Analysis of the PcrA-RNA polymerase complex reveals a helicase interaction motif and a role for PcrA/UvrD helicase in the suppression of R-loops.

Urrutia-Irazabal, I.[1], Ault, JR. [2], Sobott, F.[2], Savery, N.J.[1] and Dillingham, M.S.[1*]

[1]DNA:Protein Interactions Unit, School of Biochemistry, University of Bristol. Biomedical Sciences Building, University Walk, Bristol BS8 1TD, UK.

[2]Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT, UK.

*To whom correspondence should be addressed: mark. dillingham@bristol.ac.uk

Summary

The PcrA/UvrD helicase binds directly to RNA polymerase (RNAP) but the structural basis for this interaction and its functional significance have remained unclear. In this work we used biochemical assays and hydrogen-deuterium exchange coupled to mass spectrometry to study the PcrA-RNAP complex. We find that PcrA binds tightly to a transcription elongation complex in a manner dependent on protein:protein interaction with the conserved PcrA C-terminal Tudor domain. The helicase binds predominantly to two positions on the surface of RNAP. The PcrA C-terminal domain engages a conserved region in a lineage-specific insert within the β subunit which we identify as a helicase interaction motif present in many other PcrA partner proteins, including the nucleotide excision repair factor UvrB. The catalytic core of the helicase binds near the RNA and DNA exit channels and blocking PcrA activity *in vivo* leads to the accumulation of R-loops. We propose a role for PcrA as an R-loop suppression factor that helps to minimise conflicts between transcription and other processes on DNA including replication.

Introduction

Helicases are conserved proteins found in all kingdoms of life. They are involved in a wide variety of DNA transactions including replication, repair, recombination and transcription. Structural and bioinformatic analyses have classified helicases into several superfamilies (SF) (Gorbalenya and Koonin, 1993; Singleton et al., 2007). Amongst the best-studied systems, particularly from a mechanistic point of view, are the UvrD-like family of SF1 enzymes, which include UvrD and its close cousin Rep from *E. coli*, as well as PcrA (a UvrD orthologue) from *B. subtilis* and related organisms. Structural and biochemical analyses of the helicases in this family revealed conserved mechanisms for ATP hydrolysis, DNA translocation and DNA unwinding among the different members (Lee and Yang, 2006; Velankar et al., 1999).

The PcrA/UvrD helicase has roles in rolling circle replication, nucleotide excision repair (NER), mismatch repair (MMR) and homologous recombination (HR) (Bruand and Ehrlich, 2000; Husain et al., 1985; Lahue et al., 1989; Merrikh et al., 2015; Petit et al., 1998; Zieg et al., 1978). This adaptability is provided by the many different interaction partners, for example RepC/D, UvrB, MutL and RecA, which recruit the helicase to their respective pathways via physical interactions (Hall et al., 1998; Machón et al., 2009; Manelyte et al., 2009; Veaute et al., 2005). However, the structural basis for these interactions with different partner proteins remains largely unclear. Most recently, an interaction between PcrA/UvrD and RNAP has been identified, implying a role for PcrA in transcription, but its function is not understood (Delumeau et al., 2011; Epshtein et al., 2014; Gwynn et al., 2013). Studies in *E. coli* showed that UvrD can act as an accessory helicase which helps to minimise conflicts between the replisome and replication fork barriers including transcription complexes (Guy et al., 2009; Hawkins et al., 2019). The interaction was also reported to facilitate an alternative pathway for transcription-coupled nucleotide excision repair (TC-NER) in which UvrD removes RNAP from lesions by promoting backtracking and then helps to recruit NER factors to repair the damage (Epshtein et al., 2014; Kamarthapu and Nudler, 2015). However, the existence of this new TC-NER pathway (separate from that promoted by the canonical transcription-repair coupling factor Mfd) is debated, because high-throughput sequencing of the oligonucleotides excised during NER showed no evidence for strand-specific repair promoted by UvrD (Adebali et al., 2017a). In the Gram-(+)ve model organism *Bacillus subtilis*, PcrA was found to be enriched at highly transcribed regions of the genome and, in common with its orthologue UvrD, shown to alleviate replication-transcription conflicts (Merrikh et al., 2015). Furthermore, genetic analyses of PcrA place this helicase at the interface of replication, repair, transcription

and chromosome segregation (Moreno-del Alamo et al., 2020; Petit et al., 1998). However, the mechanistic details of how PcrA/UvrD might help to promote the replication and/or repair of DNA on actively transcribed DNA remain unclear.

One nucleic acid intermediate that impacts on each of these key DNA transactions is the R-loop; a three-stranded structure in which a single RNA strand hybridises to a DNA duplex displacing an equivalent region of ssDNA (Aguilera and García-Muse, 2012; Crossley et al., 2019). R-loops can be formed when an RNA transcript re-hybridises to template DNA behind RNA polymerase. Once formed, R-loops are relatively stable structures which require active removal by enzymes such as RNases to prevent interference with other processes on DNA and preserve genome integrity. Interestingly, recent publications have shown that UvrD, Rep, Mfd and the single-stranded DNA binding protein SSB are all important for R-loop homeostasis via unknown mechanisms (Wimberly et al., 2013; Wolak et al., 2020).

Although the function of PcrA/UvrD in transcription is still unclear, details are beginning to emerge about the interface formed between the helicase and RNAP which may provide some important clues. Crosslinking experiments revealed contacts between the 1B and 2B subdomains of the helicase and the β flap tip of the RNAP β subunit, as well as the N-terminal region of the RNAP β′ subunit (Epshtein et al., 2014). The extreme C-terminal domain (CTD) of PcrA, which is disordered or was removed to aid crystallisation in the available structures of this helicase, also plays a critical role in binding RNAP but its binding site on RNAP is unknown (Gwynn et al., 2013). A structure of the isolated CTD shows that it adopts a Tudor fold, which is connected to the main body of the helicase via a long, apparently disordered, linker (Sanders et al., 2017). This is intriguing, because the canonical TC-NER factor Mfd also employs a Tudor domain for interaction with RNAP (Deaconescu et al., 2006).

In this work, we find that PcrA interacts tightly with a transcription elongation complex *in vitro* in a manner dependent on the CTD. We show that (unlike Mfd) the CTD binds to a lineage-specific insertion domain within the β subunit of RNAP, while the main body of the helicase surrounds the RNA and DNA exit channels. The CTD engages with an exposed beta-hairpin which defines a novel helicase interaction motif also found in other PcrA partners including UvrB. We further show that PcrA efficiently unwinds DNA:RNA hybrids *in vitro* and that blocking PcrA activity *in vivo*, by either overexpressing a dominant negative mutant or by preventing PcrA-RNAP association, results in increased R-loop levels. On the basis of these observations we propose a role for PcrA in R-loop surveillance, where it acts to unwind

DNA:RNA hybrids and minimise the potentially toxic effects of transcription on DNA replication and repair (Epshtein et al., 2014; Matson, 1989; Wimberly et al., 2013).

Methods

Strain construction. All strains used in this work are listed in **Table S2**. The plasmids and primers used to create them are described in **Tables S3 and S4**, respectively. Standard techniques were used for strain construction (Harwood and Cutting, 1990). *B. subtilis* was transformed using the LM-MD method or an optimised two-step starvation procedure (Anagnostopoulos and Spizizen, 1961; Burby and Simmons, 2017) Integrations within the ectopic locus were verified by PCR and/or sequencing.

Protein expression and purification. His-tagged *Bacillus subtilis* RNA polymerase was purified as described previously from the MH5636 strain (Gwynn et al., 2013; Qi and Hulett, 1998). Wild-type and mutant *B. subtilis* PcrA and biotinylated PcrA were purified as described (Gwynn et al., 2013). Plasmids expressing PcrA ΔCTD (residues 1-653) and PcrA V448C were generated by site-directed mutagenesis of the wild type construct using the Quikchange II XL kit (Agilent Technologies). In the last gel filtration chromatography of PcrA V448C, DTT was omitted to allow efficient fluorescent labelling. Fluorescent PcrA was prepared by mixing the V448C mutant with Cy3-maleimide (GE healthcare) in a 1:2 molar ratio and incubating overnight at 4 ºC. The reaction was quenched by the addition of 5 mM DTT. The free dye was separated from the labelled protein by loading the mixture on a Superdex 200 5/150 GL. Fluorescent labelling efficiency was calculated spectrophotometrically following the manufacturer's instructions. A plasmid expressing the PcrA-CTD was generated by PCR amplification from the wild type plasmid followed by sub-cloning into the pET47b plasmid. *B. subtilis* PcrA-CTD was then purified in the same way as *Geobacillus stearothermophilus* PcrA-CTD but the gel-filtration chromatography was substituted for MonoS (GE healthcare) ion-exchange chromatography (Sanders et al., 2017).

B. subtilis uvrB, *mfd* and *rpoB* were amplified by colony PCR from *B. subtilis 168* single colonies and cloned into pET47b, pET22b and pET28a, respectively (cloning oligonucleotides available in the **Supplementary Methods** and **Table S4**). Mutations to generate plasmids for expression of UvrB E233A, RpoB E310A and RpoB E310K were made by site-directed mutagenesis of the wild type constructs using the Quikchcange II XL kit (Agilent). *B. subtilis* Mfd was purified in the same way as *E. coli* Mfd, as described previously (Chambers et al., 2003). His-tagged *B. subtillis* UvrB was expressed in BL21 (DE3) cells. Following induction at mid-log phase with 1 mM IPTG, cells were grown overnight at 18 ºC. Cell pellets were resuspended in 20mM Tris–HCl pH 7.9, 300mM NaCl, 20mM imidazole, 0.1 mM DTT

supplemented with EDTA-free cOmplete protease inhibitor cocktail (Roche) and lysed by sonication. The soluble fraction was loaded on a 5 ml HisTrap column (GE healthcare) in the same buffer and eluted with an imidazole gradient. A portion of the material was supplemented with 3C protease and the two separate samples were then dialysed against 20 mM Tris-HCl pH 7.9, 300 mM NaCl, 20 mM imidazole and 0.1 mM DTT overnight at 4 ºC. For the sample that had been cleaved with 3C, the protein was loaded again onto a HisTrap column to remove the cleaved tag and the protease. The flow through was diluted into 20 mM Tris-HCl, 1 mM DTT to reduce the salt concentration before loading onto a MonoQ column (GE healthcare) equilibrated in 20 mM Tris-HCl, 100 mM NaCl, 1 mM DTT. The protein was eluted with a NaCl gradient and fractions containing UvrB were pooled and dialysed against storage buffer (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 2 mM EDTA, 1 mM DTT and 10% glycerol) overnight at 4 ºC. The uncleaved sample was purified in the same way but without the second HisTrap chromatography step.

His-tagged *B. subtilis* RNAP β subunit was expressed and purified in the same manner as UvrB. The uncleaved sample was then further purified using a heparin column equilibrated in 20 mM Tris-HCl, 100 mM NaCl, 1 mM DTT followed by elution with a NaCl gradient. The protein was dialysed against 20 mM Tris-HCl pH 7.9, 300 mM NaCl, 2 mM EDTA, 1 mM DTT and 10% glycerol overnight at 4 ºC.

Analytical size-exclusion chromatography. Transcription elongation complexes (TEC) were formed as described previously (Sidorenkov et al., 1998) using scaffold 1 (**Table S5**) and the RNA sequence: AUCGAGAGG (IBA life sciences). Briefly, the RNA oligonucleotide was incubated with the template strand (TS) for 5 minutes at 45 °C in buffer T (50 mM Tris pH 7.9, 150 mM NaCl, 10 $MgCl_2$ and 1 mM DTT) and cooled to room temperature at a rate of 1 °C/min. Then, RNAP was added and the solution was incubated at room temperature for 10 minutes. Finally, non-template strand (NTS) was added and incubated at 37°C for 10 minutes. The final concentration of each component was: 118 nM TS, 236 nM RNA, 2.36 µM NTS and 66 nM RNAP. To form the PcrA-TEC complex, the TEC was concentrated 10-fold using Amicon Ultra-0.5 3 kDa NMWL centrifugal filters and then incubated with 4.6 µM PcrA for 10 min at room temperature. Samples were loaded onto a Superdex 200 Increase 5/150 GL (GE Healthcare) connected to an AKTA FPLC instrument (GE Healthcare) in buffer T (50 mM Tris pH 7.9, 150 mM NaCl, 10 $MgCl_2$ and 1 mM DTT) at room temperature. Samples were directly injected in the injection loop prior to the chromatography. Where necessary, samples were collected to analyse their content by SDS-PAGE.

Electrophoretic mobility shift assays TECs were assembled as above but using different DNA:RNA scaffolds containing fluorescently-labelled oligonucelotides (**Table S5**), the buffer supplemented with 2.5% (v/v) Ficoll and analysed for their ability to bind to PcrA. Wild type or fluorescent PcrA at the concentration indicated was added to the TEC at room temperature for 10 minutes before loading the samples onto a 4.5 % polyacrylamide TBE-PAGE gel. The gels were imaged using a Typhoon FLA 9500 (GE Healthcare).

Analysis of protein:protein interactions using pulldown assays *B. subtilis* cell extracts were prepared for *ex vivo* pulldowns as described previously (Gwynn et al., 2013) using strains 1A1 or IU79, the latter supplemented with 1% xylose to induce the expression of myc-tagged PcrA. 1 µM biotin-tagged proteins were used as baits on streptavidin beads, whereas 2 µM his-tagged proteins were used with Ni-NTA beads. The wash and dilution buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.1% Triton X-100) was supplemented with1 mM EDTA for experiments with the streptavidin beads and 20 mM imidazole for Ni-NTA beads. For the *in vitro* pulldown, purified β subunit was used at 2 µM concentration. For experiments that contained competitors 1.5 µM PcrA or PcrA-CTD, 1.5 µM Mfd and 3 µM UvrB were used. *In vitro* pulldowns were analysed by Coomassie staining the SDS-PAGE gel. *Ex vivo* pulldowns were analysed by Western blotting using a monoclonal anti-RNAP β subunit antibody (Abcam; ab202891) or a monoclonal anti-myc antibody (Proteintech; 67447-1-Ig). In the latter experiments, a portion of the gel was excised before transfer to the blotting membrane to confirm uniform bait levels by Coomassie staining, except where this was not possible with UvrB due to its similar molecular weight to PcrA. Images were quantified using ImageQuant (GE healthcare).

Hydrogen-deuterium exchange coupled to mass spectrometry HDX-MS experiments were performed using an automated robot (LEAP Technologies) coupled to an M-Class Acquity UPLC, HDX manager (Waters Ltd.), and Synapt G2S-i mass spectrometer. For the CTD-RNAP experiments, protein samples were prepared containing either or both 5 µM RNAP and 10 µM CTD. For the PcrA-RNAP experiments samples were prepared using 0.5 µM RNAP and/or 2 µM PcrA. 30 µl of the protein solution in sample buffer (10 mM potassium phosphate buffer pH 7.4) was added to 135 µl deuterated buffer (10 mM potassium phosphate buffer in deuterated water, pD 7.4) and incubated at 4ºC for 0.5, 2, 5 and 10 minutes. After labelling, the reaction was terminated by the addition of 50 µl of sample to 100 µl quench buffer (10 mM potassium phosphate, 2 M guanidine- HCl, pH 2.2) resulting in a final pH of approximately 2.5. 50 µl of quenched sample were loaded onto an immobilised ethylene-bridged hybrid (BEH) pepsin column (Waters Ltd.). The resulting peptides were passed through a VanGuard

BEH C18 pre- column and a C18 column (Waters Ltd.) and separated by gradient elution of 0–40% MeCN (0.1% v/v formic acid) in H_2O (0.3% v/v formic acid). For the CTD-RNAP experiments, quench and pepsin wash buffers were supplemented with 0.05% (w/v) DDM.

The UPLC was interfaced to the mass spectrometer via electrospray ionisation. HDMSE and dynamic range extension modes (Data Independent Analysis couple with IMS separation) were used to separate peptides by IMS prior to fragmentation in the transfer cell. Data were analysed using PLGS and DynamX software provided with the mass spectrometer. The restrictions applied for peptide identification in DynamX were the following: minimum intensity 1000, minimum products per amino acid 0.3, maximum sequence length 25, maximum ppm error 6, replication file threshold 4. The difference plots were generated using the in-house developed algorithm Paved (Cornwell et al., 2018) and the data represented in Graphpad Prism 7.0. Difference plots were then mapped onto structures in Chimera (Pettersen et al., 2004) and PyMOL using available structures and/or homology models generated by Swiss-model (Waterhouse et al., 2018).

Bioinformatic analysis Protein orthologues were searched for in the reference proteomes of the Protein Information Resource using the *B. subtilis* protein sequence as a query (Chen et al., 2011). Obtained sequences were manually curated to remove sequences that were substantially smaller than the *B. subtilis* sequence or that contained large gaps. The conservation logo was created using WebLogo (Crooks et al., 2004) and sequence alignment figure created with ESPript 3.0 (Robert and Gouet, 2014). *B. subtilis* proteins' homology models were created using SWISS-MODEL with the indicated PDB files (Waterhouse et al., 2018). Protein-protein interaction docking analysis was performed using Haddock 2.2 and the best structures from the clusters with the lowest HADDOCK score were compared (Van Zundert et al., 2016).

CD spectroscopy CD spectra were collected using a JASCO J-810 spectropolarimeter fitted with a Peltier temperature control (Jasco UK). 0.5 mg/ml protein samples were buffer exchanged into phosphate buffered saline (PBS; 8.2 mM NaH_2PO4, 1.8 mM KH2PO4, 137 mM NaCl and 2.7 mM KCl (pH 7.4)) by 16 hr dialysis at 4˚C using a membrane with a MWCO of 10 kDa. Using 0.25 mg/ml of protein in a 0.1 cm quartz cuvette, data were acquired across a 190-206 nm absorbance scan with a scan rate of 100 nm/min at 20ºC. Raw data was normalised to molar ellipticity (MRE ($deg.cm^2.dmol^{-1}$)) using calculation of the concentration of peptide bonds and the cell path length. A buffer only baseline was subtracted from all datasets.

Helicase assays Helicase assays were carried out essentially as described previously with the following changes (Gwynn et al., 2013). The reactions were performed at 20 ºC in 50 mM Tris pH 7.5, 50 mM NaCl, 2 mM $MgCl_2$ and 1 mM DTT and started by the addition of the protein. The list of DNA and RNA substrates used is shown in **Table S6**. The products were run on 12% acrylamide TBE gels, imaged using a Typhoon phosphorimager and quantified using ImageQuant software.

ATPase assay The ATPase activities of PcrA and UvrB were measured using an enzyme linked assay in which ATP hydrolysis is coupled to NADH oxidation (Kiianitsa et al., 2003). The assay was carried out at 25 ºC using 1 nM PcrA, 50 nM UvrB and 5 µM nucleic acids (as indicated). For measurements of PcrA ATPase activity the reaction buffer contained 50 mM Tris pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$ and 5 mM DTT. The method used for UvrB was as described previously (Webster et al., 2012).

Growth curves Single colonies were grown overnight in LB at 30 ºC. Each strain was diluted to an OD_{600} of 0.01 in LB supplemented with 1 mM IPTG and then grown at 37 ºC. Absorbance measurements were taken at the indicated times, averaged and plotted using Graphpad Prism.

RNA/DNA hybrid dot blot Single colonies were grown overnight in LB at 30 ºC. Cultures were diluted to an OD_{600} of 0.05 into fresh LB medium supplemented with 1 mM IPTG for both *B. subtilis* morning and overnight inductions. For the FL-PcrA overexpression experiment, overnight cultures were induced the next morning at the time that they were diluted. For the CTD experiments, induction was started while inoculating the single colonies for the overnight growth. Cultures were grown until an OD_{600} 1.1-1.2 and genomic DNA was purified using the DNeasy Blood and Tissue (Qiagen) or the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich) following manufacturer's instructions. gDNA was quantified by Nanodrop and a fraction of DNA was treated with 5 U RNase H (NEB) for 1h at 37 ºC. gDNA serial dilutions and the RNase treated gDNA were spotted on a positively charged Hybond-N+ nylon membrane (Amersham) using a dot-blot apparatus. The DNA was probed with the S9.6 antibody (1:1000 dilution, Millipore) in 1% BSA/TBST overnight at 4 ºC after UV-crosslinking (0.12 J/cm^2) and blocking the membrane with 5% milk/TBST for 1h at RT. An anti-mouse HRP antibody (1:10000 dilution, Santa Cruz Biotechnology) was used as secondary antibody. Images were acquired with Odyssey Fc (Li-COR Biosciences). DNA loading was calculated by staining with 0.05% methylene blue in 0.5 M sodium acetate buffer (pH 5.2) after washing the membrane with 5% acetic acid as described previously (Ko et al., 2010;

Raghunathan et al., 2018). Images were quantified using ImageQuant (GE healthcare). Error bars show the SEM of at least 3 independent experiments.

Immunoblotting *B. subtilis* cells were grown until stationary phase in LB at 37 ºC, harvested by centrifugation and the pellet was resuspended in SSC prior to sonication. OD_{600} was used to normalize the loading of samples in the SDS-PAGE gel. The gel was cut into two parts prior to transfer, and the upper portion was Coomassie stained to use as loading control and the lower portion was transferred to a PVDF membrane by electroblotting. Myc-tagged PcrA was detected using a monoclonal anti-myc antibody (Proteintech) followed by an anti-mouse-HRP goat antibody (Santa Cruz Biotechnology).

Results

PcrA binds tightly to a transcription elongation complex using its C-terminal Tudor domain. Interaction between PcrA/UvrD and RNAP is well-documented in the literature, but little is known about which nucleic acid scaffolds or transcription states are the most favourable for association. In *E. coli*, a complex between RNAP and UvrD was reported to be stable enough to coelute during size-exclusion chromatography (SEC) (Epshtein et al., 2014). In the *B. subtilis* system, although purified PcrA and RNAP do clearly interact in the absence of nucleic acids, we found that a DNA/RNA scaffold which mimics an elongation complex was required to form a sufficiently stable complex to resolve using SEC (**Figure S1A-B**). To rule out the possibility that the enhanced interaction was due simply to the helicase binding to the DNA, the complex formation was analysed further by electrophoretic mobility shift assays (EMSA) (**Figure S1C**). PcrA efficiently supershifted the transcription elongation complex (TEC; a pre-formed complex between RNAP and a DNA/RNA scaffold) whereas it was barely able to bind to the scaffold alone. Based upon EMSA, the apparent dissociation constant was ~500 nM and the interaction was dependent on the C-terminal Tudor domain of PcrA (CTD; **Figure 1A**). Removal of the CTD almost eliminated complex formation at concentrations of up to 1.5 µM of PcrA. Moreover, the purified CTD alone supershifted the TEC efficiently, albeit to a lower position in the gel. Although the UvrD CTD has been shown to possess a weak DNA binding activity (Kawale and Burmann, 2020), in our hands the PcrA CTD does not detectably bind to either single- or double-stranded DNA (Kd >> 5 µM) (Sanders et al., 2017). Therefore, this result further supports the idea that the PcrA-TEC interaction (at least with this scaffold) is substantially mediated by protein-protein interactions. Additional support for this assertion was provided by EMSA supershift experiments in which the structure of the scaffold was varied (**Figure 1B**). PcrA efficiently shifted the TEC regardless of the presence or length of the upstream and downstream DNA in the scaffold.

Competition experiments showed that full length PcrA and free CTD competed for TEC association, and there was no evidence for a ternary complex formed between PcrA, the free CTD and the TEC (**Figure. S1D**). The addition of unlabelled CTD caused a decrease in the levels of fluorescently labelled PcrA bound to the TEC rather than a further supershift. This strongly suggests that full length PcrA and the CTD bind to the same (unknown) surface of RNAP and is consistent with the idea that the TEC recruits a single PcrA molecule. Note that the ability of the free CTD to outcompete full length PcrA for RNAP association is relevant for *in vivo* experiments presented below.

The C-terminal domain of PcrA interacts with the lineage-specific insertion domain 1 (SI1) of the β subunit of RNAP. Crosslinking of UvrD to RNAP has shown that the core helicase interacts with the RNAP β-flap tip and the N-terminal region of the β′ subunit in bacteria, but no information about the CTD-interaction domain was obtained (Epshtein et al., 2014). Therefore, we pursued a complementary approach in the *B. subtilis* system, using hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) to study the complex formed between the free CTD and RNAP. This technique allows one to measure the rate of exchange of amide hydrogen atoms with deuterium atoms in the solvent. This rate is affected by hydrogen bonding, surface accessibility, protein dynamics and changes in the presence and absence of partner proteins and can therefore be used to probe potential binding interfaces or conformational changes.

We obtained good sequence coverage and redundancy for both the CTD and the major subunits of RNAP (**Table S1**). The experiment was first validated by analysing the protection pattern on the CTD. As expected, data for the CTD in the CTD-RNAP complex showed significant protection compared to the CTD alone (**Figure 1C**). To visualise the results, a homology model of the Tudor domain of *B. subtilis* PcrA was created from the structure of the closely related *G. stearothermophilus* protein (Sanders et al., 2017) (**Figure 1D**). The areas of protection, which indicate the binding site location, overlap with residues that have been shown to be important for the interaction between PcrA and RNAP through site-directed mutagenesis studies (Sanders et al., 2017). Moreover, it is mostly localised on the face of the Tudor domain that is typically expected to interact with partner proteins (Liu et al., 2015; Westblade et al., 2010). This agreement with previous studies provided confidence that the HDX-MS data was reporting correctly on the interface formed between PcrA and RNAP.

We next searched for protected peptides in the different RNAP subunits. It should be noted that, given the large size of RNAP (374 kDa) compared to the tiny Tudor domain (9.3 kDa), we did not expect to observe any protection for the large majority of the RNAP sequence if the stoichiometry of interaction is 1:1. In accordance with this expectation, only a small region of the β subunit of RNAP at approximately amino-acid position 300 showed significant protection (**Figure 2A,** with HDX protection plots for the rest of RNAP β and the other RNAP subunits in **Figure S2**). Comparison of the HDX at different timepoints showed that protection of this region increased with time after exposure to D_2O, in agreement with a protein-protein

interaction site. Fortunately, this region of interest was well covered by the LC-MS/MS analysis with overlapping peptides, and the smallest part of the polypeptide protected was the seven-residue sequence PETGEIL (**Figures 2A and 2B**). Intriguingly, despite the ubiquity of UvrD/PcrA orthologues in the bacterial domain and the strong conservation of the CTD sequence, the protected region in RNAP belongs to a lineage-specific insertion domain (SI1, also referred to as βi5 (Lane and Darst, 2010)) which differs significantly between organisms in terms of the length, sequence and precise location within the primary structure. Nevertheless, a nine-residue sequence motif (including the seven amino acids of interest) is present in most bacterial RpoB sequences despite variation in its exact position. As an example, primary structure diagrams for the *B. subtilis* and *E. coli* RpoB subunits are shown in **Figure 2B** with complete aligned sequences for the β2 and SI1 regions shown in **Figure S3**. Despite being quite divergent compared to better-conserved parts of RpoB, the two SI1 domains contain a very similar peptide motif (VDPETGEIL and IDESTGELI respectively). Although the motifs are located in different parts of the SI1 domain primary structure they occupy a broadly similar space within the global architecture of RNAP (**Figure 2C**). In *E. coli* the motif adopts a beta hairpin structure, whereas this region is locally disordered in the *B. subtilis* TEC cryoEM structure (Newing et al., 2020; Pei et al., 2020). The variability in the SI1 domain would make it extremely challenging to appreciate the conservation of this motif across different species from primary sequence alone.

Alignment of the beta-hairpin motif from *B. subtilis*-like SI1 domains demonstrated its high conservation (shown in Weblogo format in **Figure 2C**). The consensus sequence obtained from *E. coli*-like SI1 domains is very similar, consistent with the idea that they are in fact the same motif (**Figure 2C**). The most conserved and striking feature of this putative helicase interaction motif is a TGE triad which occupies the tip of a loop between two beta strands. Indeed, the glutamic acid (E310 in *B. subtilis* RpoB) is absolutely conserved across all aligned RpoB sequences. To experimentally validate the motif, we purified the β subunit of *B. subtilis* RNAP with and without substitutions at the E310 position. *In vitro* assays showed that substitution of glutamic acid with either alanine or lysine at position 310 severely disrupts pulldown of purified RpoB by PcrA (**Figure 2D**), confirming the importance of the SI1 motif for the interaction. Lack of interaction with the E308 mutants was not due to global mis-folding because both mutant proteins had equivalent CD spectra to wild type (**Figure S4**).

The helicase interaction motif is found in several other PcrA interaction partners including UvrB. In common with other SF1 helicases, PcrA/UvrD interacts with many

different partner proteins explaining its recruitment to different DNA transactions and the resulting multi-functionality of this helicase (Gilhooly et al., 2013). To investigate whether the helicase interaction motif we had identified in RpoB played a wider role in mediating PcrA interactions we searched for it within the *Bacillus subtilis* proteome using PROSITE with the query [VIL]-D-X-X-T-G-E-[VIL]-[VILT], where X is any amino acid and the square brackets indicate ambiguity. This search retrieved seven proteins (only ~0.3 are expected by chance), four of which (RpoB, UvrB, YxaL and YwhK) are known to interact with PcrA (**Figure 3A**). Importantly, the structure of UvrB and high-confidence homology models of YxaL and YwhK all show that, as is the case for the beta subunit of *E. coli* RNAP, the conserved TGE triad is located at the tip of a surface loop flanked by β sheets (**Figure 3B**). There is currently no evidence for interaction between PcrA and the other three proteins (QueA, RplX and YtzB) which may now be considered as putative PcrA interactors. When this same motif search was applied more broadly across the entire SwissProt database it returned only 607 sequences, of which 326 and 42 were bacterial RpoB and UvrB sequences respectively, demonstrating its very high specificity for returning PcrA interactors.

YxaL and YwhK are β propeller proteins that have been identified as PcrA interactors using yeast two-hybrid and biochemical assays (Noirot-Gros et al., 2002). The fragment of YxaL found to interact with PcrA in the yeast two-hybrid assay is consistent with the location of the newly identified motif. These two proteins are found in a very limited subset of bacterial proteomes and their cellular function is obscure. Therefore, we next decided to focus on the interaction of PcrA with UvrB; a ubiquitous Superfamily 2 helicase that is a key component of the nucleotide excision repair apparatus (Kisker et al., 2015). Importantly, UvrB has already been shown to bind to the C-terminal domain of *E. coli* UvrD (Manelyte et al., 2009) and this interaction is, therefore, potentially analogous to that formed between the CTD and RNAP β. To validate the predicted PcrA binding site on UvrB, we performed pulldown experiments using his-tagged UvrB as bait and myc-tagged PcrA in *B. subtilis* extracts as prey (**Figure 3C**). These experiments confirmed the expected interaction between PcrA and wild type UvrB protein. Furthermore, and in agreement with our hypothesis, substitution of the glutamic acid in the UvrB TGE motif (E233) severely disrupted the interaction with PcrA while having no effect on the DNA-stimulated ATPase activity of UvrB (**Figure S5**).

Inspection of the domain architecture of UvrB reveals that the PcrA interaction motif is located within domain 2, which also interacts with UvrA (Pakotiprapha et al., 2012; Truglio et al., 2004) (**Figure S6A**). However, because amino acid substitutions reported to disrupt UvrA-

UvrB interaction are not within the motif itself, binding of UvrA to UvrB would not necessarily preclude binding of PcrA/UvrD. Interestingly, a high scoring docking model for the interaction between the PcrA-CTD and UvrB using HADDOCK (Van Zundert et al., 2016) suggests that the complex is stabilised by an electrostatic interaction between UvrB E233 and the PcrA K727 residue, which is known to be critical for interaction between PcrA and RNAP (**Figure S6B**). The docked complex shows how PcrA might be recruited to sites of damage after the damaged oligonucleotide has been nicked, as DNA near the docking site is accessible. A conservation logo for the PcrA interaction site on UvrB was created from reference proteomes (**Figure S6C**). The consensus sequence is highly similar to that obtained for the RNAP interaction although, interestingly, the glutamate residue shows somewhat greater variability.

Our results predict that the interaction between the PcrA C-terminal Tudor domain and UvrB is equivalent in molecular terms to its interaction with RNA polymerase. A corollary is that UvrB and RNA polymerase might be engaged in a simple competition for the Tudor domain of PcrA. To test this, we performed pulldown of RNA polymerase using PcrA as a bait and titrated the cell extract (i.e. the prey proteins) with increasing concentrations of purified recombinant UvrB (**Figure 3D-E**). As expected, introduction of UvrB significantly reduced the amount of RNA polymerase that was recovered from the cell extract. As has been noted previously, other RNAP interactors also bind to the RpoB subunit via Tudor domains, including the transcription-repair coupling factor Mfd. Using the same pulldown assay, titration of Mfd did not diminish the yield of RNAP (**Figure 3D-E**). This strongly suggests that, although both proteins contain Tudor domains for interaction with RpoB, PcrA and Mfd bind at different physical locations and without competing. This is in agreement with recently-published observations for the *E. coli* system (Kawale and Burmann, 2020).

The helicase core of RNAP interacts with RNAP near the RNA and DNA exit channels. Despite the apparently critical role played by the CTD in RNAP interactions, previous work in live *E. coli* and *B. subtilis* cells has found at most mild phenotypes associated with deletion of the C-terminal domain of UvrD and PcrA, respectively (Manelyte et al., 2009; Merrikh et al., 2015; Sanders et al., 2017). This led us to consider the possibility that the N-terminal region of PcrA, including the helicase core, also contributes significantly to the interaction with RNA polymerase. This view is consistent with crosslinking studies of UvrD-RNAP which implicated the 1B and 2B subdomains of UvrD in interactions with the RNAP β-flap tip and the N-terminal

region of the β′ subunit (Epshtein et al., 2014). Note that these structural features surround the RNA and DNA exit channels.

To explore interaction between full length PcrA and RNAP we again turned to HDX-MS, obtaining good coverage and redundancy values for the various constituent polypeptides (**Table S1**). PcrA showed HDX protection in two main regions upon interaction with RNAP: the 2A domain and the CTD (blue and magenta highlights respectively; **Figure 4A** and **S7A**). Note that peptides from the long flexible linker between these two regions were not detected above the confidence threshold and so we have no data for that part of the protein. The protection pattern was visualised on a homology model of the active DNA bound state of PcrA (**Figure 4B**) (Lee and Yang, 2006; Ordabayev et al., 2019; Velankar et al., 1999), in which the CTD is disordered. Most of the protected regions outside of the CTD prominently cluster in the 2A domain and on the side of the helicase which also binds to DNA. The crosslinking positions identified in domains 1B and 2B in previous work are on the same DNA-facing side of the helicase and are therefore broadly consistent with these data (Epshtein et al., 2014).

The HDX protection afforded to the RNAP subunits in the presence of full length PcrA was more extensive than in the presence of the CTD, and identified important regions in the two largest subunits, β and β′ (butterfly plots for the rest of the subunits are shown in **Figure S8**). In agreement with our data for the CTD-RNAP interaction, the SI1 region showed protection that we therefore assign to the interaction with the PcrA Tudor domain (**Figure 5A**, green shadow and **Figure 6B**). The other protected region within the β subunit corresponds to the β flap tip, as identified by crosslinking in the *E. coli* system (Epshtein et al., 2014)(**Figure S7B**). The β′ subunit also showed two protected clusters in the N-terminal and central regions (**Figure 5B** and **S7C**). The recent structure of the *B. subtilis* transcription elongation complex allowed us to visualise the results (Newing et al., 2020) (**Figure 5C**) and reveals that protected regions of RNAP surround the RNA and DNA exit channels. Unfortunately, the zinc finger domain of the β′ subunit was not covered by our dataset but, given its close proximity to the protected regions we have observed, as well as the crosslinking analysis of the *E. coli* RNAP-UvrD complex (Epshtein et al., 2014), it is likely to be involved in the interaction too. Taken together, our data suggested to us that the PcrA helicase might be acting close to the RNA and DNA exit channels and are consistent with our earlier conclusion that Mfd and PcrA bind to different parts of RNAP.

PcrA is able to unwind DNA-RNA hybrids by translocating on ssDNA. It is well established that PcrA/UvrD is a 3′>5′ ssDNA translocase and helicase that binds preferentially to 3′-ssDNA overhangs to initiate unwinding of flanking DNA duplexes. However, UvrD has also been shown to unwind DNA-RNA hybrids efficiently (Matson, 1989) although we are not aware of a cellular function having been ascribed to this activity. Moreover, no data is available for the enzyme's activity on RNA duplexes. Therefore, given the apparent recruitment of PcrA close to the RNA and DNA exit channels of RNAP, we next investigated the relative helicase activity displayed by PcrA against all possible combinations of a 3′-tailed duplex that can be formed by DNA and RNA oligonucleotides (**Figure 6A**). PcrA was able to efficiently unwind the two substrates in which the 3′-ssDNA tail was formed from DNA, including an all-DNA duplex and a DNA-RNA hybrid (**Figure 6B**). In contrast, the enzyme was unable to unwind either substrate containing a 3′-ssRNA tail (either an RNA-DNA hybrid or an RNA duplex), even at elevated concentrations. Unwinding of both the DNA duplex and the RNA-DNA hybrid was absolutely dependent on the presence of the 3′-ssDNA tail (**Figure S9A-B**), implying that it is initiated by enzymes binding to ssDNA and moving in the 3′>5′ direction. This idea was further corroborated by investigating the nucleic acid-dependence of the ATPase activity (**Figure 6C**). In the absence of DNA or RNA, the basal ATPase activity of PcrA was low (~1 s^{-1}). However, it was stimulated approximately 100-fold in the presence of poly(dT) single-stranded DNA. In contrast, poly(U) single-stranded RNA had no effect on the observed ATPase activity. Similar results were obtained with short DNA and RNA oligonucleotides containing mixed base sequences (data not shown). These experiments demonstrate that PcrA helicase can efficiently unwind DNA-RNA hybrids but can only do so by ATP-dependent translocation on the DNA strand. Together with the HDX experiments localising PcrA to the RNA and DNA exit channels, we hypothesised that PcrA might act to unwind R-loops formed during transcription.

PcrA helicase suppresses R-loop levels *in vivo*. Both PcrA and UvrD have been shown to interact with RNAP and to alleviate replication-transcription conflicts (in *B. subtilis* and *E. coli* respectively) (Guy et al., 2009; Hawkins et al., 2019; Merrikh et al., 2015). However, the precise role for this interaction remains unknown and study of the system is frustrated by the essentiality of both PcrA and RNA polymerase, making phenotypic analysis more complex. To overcome this difficulty, we first exploited a dominant negative mutant of PcrA/UvrD in which a glutamate in the Walker B motif is substituted to allow DNA and ATP binding, but not hydrolysis and concomitant DNA translocation (Brosh and Matson, 1995). Studies in many

Additional Strand Catalytic E (ASCE)-class NTPases identify this residue as the catalytic base which accepts a proton from water, thereby promoting nucleophilic attack at the γ-phosphate and enabling ATP hydrolysis (Thomsen and Berger, 2008). Substitution of this glutamate residue with glutamine prevents ATP hydrolysis, trapping the enzyme in an ATP-bound or transition-like state (Hirano and Hirano, 2004; Soultanas et al., 1999).

Induction of PcrA-E224Q from an ectopic locus in otherwise wild type cells caused severe growth defects which became most apparent as the cells entered exponential phase (**Figure S10A**). This defect was not apparent when wild type PcrA was expressed from the same locus and is therefore due to the E224Q mutation as opposed to the effects of overexpression *per se* or the elevated cellular PcrA concentration that results. Interestingly, delaying the induction of PcrA until the onset of exponential phase (see **Methods** and **Figure S10A** for details) resulted in a much smaller effect, presumably because a lower concentration of mutant PcrA had accumulated in early exponential phase. This may suggest that the toxic effect of dysfunctional PcrA is mainly felt during rapid growth. Removal of the CTD from the dominant negative mutant did not relieve the toxicity we observed. This could indicate that PcrA lacking the CTD can still interact weakly with RNAP (as we have discussed above) or that the toxic effect we observe arises from another role of this multifunctional enzyme that does not require the CTD. We also tested whether removal of *mfd* (that has been shown to counteract PcrA essentiality (Moreno-del Alamo et al., 2020)) could relieve the mutant's toxicity but found that the deletion had no effects on the growth defect (**Figure S10B**).

To test our hypothesis that PcrA unwinds R-loops, we next measured DNA-RNA hybrid levels in these PcrA overexpression strains using the S9.6 antibody (Boguslawski et al., 1986). In this experiment, genomic DNA was purified, spotted onto a dot blot membrane and probed for both DNA-RNA hybrids (S9.6) and total DNA (methylene blue staining) (**Figure 6D**). The relative hybrid to DNA level was then normalised to a control experiment in which there was no gene in the ectopic expression cassette. Wild type PcrA overexpression did not alter relative R-loop levels, whereas overexpression of the E224Q mutant led to a significant increase (~2.7 fold). This increase is comparable to the effect of deleting known R-loop suppression factors from bacterial cells such as *rnhC* or inhibiting *rho* using BCM in *E. coli* (Raghunathan et al., 2018) (see also **Figure 6G**).

With the aim of testing whether the interaction of PcrA with RNAP is important for R-loop regulation, we next overexpressed the CTD of WT PcrA or a CTD mutant (K727A) that does

not interact with RNAP (Sanders et al., 2017). Based upon *in vitro* TEC supershifts (**Figure S1D**) and *ex vivo* pulldown assays (**Figure 6F and S11A**), we knew that the free CTD inhibited the interaction between full length PcrA and RNAP. Therefore, we reasoned that overexpression of the CTD in cells should block endogenous wild type PcrA from being recruited by RNAP. The relative R-loop levels in strains overexpressing the WT or mutant CTD were measured by dot-blot as before (**Figure 6G** and **Figure. S11B-C**). Overexpression of the WT CTD caused a significant increase in R-loop levels compared to WT cells or the empty expression cassette control. This increase approached the levels associated with an *rnhC* deletion strain. However, replacing the wild type CTD with a CTD mutant that fails to interact with RNAP reduced R-loop levels to a value not significantly different to the strain with an empty expression cassette. This supports the idea that the activity of PcrA suppresses R-loop formation and that this effect is dependent upon the ability of the CTD Tudor domain to interact with RNAP and/or other partner proteins. Finally, to determine whether this function of PcrA is conserved in other bacteria, we compared R-loop levels in WT and Δ*uvrD E. coli* strains (where deletion of *uvrD* is not lethal). We observed that loss of UvrD caused a significant increase in R-loops levels in the cell (**Figure S11D-E**).

Discussion

The bacterial cell displays a remarkable ability to maintain genomic stability during rapid growth when many different DNA and RNA transactions occur simultaneously on the same chromosome. The UvrD-like family of helicases play a critical role at the interface of such pathways, orchestrating DNA repair as damage arises and helping to minimise conflicts that would otherwise occur between DNA replication and transcription (Dillingham, 2011). Several reports have identified physical or functional interactions between PcrA/UvrD and RNA polymerase which could underpin its emerging role in managing conflicts between transcription and replication or repair processes.

We showed here that PcrA binds tightly to a pre-formed TEC *in vitro* to form a complex that is stable during gel filtration, in contrast to complexes formed between free PcrA and RNAP. The complex was unaffected by removal of the upstream or downstream DNA duplex from the TEC, was still formed between the CTD (which lacks the helicase domains) and the TEC, and was not formed when a truncated form of PcrA lacking the CTD was used. These observations suggest that the increased stability associated with the PcrA:TEC complex is not simply due to the helicase being able to engage with DNA/RNA that flanks the polymerase. Instead, the RNA polymerase might adopt a conformation in the TEC that favours the recruitment of PcrA via the CTD.

HDX-MS analysis confirmed that RNAP interacted with a face of the PcrA Tudor fold that is frequently observed to form protein interactions in other systems (Musselman et al., 2012; Ruthenburg et al., 2007). The site for this PcrA-CTD interaction on the RNA polymerase was identified as a short peptide motif in the β subunit that sits within the SI1 domain. This is the first protein-protein interaction reported for the enigmatic SI1 domain, which is a lineage-specific insert in the β subunit whose function remains mysterious (Newing et al., 2020). This interaction site is distinct from that bound by the TC-NER factor Mfd, which also engages with the beta subunit via a Tudor domain but at the β1 domain. In agreement with this and other data (Kawale and Burmann, 2020), we also found that PcrA and Mfd did not compete for RNAP in a physical interaction assay.

The short peptide motif to which PcrA binds is highly conserved within *Bacillus*-like SI1 domains but does not obviously align with β subunit sequences from more distantly-related organisms. Indeed, because PcrA/UvrD is ubiquitous and its CTD so highly-conserved, we were initially surprised to find that the interaction with RNAP mapped to a highly variable

domain. This paradox was resolved using bioinformatics, which revealed that the same motif is present widely in bacterial RNAPs, but that its positioning within the SI1 domain is different depending on the lineage. Relative to other more universal features in RpoB sequences, the interaction motif appears later in *B. subtilis*-like domains (which are representative of Firmicutes) than it does in *E. coli*-like domains (which are representative of Proteobacteria; see **Figure S3**). Note however that we did find examples of bacterial RpoB species for which we could find no clear helicase-interaction motif present in the SI1 domain. Our sequence searches proved rewarding as they revealed the presence of the same motif in other known PcrA-interacting proteins, including UvrB, YxaL and YwhK. Moreover, this helicase interaction motif is also present in several poorly-characterised proteins which might be previously unidentified interaction partners for PcrA. Where a structure is available, the motif always adopts a beta-hairpin with the highly conserved TGE triad protruding towards the solvent. The corollary of our observation that RNAP and UvrB employ the same structural element to recruit PcrA is that they might compete in the cell for the PcrA/UvrD helicase. This idea is supported directly by competition pulldown assays, and is consistent with previous work in which the PcrA/UvrD CTD was shown to be important for interaction with both RNAP and UvrB (Manelyte et al., 2009; Sanders et al., 2017). However, the helicase interaction motif was absent from some known PcrA partners including the mismatch repair factor MutL. This is in agreement with recent work which showed that MutL binds and activates PcrA/UvrD via the 2B domain, rather than the CTD (Ordabayev et al., 2018, 2019).

PcrA/UvrD, along with ppGpp, has been suggested to participate in an Mfd-independent transcription-coupled DNA repair pathway (Epshtein et al., 2014; Pani and Nudler, 2017). In this pathway, UvrD is thought to promote backtracking of RNAP stalled at bulky DNA lesions and then recruits NER factors including UvrB to repair the damage. We find here that RNAP and UvrB compete for PcrA interaction using the same conserved motif, meaning it is unclear how UvrB would be recruited to the UvrD:RNAP complex. Moreover, the role of UvrD in promoting strand-specific repair of DNA lesions remains controversial because a genome-wide analysis of cyclobutane pyrimidine dimers showed no evidence for Mfd-independent TC-NER and also that the levels of ppGpp do not change the repair rate in *E. coli* (Adebali et al., 2017a, 2017b). Our data lead us to propose an alternative idea for why PcrA/UvrD interacts with RNA polymerase.

The HDX-MS data with full length PcrA showed that the core helicase domains interact with multiple elements surrounding the RNA and DNA exit channels, for instance the β-flap tip and

the N-terminal part of the β′ subunit, as has been shown previously by crosslinking-MS data for the complex between *E. coli* RNAP and UvrD (Epshtein et al., 2014). Our HDX-MS data also identified an interaction with the SI1 domain which we can assign to the PcrA CTD. The exit channels for nucleic acids are distant from the SI1 domain, but the long linker between the helicase core and the CTD is sufficiently long to accommodate both interaction patches. Interestingly, the length of this linker varies in different PcrA orthologues in a manner that reflects the distance between the RNAP exit channels and SI1 domains in each system. The precise details of the PcrA-RNAP interaction will await a high-resolution structure of the complex, but it is possible to build a speculative physical model which satisfies both our own HDX data and the crosslinking study of the *E. coli* system. (**Figure S12**)

The involvement of the main body of the helicase in RNAP interactions may also help to explain the phenotypes associated with CTD deletion. Given the apparently key role played by the PcrA/UvrD CTD as a protein interaction hub, it is surprising that its complete removal has little effect on nucleotide excision repair or replication-transcription conflicts in *E. coli*. Similarly, in *B. subtilis* where PcrA is recruited to highly transcribed regions of the genome to facilitate replication through active transcription units, the CTD seems to be dispensable (Merrikh et al., 2015). A possible explanation lies in the fact that PcrA-RNAP association also involves the main body of the helicase including the core motor domain 2A, whose activity is largely unchanged by removal of the CTD. Perhaps these residual interactions allow PcrA to function effectively *in vivo*, albeit with a weaker affinity for its partner proteins.

The location of the helicase core close to the RNA exit channel led us to consider roles for PcrA in manipulating the transcript. Due to the enzyme's inability to translocate on RNA or to unwind duplex RNAs, we hypothesised that PcrA might act to unwind DNA/RNA hybrids that are formed upstream of the transcription bubble. This idea was supported by experiments in *Bacillus subtilis* showing that R-loop concentration in cells was increased by 2-3 fold by overexpressing dominant negative PcrA or free CTD (which blocks PcrA-RNAP interaction). This strongly suggests that an activity of wild type PcrA, which is dependent upon interactions made by its CTD, is important for suppressing R-loops. Based on our observations, we propose that PcrA/UvrD acts as an R-loop surveillance factor, unwinding RNA-DNA hybrids that form behind the transcription elongation complex (**Figure 7**). The well-documented 3′>5′ polarity of PcrA/UvrD helicase (Bird et al., 1998), together with its inability to move along RNA strands, suggests two possible models for the unwinding of R-loops. In the first "co-directional" model, the helicase translocates on the template strand and in the same direction as the RNA

polymerase, unwinding R-loops as they form. Another possibility (the "backtracking model") is that the helicase translocates on the non-template strand of the R-loop and in the opposite direction to transcription, indirectly unwinding the DNA/RNA hybrid by making the polymerase backtrack (Epshtein et al., 2014). In this scenario the lid domain of RNAP, rather than the helicase itself, would act as the "ploughshare" to separate the two strands. We favour the first model because our HDX protection data places the 2A domain of PcrA (which is at the front of the moving helicase) in direct contact with the RNA exit region, suggesting that the helicase translocates towards the RNA polymerase. However, we note also that previous observations of UvrD-dependent backtracking of RNAP are more consistent with the second model (Epshtein et al., 2014; Sanders et al., 2017).

A role for PcrA in alleviating R-loops can explain why the helicase is enriched at highly transcribed rRNA and tRNA genes that have been shown to be prone to R-loop formation in yeast and human cells (Chan et al., 2014; Chen et al., 2017). Indeed, recent work in *E. coli* has shown that *uvrD* is synthetically lethal with *rnhA*, in agreement with a role for this helicase in unwinding R-loops that may be distinctive from those dealt with by other R-loop suppression factors (Wolak et al., 2020). An R-loop suppression function might also explain why loss of PcrA leads to hyper-recombination and the formation of RecFOR- and RecA-dependent toxic recombination intermediates (Arthur and Lloyd, 1980; Moreno-del Alamo et al., 2020; Petit and Ehrlich, 2002; Veaute et al., 2005). The displaced ssDNA within an R-loop may act as a "frustrated" substrate for ssDNA gap repair, resulting in RecA recruitment and strand exchange but failing to resolve correctly into repaired products because of the presence of the R-loop and TEC. These ideas will be the subject of future work.

Acknowledgements

This work was funded by the European Union via the DNAREPAIRMAN innovative training network (Urrutia-Irazabal, Savery and Dillingham) and by a BBSRC ALERT grant (Ault and Sobott; BB/M012573/1). We are grateful to Harry Thompson for technical assistance with CD measurements, to Peter McGlynn, Heath Murray, Marie-Agnès Petit, Houra Merrikh and Christopher Merrikh for advice and the sharing of plasmids and strains used in this work, and to Peter Lewis for generously sharing the co-ordinates of the *B. subtilis* transcription elongation complex ahead of public release.

References

Adebali, O., Chiou, Y.-Y., Hu, J., Sancar, A., and Selby, C.P. (2017a). Genome-wide transcription-coupled repair in Escherichia coli is mediated by the Mfd translocase. Proc. Natl. Acad. Sci. *114*, E2116–E2125.

Adebali, O., Sancar, A., and Selby, C.P. (2017b). Mfd translocase is necessary and sufficient for transcription-coupled repair in Escherichia coli. J. Biol. Chem. jbc.C117.818807.

Aguilera, A., and García-Muse, T. (2012). R Loops: From Transcription Byproducts to Threats to Genome Stability. Mol. Cell *46*, 115–124.

Anagnostopoulos, C., and Spizizen, J. (1961). Requirements for Transformation in Bacillus Subtilis. J. Bacteriol. *81*, 741–746.

Arthur, H.M., and Lloyd, R.G. (1980). Hyper-recombination in uvrD mutants of Escherichia coli K-12. MGG Mol. Gen. Genet. *180*, 185–191.

Bird, L.E., Brannigan, J.A., Subramanya, H.S., and Wigley, D.B. (1998). Characterisation of Bacillus stearothermophilus PcrA helicase: Evidence against an active rolling mechanism. Nucleic Acids Res. *26*, 2686–2693.

Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L., and Carrico, R.J. (1986). Characterization of monoclonal antibody to DNA · RNA and its application to immunodetection of hybrids. J. Immunol. Methods *89*, 123–130.

Brosh, R.M., and Matson, S.W. (1995). Mutations in motif II of Escherichia coli DNA helicase II render the enzyme nonfunctional in both mismatch repair and excision repair with differential effects on the unwinding reaction. J. Bacteriol. *177*, 5612–5621.

Bruand, C., and Ehrlich, S.D. (2000). UvrD-dependent replication of rolling-circle plasmids in Escherichia coli. Mol. Microbiol. *35*, 204–210.

Burby, P.E., and Simmons, L.A. (2017). MutS2 promotes homologous recombination in Bacillus subtilis. J. Bacteriol. *199*, 1–13.

Chambers, A.L., Smith, A.J., and Savery, N.J. (2003). A DNA translocation motif in the bacterial transcription-repair coupling factor, Mfd. Nucleic Acids Res. *31*, 6409–6418.

Chan, Y.A., Aristizabal, M.J., Lu, P.Y.T., Luo, Z., Hamza, A., Kobor, M.S., Stirling, P.C., and Hieter, P. (2014). Genome-Wide Profiling of Yeast DNA:RNA Hybrid Prone Sites with

DRIP-Chip. PLoS Genet. *10*.

Chen, C., Natale, D.A., Finn, R.D., Huang, H., Zhang, J., Wu, C.H., and Mazumder, R. (2011). Representative Proteomes: A Stable, Scalable and Unbiased proteome set for sequence analysis and functional annotation. PLoS One *6*, 1–9.

Chen, L., Chen, J.Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., et al. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Mol. Cell *68*, 745-757.e5.

Cornwell, O., Radford, S.E., Ashcroft, A.E., and Ault, J.R. (2018). Comparing Hydrogen Deuterium Exchange and Fast Photochemical Oxidation of Proteins: a Structural Characterisation of Wild-Type and ΔN6 β2-Microglobulin. J. Am. Soc. Mass Spectrom. *29*, 2413–2426.

Crooks, G., Hon, G., Chandonia, J., and Brenner, S. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188–1190.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019). R-Loops as Cellular Regulators and Genomic Threats. Mol. Cell *73*, 398–411.

Deaconescu, A.M., Chambers, A.L., Smith, A.J., Nickels, B.E., Hochschild, A., Savery, N.J., and Darst, S.A. (2006). Structural basis for bacterial transcription-coupled DNA repair. Cell *124*, 507–520.

Delumeau, O., Lecointe, F., Muntel, J., Guillot, A., Guédon, E., Monnet, V., Hecker, M., Becher, D., Polard, P., and Noirot, P. (2011). The dynamic protein partnership of RNA polymerase in Bacillus subtilis. Proteomics *11*, 2992–3001.

Dillingham, M.S. (2011). Superfamily I helicases as modular components of DNA-processing machines. Biochem. Soc. Trans. *39*, 413 LP – 423.

Epshtein, V., Kamarthapu, V., McGary, K., Svetlov, V., Ueberheide, B., Proshkin, S., Mironov, A., and Nudler, E. (2014). UvrD facilitates DNA repair by pulling RNA polymerase backwards. Nature *505*, 372–377.

Gilhooly, N.S., Gwynn, E.J., and Dillingham, M.S. (2013). Superfamily 1 helicases . Front. Biosci. (Schol. Ed). *5*, 206–216.

Gorbalenya, A.E., and Koonin, E. V. (1993). Helicases: amino acid sequence comparisons

and structure-function relationships. Curr. Opin. Struct. Biol. *3*, 419–429.

Guy, C.P., Atkinson, J., Gupta, M.K., Mahdi, A.A., Gwynn, E.J., Rudolph, C.J., Moon, P.B., van Knippenberg, I.C., Cadman, C.J., Dillingham, M.S., et al. (2009). Rep Provides a Second Motor at the Replisome to Promote Duplication of Protein-Bound DNA. Mol. Cell *36*, 654–666.

Gwynn, E.J., Smith, A.J., Guy, C.P., Savery, N.J., McGlynn, P., and Dillingham, M.S. (2013). The Conserved C-Terminus of the PcrA/UvrD Helicase Interacts Directly with RNA Polymerase. PLoS One *8*, e78141.

Hall, M.C., Jordan, J.R., and Matson, S.W. (1998). Evidence for a physical interaction between the Escherichia coli methyl-directed mismatch repair proteins MutL and UvrD. EMBO J. *17*, 1535–1541.

Harwood, C., and Cutting, S. (1990). Molecular Biological Methods for Bacillus (New York, NY: Wiley).

Hawkins, M., Dimude, J.U., Howard, J.A.L., Smith, A.J., Dillingham, M.S., Savery, N.J., Rudolph, C.J., and McGlynn, P. (2019). Direct removal of RNA polymerase barriers to replication by accessory replicative helicases. Nucleic Acids Res. 1–14.

Hirano, M., and Hirano, T. (2004). Positive and negative regulation of SMC-DNA interactions by ATP and accessory proteins. EMBO J. *23*, 2664–2673.

Husain, I., Houten, B.V.A.N., Thomas, D.C., and Abdel-monemt, M. (1985). Effect of DNA polymerase I and DNA helicase II on the turnover rate of UvrABC excision nuclease. Proc Natl Acad Sci U S A *82*, 6774–6778.

Kamarthapu, V., and Nudler, E. (2015). Rethinking transcription coupled DNA repair. Curr. Opin. Microbiol. *24*, 15–20.

Kawale, A.A., and Burmann, B.M. (2020). UvrD helicase–RNA polymerase interactions are governed by UvrD's carboxy-terminal Tudor domain. Commun. Biol. *3*, 607.

Kiianitsa, K., Solinger, J.A., and Heyer, W.D. (2003). NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. Anal. Biochem. *321*, 266–271.

Kisker, C., Kuper, J., and Houten, B. Van (2015). Prokaryotic Nucleotide Excision Repair.

Cold Spring Harb. Perspect. Biol. *5*, 1–18.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839–843.

Lahue, R.S., Au, K.G., and Modrich, P. (1989). DNA mismatch correction in a defined system. Science (80-.). *245*, 160–164.

Lane, W.J., and Darst, S.A. (2010). Molecular Evolution of Multisubunit RNA Polymerases: Structural Analysis. J. Mol. Biol. *395*, 686–704.

Lee, J.Y., and Yang, W. (2006). UvrD Helicase Unwinds DNA One Base Pair at a Time by a Two-Part Power Stroke. Cell *127*, 1349–1360.

Liu, B., Zuo, Y., and Steitz, T.A. (2015). Structural basis for transcription reactivation by RapA. Proc. Natl. Acad. Sci. *112*, 2006–2010.

Machón, C., Lynch, G.P., Thomson, N.H., Scott, D.J., Thomas, C.D., and Soultanas, P. (2009). RepD-mediated recruitment of PcrA helicase at the Staphylococcus aureus pC221 plasmid replication origin, oriD. Nucleic Acids Res. *38*, 1874–1888.

Manelyte, L., Guy, C.P., Smith, R.M., Dillingham, M.S., McGlynn, P., and Savery, N.J. (2009). The unstructured C-terminal extension of UvrD interacts with UvrB, but is dispensable for nucleotide excision repair. DNA Repair (Amst). *8*, 1300–1310.

Matson, S.W. (1989). Escherichia coli DNA helicase II (uvrD gene product) catalyzes the unwinding of DNA.RNA hybrids in vitro. Proc. Natl. Acad. Sci. *86*, 4430–4434.

Merrikh, C.N., Brewer, B.J., and Merrikh, H. (2015). The B. subtilis Accessory Helicase PcrA Facilitates DNA Replication through Transcription Units. PLoS Genet. *11*, 1–25.

Moreno-del Alamo, M., Torres, R., Manfredi, C., Ruiz-Masó, J.A., del Solar, G., and Alonso, J.C. (2020). Bacillus subtilis PcrA Couples DNA Replication, Transcription, Recombination and Segregation. Front. Mol. Biosci. *7*, 1–16.

Musselman, C.A., Lalonde, M.E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. Nat. Struct. Mol. Biol. *19*, 1218–1227.

Newing, T.P., Oakley, A.J., Miller, M., Dawson, C.J., Brown, S.H.J., Bouwer, J.C., Tolun, G., and Lewis, P.J. (2020). Molecular basis for RNA polymerase-dependent transcription

complex recycling by the helicase-like motor protein HelD. Nat. Commun. *11*, 6420.

Noirot-Gros, M.F., Soultanas, P., Wigley, D., Ehrlich, S., Noirot, P., and Petit, M.A. (2002). The beta-propeller protein YxaL increases the processivity of the PcrA helicase. Mol. Genet. Genomics *267*, 391–400.

Ordabayev, Y.A., Nguyen, B., Niedziela-Majka, A., and Lohman, T.M. (2018). Regulation of UvrD Helicase Activity by MutL. J. Mol. Biol.

Ordabayev, Y.A., Nguyen, B., Kozlov, A.G., Jia, H., and Lohman, T.M. (2019). UvrD helicase activation by MutL involves rotation of its 2B subdomain. Proc. Natl. Acad. Sci. U. S. A. *116*, 16320–16325.

Pakotiprapha, D., Samuels, M., Shen, K., Hu, J.H., and Jeruzalmi, D. (2012). Structure and mechanism of the UvrA-UvrB DNA damage sensor. Nat. Struct. Mol. Biol. *19*, 291–298.

Pani, B., and Nudler, E. (2017). Mechanistic Insights into Transcription Coupled DNA Repair. DNA Repair (Amst). *56*, 42–50.

Pei, H., Hilal, T., Chen, Z.A., Huang, Y., Gao, Y., Said, N., Loll, B., Rappsilber, J., Belogurov, G.A., Artsimovitch, I., et al. (2020). The δ subunit and NTPase HelD institute a two-pronged mechanism for RNA polymerase recycling. Nat. Commun. *11*, 6418.

Petit, M.A., and Ehrlich, D. (2002). Essential bacterial helicases that counteract the toxicity of recombination proteins. EMBO J. *21*, 3137–3147.

Petit, M., Dervyn, E., Rose, M., Entian, K., McGovern, S., Ehrlich, S.D., and Bruand, C. (1998). PcrA is an essential DNA helicase of Bacillus subtilis fulfilling functions both in repair and rolling-circle replication. Mol. Microbiol. *29*, 261–273.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Qi, Y., and Hulett, F.M. (1998). PhoP~P and RNA polymerase σA holoenzyme are sufficient for transcription of Pho regulon promoters in Bacillus subtilis: PhoP~P activator sites within the coding region stimulate transcription in vitro. Mol. Microbiol. *28*, 1187–1197.

Raghunathan, N., Kapshikar, R.M., Leela, J.K., Mallikarjun, J., Bouloc, P., and Gowrishankar, J. (2018). Genome-wide relationship between R-loop formation and antisense

transcription in Escherichia coli. Nucleic Acids Res. *46*, 3400–3411.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, 320–324.

Ruthenburg, A.J., Li, H., Patel, D.J., and David Allis, C. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nat. Rev. Mol. Cell Biol. *8*, 983–994.

Sanders, K., Lin, C.-L., Smith, A.J., Cronin, N., Fisher, G., Eftychidis, V., McGlynn, P., Savery, N.J., Wigley, D.B., and Dillingham, M.S. (2017). The structure and function of an RNA polymerase interaction domain in the PcrA/UvrD helicase. Nucleic Acids Res. *45*, 3875–3887.

Sidorenkov, I., Komissarova, N., and Kashlev, M. (1998). Crucial role of the RNA:DNA hybrid in the processivity of transcription. Mol. Cell *2*, 55–64.

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.

Soultanas, P., Dillingham, M.S., Velankar, S.S., and Wigley, D.B. (1999). DNA binding mediates conformational changes and metal ion coordination in the active site of PcrA helicase. J. Mol. Biol. *290*, 137–148.

Thomsen, N.D., and Berger, J.M. (2008). Structural frameworks for considering microbial protein- and nucleic acid-dependent motor ATPases: MicroReview. Mol. Microbiol. *69*, 1071–1090.

Truglio, J.J., Croteau, D.L., Skorvaga, M., DellaVecchia, M.J., Theis, K., Mandavilli, B.S., Van Houten, B., and Kisker, C. (2004). Interactions between UvrA and UvrB: The role of UvrB's domain 2 in nucleotide excision repair. EMBO J. *23*, 2498–2509.

Veaute, X., Delmas, S., Selva, M., Jeusset, J., Le Cam, E., Matic, I., Fabre, F., and Petit, M.A. (2005). UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in Escherichia coli. EMBO J. *24*, 180–189.

Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S., and Wigley, D.B. (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell *97*, 75–84.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T.,

De Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res. *46*, W296–W303.

Webster, M.P.J., Jukes, R., Zamfir, V.S., Kay, C.W.M., Bagnéris, C., and Barrett, T. (2012). Crystal structure of the UvrB dimer: Insights into the nature and functioning of the UvrAB damage engagement and UvrB-DNA complexes. Nucleic Acids Res. *40*, 8743–8758.

Westblade, L.F., Campbell, E.A., Pukhrambam, C., Padovan, J.C., Nickels, B.E., Lamour, V., and Darst, S.A. (2010). Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. Nucleic Acids Res. *38*, 8357–8369.

Wimberly, H., Shee, C., Thornton, P.C., Sivaramakrishnan, P., Rosenberg, S.M., and Hastings, P.J. (2013). R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. Nat. Commun. *4*, 1–10.

Wolak, C., Ma, H.J., Soubry, N., Sandler, S.J., Reyes-lamothe, R., and Keck, J.L. (2020). Interaction with single-stranded DNA-binding protein localizes ribonuclease HI to DNA replication forks and facilitates R-loop removal. Mol. Microbiol. 0–3.

Zieg, J., Maples, V.F., and Kushner, S.R. (1978). Recombination levels of Escherichia coli K-12 mutants deficient in various replication, recombination, or repair genes. J. Bacteriol. *134*, 958–966.

Van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P.L., Karaca, E., Melquiond, A.S.J., Van Dijk, M., De Vries, S.J., and Bonvin, A.M.J.J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J. Mol. Biol. *428*, 720–725.

Figure 1

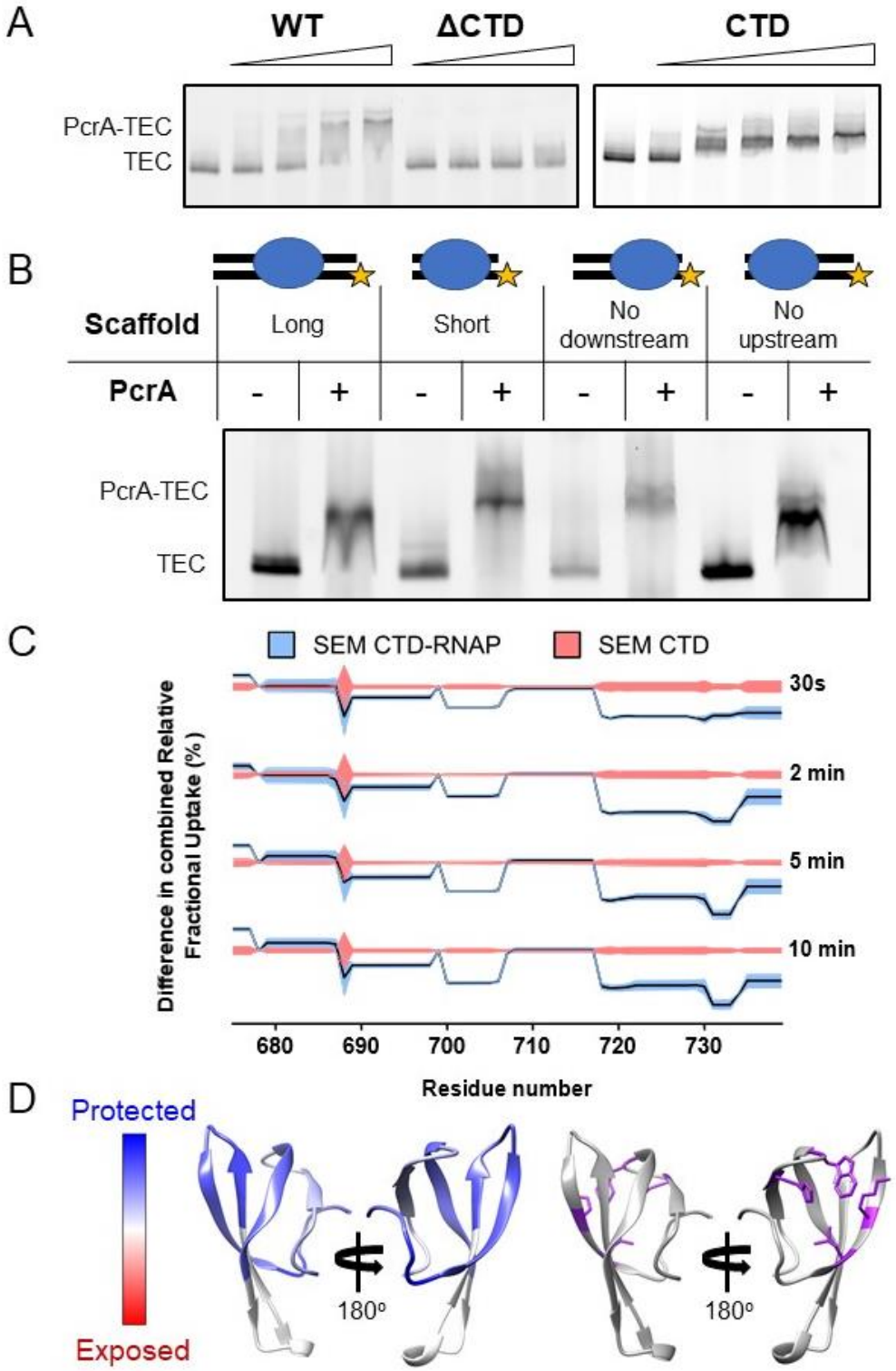

Figure 1. Interactions between PcrA and a transcription elongation complex are mediated by protein-protein interactions involving the PcrA-CTD (A) EMSA supershift assays to monitor association of PcrA with a TEC. The PcrA-CTD is necessary and sufficient for stable formation of the PcrA-RNAP complex. WT PcrA and ΔCTD PcrA were titrated from 0.25μM

to 1.5 µM. The PcrA-CTD was titrated from 0.5 µM to 3 µM. (B) EMSA supershift assay showing that binding of PcrA is not dependent on the presence of upstream or downstream DNA in the TEC. The star indicates the position of the fluorescent label at the 5′ end of the template strand of the scaffold. PcrA was used at 1 µM. The oligonucleotides used to assemble the scaffolds are shown in Table S5. (C) Relative HDX measured for the PcrA-CTD in the CTD-RNAP complex compared to CTD alone. The black line shows the differential relative uptake and the pink and blue shadowing, the SEM of the CTD and CTD-RNAP conditions, respectively. The four offset traces show different exchange times. Negative uptake values over the CTD baseline with non-overlapping shadowing show protected amino acid regions on the CTD when it is in complex with RNAP. Note the key regions of the CTD (aa ~690-705 and aa ~720-Ct; shaded yellow) are protected upon binding RNAP. (D) Left - homology model of the PcrA-CTD showing regions protected from HDX in the complex with RNAP (dark blue). Right - amino acids known to be important for interaction with RNAP (purple residues).

Figure 2

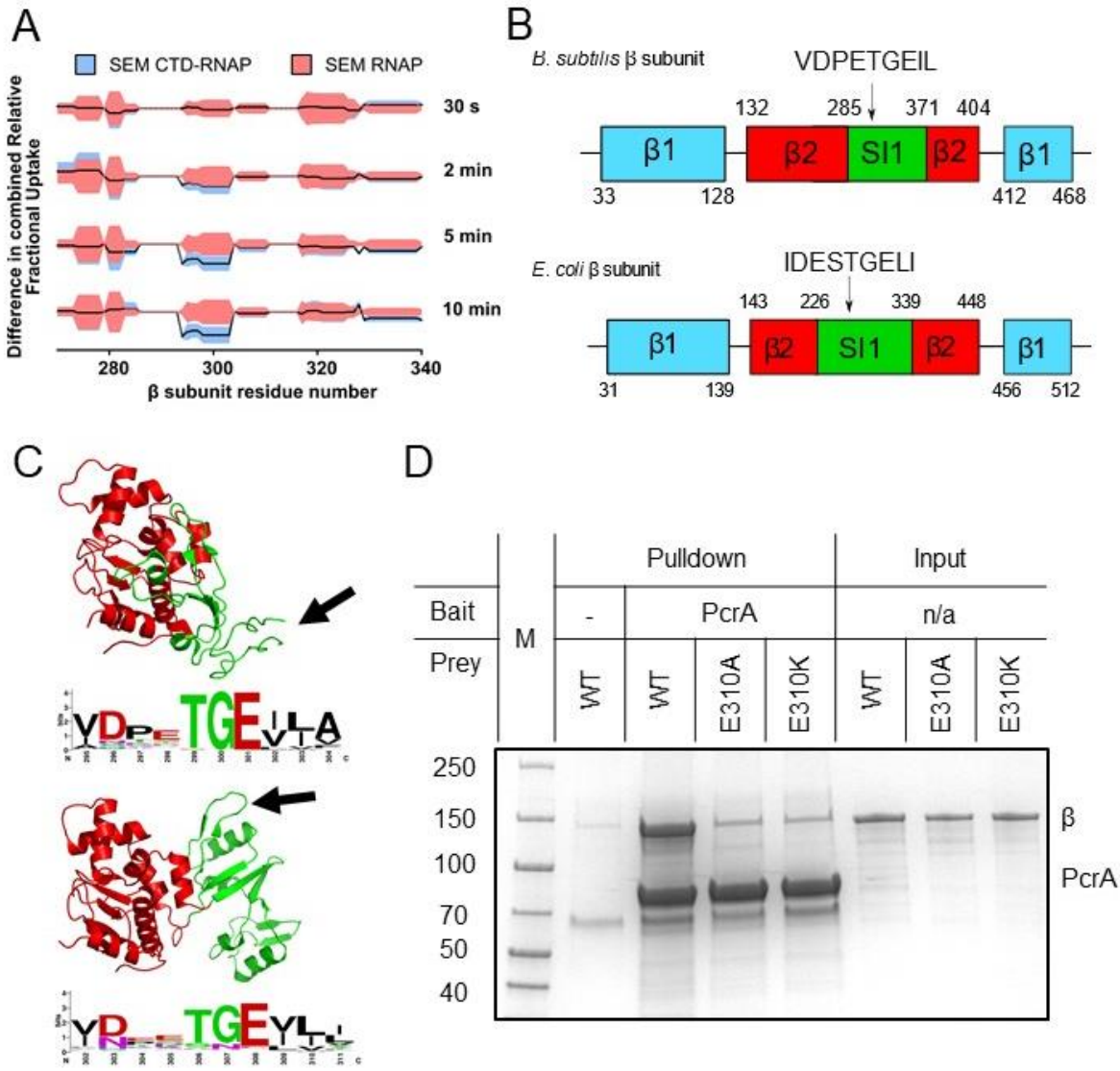

Figure 2. The PcrA-CTD binds to a conserved motif in the SI1 domain of RNAP (A) Relative HDX measured for a region of the RNAP β subunit (residue numbers on x axis) within the CTD-RNAP complex (blue) compared to RNAP alone (red). A small region of RpoB (at amino-acid positions around ~300) becomes significantly protected by interaction with the PcrA CTD as the exchange time becomes longer. (B) The protected region maps to a conserved motif in the SI1 domain of *B. subtilis* RpoB. This region is organised differently in *E. coli* RpoB, but the same conserved amino acid motif appears in a slightly different position in the structure (black arrow). (C) Structure of the *B.subtilis* (upper panel) and *E. coli* (lower panel) β2 (red) - SI1(green) domains indicating the beta-loop structure containing a putative interaction motif at the tip (black arrows). This sequence is well-conserved in bacterial RNA polymerases and the consensus sequence is shown in weblogo format beneath each structure.

(D) *In vitro* pulldown of RpoB using PcrA as a bait (see Methods for details). Mutation of the conserved glutamate (E301) in the putative helicase interaction motif dramatically reduces RpoB pulldown.

Figure 3

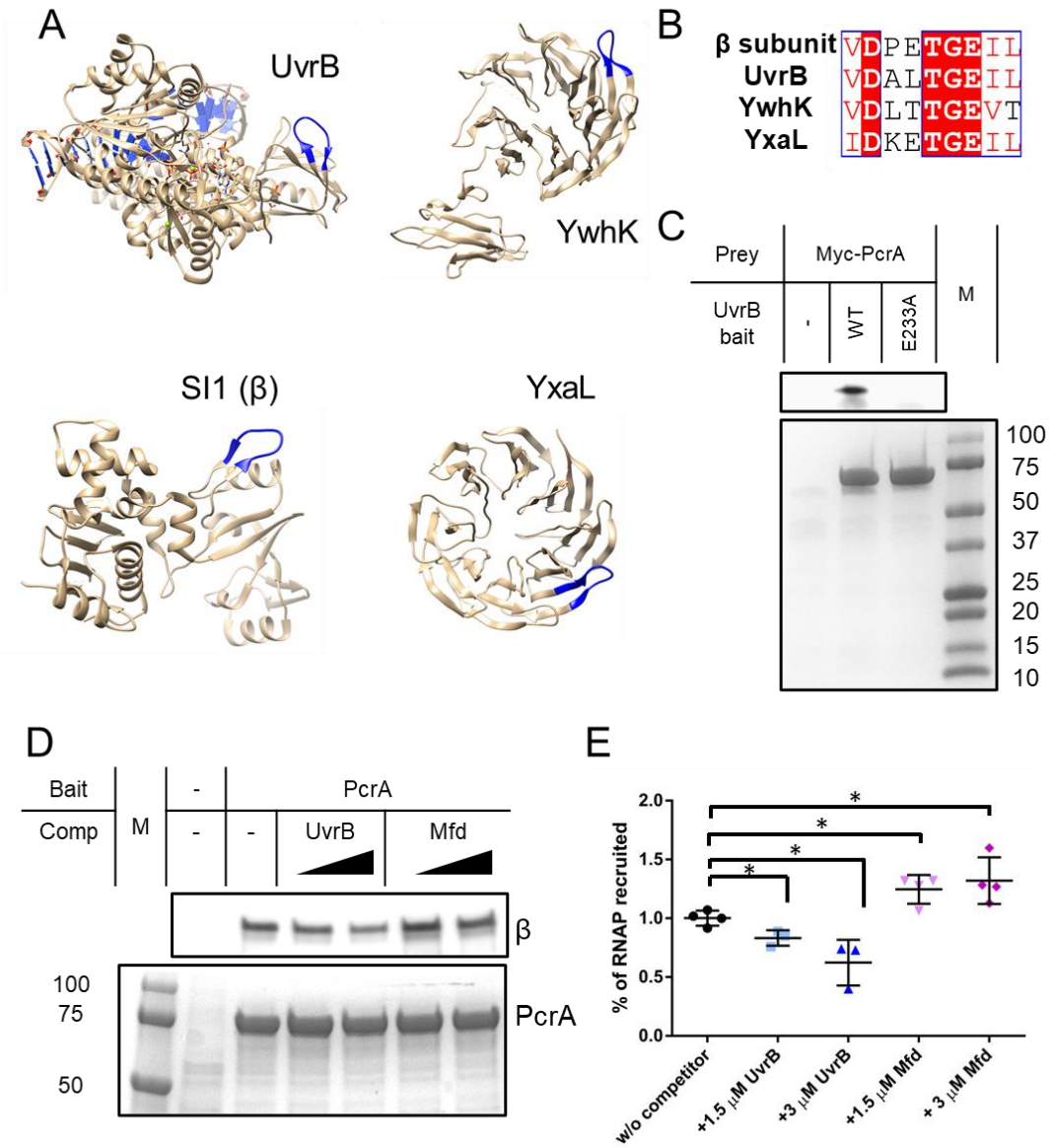

Figure 3. Many PcrA partner proteins contain the helicase interaction motif (A) Putative helicase interaction motif (blue) in known PcrA interaction partners from *B. subtilis*. A beta hairpin structure (blue) is formed by the interaction motif in each of the proteins. Two of the structures are homology models as indicated. (B) Sequences of putative helicase interaction motifs in four known PcrA partner proteins. (C) Pulldown of Myc-tagged PcrA from *B. subtilis* cell extracts using UvrB as bait (for details see the Methods). Mutation of the conserved glutamate (E233A) in the putative helicase interaction motif dramatically reduces PcrA pulldown. PcrA was detected using an anti-Myc antibody (upper gel). Equivalent loading of WT and mutant UvrB was confirmed by Coomassie staining (lower gel). (D) and (E) Pulldown of RNAP from *B. subtilis* cell extracts using biotinylated PcrA as bait. Where

indicated the prey was supplemented with purified UvrB or Mfd. Free UvrB, but not Mfd, competes for the interaction site formed between PcrA and RNA polymerase. Error bars show the SEM of three independent experiments. Two-tailed Student's t test determined statistical significance (*p value < 0.05).

Figure 4

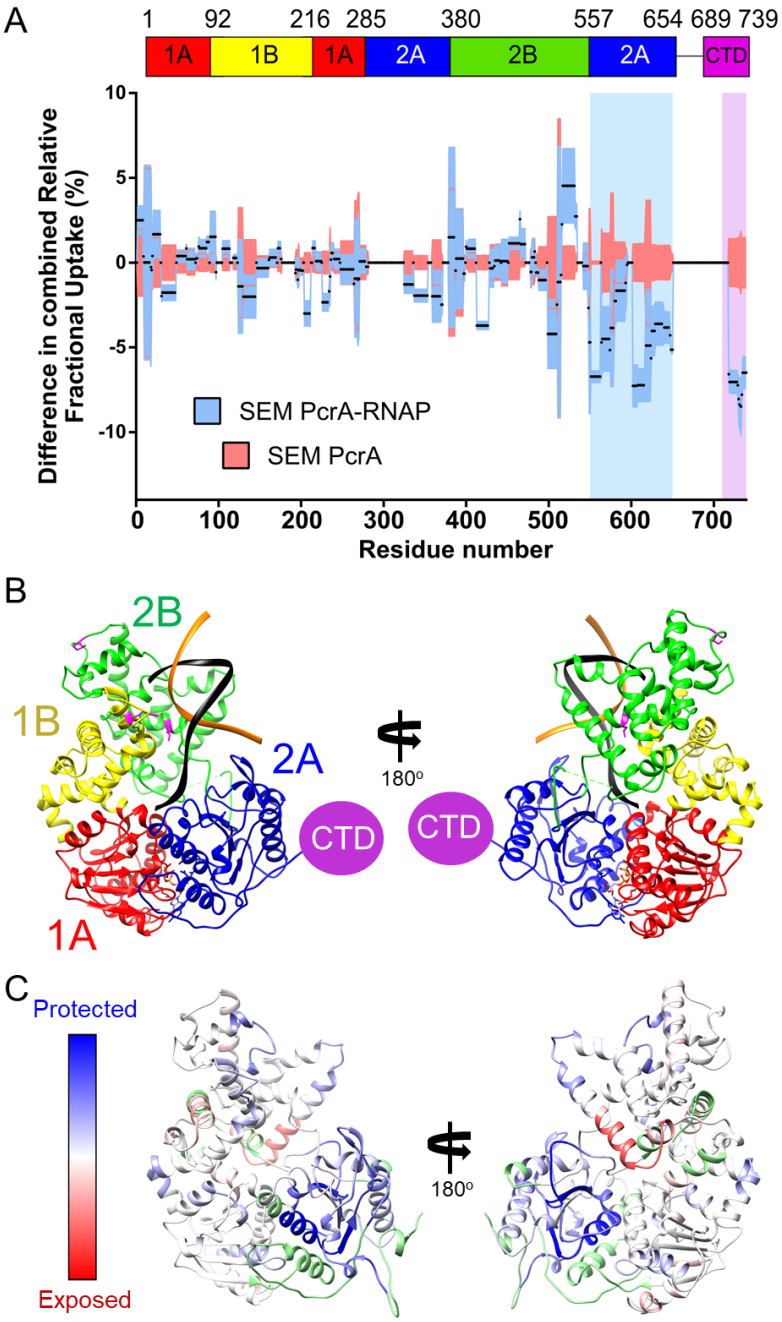

Figure 4. The PcrA helicase core is protected by interaction with RNA polymerase (A) Relative HDX measured for full length PcrA (residue numbers on x axis) in the PcrA-RNAP complex (blue) compared to PcrA alone (red). The magenta rectangle highlights strong protection of the CTD afforded by interaction with RNAP as expected based on results presented earlier in this manuscript. The blue rectangle highlights a second region of strong protection within the 2A domain of PcrA. (B) The PcrA helicase core (homology model from PDB: 3PJR) showing domain organisation. The DNA substrate is shown in black and orange

and the CTD (which is disordered in this structure) is indicated as a purple circle. (C) The same structure showing the mapping of the HDX-protection data (bottom; blue indicates protection in the complex, red indicates exposure and green indicates a lack of data). Note that the HDX-protection data maps largely to one face of the helicase within domains 2A and, to a lesser extent, 2B.

Figure 5

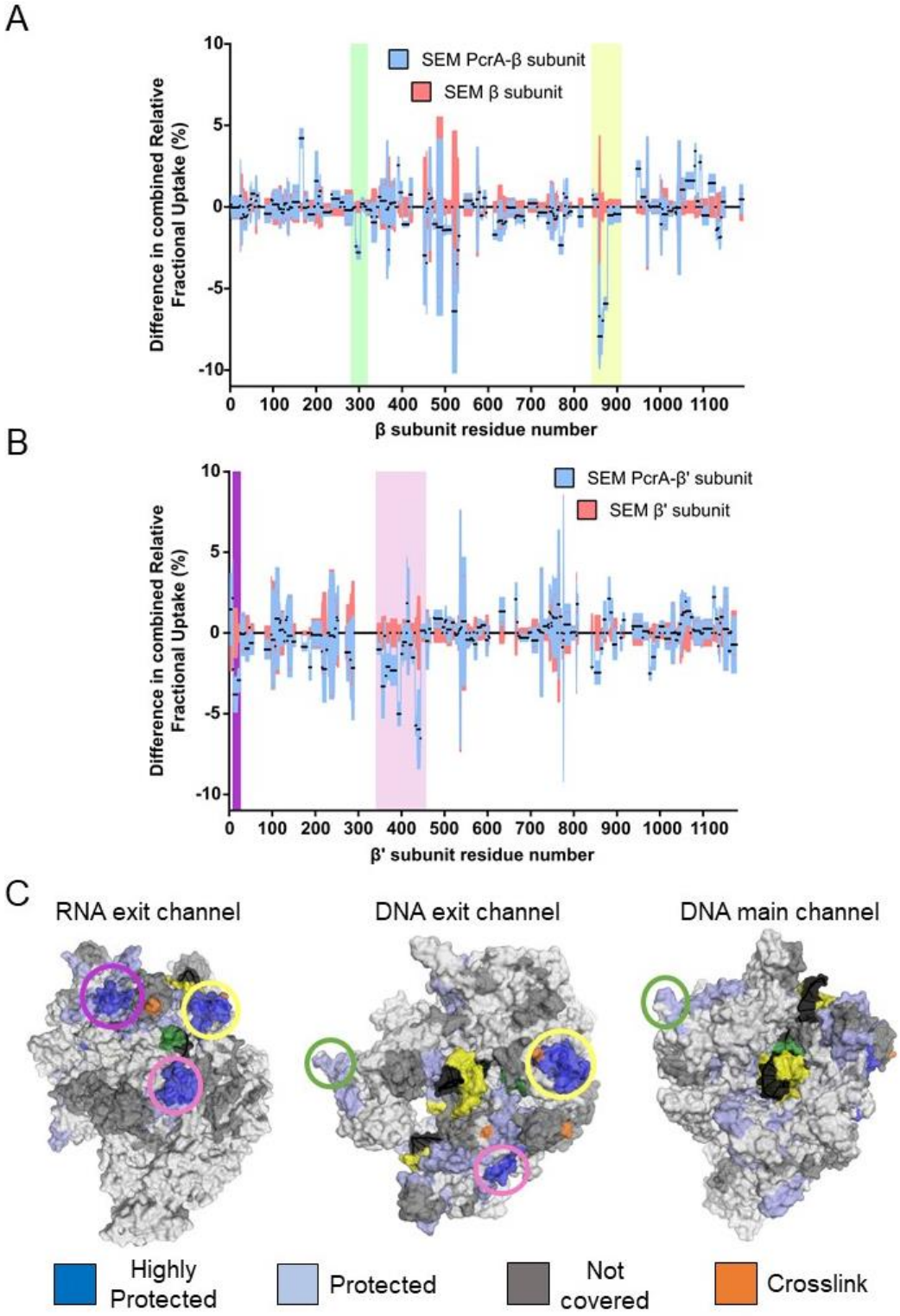

Figure 5. The PcrA helicase core binds to RNAP close to the RNA and DNA exit channel (A) Differential protection plot for the β subunit of RNAP showing the differential relative uptake at 10 minutes of exposure to deuterium for a PcrA-RNAP complex (blue)

compared to RNAP alone (red). Green and yellow rectangles highlight regions of RpoB that are significantly protected in the complex state. (B) Differential protection plot for the β′ subunit of RNAP showing the differential relative uptake at 10 minutes of exposure to deuterium for a PcrA-RNAP complex compared to RNAP alone. Purple and pink rectangles highlight regions of RpoC that are significantly protected in the complex state. (C) Three views of the *B. subtilis* TEC (PDB: 6WVJ) showing regions protected by interaction with PcrA. DNA is shown in yellow and black. RNA is green. Coloured circles correspond to the protected regions highlighted in panels A and B. The green circle highlights the helicase interaction motif already identified as the site of binding of the PcrA-CTD. Note that the other major protection sites surround the DNA/RNA exit channels. The crosslinking sites (orange) are from reference (Epshtein et al., 2014).

Figure 6

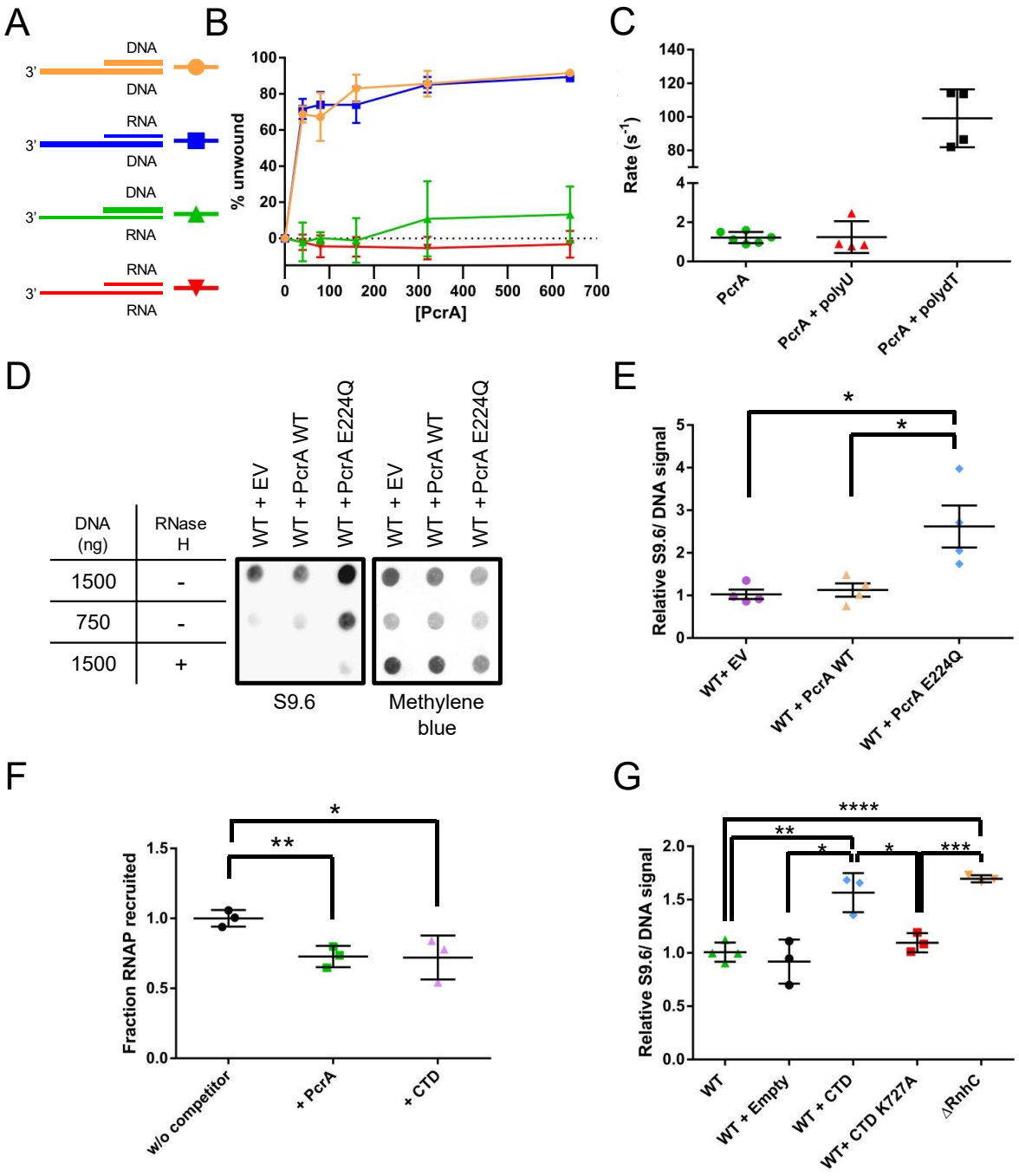

Figure 6. PcrA unwinds DNA-RNA hybrids *in vitro* and supresses R-loops *in vivo* (A) DNA and RNA substrates used for helicase assays. Thick lines represent DNA strands and thin lines, RNA strands. The oligonucleotides used to form these substrates are shown in Supplementary Figure 6. (B) Quantification of unwinding as a function of PcrA concentration for the 3′-tailed substrates shown in panel A. The substrate is only efficiently unwound if the longer of the two nucleic acids strands is DNA. Error bars show the standard deviation of at

least 3 independent experiments. (C) The ATPase activity of PcrA is strongly stimulated by single-stranded DNA but not single-stranded RNA. Error bars show the standard deviation of at least 3 independent experiments. (D) Anti R-loop antibody (S9.6) dot blot for nucleic acid samples purified from three strains of *B. subtilis*. These strains contain an integrated expression cassette for either wild type PcrA or a dominant negative form of PcrA (E224Q). The control strain (EV) contains an integrated but empty expression cassette. The S9.6 signal is normalized using methylene blue as a stain for all DNA. Note the high S9.6 signal for the strain expressing PcrA E224Q. (E) Quantification of four independent repeats of the experiment shown in (c). Error bars show the SEM. Expression of a dominant negative form of PcrA increases R-loop content (relative to DNA) in *B. subtilis* by ~2.5 fold. (F) Quantification of pulldown experiments of RNAP from *B. subtilis* cell extracts using biotinylated PcrA as bait and supplemented with purified PcrA WT or CTD. Addition of the CTD competes with WT PcrA to bind to RNAP. Error bars show the SEM of three independent repeats. (G) Relative R-loop levels in strains of *B. subtilis* expressing free CTD, a CTD mutant that interacts weakly with RNAP, or with a control expression cassette. A Δ*rnhC* strain is shown as a control for elevated R-loop levels. Error bars show the SEM of at least 3 independent experiments. In all panels the statistical significance was determined using two-tailed Student's t test (*p value < 0.05, **p value < 0.01, ***p value < 0.001, ****p value < 0.0001).

Figure 7

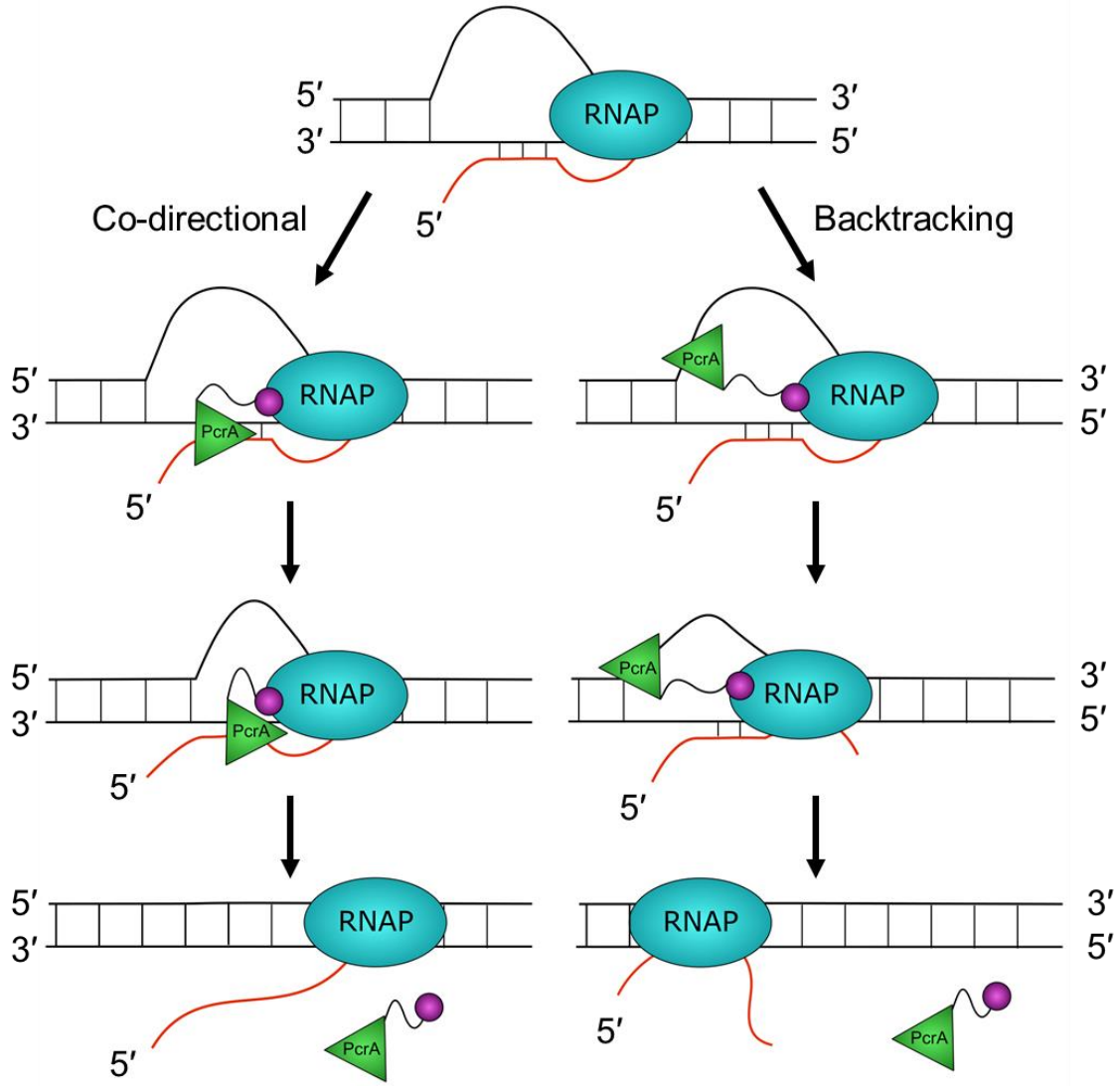

Figure 7. Working models for PcrA-dependent R-loop suppression during transcription. R-loops form during transcription but lead to genomic instability and are therefore targeted for removal by PcrA. In the co-directional model, PcrA helicase (green) interacts with an R-loop associated RNAP (blue) via the PcrA-CTD (magenta). It then engages the template DNA strand before translocating in the 3′>5′ direction and directly unwinding the DNA:RNA hybrid. In the backtracking model, PcrA binds to the displaced (non-template) strand behind RNAP and pulls it backwards, thereby unwinding the R-loop indirectly (see main text for further details and discussion).