for 16 h at 17°C with constant shaking. All subsequent procedures were performed at 4°C. Cells were harvested by centrifugation and resuspended in 100 ml of buffer A (50 mM Tris–HCl, pH 8.0, 500 mM NaCl, 1 mM dithiothreitol (DTT), 10% sucrose and 20 mM imidazole) containing two protease inhibitor cocktail tablets (Roche). Lysozyme was added to a concentration of 1 mg/ml and the suspension was incubated for 1 h. The lysate was sonicated to reduce viscosity and the insoluble material was pelleted by centrifugation at 38 000 *g* for 30 min. The supernatant was mixed for 1 h with 4 ml of Ni-NTA agarose resin (Qiagen) that had been equilibrated with buffer A. The resin was recovered by centrifugation and resuspended in 30 ml of buffer A. This wash step was repeated twice. The resin was again recovered by centrifugation, resuspended in 25 ml of buffer B (50 mM Tris–HCl, pH 8.0, 500 mM NaCl, 10% glycerol) containing 20 mM imidazole and poured into a column. The resin was washed with 25 ml of 3 M KCl, followed by buffer B containing 20 mM imidazole and buffer B containing 50 mM imidazole until no protein was eluted as detected by Bradford reagent. The bound material was eluted with buffer C (50 mM HEPES–NaOH, pH 7.5, 300 mM KCl and 10% glycerol) containing 500 mM imidazole. The sample was diluted to reduce KCl to 150 mM immediately before loading onto a HiTrapQ ion exchange column, which was developed with a linear gradient of 0 to 1 M KCl in buffer D (50 mM HEPES–NaOH, pH 7.5, 10% glycerol and 4 mM DTT). Lhr-(1–856) eluted at ∼400 mM KCl. Peak fractions were pooled and subjected to gel filtration through a Superdex-200 column equilibrated in buffer E (10 mM HEPES–NaOH, pH 7.5, 300 mM KCl, 10% glycerol and 1 mM Tris(2-carboxyethyl)phosphine (TCEP)). Peak fractions were pooled, concentrated by centrifugal ultrafiltration and frozen. Final yield was ∼20 mg of protein.

SeMet Lhr-(1–856)

pET16b-Lhr-(1–856) plasmid was transformed into *E. coli* B834(DE3) cells. A single transformant was inoculated into 10 ml of LB medium containing 100 μg/ml ampicillin and incubated for 8 h at 37°C. The bacteria were harvested by centrifugation and then resuspended in 200 ml of complete LeMaster medium containing 100 μg/ml ampicillin and 50 μg/ml SeMet (selenomethionine). After overnight incubation, the culture was diluted to an A_{600} of 0.1 in 4 l of fresh LeMaster medium (+ampicillin +SeMet) and growth was continued at 37°C with constant shaking until the A_{600} reached 0.6. The cultures were chilled on ice for 1 h, then adjusted to 2% (v/v) ethanol and 0.3 mM isopropyl-β-D-thiogalactopyranoside and incubated for 16 h at 17°C with constant shaking. SeMet-substituted Lhr-(1–856) was purified from a soluble bacterial extract as described above for the native protein.

Crystallization

Crystals of native and SeMet-substituted Lhr-(1–856) in complex with a 16-mer DNA oligonucleotide 5′-GACCTGCTGCGGATTG were grown by the hanging drop vapor diffusion method at 22°C. A mixture of Lhr-(1–856) (91 μM native or 97 μM SeMet), 109 or 116 μM 16-mer DNA (1:1.2 protein to DNA ratio), 5 mM $MgCl_2$ and 2 mM AMPPNP was incubated for 20 min at 22°C. Aliquots (1 μl) were then mixed with 1 μl of reservoir buffer containing 21% or 22% (w/v) PEG3350, 0.2 M NaF, 0.1 M Bis-Tris Propane (pH 7.0–7.2). Crystals grew to their full size within 3 to 5 days. Crystals were harvested, cryoprotected by transfer to a solution containing 20% (w/v) PEG3350, 25% glycerol, 0.3 M KCl, 0.1 M Bis-Tris Propane (pH 7.0) and then flash-frozen in liquid nitrogen.

X-ray diffraction and structure determination

Diffraction data for a SeMet-Lhr-(1–856) crystal were collected at the Se anomalous peak wavelength at the Argonne National Laboratory beamline ID-24-E equipped

with ADSC Q315 3 × 3 CCD detector. Due to the radiation sensitivity of the crystal, we collected two datasets on two different sections of the same crystal. Data were integrated, merged and scaled using HKL2000. Positions of heavy atoms were calculated using HKL2map (21) and used to obtain initial maps and a model in Phenix.Autosol (22) at 3.5 Å resolution cut-off. The crystal belonged to space group $P2_1$ with unit cell dimensions consistent with two protomers per asymmetric unit, assuming a solvent content of ∼59%. The analysis of the Patterson function in Phenix.Xtriage revealed a significant off-origin peak that was ∼57% of the origin peak, indicating pseudo translational symmetry of the crystal. The initial model was build using Phenix.Autobuild and iteratively refined in Phenix and rebuilt in Coot (23) to 2.8 Å resolution. The maps showed clear density for most of the two Lhr-(1–856) protomers, two DNA oligonucleotides and two AMPPNP nucleotides. The model was refined to R_{work}/R_{free} values of 0.295/0.355. Further refinement was not pursued as a better quality dataset for a native crystal became available.

Diffraction data for native Lhr-(1–856) were collected at the Argonne National Laboratory beamline ID-24-C equipped with Pilatus 6M-F detector. The native Lhr-(1–856) crystal belonged to space group C2 and diffracted to 2.8 Å resolution. Unit cell dimensions were consistent with one protomer per asymmetric unit, assuming a solvent content of ∼53%. Phases were determined by molecular replacement in Phenix.Phaser using the protomer A of the SeMet-Lhr-(1–856) structure as a search model. The structure was iteratively refined in Phenix and adjusted manually in Coot. The final protein model comprised a 772-aa polypeptide punctuated by six gaps: from residues 70 to 74, 301 to 326, 490 to 495, 544 to 549, 560 to 587 and 725 to 728. There was no visible electron density for the N-terminal 5 aa of Lhr-(1–856) or for the C-terminal 4 aa. The structure model included AMPPNP•Mg^{2+} and a continuous 13-nt segment of the input 16-mer DNA oligonucleotide. The final 2.8 Å model was refined to R_{work}/R_{free} values of 19.5/24.2. Diffraction data and refinement statistics are compiled in Supplementary Table S1.

Lhr-(1–856) mutants

Alanine mutations were introduced into the Lhr-(1–856) open reading frame (ORF) by two-stage overlap extension polymerase chain reaction with mutagenic primers and the mutated genes were inserted into plasmid pET16b. The plasmid inserts were sequenced to exclude the acquisition of unwanted coding changes during amplification or cloning. Recombinant wild-type (WT) Lhr-(1–856) and Lhr-(1–856)-Ala proteins were produced in *E. coli* BL21(DE3) and purified from soluble extracts by nickel-agarose and Diethyaminoethyl (DEAE)-Sephacel chromatography (11). The DEAE protein preparations—in buffer containing 50 mM Tris–HCl, pH 8.0, 500 mM NaCl, 1 mM DTT, 1 mM ethylenediaminetetraaceticacid (EDTA), 0.05% Triton X-100 and 10% glycerol—were aliquoted and stored at −80°C and thawed immediately prior to use in enzyme assays. Total protein concentrations of the DEAE preparations were measured by using the Bio-Rad dye reagent with bovine serum albumin (BSA) as the standard. The concentrations of the Lhr-(1–856) polypeptides were determined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of 5 μg aliquots of each protein preparation in parallel with 1.25, 2.5, 5 and 10 μg of BSA standards. The gel was stained with Coomassie blue, and the extent of dye binding to Lhr-(1–856) and BSA polypeptides was quantified using Chemidoc XRS imager and ImageJ. The concentrations of the Lhr-(1–856) proteins were calculated by interpolation to the BSA standard curve.

ATPase assay

Reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 7.0, 50 mM NaCl, 1 mM DTT, 1 mM $CaCl_2$, 1 mM [α^{32}P]ATP, DNA cofactor (heat-denatured salmon sperm DNA or 59-mer DNA oligonucleotide as specified) and Lhr-(1–856) proteins as specified, were incubated for 30 min at 37°C. Reactions were quenched by adding 2 μl of 5 M formic acid. An aliquot (2 μl) of the mixture was spotted on a polyethyleneimine (PEI)-cellulose thin layer chromatography (TLC) plate. Ascending TLC was performed with 0.45 M ammonium sulfate as the mobile phase. The extent of conversion of [α^{32}P]ATP to [α^{32}P]ADP was quantified by scanning the TLC plate with a Fujix BAS2500 imager.

Helicase assay

A 5′ ^{32}P-labeled RNA strand was prepared by reaction of a synthetic 24-mer RNA oligonucleotide with T4 polynucleotide kinase (Pnk) and [γ^{32}P]ATP. The kinase reaction mixture was heated to 95°C to inactivate T4 Pnk. The labeled RNA was then annealed to a 3-fold excess of complementary 39-mer DNA strand to form a 3′ tailed duplex RNA:DNA helicase substrate. The duplex was purified by electrophoresis through a native 12% polyacrylamide gel and eluted from an excised gel slice by overnight incubation at 4°C in 400 μl of 10 mM Tris–HCl, pH 6.8, 1 mM EDTA and 50 mM NaCl. Helicase reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 7.0, 50 mM NaCl, 1 mM DTT, 5 mM $CaCl_2$, 50 nM (0.5 pmol) ^{32}P-labeled RNA:DNA duplex and Lhr-(1–856) proteins as specified were pre-incubated for 10 min at room temperature. The unwinding reaction was initiated by the addition of 1 mM ATP and 10-fold excess of an unlabeled RNA oligonucleotide identical to the labeled strand of the helicase substrate. Addition of an excess of unlabeled strand was necessary to prevent the spontaneous reannealing of the unwound ^{32}P-labeled RNA strand. The reaction mixtures were incubated for 30 min at 37°C and then quenched by adding 1.5 μl of a solution containing 3.5% SDS, 80 mM EDTA, 10 μg proteinase K. After protease digestion for 5 min at 37°C, the mixtures were supplemented with 4 μl of 50% glycerol, 0.3% bromophenol blue. The reaction products were analyzed by electrophoresis through a 15 cm 10% polyacrylamide gel in 89 mM Tris-borate, 2.5 mM EDTA. The products were visualized by autoradiography and quantified by scanning the gel with a Fujix BAS2500 imager.

Streptavidin displacement assay of Lhr-(1–856) translocation on DNA

A synthetic 34-mer oligodeoxynucleotide containing a Biotin-ON inter-nucleotide spacer at the fourth position from the 5′ end was 5′ ^{32}P-labeled by reaction with T4 Pnk and [γ^{32}P]ATP. The labeled DNA was purified by electrophoresis through a 18% native polyacrylamide gel and eluted from an excised gel slice by overnight incubation at 4°C in 400 μl of 10 mM Tris–HCl, pH 7.5, 1 mM EDTA and 50 mM NaCl. Translocation reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 7.0, 50 mM NaCl, 1 mM DTT, 1 mM $CaCl_2$, 1 mM ATP, 100 nM (1 pmol) biotinylated ^{32}P-labeled DNA and 4 μM streptavidin (Sigma) were pre-incubated at room temperature for 10 min to form streptavidin–DNA (SA–DNA) complexes. The mixtures were supplemented with 40 μM free Biotin (Fisher) and the translocation reactions were initiated by adding Lhr-(1–856). After incubation for 15 min at 37°C, the reactions were quenched by addition of 3 μl of a solution containing 200 mM EDTA, 0.6% SDS, 25% glycerol and 20 μM unlabeled single-stranded 59-mer DNA. The reaction products were analyzed by electrophoresis through a 15-cm native 18% polyacrylamide gel containing 89 mM Tris-borate and 2.5 mM EDTA. The free ^{32}P-labeled 34-mer DNA and the SA–DNA complexes were visualized and quantified by scanning the gel with a Fujix BAS-2500 imaging apparatus.

RESULTS

Structure of Lhr-(1–856)

We grew crystals of native and SeMet-substituted Lhr-(1–856) that had been pre-incubated with Mg^{2+}, AMPPNP (a non-hydrolysable ATP analog) and a 16-mer ssDNA oligonucleotide. The SeMet-Lhr-(1–856) crystals diffracted X-rays to 2.8 Å resolution and were in space group $P2_1$. Se-SAD phases at 3.5 Å resolution were used to derive an initial model that was used in turn for molecular replacement to derive a model for native Lhr-(1–856), crystals of which diffracted to 2.8 Å, were in space group C2 and contained one Lhr-(1–856) polypeptide per asymmetric unit. The electron density map was of sufficient quality to allow placement of AMPPNP•Mg^{2+} and a 13-nt segment of the ssDNA strand.

A stereo view of the refined 2.8 Å structure model (R_{work}/R_{free} = 0.195/0.242) (Supplementary Table S1) is shown in Figure 1B and is colored according to the linear domain organization of the polypeptide, depicted in cartoon form in Figure 1A. The secondary structure elements are aligned to the aa sequence in Figure 1C. As expected, the N-terminal segment comprises two RecA-like domains (domain 1 from aa 8–230; domain 2 from aa 231–435), characteristic of SF2 helicases, that together form the ATP-binding site at their interface (Figure 1B). The tandem RecA domains are punctuated by disordered segments in surface loops, from aa 70 to 74 in domain 1 and aa 301 to 326 in domain 2 (Figure 1C). The central segment from aa 436 to 529 comprises a WH domain, which could not have been surmised from a BLAST search with the Lhr primary structure. However a DALI search (24) with the Lhr crystal structure recovered the archaeal SF2 helicase *Archaeoglobus fulgidus* Hel308 as a top hit (pdb 2P6R; Z score 24.4; rmsd of 4.4 Å at 440 Cα positions with 23% aa identity) (Supplementary Table S2). The homology between Lhr and Hel308 embraces the RecA domains. Hel308 also has a WH domain flanking its tandem RecA modules; yet, whereas the Lhr and Hel308 WH domains have a similar fold, they share only six positions of aa identity. The Lhr WH domain is punctuated by a disordered surface loop from aa 490–495 (Figure 1C).

The Lhr C-terminal segment from aa 530 to 856 is a novel fold, composed of 11 α-helices, two 3_{10} helices, a π helix and a five-stranded β-sheet (Figure 1B and C), that has no significant similarity to any known protein, as determined by a DALI search of the PDB with the C-domain, confirming our prediction that mycobacterial Lhr inaugurates a new helicase family within superfamily-2. The C-domain includes three disordered segments: from aa 544 to 549, 560 to 587 and 725 to 728 (Figure 1C).

Each of the four domains packs closely against two other flanking domains in the Lhr tertiary structure. To wit: domain 1 interfaces with domain 2 and the WH domain; domain 2 is flanked by domain 1 and the C-domain; the C-domain abuts domain 2 and the WH domain; and the WH domain interfaces with the C-domain and domain 1 (Figure 1B). The ssDNA binds in an extended conformation across the anterior surface of the Lhr protein, as shown in Figure 1B. All four protein domains contribute to an extensive DNA interface, as will be discussed below.

DNA binding

The refined structure includes a 13-mer ssDNA segment, to which we will refer with nucleotides numbered from 3′ to 5′ as: 3′-$T^1A^2G^3G^4C^5G^6T^7C^8G^9T^{10}C^{11}C^{12}A^{13}$. The 3′ end of the ssDNA is at the left periphery of the Lhr structure; the 5′ end of the ssDNA is at the right edge (Figure 1B). The ssDNA binds in a crescent-shaped groove at the interface between the first RecA domain and the WH domain and tracks rightward into a groove between the second RecA domain and the C-domain (Figures 1B and 2C). A salt bridge between the Arg280 and Glu550 side chains (shown in stick models in Figure 1B; see also Figure 7 below) creates a kissing interaction between domain 2 and the C-domain that closes a clamp around the ssDNA (highlighted in the protein surface model in Figure 2C). The aperture of the clamp suffices for threading of ssDNA but is too narrow to permit passage of a duplex nucleic acid. The aperture effectively divides the ssDNA into two regions: a 3′ 9-mer segment $T^1A^2G^3G^4C^5G^6T^7C^8G^9$ on the left side of the aperture; and a 5′ segment $T^{10}C^{11}C^{12}A^{13}$ on the right side of the aperture. We regard the left-side ssDNA in the groove lined by domain 1, the WH domain and the left side of the C-domain (Figure 2C) as synonymous with the loading/tracking DNA single strand on which Lhr translocates 3′-to-5′ before encountering a duplex to unwind (11). This loading strand groove is surrounded by positive electrostatic surface potential (Figure 2D) conducive to nucleic acid interactions.

The electron density map for the loading strand was of sufficient quality to clearly assign the $A^2G^3G^4C^5G^6T^7C^8$ sequence (see Supplementary Figure S1 for an omit map

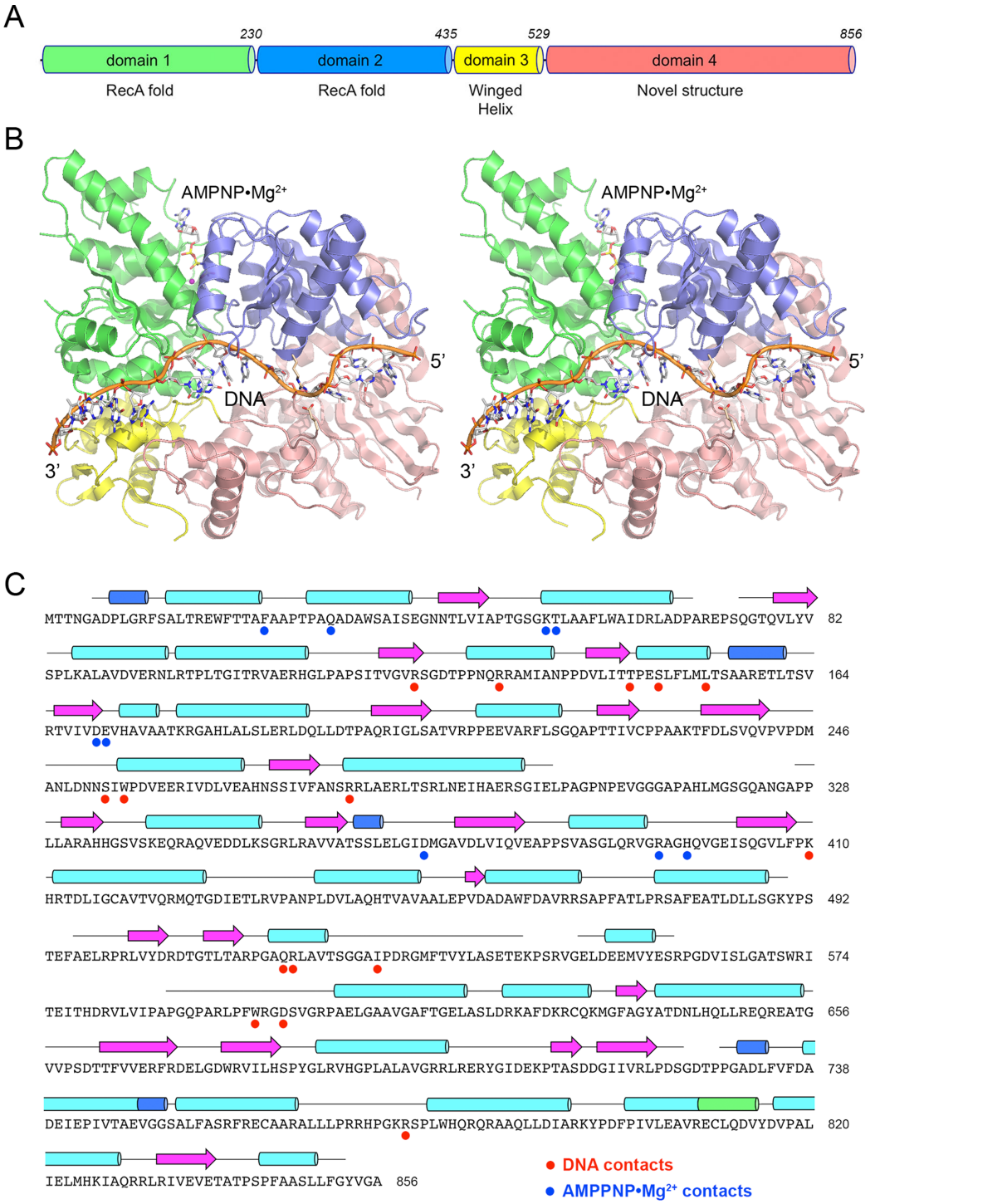

Figure 1. Structure of Lhr-(1–856). (**A**) Lhr-(1–856) consists of four structural domains, depicted as horizontal cylinders. Domains 1 (aa 1–230; colored green) and 2 (aa 231–435; blue) are RecA-like folds. Domain 3 (aa 436–529; yellow) is a WH module. Domain 4 (aa 530–856; colored salmon) is a novel structural module. (**B**) Stereo view of the Lhr-(1–856) tertiary structure, with the protein depicted as a cartoon model, colored by domain as in panel A. The 13-mer DNA is rendered as a stick model with a cartoon trace through the phosphodiester backbone; the 3′ and 5′ ends are indicated. AMPPNP is shown as a stick model. Mg^{2+} is a magenta sphere. Two amino acids forming an Arg-Glu salt bridge between domains 2 and 4 are shown as stick models with beige carbons. (**C**) Primary and secondary structure. The secondary structure elements of the Lhr-(1–856) crystal structure are displayed above the aa sequence, with β-strands as magenta arrows, α-helices as cyan cylinders, 3_{10} helices as blue cylinders and a π helix as a green cylinder. Disordered segments of the tertiary structure are indicated by gaps in the line connecting the secondary structure element. Amino acids that make side chain contacts to DNA or AMPPNP•Mg^{2+} are denoted by red dots and blue dots, respectively.

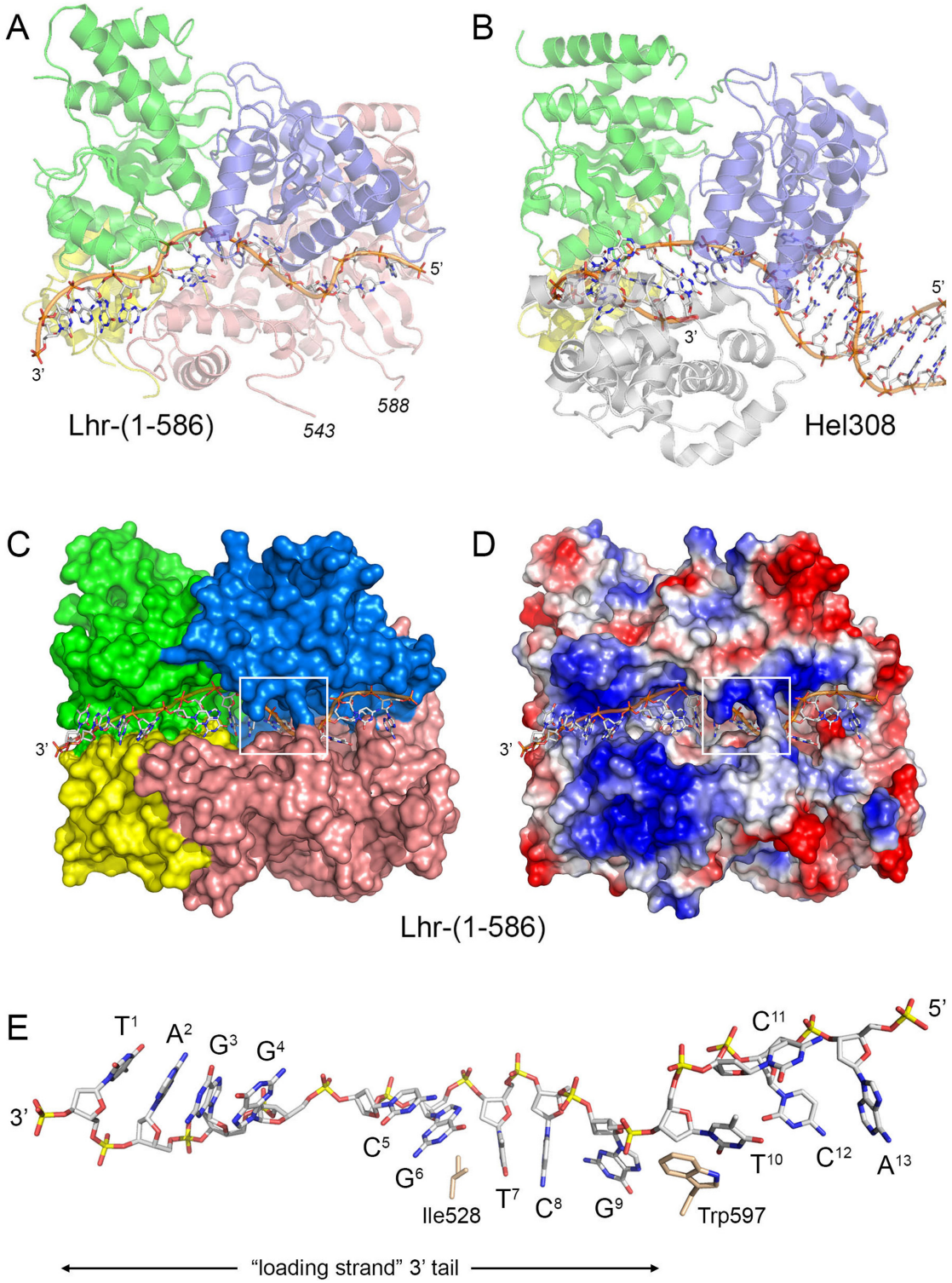

Figure 2. ssDNA binding and comparison to Hel308. (**A** and **B**) The structure of Lhr-(1–856) bound to ssDNA (panel A) is compared to that of *Archaeoglobus* Hel308 in complex with a partially unwound 3′-tailed DNA duplex (panel B; pdb 2P6R). The two structures were aligned with respect to their tandem RecA domains and are offset horizontally. The proteins are depicted as cartoon models, with shared domains 1, 2 and 3 colored as in Figure 1. The C-domain of Lhr-(1–586) (colored salmon) is unrelated to the C-domain of Hel308 (colored gray) with respect to structure and position *vis à vis* the other domains. The DNAs are depicted as stick models with cartoon traces through the phosphodiester backbones. (**C**) Surface model of the Lhr-(1–856) protein, colored by domain, with DNA depicted as a stick/cartoon model. The white box highlights an osculating contact between domains 2 and 3 (via an Arg280-Glu550 salt bridge) that forms an aperture through which the ssDNA threads. (**D**) Surface electrostatic model of Lhr-(1–856), generated in Pymol, in the same orientation as in panel C. (**E**) Conformation of the 13-mer ssDNA bound to Lhr-(1–856), rendered as a stick model with gray carbons and yellow phosphate atoms. The nucleobases are numbered from 3′ to 5′. Two amino acid side chains that intercalate between vicinal bases are shown as stick models with beige carbons.

of the $A^2G^3G^4C^5G^6$ segment). All of the purine nucleosides of the ssDNA were in an *anti* conformation, except for G^4, which was in the *syn* conformation (Supplementary Figure S1 and Figure 2E). The ssDNA of the loading strand is partitioned by the protein into three distinct base-stacked segments, 3′-$T^1A^2G^3G^4$, C^5G^6 and T^7C^8 (Figure 2E), by virtue of insertion of protein elements between the segments, splaying apart the nucleosides at the segment margins. For example, the Ile528 side chain intercalates between G^6 and T^7 (Figure 2E). Also, the G^9 and T^{10} nucleosides on either side of the aperture are splayed apart around a kink in the phosphodiester backbone, with the T^{10} nucleoside held in place by stacking on the Trp597 indole ring (Figure 2E).

It was instructive to compare the nucleic acid bound structure of Lhr-(1–856) to that of *Archaeoglobus* Hel308 in complex with a partially unwound 3′-tailed DNA duplex (25). The two structures were aligned with respect to their tandem RecA domains and are offset horizontally in Figure 2A and B. Whereas, the respective WH modules are positioned similarly, the C-domain of Lhr-(1–856) (Figure 2A, colored salmon) is unrelated structurally and spatially to the C-domain of Hel308 (Figure 2B, colored gray). The trajectory across the RecA domains of the tracking 3′ ssDNA 6-nucleotide segment at the ssDNA:dsDNA junction in Hel308 is roughly analogous to that in Lhr-(1–856), although the protein–DNA contacts are distinct for the two helicases. In Hel308, an Arg252–Asp589 salt bridge establishes a kissing contact between domain 2 and the Hel308 C-domain that forms a clamp around the first single-stranded nucleotide at the ssDNA:dsDNA junction. This Hel308 feature is analogous to the Arg280-Glu550 aperture seen in Lhr-(1–856). Moreover, the first paired tracking strand base on the duplex side of the Hel308 aperture stacks on a phenylalanine side chain of a β-hairpin wedge in domain 2 that drives melting of the first two base-pairs of the ssDNA:dsDNA junction (25). The analogous 'bookend' aromatic stack to the corresponding base in Lhr-(1–856) is made by Trp597 of the C-domain. Indeed, Lhr-(1–856) conspicuously lacks in its domain 2 a counterpart of the Hel308 β-hairpin wedge. These comparisons to Hel308 prompt the surmise that T^{10} in the Lhr-(1–856) structure corresponds to the first paired tracking strand nucleotide of the duplex at a ss/ds junction. That the C^{11}, C^{12} and A^{13} nucleobases are not stacked on T^{10} could simply reflect the absence of a complementary duplex-forming strand in our structure. In the same vein, it is noteworthy that the disordered gaps in the Lhr-(1–856) C-domain, with one gap starting at aa 543 and another gap ending at aa 588 (Figure 2A), are situated right below the proposed duplex-forming portion of the bound ssDNA. We speculate that the missing parts of the C-domain will interact with the duplex portion of the Lhr helicase substrate.

Lhr-(1–856) interactions with the loading strand

The aa in Lhr-(1–586) that make side chain contacts to the 13-mer ssDNA are denoted by red dots under the primary structure in Figure 1C. A detailed stereo view of the protein interactions with the loading strand 3′ tail segment $T^1A^2G^3G^4C^5G^6T^7C^8G^9$ is shown in Figure 3. Each Lhr domain makes region-specific DNA contacts as follows. The WH domain engages the 3′-most segment, via H-bonds from Gln518 to the A^2pG^3 phosphate and from Arg519 to the G^4 nucleobase O6 and N7 atoms. (The Arg519 contacts likely stabilize the G^4 base in its *syn* nucleoside conformation.) Gly525 receives a H-bond to its main chain carbonyl from the G^6 nucleobase N2 atom. Ile528 intercalates between and makes van der Waals contacts with the G^6 and T^7 bases (Figure 3). Domain 1 aa make serial H-bonds or electrostatic contacts to the DNA phosphates at the G^3pG^4 (via Arg122 and Arg131), G^4pC^5 (Arg122), C^5pG^6 (Thr145-Oγ, Ser148-Oγ and Gly124 amide) and G^6pT^7 (Lys86 amide) steps. In addition, Leu151 makes a van der Waals contact to the G^4 sugar O5′ atom. Domain 2 then makes H-bonds to the phosphates at the T^7pC^8 (via Arg279 and Gly335 amide) and C^8pG^8 (Arg279 amide) steps (Figure 3). C-domain residue Arg777 contacts the C^5 base O2 atom.

AMPPNP binding and ATPase active site

The aa in Lhr-(1–856) that make direct or water-mediated side chain contacts to AMPPNP•Mg^{2+} are denoted by blue dots under the primary structure in Figure 1C. The AMPPNP site, shown in stereo in Figure 4, is formed by the classic SF2 motifs. The Q motif (^{24}FAAPTPAQ^{31}) engages the adenine nucleobase, which adopts the *anti* nucleoside conformation. Phe24 stacks on the purine ring. Gln31 makes hydrogen bonds to adenine N7 and N6 that likely account for Lhr's specificity for hydrolyzing ATP/dATP (11). Motif I (GSGK^{54}T^{55}; the P-loop or Walker A-box) coordinates the AMPPNP β and γ phosphates via Lys54, while Thr55 makes a water-mediated contact to the Mg^{2+} ion that bridges the β and γ phosphates. Asp170 and Glu171 in motif II (^{170}DEVH^{173}, the Walker B-box) coordinate the Mg^{2+} ion via waters. The Mg^{2+} coordination complex contains five modeled constituents (three waters and the two phosphate oxygens) and the geometry is imperfect for a typical octahedral Mg^{2+} complex; this might reflect the fact that Ca^{2+} is the preferred metal cofactor for ATP hydrolysis by Lhr (11) and Ca^{2+} can accommodate a seven-membered coordination sphere with pentagonal bipyramidal geometry. Motif VI (^{390}QRVGRAGH^{397}) constituents Arg394 and His397 coordinate the ATP γ phosphate, while Gln390, located near the γ phosphate, is predicted to coordinate a water nucleophile for NTP hydrolysis. Motif III ^{205}SAT^{207} makes bridging contacts from Ser205 to motif II (His173) and from Thr207 to motif VI (Gln390). In other DEx(H/D) helicases, motif III couples NTP hydrolysis to mechanical work (26–31). In addition, Asp369 coordinates the ATP ribose 3′-OH and motif VI His397.

Mutational analysis of the 3′ loading strand interface

Guided by the crystal structure, we conducted an alanine scan of Lhr-(1–856), targeting 9 aa side chains that contact the 3′ loading strand segment of the ssDNA: Arg122, Arg131, Thr145, Ser148 and Leu151 in domain 1; Arg279 in domain 2; Gln518 and Ile528 in domain 3; Arg777 in domain 4. The Ala-mutants were produced in *E. coli* in parallel with WT Lhr-(1–856) and isolated from soluble extracts by Ni-affinity and DEAE chromatography (Supple-

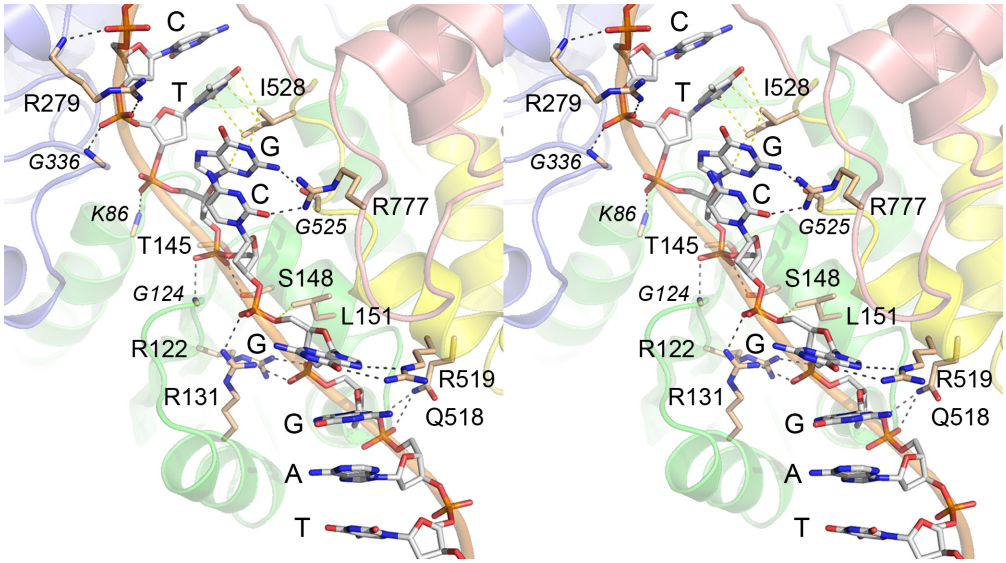

Figure 3. Interactions of Lhr-(1–856) with the DNA loading strand. Stereo view of the 3′ segment of the ssDNA strand (3′-$T^1A^2G^3G^4C^5G^6T^7C^8$), rendered as a stick/cartoon model with gray carbons and labeled by nucleobase. Lhr-(1–856) is a cartoon model, colored by domain, with interfacial aa shown as stick models with beige carbons. Hydrogen bond contacts between protein and DNA are denoted by black dashed lines; van der Waals contacts are indicated by yellow dashed lines.

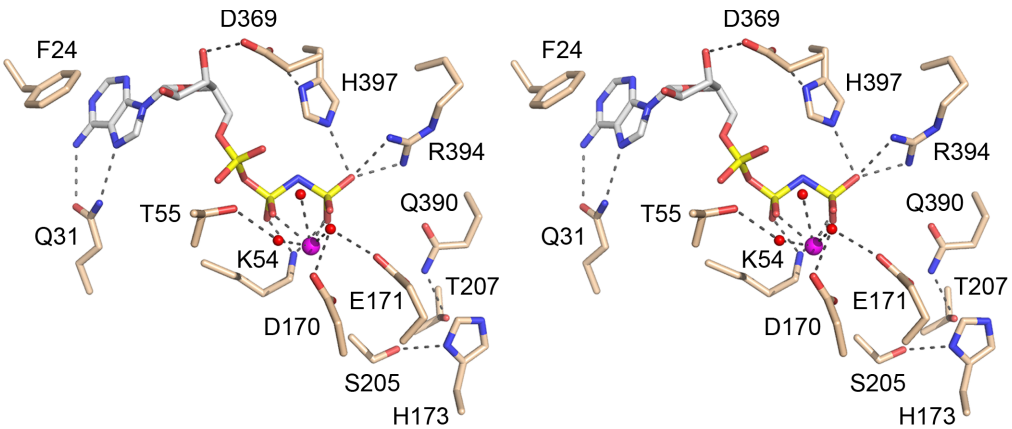

Figure 4. AMPPNP binding and ATPase active site. Stereo view with aa depicted as stick models with beige carbons; AMPPNP as a stick model with gray carbons, Mg^{2+} as a magenta sphere and metal-bound waters as red spheres. Atomic interactions are indicated by black dashed lines.

mentary Figure S2). Their characterization entailed assaying: (i) ssDNA-dependent ATPase activity (Figure 5A); (ii) helicase activity on an RNA:DNA duplex with a 3′ ssDNA tail (Figure 5B); 3′-to-5′ translocation along ssDNA, resulting in displacement of a SA–biotin complex at the 5′ end (Figure 5C); and (iv) the dependence of ATP hydrolysis on ssDNA concentration, as a functional gauge of ssDNA binding (11) (Supplementary Figure S3A).

Titrations of the Lhr-(1–856) proteins for ATP hydrolysis in the presence of denatured DNA revealed that all of the DNA site mutants retained ATPase activity (Figure 5A). Eight of the mutants had specific activity similar to the WT (i.e. within a factor of two). The exception was T145A, which was one-third as active as WT Lhr-(1–856).

RNA:DNA helicase activity was assayed in a single-turnover format (11) as a function of input Lhr-(1–856). The substrate was prepared by annealing a ^{32}P-labeled 24-mer RNA strand to a complementary unlabeled DNA strand with a 15-nt 3′ dT tail as the loading strand (Figure 5B). Mutational effects fell into three categories. Benign mutants L151A, R131A and Q518A had helicase activity similar to WT. Modest defects were apparent for mutants R122A, S148A and R777A, reflected in a lower extent of unwinding at saturating enzyme. By contrast, mutants T145A, R279A and I528A were grossly defective for duplex unwinding (Figure 5B).

Helicase defects can reflect problems with protein translocation along ssDNA, strand separation at the single-strand:duplex junction or progression of the enzyme through the duplex segment. In order to discriminate defects in ssDNA movement versus strand displacement, we employed a SA displacement assay described previously (11), whereby Lhr-(1–856) engages a 5′ ^{32}P-labeled 34-mer ssDNA biotinylated near the 5′ end and complexed with SA (Figure 5C). ATPase-coupled translocation of Lhr-(1–856) toward the 5′ end mechanically disrupts the SA–biotin in-

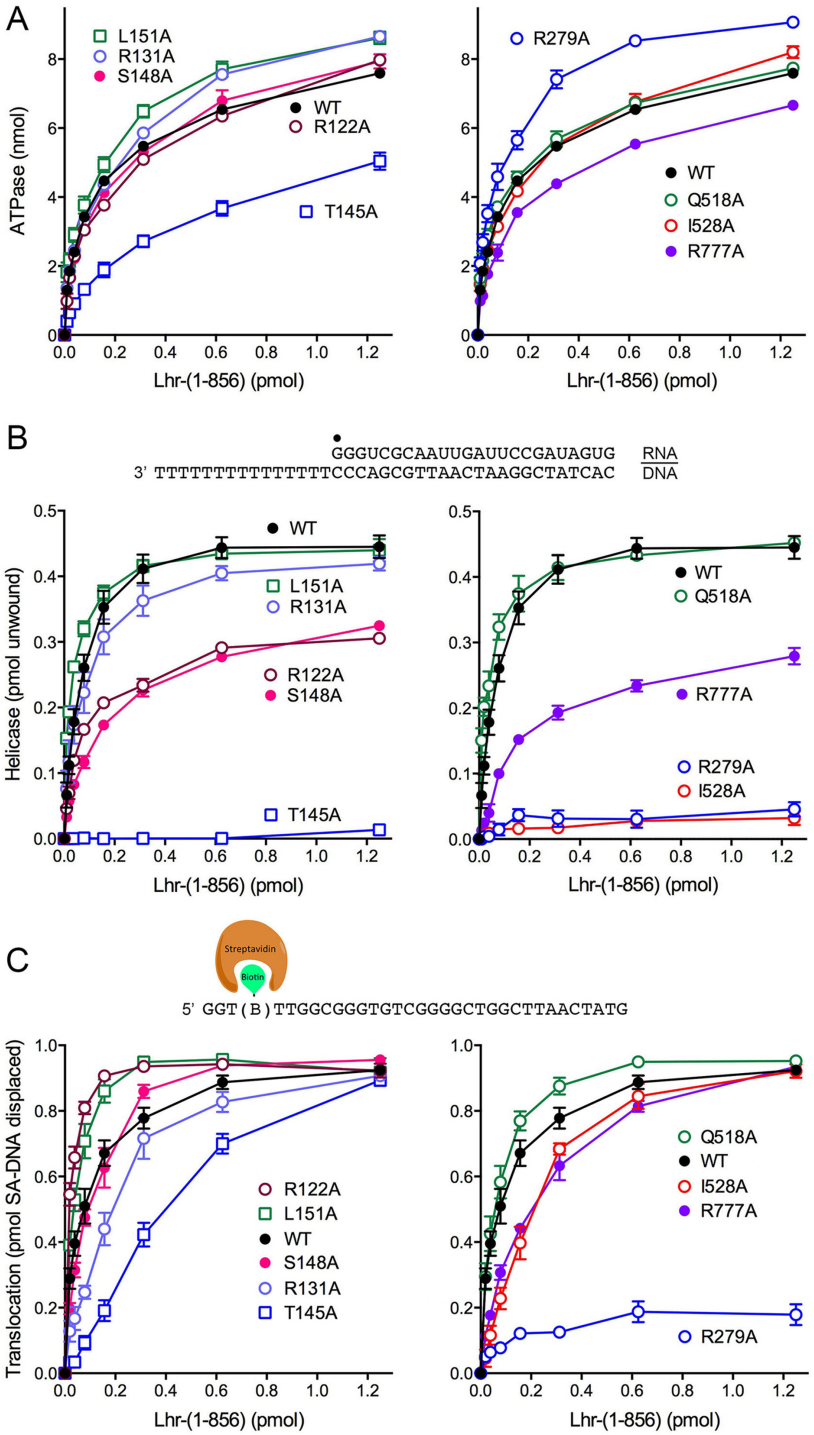

Figure 5. Mutational analysis of the loading strand interface. (**A**) ATPase activity. Reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 7.0, 50 mM NaCl, 1 mM $CaCl_2$, 1 mM DTT, 1 mM [α^{32}P]ATP (10 nmol ATP), 100 ng heat-denatured salmon sperm DNA and Lhr-(1–856) proteins as specified were incubated at 37°C for 30 min. The extents of [^{32}P]ADP formation are plotted as a function of input enzyme. Each datum represents the average of three independent enzyme titration experiments ± SEM. For the sake of clarity, the data are plotted in two separate graphs, with domain 1 mutants in the left panel and domain 2, 3 and 4 mutants in the right panel. (**B**) Helicase activity. Reaction mixtures (10 μl) contained 20 mM Tris–HCl, pH 7.0, 50 mM NaCl, 5 mM $CaCl_2$, 1 mM DTT, 1 mM ATP, 50 nM (0.5 pmol) RNA:DNA hybrid substrate (depicted at the top, with the 5′ ^{32}P label denoted by •) and Lhr-(1–856) proteins as specified. The products were analyzed by native PAGE. The extents of duplex unwinding [(ssRNA)/(ssRNA + tailed RNA:DNA duplex)] are plotted as a function of input enzyme. Each datum represents the average of three separate enzyme titrations ± SEM. (**C**) Translocase activity. Translocation was assayed by displacement of SA from a SA–(biotin)–ssDNA complex (depicted at the top) as described under 'Materials and Methods' section. Complete translocase reaction mixtures (10 μl) contained 20 mM Tris–HCl, pH 7.0, 50 mM NaCl, 1 mM $CaCl_2$, 1 mM DTT, 1 mM ATP, 100 nM (1 pmol) SA–DNA complex and Lhr-(1–856) proteins as specified. The products were analyzed by native PAGE. Extents of translocation [(free DNA)/(SA–DNA + free DNA)] are plotted as a function of input enzyme. Each datum represents the average of three separate enzyme titrations ± SEM.

teraction such that, in the presence of excess free biotin, the released SA is sequestered and the released ^{32}P-labeled ssDNA is separated from residual SA–DNA complex by native gel electrophoresis (11). Again, the mutational effects fell into three classes. Proteins R122A, L151A, S148A and Q518A were as active (or slightly more active) than WT as translocases on an enzyme concentration basis. Mutants R131A, I528A, R777A and T145A had lower than WT specific activity (evinced by a shift to the right in the titration curve) but attained the same SA displacement endpoint as did WT Lhr-(1–856) at a level of input protein corresponding to the level of input SA–DNA. By contrast, R279A was uniquely grossly defective as a translocase (Figure 5C).

A 59-mer ssDNA oligonucleotide shown previously to activate ATP hydrolysis (11) was titrated against a fixed concentration of WT and mutant Lhr-(1–856) proteins in ATPase assays as a functional readout of ssDNA binding (Supplementary Figure S3A). The titration profiles for the ssDNA of eight of the loading strand site mutants were similar to that of WT Lhr-(1–856). This group included all three helicase-defective mutants: T145A, R279A and I528A. The exception was R122A, for which the titration curve was shifted to the right and the apparent affinity was half that of WT.

Taken together, these results highlight distinct contributions of the three aa at the loading strand interface that are essential for RNA:DNA helicase activity. Arg297 is critical for coupling ATP hydrolysis to movement on the ssDNA loading strand. Thr145 and Ile528 are not required for ssDNA translocation, but are essential for duplex nucleic acid unwinding.

Mutational analysis of the ATP site

Alanine mutations were introduced in lieu of ATP site residues Phe24 in the Q motif, Ser205 and Thr207 in motif III, and Asp369. [We eschewed mutating the aa in motifs I, II and VI that coordinate the ATP phosphates and metal cofactor, in light of prior studies of many exemplary nucleic acid-dependent NTPases/helicases that showed these aa are essential for NTP hydrolysis.] The recombinant Lhr-(1–856) ATP site mutants (Supplementary Figure S2) were surveyed for the same set of activities described above for the ssDNA loading site mutants. The ATPase specific activities of the F24A, S205A, T207A and D369A proteins were 18, 50, 21 and 45% of the WT activity, respectively (Figure 6A). F24A displayed reduced helicase activity *vis à vis* WT (Figure 6B) proportional to its reduced ATPase. By contrast, S205A, T207A and D369A were grossly defective as helicases, out of proportion to the effect of these mutations on ATP hydrolysis. S205A, F24A, T207A and D369A displayed reduced translocase activity in the SA displacement assay (Figure 6C), again in rough accord with their reduced ATPase activities. The ssDNA titration profiles in the ATPase assay indicated that these ATP site mutations did not affect affinity for the ssDNA cofactor (Supplementary Figure S3B). Because the F24A mutation concordantly affected all ATPase-dependent steps surveyed (at 1 mM ATP concentration), and in light of the fact that Phe24 makes a π stack on the adenine nucleobase, we suspected that the loss of the aromatic side chain might affect ATP affinity. Thus, we gauged the rate of ATP hydrolysis as a function of ATP concentration (Figure 6D). The kinetic parameters for WT Lhr-(1–856) (apparent K_m of 0.36 mM ATP and k_{cat} of 20.4 s^{-1}) and the F24A mutant (K_m 1.78 mM ATP; k_{cat} 10.6 s^{-1}) highlighted that the phenylalanine π stack enhances ATP affinity by 5-fold. We also assessed kinetic parameters for D369A, in light of its atomic contacts with the 3′-OH of the ribose sugar, and determined a K_m of 0.76 mM ATP and k_{cat} of 16.1 s^{-1} (Figure 6D).

Lhr-(1–856) interface with the 5′ DNA segment

A detailed stereo view of the protein interactions with the 5′ segment $C^8G^9T^{10}C^{11}C^{12}A^{13}$ is shown in Figure 7. Domain 2 aa make serial H-bonds or electrostatic contacts to the DNA phosphates at the $C^{11}pC^{12}$ (via Trp255 and Lys410) and $C^{12}pA^{13}$ (via Ser253) steps. The C-domain residue Trp597 stacks on the T^{10} nucleoside and likely demarcates the ss/duplex junction, as discussed above. Asp600 makes a hydrogen bond to the C^8 base and a van der Waals contact to the G^9 base (Figure 7). The salt bridge between Arg280 in domain 2 and Glu550 in the C-domain forms an aperture around the DNA at the imputed ss/ds junction.

Mutational analysis of the interface with the 5′ DNA segment

We extended the alanine scan of Lhr-(1–856) to the five aa side chains that contact the 5′ DNA segment: Ser253, Trp255, Lys410, Trp597 and Asp600. We also targeted the Glu550 and Arg280 constituents of the aperture salt bridge. SDS-PAGE analysis of the recombinant Ala-mutants purified from soluble bacterial lysate is shown in Supplementary Figure S3. (Note: attempts to produce an R280A mutant in *E. coli* resulted in insoluble protein.) Mutants S253A, W255A, K410A, W597A and E550A hydrolyzed ATP with specific activity similar to WT enzyme. Mutant D600A was 40% as active in ATP hydrolysis as WT Lhr-(1–856) (Figure 8A).

The RNA:DNA helicase activities of mutants S253A, W255A, K410A and E550A were similar to WT (Figure 8B). D600A displayed reduced RNA:DNA unwinding, reflected in a right-shift in the titration profile and a lower extent of unwinding at the highest level of input enzyme. By contrast, W597A was inactive as a helicase.

The D600A and W597A mutants with compromised RNA:DNA helicase activity were tested for DNA translocation via SA displacement (Figure 8C). D600A displayed reduced translocase activity on a per enzyme basis, echoing the effect of D600A on duplex unwinding. W597A was wholly inactive as a translocase. These results collectively indicate that the individual domain 2 side chain contacts to the 5′ DNA segment (via Ser253, Trp255 and Lys410) are not important *per se* for Lhr-(1–856) activity. C-domain contacts are functionally relevant, especially the stacking of Trp597 on the T^{10} nucleoside, loss of which uncouples ATP hydrolysis from DNA translocation and therefore duplex unwinding.

DISCUSSION

The crystal structure of Lhr-(1–856) in a ternary complex with AMPPNP•Mg^{2+} and ssDNA establishes Lhr as the

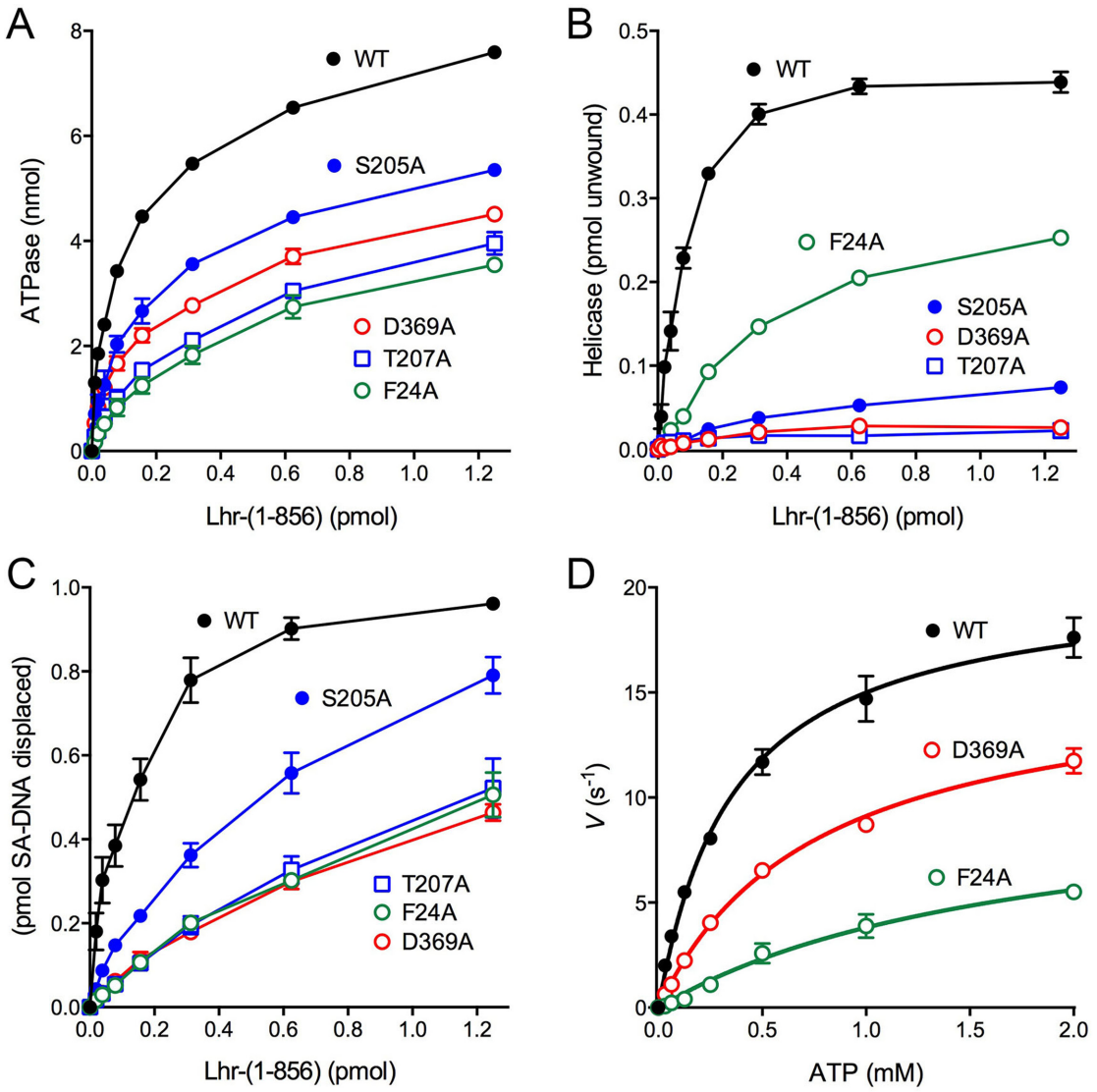

Figure 6. Mutational analysis of the ATP site. (**A**) ATPase activity was assayed as described in Figure 5. The extents of [^{32}P]ADP formation are plotted as a function of the indicated input Lhr-(1–856) proteins. Each datum represents the average of three independent enzyme titration experiments ±SEM. (**B**) Helicase activity was assayed as described in Figure 5. The extents of duplex unwinding are plotted as a function of input enzyme. Each datum represents the average of three separate enzyme titrations ±SEM. (**C**) Translocase activity was assayed as described in Figure 5. Extents of SA–DNA displacement are plotted as a function of input enzyme. Each datum represents the average of three separate enzyme titrations ±SEM. (**D**) ATP concentration dependence of ATP hydrolysis. Reaction mixtures (40 μl) containing 20 mM Tris–HCl, pH 7.0, 50 mM NaCl, 1 mM $CaCl_2$, 1 mM DTT, 100 ng of heat-denatured salmon sperm DNA, 200 nM Lhr-(1–586) (WT, F24A or D369A) and varying concentrations of [α^{32}P]ATP were incubated at 37°C. Aliquots (10 μl) were withdrawn at 10, 20 and 30 s (for 31.3, 62.5, 125 and 250 μM ATP) or 15, 30 and 45 s (for 1.0 and 2.0 mM ATP) and quenched immediately with 2 μl of 5 M formic acid. The extents of ATP hydrolysis were plotted as a function of time for each ATP concentration and the initial rates were derived by linear regression analysis in Prism. The initial rates (pmol•s^{-1}) were divided by the molar amount of input enzyme (2 pmol) to obtain a turnover number V (s^{-1}), which is plotted in the figure as a function of ATP concentration. Each datum is the average of three separate time course experiments ±SEM. A non-linear regression curve fit of the data to the Michaelis–Menten equation (in Prism) is shown.

founding member of a new family of 3′-to-5′ nucleic acid helicase distinguished by its domain architecture, mode of DNA binding and preference for unwinding RNA:DNA duplexes with a 3′ DNA tail. The enzyme comprises two N-terminal RecA-like modules (domains 1 and 2), a central WH domain, and a structurally unique C-terminal domain. As is typical for SF2 helicases, the two RecA-like modules form an active site for ATP binding and hydrolysis triggered by ssDNA. The 16-mer sized ssDNA ligand employed for co-crystallization corresponds to a minimized effective activator of ATP hydrolysis by Lhr *in vitro* and a minimized effective 3′ tail length for duplex unwinding (11). The 3′ segment of the ssDNA observed in the Lhr structure, which we take to be the DNA 3′ tail (i.e. loading/tracking strand) of the helicase substrate, binds at its 3′ end in a groove at the interface between the domain 1 and the WH domain and then feeds into a groove between the domain 2 and the C-domain. At the transition point, the ssDNA threads through a narrow aperture formed by a salt bridge between domain 2 and the C-domain. The nucleoside on

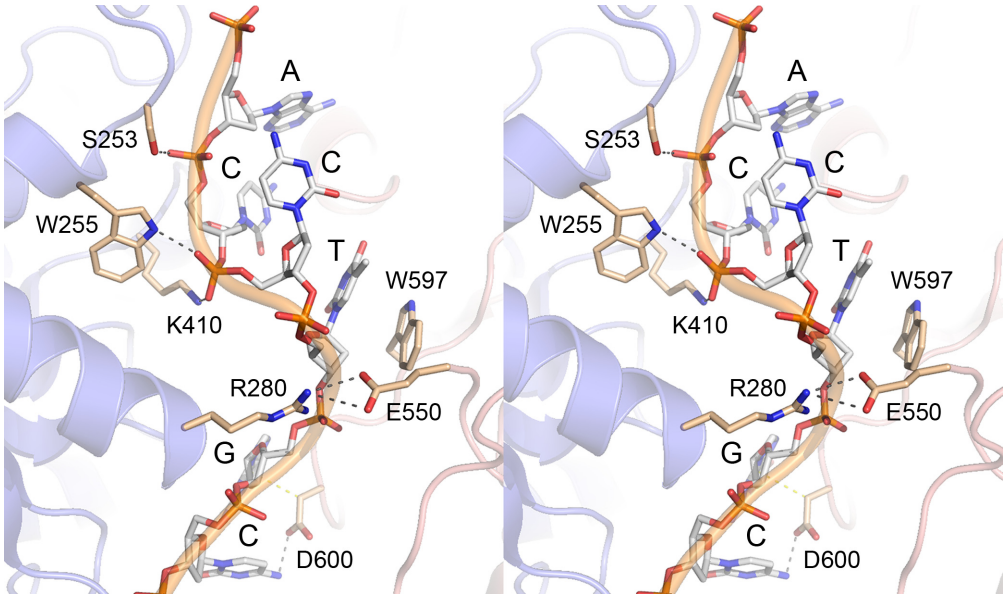

Figure 7. Interactions of Lhr-(1–856) with the 5' DNA segment. Stereo view of the 5′ segment of the ssDNA strand ($C^8G^9T^{10}C^{11}C^{12}A^{13}$), rendered as a stick/cartoon model with gray carbons and labeled by nucleobase. Lhr-(1–856) is a cartoon model, colored by domain, with interfacial aa shown as stick models with beige carbons. Hydrogen bond contacts between protein and DNA, and the Arg280-Asp550 salt bridge, are denoted by black dashed lines. A van der Waals contact of Asp600 to the G^{10} base is indicated by a yellow dashed line.

the 5′ side of the aperture is stacked on the indole ring of residue Trp597 of the C domain, which we propose demarcates the ss:ds junction of the helicase substrate. The flanking 5′ DNA segment DNA lies in a groove between domain 2 and the C-domain.

The structure of the DNA interface offered clues to the mechanism of unidirectional translocation and duplex unwinding that we pursued by alanine scanning mutagenesis of the aa in all four domains that make side-chain interactions with ssDNA. The most salient findings were that mutants T145A, R279A, I528A and W597A retained ssDNA-dependent ATPase activity but were unable to unwind a 3′ DNA-tailed RNA:DNA duplex. The helicase-dead mutants fell into two classes. R279A and W597A were defective for 3′-to-5′ translocation on ssDNA (as gauged by SA displacement), signifying that Arg279 and Trp597 are needed to couple ATP hydrolysis to mechanical work. By contrast, T145A and I538A retained translocase activity, implying that these mutants behave like ATP-driven motors on ssDNA but are unable to progress into and/or unwind the duplex portion of the helicase substrate. How can we rationalize these mutational effects?

Chemomechanical coupling by ssDNA translocating helicases is driven by iterative protein domain motions synchronized to the steps of the ATP hydrolytic cycle (1,32). ATP hydrolysis is mechanically productive when the protein domain motions ratchet the DNA tracking strand forward in one direction, typically by a single-nucleotide step. The step motion of the tracking strand can entail the flipping of nucleobases between binding pockets on the protein or between aa 'bookends' that encompass segments of the tracking strand. For translocation to proceed efficiently, the DNA ought not to be able to slide backwards, i.e. there needs to be a physical barrier. In the event that barriers to backsliding are lost, the enzyme might persist in catalyzing ATP hydrolysis accompanied by mechanically futile cycles of forward DNA movement and backsliding. If protein–DNA contacts that drive the DNA movement are lost, the ATP hydrolytic cycle might proceed with no forward motion at all.

The Lhr structure highlights discrete bookended regions of the ssDNA strand: (i) the $T^{10}C^{12}A^{13}$ trinucleotide, bookended on its 3′ side by Trp597; (ii) the $T^7C^8G^9$ trinucleotide, bracketed by Ile528 on the 3′ side and Asp600 on the 5′ side; (iii) the C^5C^6 dinucleotide delimited by Leu151 on its 3′ end and Ile528 on the 5′ side; and (iv) $T^1A^2G^3G^4$, anchored at its 5′ side to Arg122 and Arg131. These DNA segments are fulcra for protein-mediated structural transitions of the sugar–phosphate backbone (Figure 2E). We speculate that the power stroke of ATP-driven translocation entails passage of the T^{10} nucleoside over the C-domain loop that includes bracketing residues Trp597 and Asp600, so that T^{10} occupies the position previously taken by G^9. For this to occur, the narrow aperture between domains 2 and the C-domain must open up transiently. (Indeed, it is likely that the aperture adopts an open conformation during the initial binding of the enzyme to the ssDNA loading strand.) The fact that mutating Glu550 that forms the aperture salt bridge did not affect ATPase or helicase activity suggests that: (i) features of the interface between domain 2 and the C-domain other than the salt bridge suffice to support Lhr activities; or (ii) the Arg280 component of the aperture suffices in the absence of Glu550 (a hypothesis we could not test because of difficulties expressing R280A protein). In any event, the compelling inference from the inability of W597A to couple ATP hydrolysis to translocation is that the bookend stack of the indole on the 5′ DNA segment is an essential ratchet that ensures forward movement.

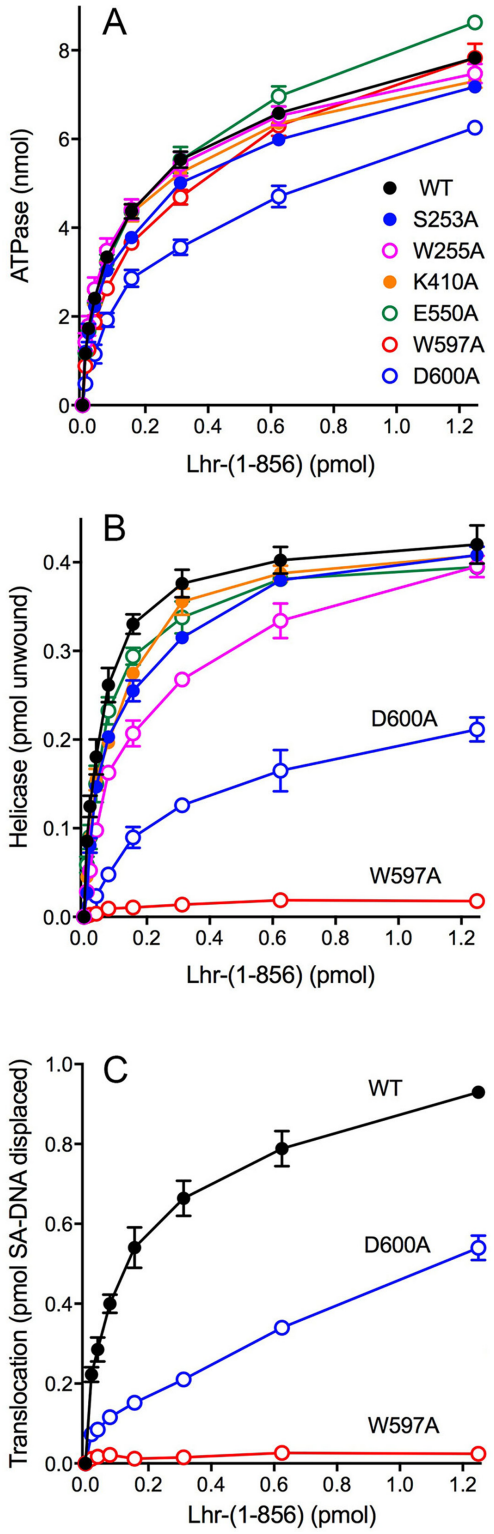

Figure 8. Effect of alanine mutations at the Lhr interface with the 5′ DNA segment. (**A**) ATPase activity was assayed as described in Figure 5. The extents of [^{32}P]ADP formation are plotted as a function of the indicated input Lhr-(1–586) proteins. (**B**) Helicase activity was assayed as described in Figure 5. The extents of duplex unwinding are plotted as a function of input enzyme. (**C**) Translocase activity was assayed as described in Figure 5. Extents of SA–DNA displacement are plotted as a function of input enzyme. Each datum in the graphs represents the average of three separate enzyme titrations ±SEM.

Arg279, which is also implicated in coupling ATP hydrolysis to translocation, contacts sequential backbone phosphates at the T^{7}pC^{8}pG^9 steps. We speculate that an ATP-driven domain 2 movement involving Arg279 is part of the power stroke alluded to above that propels the ssDNA over the Trp597-Asp600 bracket loop. Coupled to that DNA movement, we envision that the T^7 nucleoside passes over the Ile528 wedge (Figure 2E) to occupy the position previously filled by G^6. These events would collectively elicit a 1-nt movement of the enzyme in the 3′-to-5′ direction along the ssDNA. The Ile528 bookend is not essential for translocation, as gauged by SA displacement, but is critical for duplex unwinding. It is conceivable that disruption of the first base pair at the ss:ds junction, entailing re-setting the position of DNA versus Trp597, requires a second point of constraint or fixation of the ssDNA loading strand that is provided by Ile528. Thr145, the other residue that is needed for unwinding but not translocation, coordinates the C^{5}pG^6 phosphate adjacent to Ile528 and might perform a similar function during unwinding.

The structural and biochemical studies reported here provide a framework for further interrogation of the mechanism and biological activities of Lhr, as the exemplar of a conserved helicase family found in diverse bacterial taxa. From a structural perspective, the outstanding challenges are to capture snapshots of Lhr-(1–856) bound to a 3′ DNA-tailed RNA:DNA duplex (its preferred helicase substrate) and to illuminate the conformational changes in the protein–DNA complex at different stages in the ATP cycle. With respect to Lhr physiology, we have constructed an *M. smegmatis lhr*Δ strain with an unmarked deletion of the entire *lhr* gene and found that it grows as well as WT *M. smegmatis* under laboratory conditions; preliminary evidence indicates that *lhr*Δ is defective in certain aspects of DNA repair. Our collection of biochemically characterized Lhr mutants will enable a genetic analysis of the Lhr activities that are relevant to DNA repair.

DATA AVAILABILITY

Structural coordinates have been deposited in Protein Data Bank under accession code: 5V9X.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

U.S. National Institutes of Health Grants [AI64693, P30CA008748]. Funding for open access charge: National Institutes of Health Grant [AI64693].
Conflict of interest statement. None declared.

REFERENCES

1. Singleton,M.R., Dillingham,M.S and Wigley,D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, **76**, 23–50.
2. Sinha,K.M., Unciuleac,M.C., Glickman,M.S. and Shuman,S. (2009) AdnAB: a new DSB-resecting motor-nuclease from Mycobacteria. *Genes Dev.*, **23**, 1423–1437.

3. Unciuleac,M.C. and Shuman,S. (2010) Characterization of the mycobacterial AdnAB DNA motor provides insights to the evolution of bacterial motor-nuclease machines. *J. Biol. Chem.*, **285**, 2632–2641.
4. Gupta,R., Unciuleac,M.C., Shuman,S. and Glickman,M.S. (2017) Homologous recombination mediated by the mycobacterial AdnAB helicase without end resection by the AdnAB nucleases. *Nucleic Acids Res.*, **45**, 762–774.
5. Gupta,R., Barkan,D., Redelman-Sidi,G., Shuman,S. and Glickman,M.S. (2011) Mycobacteria exploit three genetically distinct DNA double-strand break repair pathways. *Mol. Microbiol.*, **79**, 316–330.
6. Sinha,K.M., Stephanou,N.C., Gao,F., Glickman,M.S. and Shuman,S. (2007) Mycobacterial UvrD1 is a Ku-dependent DNA helicase that plays a role in multiple DNA repair events, including double-strand break repair. *J. Biol. Chem.*, **282**, 15114–15125.
7. Sinha,K.M., Glickman,M.S. and Shuman,S. (2009) Mutational analysis of Mycobacterium UvrD1 identifies functional groups required for ATP hydrolysis, DNA unwinding, and chemomechanical coupling. *Biochemistry*, **48**, 4019–4030.
8. Sinha,K.M., Stephanou,N.C., Unciuleac,M.C., Glickman,M.S. and Shuman,S. (2008) Domain requirements for DNA unwinding by mycobacterial UvrD2, an essential DNA helicase. *Biochemistry*, **47**, 9355–9364.
9. Yakovleva,L. and Shuman,S (2012) *Mycobacterium smegmatis* SftH exemplifies a distinctive clade of superfamily II DNA-dependent ATPases with 3′ to 5′ translocase and helicase activities. *Nucleic Acids Res.*, **40**, 7465–7475.
10. Ordonez,H., Unciuleac,M. and Shuman,S. (2012) *Mycobacterium smegmatis* RqlH defines a novel clade of bacterial RecQ-like DNA helicases with ATP-dependent 3′-5′ translocase and duplex unwinding activities. *Nucleic Acids Res.*, **40**, 4604–4614.
11. 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.
12. Thakur,M., Kuma,M.B. and Muniyappa,K. (2016) *Mycobacterium tuberculosis* UvrB Is a robust DNA-stimulated ATPase that also possesses structure-specific ATP-dependent DNA helicase activity. *Biochemistry*, **55**, 5865–5883.
13. Thakur,R.S., Desingu,A., Basavaraju,S., Subramanya,S., Rao,D.N. and Nagaraju,G. (2014) *Mycobacterium tuberculosis* DinG is a structure-specific helicase that unwinds G4 DNA: implications for targeting G4 DNA as a novel therapeutic approach. *J. Biol. Chem.*, **289**, 25112–25136.
14. Zegeye,E.D., Balasingham,S.V., Laerdahl,J.K., Homberset,H., Kristiansen,P.E. and Tønjum,T. (2014) Effects of conserved residues and naturally occurring mutations on *Mycobacterium tuberculosis* RecG helicase activity. *Microbiology*, **160**, 217–227.
15. Thakur,R.S., Basavaraju,S., Khanduja,J.S., Muniyappa,K. and Nagaraju,G. (2015) *Mycobacterium tuberculosis* RecG protein but not RuvAB or RecA protein is efficient at remodeling the stalled replication forks: implications for multiple mechanisms of replication restart in mycobacteria. *J. Biol. Chem.*, **290**, 24119–24139.
16. Zhang,H., Zhang,Z., Yang,J. and He,Z.G. (2014) Functional characterization of DnaB helicase and its modulation by single-stranded DNA binding protein in *Mycobacterium tuberculosis*. *FEBS J.*, **281**, 1256–1266.
17. Balasingham,S.V., Zegeye,E.D., Homberset,H., Rossi,M.L., Laerdahl,J.K., Bohr,V.A. and Tønjum,T. (2012) Enzymatic activities and DNA substrate specificity of Mycobacterium tuberculosis DNA helicase XPB. *PLoS One*, **7**, e36960.
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.
19. Boshoff,H.I.M., Reed,M.B., Barry,C.E. and Mizrahi,V. (2003) DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. *Cell*, **113**, 183–193.
20. 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.
21. Pape,T. and Schneider,T.R. (2004) HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.*, **37**, 843–844.
22. Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K. and Terwilliger,T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr.*, **D58**, 1948–1954.
23. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr.*, **D60**, 2126–2132.
24. Holm,L., Kaariainen,S., Rosenstrom,P. and Schenkel,A.2008. Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**, 1780–2781.
25. Büttner,K., Nehring,S. and Hopfner,K.P. (2007) Structural basis for duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.*, **14**, 647–652.
26. Pause,A. and Sonenberg,N. (1992) Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J.*, **11**, 2643–2654.
27. Gross,C.H. and Shuman,S. (1998) The nucleoside triphosphatase and helicase activities of vaccinia virus NPH-II are essential for virus replication. *J. Virol.*, **72**, 4729–4736.
28. Schwer,B. and Meszaros,T. (2000) RNA helicase dynamics in pre-mRNA splicing. *EMBO J.*, **19**, 6582–6591.
29. Martin,A., Schneider,S. and Schwer,B. (2002) Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. *J. Biol. Chem.*, **277**, 17743–17750.
30. Papanikou,E., Karamanou,S., Baud,C., Sianidis,G., Frank,M. and Economou,A. (2004) Helicase motif III in SecA is essential for coupling preprotein binding to translocation ATPase. *EMBO Rep.*, **5**, 807–811.
31. Rocak,S., Emery,B., Tanner,N.K. and Linder,P. (2005) Characterization of the ATPase and unwinding activities of the yeast DEAD-box protein Has1p and the analysis of the roles of the conserved motifs. *Nucleic Acids Res.*, **33**, 999–1009.
32. Yang,W. (2010) Lessons learned from UvrD helicase: mechanism for directional movement. *Annu. Rev. Biophys.*, **39**, 367–385.