Published online 18 April 2023

Nucleic Acids Research, 2023, Vol. 51, No. 11 ***5527–5546***
https://doi.org/10.1093/nar/gkad263

RecA and SSB genome-wide distribution in ssDNA gaps and ends in *Escherichia coli*

Phuong Pham [1], Elizabeth A. Wood[2], Michael M. Cox [2] and Myron F. Goodman [1,*]

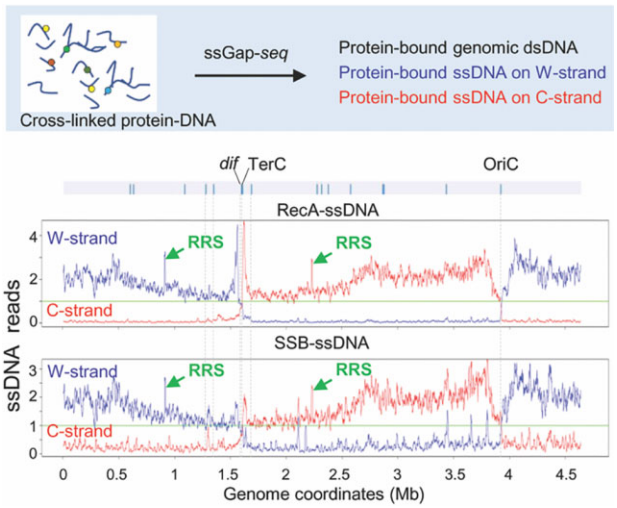
[1]Departments of Biological Sciences and Chemistry, University of Southern California, Los Angeles, CA 90089-2910, USA and [2]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706-1544, USA

Received December 06, 2022; Revised March 29, 2023; Editorial Decision March 29, 2023; Accepted March 30, 2023

ABSTRACT

Single-stranded DNA (ssDNA) gapped regions are common intermediates in DNA transactions. Using a new non-denaturing bisulfite treatment combined with ChIP-seq, abbreviated 'ssGap-seq', we explore RecA and SSB binding to ssDNA on a genomic scale in *E. coli* in a wide range of genetic backgrounds. Some results are expected. During log phase growth, RecA and SSB assembly profiles coincide globally, concentrated on the lagging strand and enhanced after UV irradiation. Unexpected results also abound. Near the terminus, RecA binding is favored over SSB, binding patterns change in the absence of RecG, and the absence of XerD results in massive RecA assembly. RecA may substitute for the absence of XerCD to resolve chromosome dimers. A RecA loading pathway may exist that is independent of RecBCD and RecFOR. Two prominent and focused peaks of RecA binding revealed a pair of 222 bp and GC-rich repeats, equidistant from *dif* and flanking the Ter domain. The repeats, here named RRS for replication risk sequence, trigger a genomically programmed generation of post-replication gaps that may play a special role in relieving topological stress during replication termination and chromosome segregation. As demonstrated here, ssGap-seq provides a new window on previously inaccessible aspects of ssDNA metabolism.

GRAPHICAL ABSTRACT



INTRODUCTION

Single stranded (ss) DNA in the *Escherichia coli* genome (gDNA) arise as transient intermediates during 3R (replication, recombination, repair) pathways (1). Studies of ssDNA gapped regions interspersed within the genome have a rich history spanning fifty-five years and counting. In 1968, two classic papers were published that portrayed the presence and properties of ssDNA gaps. A paper by Rupp and Howard-Flanders showed that gaps formed in cells irradiated with UV light (2). It was proposed that the presence of UV-induced thymidine dimers could impede DNA synthesis thus generating ssDNA gaps located in between replication-blocking T-T dimers and regions where DNA synthesis had resumed downstream (2). As established by Okazaki and colleagues (3), short nascent DNA fragments were observed in cells dividing exponentially. These fragments, termed Okazaki fragments arise as replication intermediates during lagging strand synthesis. Ongoing reinitiation of Okazaki fragments, which involves cycling of the replisome, can generate transient ssDNA gaps on the lagging strand (4,5).

*To whom correspondence should be addressed. Tel: +1 213 740 5190; Fax: +1 213 821 1138; Email: mgoodman@usc.edu

In the ensuing years, ssDNA discontinuities have been found to occur on both leading and lagging strands. ssDNA nicks and gaps are attributable to a multiplicity of processes including DNA recombination, repair of DNA lesions by nucleoside excision repair (NER), base excision repair (BER), and postreplication mismatch repair (MMR) pathways (6–8), and non-3R mechanisms such as transient arrest of replication forks at DNA damage sites, followed by replication-restart (4,9). Gaps can also arise when the polymerase undergoes lesion-skipping, leaving the template lesion behind in a post-replication gap (10). Gap sizes can range from about 1 to a few nt (BER) (11), 12–13 nt (NER) (12), ∼50–2000 nt (MMR) (8), and an indeterminate size (post-replication gaps). Several non-exclusive ways may be used to fill in gDNA gaps, including translesion DNA synthesis (9,13), to more involved mechanisms such as template switching (14), which may also involve replication-fork regression, and the oft-used RecA–RecFOR recombination repair pathway (15,16).

Double stranded (ds)DNA ends may occasionally result from dsDNA breaks. However, dsDNA ends can also occur during replication fork reversal and from head-to-tail fork collisions (17). More commonly, replisome encounters with unrepaired ssDNA gaps results in the generation of a one-sided double-stranded break. This can occur following replication fork collapse if a replisome encountered the template discontinuity during an ensuing replication cycle (9). The repair of each of these types of dsDNA breaks is carried out principally by RecBCD, a heterotrimeric complex containing helicase and dsDNA exonuclease activities that generate lengthy 3′-ssDNA ends. RecBCD loads RecA onto the 3′ ssDNA ends to carry out recombination repair (18).

RecA recombinase and SSB (ssDNA binding) proteins are essential for the processing of ssDNA gaps and ends in each of the 3R pathways. To date, the genome-wide distribution of RecA and SSB on ss gDNA is unknown. Recently, we have developed a method using non-denaturing bisulfite treatment of gDNA, in conjunction with whole genome sequencing, to map ssDNA regions to near single-nt resolution (19). In this paper, we further develop this method by combining it with ChIP-seq to create an approach abbreviated as 'ssGap-seq', allowing us to measure the genome-wide distributions of RecA and SSB bound to ssDNA. We have mapped the binding of each protein to ssDNA on leading- and lagging W- and C-strands in cells dividing exponentially, in the presence and absence of UV light, and in stationary phase cells. New observations stemming from this analysis include: (i) the assembly of RecA and SSB on ssDNA occurs predominantly on the lagging strand in mid-log phage cells, a bias which is absent during growth in stationary phase; (ii) the binding profiles of RecA and SSB tend to coincide globally with departures confined to localized ssDNA regions; (iii) a notable departure is observed in the large replication termination region (∼400 kb) where RecA binding is strongly favored; (iv) SSB–ssDNA binding is strongly favored on the leading strand in 230 nt regions surrounding *Ter*A, B, C, D sites, which could imply the formation of ssDNA gaps caused by replication fork fusion owing to the presence of Tus-*Ter* complexes. We also examined how the presence or absence of individual 3R genes (*recA, recB, recF, recG, recO, mutS, radD, mfd, xerD)* influence the distribution of RecA and SSB on genomic ssDNA. An unanticipated finding is that although loading of RecA on ssDNA for the most part requires the presence of RecFOR and RecBCD, there is considerable loading of RecA that does not require RecFOR or RecBCD complexes. Lastly, we report on dynamic changes in genome-wide distribution RecA–ssDNA and SSB–ssDNA after UV-irradiation in wild-type *E. coli* and in cells lacking RecB and RecO.

Going further afield, we note that a combination of computational and experimental analyses has revealed macroscopic genomic features including chromosomal condensation and the boundaries of major macrodomains (20,21). We have used our data showing the persistence of ssDNA gaps located proximal to the Ter macrodomain boundary to speculate on the possible role of a nearly perfect inverted repeat sequence in blocking or inhibiting replication fork progression, which we've termed 'Replication-Risk Sequence' (RRS). These RRS are 1.3 Mb apart and are symmetrically arranged around *dif*, 650 kb to either side. More, generally, our experimental approach is not restricted to *E. coli* but can be used more expansively to determine ss gDNA interaction landscapes in other bacterial and eukaryote genomes.

MATERIALS AND METHODS

Materials

Wild-type MG1655 *E. coli*, isogenic *recA730*, *recAΔC17*, *ssb-113*, deletion mutants (Δ*mfd,* Δ*mutS,* Δ*radD,* Δ*recA,* Δ*recB,* Δ*recF,* Δ*recO,* Δ*recG,* Δ*sbcB (exo1),* Δ*scbC,* Δ*uvrD,* Δ*xerD,* Δ*recB*Δ*recF,* Δ*recB*Δ*recO*) and polyclonal chicken anti-RecA and anti-SSB antibodies were from our labs. PrecipHen (Agarose coupled Goat anti-Chicken IgY beads) was purchased from Aves Labs, inc (Davis, CA). Polyclonal rabbit anti-RecA antibody (ab63797) was purchased from Abcam, and Protein G Dynabeads was from ThermoFisher. Monarch PCR & DNA clean-up kit was purchased from New England Biolabs. DNase-free RNAse A (100 mg/ml) was from Qiagen, Proteinase K, sodium metabisulfite and hydroquinone were from Sigma-Aldrich. xGen Methyl-Seq Library kits were purchased from Integrated DNA Technologies, IA.

Preparation of *E. coli* cells for immunoprecipitation

MG1655 wild-type and mutant (except for *ssb-113*) cells were grown in LB broth in the presence of 0.2% glucose at 37°C. *ssb-113* cells were grown at 30°C. Mid-log phase cells were harvested at OD_{600} ∼0.5. Cells from an overnight culture (OD_{600} ∼ 3.0) were collected as stationary cells. Cells were harvested in 50 ml Falcon tubes by centrifugation at 4000 × g for 7 min and resuspended in an equal volume of 1× PBS buffer (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 2.7 mM KCl, 137 mM NaCl, pH 7.4) at room temperature. For Chromatin Immunoprecipitation (ChIP), cells were crosslinked by addition of formaldehyde to a final concentration of 1%. After incubation at room temperature for 10 min, crosslinking was stopped by addition of Glycine (final concentration 125 mM) and incubated for 10 min at room temperature. Crosslinked cells were washed twice with

ice-cold 1× TBS (50 mM Tris–HCl, 150 mM NaCl, pH 7.5) and cell pellets were stored at –80 °C.

To prepare UV-irradiated cells, mid-log phase cells were harvested by centrifugation at 4000 × g for 7 min and resuspended in an equal volume of ice-cold 100 mM $MgSO_4$. Four sterile plastic Petri dishes, each containing 20 ml of cells were irradiated with 254 nm UV light at 100 J/m^2 using a CL-1000 UV crosslinker (UVP, Inc., Upland, CA). Immediately after UV-treatment, the irradiated cells were transferred to a 250 ml centrifuge tube containing 80 ml of prewarmed (37°C) LB + 0.2% glucose. After incubation at 37°C for periods of 1, 5, 10 or 20 min, the cells were centrifuged at 4000 × g for 7 min at 4°C and resuspended in 80 ml of 1X PBS (room temperature). Crosslinking of UV-irradiated cells was carried out as described above for non-irradiated cells.

ssGap-sequencing

The ssGap-seq protocol used as a combination of a standard ChIP protocol (22), involving a non-denaturing bisulfite treatment of immunoprecipitated DNA, and whole genome next-gen sequencing. Treatment of DNA by sodium bisulfite under non-denaturing condition (37 °C for 18 h) results in the efficient conversion of C → U in ssDNA (∼95%), with little C deamination in dsDNA (∼ 5%) (19). The large difference in the efficiency of bisulfite-induced C deamination in ssDNA *vs*. dsDNA allows for the direct identification of ChIP-seq reads derived from ssDNA, which contain long tracks of consecutive C → T (or G → A) conversions, and ChIP-seq reads derived from dsDNA (19).

Chromatin shearing by sonication

A crosslinked cell pellet (prepared from 40 ml at OD_{600} ∼ 0.5) was resuspended in 2 ml of freshly prepared ChIP lysis buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA, 1.0% Triton X-100, 1.0% sodium deoxycholate and 2× Roche's Complete Protease inhibitor cocktail). The chromatin was sheared to an average size of ∼200–300 bp using a bath sonicator (Covaris S2) in a milliTUBE 2 ml tube for 8 cycles of 60 s each (8 min total sonication duration) with settings of duty cycle 20%, intensity 8, cycles/bust 200. The temperature of the sonicator water bath was held between 5 and 8 °C. After sonication, the cell lysates were centrifuged for 5 min at 16 000 × g at 4 °C to remove insoluble material.

Immunoprecipitation (IP)

To a 1.5 ml Eppendorf tube containing 250 μl of ChIP lysis buffer + 250 μl of a freshly prepared cell lysate, 5 μl of a purified anti-RecA or anti-SSB antibody was added. Incubation was then carried out at 4°C on a rotator overnight. The following morning, 100 μl of PrecipHen (Agarose coupled Goat anti-Chicken IgY beads), prewashed with the lysis buffer, was added to IP tubes (for chicken anti-RecA or anti-SSB antibodies), and incubated for 2 h at room temperature. For IP tubes containing rabbit anti-RecA antibody, Protein G Dynabeads were used instead of PrecipHen. The IP beads were sequentially washed (2 times each) with 1 ml of ChIP lysis buffer, wash buffer #1 (20 mM Tris, pH 8.0, 500 mM NaCl, 1 mM EDTA, 1.0% Triton X-100, 1.0% sodium deoxycholate), wash buffer #2 (20 mM Tris, pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% sodium deoxycholate, 0.1% Tween-20) and TE buffer (10 mM Tris, pH 8.0, 1 mM EDTA). Elution was carried out by incubation of the IP beads with 250 μl elution buffer (0.1 M $NaHCO_3$, 1% SDS) for 5 min. Following a second elution, the IP eluates were combined (500 μl) and transferred to an Amicon Ultra 10 K Centrifugal Filter Device (Millipore) and washed 3 times with TE buffer. Trace contaminants of SDS and sodium bicarbonate were then removed using a Bio-spin P6 column (Bio-Rad, CA). The IP-eluted chromatin (∼40–50 μl in 10 mM Tris, pH 8.0) was used either in a non-denaturing bisulfite treatment or was stored at –20°C.

Non-denaturing bisulfite treatment

A non-denaturing bisulfite treatment was carried out by incubating the IP-eluted chromatins with 460 μl of a freshly prepared solution of 5 M sodium bisulfite and 20 mM hydroquinone at 37°C overnight for 18 h (19). The following morning, the bisulfite-treated chromatin was washed 4 times with water using Amicon Ultra 10 K Centrifugal Filter Device (Millipore) followed by desulphonation in a 0.3 M NaOH solution for 20 min at room temperature. After removal of NaOH with an Amicon Centrifugal Filter Device by washing 4 times with TE buffer, the DNA was subjected a buffer exchange to 10 mM Tris pH 8.0, using Bio-spin P6 column (Bio-Rad, CA).

Reverse crosslinking and purification of IP DNA

To each tube of bisulfite-treated chromatins, 1 μl of RNAse A (10 μg/μl) was added and incubated at 37°C for 10 min to degrade cellular RNAs. Reverse crosslinking and protein digestions were carried out by incubation of RNAse A-treated chromatins (total volume 200 μl), at an elevated temperature (65°C for 4 h) in the presence of proteinase K (0.25 mg/ml) and 1% SDS. IP DNA was purified using a Monarch PCR & DNA clean-up kit, eluted in 20 μl of EB buffer (10 mM Tris, pH 8.5) and stored at –20 °C.

Illumina library preparation and genome sequencing

To prepare an Illumina's sequencing library, 10 μl of purified IP was used as input DNA in xGen Methyl-Seq Library kit (Integrated DNA Technologies, IA). The library was prepared according to the company's protocol with 16 PCR amplification cycles. This library preparation kit relies on enzymatic attachment of adaptors directly on bisulfite-treated ssDNA by the company's Adaptase technology. Each library was purified using two rounds of an SPRI bead clean up, quantified by Qubit and qPCR, and subjected to Illumina's sequencing (2 × 150 bp paired-ends) on a MiniSeq using Mid Output Reagent Kits (300-cycles).

Melting analysis of RRS-lysO and RRS-dusC sequences

RRS-*lysO*, RRS-*dusC* and a control sequence were amplified from MG1655 genomic DNA by PCR (35 cycles, 98°C

– 10 s, 50°C – 10 s, 72°C – 30 s) using Q5 High-fidelity master mix (New England Biolabs). Forward and reverse PCR primers contain 10–11 additional bases at each end of RRS-*lysO* and RRS-*dusC*. The control sequence represents a 244 bp region located 60 bp upstream of RRS-dusC. Amplicons (244 bp) were purified by a Monarch PCR & DNA clean-up kit, eluted in EB buffer (10 mM Tris, pH 8.5). Melting kinetics were measured in Q-qPCR instrument (QuantaBio). For each amplicon, 2 μl (50 ng) of DNA was mixed with 8 μl of SparQ qPCR master mix (QuantaBio), heated at 95°C for 2 min and quickly cooled down to 40°C for 1 min. SYBR Green fluorescence was monitored as the temperature was increased from 70°C to 95°C. Melting peaks were calculated by plotting changes in SYBR Green fluorescence over time ($-dF/dt$) versus temperature.

Sequencing data analysis

The Illumina sequencing data were analyzed using the CLC Genomic Workbench (v. 21.0) (Qiagen) as described previously (19). High quality sequencing reads were imported, 25 nt were trimmed from the 5'-end and 20 nt from the 3′-end, and then filtered to select for approximately uniform lengths of 101–105 nt. To identify all genomic DNA (both dsDNA and ssDNA) that was bound to RecA or SSB, ChIP-seq reads were mapped to an MG1655 reference genome (Accession number NC_000913) using a bisulfite sequencing (BS-map) alignment algorithm, which maps all reads regardless of whether or not C → T conversion occurred in individual reads. A BS-map alignment was carried out by a directional protocol with a length fraction setting of 0.8 and similarity fraction setting of 0.9 (19). Reads that mapped non-specifically (i. e., reads aligned at more than one genome position with equally good scores) were ignored and not used for analysis. Non-specific match reads often represent genes with multiple copies in contained in the *E. coli* chromosome, such as seven copies of ribosomal 23S, 16S and 5S, numerous insertion sequences, Rhs elements and *tufA* and *tufB* genes. All mapped duplicated ChIP-seq reads that were generated by PCR amplification during the library preparation were removed using the 'duplicate reads removal' tool.

To identify ChIP-seq reads that represent ssDNA bound to RecA or SSB on the W-strand or on the C-strand, the reads were aligned and mapped to CT-converted (or GA-converted) reference genomes using a standard 'Map reads to reference' tool in the CLC program, as described previously (19). Some reads in genomic regions with low GC content mapped to the CT-converted (or GA-converted) reference genomes, but do not represent ssDNA. These reads were removed by their alignment to the unconverted MG1655 reference genome. In addition, duplicated ChIP-seq reads and reads that map to non-specific regions (i.e. ribosomal 23S, 16S and 5S, numerous insertion sequences, Rhs elements and *tufA* and *tufB* genes) were also removed from the final maps.

Sequencing reads having their entire length aligned and mapped to the CT-converted or GA-converted reference genomes were used to analyze the spatial distribution of RecA–ssDNA and SSB–ssDNA on W- or C-chromosomal strands, respectively. Hot genomic regions with enrichment of RecA–ssDNA (or SSB–ssDNA) on W- and C-strands were identified by the 'Whole Genome Coverage Analysis' tool in the CLC Genomic Workbench, using 100 bp as the minimum length and 0.01 as the 'P-value threshold'. This tool identifies regions with RecA–ssDNA (or SSB–ssDNA) 'unexpectedly' enriched in coverage. This analysis involves the examination of the coverage in each of the positions in the RecA–ssDNA (or SSB–ssDNA) read mappings and identifies those having coverage in the upper tails (1%) of a Poisson distribution. Regions with consecutive positions marked as having high RecA–ssDNA (or SSB–ssDNA) coverage are designated as 'hot' regions.

Statistical analysis

Mapping coverages with detailed per-base coverage information for BS-seq maps and ssDNA maps for W- and C-strands were exported from the CLC Workbench program in a tab delimited format. Percentages of ssDNA bound to RecA and SSB, normalized coverages of RecA–ssDNA and SSB–ssDNA on W- and C-strands, and enrichment at specific genome regions were computed using custom-written R scripts. Enrichments of RecA–ssDNA and SSB–ssDNA at a specific region on leading or lagging strands were calculated as ratios of average ssDNA read coverages in the region divided by the average coverage for the entire leading strand or lagging strand. The following genome coordinates for the *Ter* and *dif* regions were used for analysis of RecA–ssDNA and/or SSB–ssDNA enrichment: *Ter*A (1341630..1341865), *Ter*B (1684142..1684372), *Ter*C (1609021..1609255), *Ter*D (1279446..1279676), *dif*-Left (1568479..1542555), *dif*-Right (1610176..1634275), RPS-*lys*O (913842..915206) and RPS-*dus*C (2229192..2231691). For a comparative analysis of relative genome-wide distributions of RecA and SSB, the normalized RecA–ssDNA and SSB–ssDNA read coverages were computed at individual positions by taking moving averages within a 10 kb window. The differences in normalized coverage between RecA–ssDNA and SSB–ssDNA at individual positions were calculated using custom-written R-scripts. Differences between RecA–ssDNA and SSB–ssDNA are considered significant if they are greater than two standard deviations (2σ). CLC workbench and R-programming were used to generate whole genome ChIP-seq coverage graphs.

RESULTS

Profiles of RecA- and SSB-bound genomic ssDNA and dsDNA

We have developed a method that we call ssGap-seq, which combines ChIP-seq with non-denaturing sodium bisulfite treatment to measure genome-wide protein interactions with dsDNA and ssDNA (Figure 1A, Supplementary Figure S1). The method is based on our recently described procedure that permits genome-scale definition of the ssDNA landscape (19). ssGap-seq is carried out first by shearing formaldehyde-crosslinked chromatin to 200–300 bp fragments. Following standard chromatin immunoprecipitation (ChIP) protocols using an antibody specific for a protein of interest (22), the eluted ChIP DNA is then incubated with

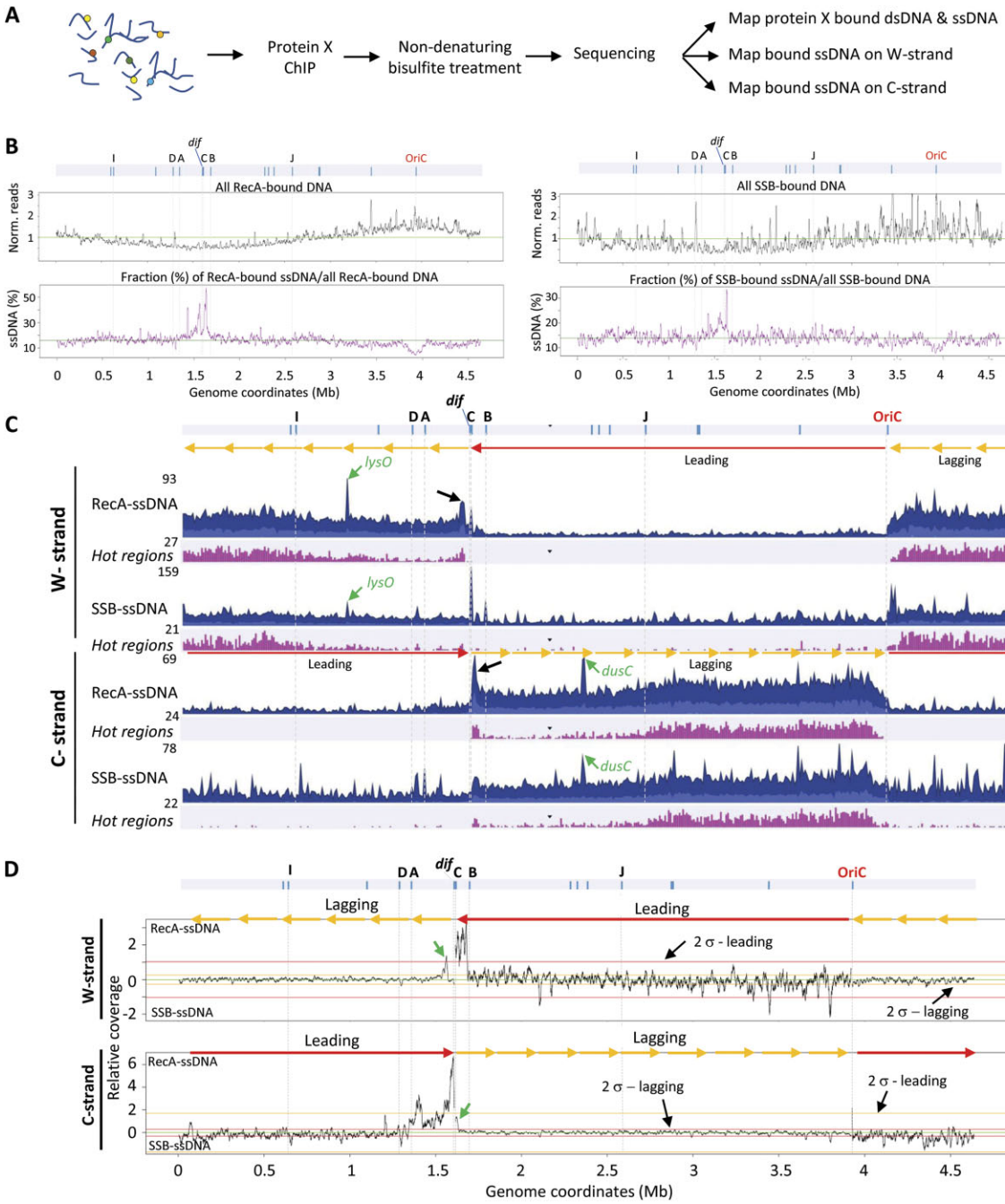


Figure 1. Analysis of RecA and SSB association with genomic dsDNA and ssDNA. (**A**) The ssGap-seq assay combines a standard ChIP protocol with non-denaturing bisulfite treatment of ChIP DNA and next-gen sequencing to identify proteins bound to ds- and ss-gDNA. (**B**) Genomic distribution of RecA- and SSB-bound gDNA (top panels) and fraction (%) of bound ssDNA relative to total bound DNA (bottom panel) in exponentially growing wild-type MG1655 cells. The distribution of RecA-bound DNA and SSB-bound DNA are shown as normalized read coverages. The solid green line (value = 1) represents the average genome coverage. For fractions of bound ssDNA/all bound DNA graphs, the solid green line represents the average percentage of bound ssDNA. Each data point represents a moving average in a 10 kb window. (**C**) Genome distribution of RecA–ssDNA and SSB–ssDNA in log phase wild-type cells. Mapped RecA–ssDNA and SSB–ssDNA on W- and C-strands are shown as aggregated coverage graphs with average coverage values at each genome position shown in blue and maximum coverage values shown in dark blue. The Y-axis gives the number of mapped RecA–ssDNA or SSB–ssDNA reads. The green arrows indicate the peaks that map to the lagging-strand at *lysO* and *dusC* loci. Black arrows denote RecA–ssDNA peaks flanking *dif*. A bar graph (pink) gives the distribution of 'hot' RecA–ssDNA and SSB–ssDNA regions, that are at least 100 bp long and have significantly higher coverage ($P < 0.01$). The height of each bar represents the number of hot regions located within a 10 kb genomic segment. The leading and lagging strands for the two replication forks initiated at OriC are depicted as red and orange arrows, respectively. (**D**) Difference in normalized coverages for RecA–ssDNA and SSB–ssDNA on the leading and lagging strands. The data represent the normalized coverage of RecA–ssDNA minus the normalized coverage of SSB–ssDNA along the gDNA. Lines representing two standard deviations (2 σ) of the differences for the leading and lagging strands are indicated. Green arrows located near the *dif* site indicate peaks having significantly more RecA–ssDNA compared to SSB–ssDNA on the lagging strand. The positions of the replication origin (OriC), *dif* and 14 *Ter* sites (57) are shown as tick marks at the top. *Ter*A, *Ter*B, *Ter*C, *Ter*D, *Ter*I and *Ter*J are labeled as A, B, C, D, I and J, respectively.

sodium bisulfite under non-denaturing conditions to convert C to U on protein-bound ssDNA, but not dsDNA. A high-throughput sequencing and bioinformatic analysis is used to identify protein-bound ssDNA as ChIP-seq reads containing C to T conversions (W-strand) and G to A conversions (C-strand) (19). We can then determine what fraction of the pulled down reads are ssDNA.

The profiles of RecA- and SSB-bound to total gDNA (dsDNA + ssDNA) in wild type cells grown to mid-log phase exhibit a gradient, in which the total DNA bound and pulled down by RecA and by SSB is greatest at the origin of replication (OriC), and gradually decreases moving away from OriC, with minimal binding of RecA and SSB occurring in replication terminal Ter region (Figure 1B, two top panels). The normalized values of DNA bound by RecA and SSB are approximately 1.5-fold higher at OriC, and 2-fold lower at *Ter*C compared to the genome averages (green lines with value = 1.0, Figure 1B, two top panels).

In contrast to the protein binding profiles to the total gDNA, the fraction of RecA- and SSB-bound ssDNA, calculated as the percentage of ssDNA reads on W- and C-strands divided by total bound ChIP-seq reads at individual positions along the genomes, show no apparent polarity in between OriC and *Ter*C. Throughout the genome, about 18% of the DNA pulled down by RecA is ssDNA, while about 13% of the DNA pulled down by SSB is ssDNA (Table 1). Most of pulled-down DNA are dsDNA, 82% for RecA and 87% for SSB (Table 1). There are two major deviations from this pattern. In ∼80 kb regions on both sides of OriC, the fraction of DNA pulled down by both RecA and SSB proteins that is ssDNA declines substantially (Figure 1B, two lower panels) suggesting that gaps available for binding of these two proteins are constrained. In contrast, in the Ter region between *Ter*D and *Ter*B, major peaks of RecA and SSB-bound ssDNA are seen, particularly around *Ter*C. Outside these two regions, the percentage of ssDNA bound remains approximately uniform with small fluctuations from the average fraction of ssDNA in the pull downs (Figure 1B, two lower panels). These marked differences in fractions of bound ssDNA at OriC and Ter regions in mid-log phase cells may reflect different processes occurring during circular chromosomal duplication, such as initiation and assembly of replisomes at OriC, repair of ssDNA gaps/ends, or perhaps overamplification during replication fork fusions that could be occurring in the vicinity of *Ter*C (23,24).

There is a considerable background of dsDNA reads in the RecA and SSB pull downs in most genomic regions. As seen in the top panels of Figure 1B, the total reads in the pull downs track closely with the amount of total DNA present at particular genomic regions during exponential growth—more near the origin and less near the terminus. This indicates that the dsDNA reads reflect transient and perhaps weak interactions of both proteins with random segments of genomic dsDNA. During exponential growth, only about 1.3% of gDNA is single stranded at any given moment (19), so that ssDNA is highly enriched in the RecA and SSB pull downs. The vast majority of both proteins are not bound to ssDNA under normal growth conditions (Table 1, Figure 1B). Both have a limited capacity to bind to dsDNA, particularly when that DNA is supercoiled (25,26). Nevertheless, transient interactions of RecA and SSB can occur with genomic dsDNA regions that are not actively involved in replication. Therefore, the fraction of reads that are ssDNA in the pull downs will be a complex function of the concentrations of the proteins in the cell, the fraction of the genome that is ssDNA at any moment, the relative affinities of each protein for ssDNA and supercoiled gDNA, and (at least for RecA) the availability of appropriate loader proteins to effect binding to SSB-coated ssDNA. RecA could also pull down dsDNA into which RecA filaments extended near the boundaries of gaps (27,28) or dsDNA involved in RecA-mediated pairing reactions (29). SSB may also pull down dsDNA as a result of the many interactions SSB has with other DNA binding proteins (1). Important effects relevant to DNA metabolism will thus be best reflected in significant increases or decreases in the fraction of pulled down reads that are ssDNA. Also, *E.* coli mutants (e.g. *recA730* (30)) or UV-irradiation, which induce the SOS response leading to changes in the cellular concentrations of RecA and/or SSB, could affect the proportion of bound ssDNA. As seen below, the patterns of increases and decreases in the ssDNA read fraction found in pull downs, in mutant strains where effects on RecA binding to ssDNA are well characterized, indicate that this parameter provides a measure of increases and decreases in RecA or SSB binding to genomic ssDNA.

The genomic profiles of RecA- and SSB-bound DNA and fractions of pull downs present as ssDNA for mutant strains are shown in Supplementary Figures S2 to S20. In mid log-phase cells, the average fraction of ssDNA in the RecA pull downs is 17.8%, with the fraction in the SSB pull downs being 13% (Table 1). As might be expected, ssDNA in the pull downs is ∼3-fold lower during growth in stationary phase where replication is limited, 3% for RecA and 6% for SSB. UV-irradiated log-phase cells show a sizable increase in the ssDNA reads found in both protein pull downs, 54.7% for RecA and 71.4% for SSB at 20 min post UV (Table 1), as larger fractions of both protein pools become engaged in DNA repair activities.

We have measured the binding of RecA and SSB for mutant strains including deletions of RecA; proteins involved in RecA loading on ssDNA (RecB, RecF, and RecO) (9); helicases (RecG, UvrD) (31,32); nucleases (SbcB-exo1 and SbcC) (33); a mismatch repair protein MutS (34); a transcription-coupled repair protein Mfd (35); a RecA auxiliary protein RadD (36); a site-specific recombinase XerD (37), and the replication termination protein Tus (38). There are also two RecA mutants, SOS-constitutive *recA730* (E38K) (30), and *recA*Δ*C17* (C-terminal 17 amino-acid deletion) (39), and a temperature sensitive SSB mutant *ssb-113* (P176S) (40).

The deletion of proteins known to be involved in the assembly of RecA on ssDNA (Δ*recB*, Δ*recF*, Δ*recO*, Δ*recB*Δ*recF*, Δ*recB*Δ*recO*) result in an ∼2-fold reduction in the ssDNA found in the RecA pull downs (ranging between 7.4 and 12%, Table 1). As already mentioned, only about 1.3% of the gDNA is ssDNA under normal growth conditions, suggesting that RecA binds to some cellular ssDNA even in the absence of the RecFOR or RecBCD loading systems. In contrast, the fraction of ssDNA reads found in the SSB pull downs is relatively insensitive to these

Table 1. Analysis of RecA and SSB interaction with genomic dsDNA and ssDNA in wild-type and mutant *E. coli*

	RecA ChIP		SSB ChIP		
Cells	Percent of bound-ssDNA/ all bound DNA[a]	Lag: Lead ratio[b]	Percent of bound-ssDNA/ all bound DNA	Lag: lead ratio	RecA and SSB overlapping hot regions (%)[c]
Mid-log					
Wild-type	17.8 ± 4.3	23.4 ± 2.7	13.0 ± 2.2	6.2 ± 0.6	49
Δ*mfd*	17.3 ± 0.5	21.3 ± 2.2	14.0 ± 0.4	5.4 ± 0.1	44
Δ*mutS*	15.0 ± 1.5	16.9 ± 2.2	8.6 ± 0.9	5.7 ± 0.3	25
Δ*radD*	6.4 ± 0.1	25.8 ± 4.9	12.7 ± 0.4	10.8 ± 0.3	71
Δ*recA*			9.4 ± 0.1	2.9 ± 0.1	
Δ*recB*	7.4 ± 0.1	8.8 ± 0.5	13.7 ± 0.2	2.7 ± 0.1	42
Δ*recF*	8.9 ± 0.2	10.1 ± 0.8	13.6 ± 0.4	7.1 ± 0.2	53
Δ*recG*	8.3 ± 1.1	6.1 ± 0.1	13.2 ± 1.4	4.7 ± 0.3	54
Δ*recO*	12.1 ± 0.2	8.6 ± 0.8	20.5 ± 0.4	4.5 ± 0.1	70
Δ*recB*Δ*recF*	11.4 ± 0.3	7.9 ± 1.0	14.8 ± 0.5	5.4 ± 0.4	56
Δ*recB*Δ*recO*	12.1 ± 0.2	6.6 ± 1.1	15.8 ± 0.5	5.0 ± 0.4	61
Δ*sbcB (exo1)*	12.1 ± 0.3	18.8 ± 3.8	11.6 ± 0.3	7.9 ± 0.3	40
Δ*sbcC*	10.7 ± 0.1	19.7 ± 3.4	11.4 ± 0.2	6.8 ± 0.1	36
Δ*tus*	8.7 ± 0.2	17.1 ± 1.4	9.9 ± 0.2	8.0 ± 0.1	41
Δ*uvrD*	13.6 ± 2.4	20.2 ± 2.8	11.5 ± 0.7	6.8 ± 0.1	26
recA730	11.9 ± 0.2	11.0 ± 1.4	12.9 ± 0.5	4.9 ± 0.1	26
*recA*Δ*C17*	15.4 ± 0.2	19.5 ± 2.1	13.0 ± 0.4	5.7 ± 0.1	52
ssb-113	31.3 ± 0.1	4.2 ± 0.1	35.1 ± 0.3	3.5 ± 0.2	78
Stationary					
Wild-type	3.0 ± 0.1	1.9 ± 0.1	6.1 ± 0.1	1.0 ± 0.1	26
UV-irradiated					
WT - 1 min post UV	3.9 ± 0.1	7.4 ± 0.1	8.4 ± 0.1	2.5 ± 0.1	16
WT - 5 min post UV	19.1 ± 0.1	1.8 ± 0.1	50.4 ± 0.3	1.5 ± 0.1	61
WT - 10 min post UV	24.4 ± 0.2	1.6 ± 0.1	45.1 ± 0.1	1.5 ± 0.1	71
WT - 20 min post UV	54.7 ± 0.1	1.5 ± 0.1	71.4 ± 0.2	1.4 ± 0.1	50
Δ*recB - 1 min post UV*	3.6 ± 0.1	6.1 ± 0.3	9.6 ± 0.2	1.8 ± 0.1	34
Δ*recB - 20 min post UV*	26.1 ± 0.1	1.2 ± 0.1	49.8 ± 0.1	1.3 ± 0.1	76
Δ*recO - 1 min post UV*	3.8 ± 0.1	7.7 ± 0.1	13.8 ± 0.1	3.6 ± 0.1	71
Δ*recO - 20 min post UV*	14.6 ± 0.1	2.0 ± 0.1	47.8 ± 0.3	1.9 ± 0.1	91

[a]Percent (%) of RecA-bound or SSB-bound ssDNA relative to all RecA-bound or SSB-bound DNA was calculated as the fraction of ssDNA ChIP-seq reads on the W- and C-strands relative to the total number of mapped ChIP-seq reads. [b] Ratio of ssDNA bound by RecA or SSB on lagging vs leading strands. Data are expressed as the mean ± standard deviation for two *E. coli* replichores. Values for mid-log phase *wild-type*, Δ*uvrD* and Δ*mutS* cells were calculated from at least two independent experiments. [c] Percentage of hot RecA–ssDNA regions that overlap with hot SSB–ssDNA regions.

Δ*rec* mutants, showing either unchanged binding frequencies or small increases in binding (Table 1). Additional genes that significantly reduce the fraction of ssDNA in the RecA pull downs are Δ*radD* (6.4%) and Δ*tus* (8.7%). The Δ*sbcB* (12.1%) and Δ*sbcC* (10.7%) mutants produce smaller declines in RecA binding to ssDNA.

RecA and SSB interact with ssDNA predominantly on lagging strands

The *E. coli* chromosome is divided into two oppositely replicating replichores, one replicating clockwise on the right side of OriC and other counterclockwise on the left side of OriC (41). Previous studies have suggested that the majority of fork fusions of two replichores occur near an arithmetic mid-point of the chromosome at the *Ter*C site (42,43). Genome-wide distribution of RecA–ssDNA and SSB–ssDNA on W- and C-strands for mid-log phase cells showed approximately equal numbers of mapped RecA–ssDNA (or SSB–ssDNA) on both strands. RecA–ssDNA and SSB–ssDNA exhibited a strong preference for lagging strands (Figure 1C, Supplementary Figures S2B to S19B). The ratios of lagging/leading strand occupancies were 23.4-fold for RecA–ssDNA and 6.2-fold for SSB–ssDNA (Table 1, mid-log Wild-type). However, where chromosomal replication is minimal in stationary phase, leading/lagging strand bias was reduced to ∼1.9-fold for RecA, and is absent for SSB (Table 1, Stationary Wild-type, Supplementary Figure S20B). Preferential binding of SSB to the lagging strand during growth in log phase is likely attributable primarily to binding within ssDNA gaps generated during discontinuous synthesis of Okazaki fragments. For RecA, double strand break repair is triggered by replisome encounters with template strand discontinuities. Due to the patterns of RecBCD processing of double strand ends which will ultimately load RecA onto the strand with a 3′ end proximal to the downstream disrupted replisome, this will always lead to RecA loading onto either the lagging strand template or onto its sequence-equivalent strand that is complementary to the leading strand template. Double strand break repair is thus expected to produce a lagging strand bias in RecA binding. The repair of post-replication gaps will also produce a lagging strand bias if most of these gaps are generated during lagging strand discontinuous DNA synthesis. Given the reduction in the lagging/leading strand bias seen in *recB, recO* and *recF* mutants, both processes contribute to the observed bias. However, the bias is not eliminated, even in strains lacking both RecF and RecB. If RecA was unable

to load onto ssDNA in the absence of RecB and RecF/O, then the RecA–ssDNA signals would represent background noise and thus be expected to be distributed equally on both leading and lagging strands. The presence of a strong RecA lagging strand bias suggests the existence of processes that load RecA protein into gaps even without the RecBCD and RecFOR systems. Notably, RecA-dependent homologous recombination in the absence of RecFOR and RecBCD has been observed in *E. coli* (44,45).

In wild type cells, the genome distributions of RecA bound to ssDNA determined by ssGap-seq gave essentially the same results (i.e. same genomic profiles of RecA–ssDNA on W- and C-strands) using two different antibodies against RecA, chicken anti-RecA and the widely used commercially available rabbit anti-RecA (46,47) (Supplementary Figure S21).

RecA–ssDNA is strongly enriched relative to SSB–ssDNA in the Ter region

In this subsection we perform a coarse-grained analysis that identifies extended gDNA regions in which RecA–ssDNA is strongly favored over SSB–ssDNA in exponentially growing cells (Figure 1D). Normalized coverages of RecA–ssDNA and SSB–ssDNA were computed separately for leading and lagging strands using 10 kb windows, with differences between them plotted in Figure 1D. For the leading strands, there is a clear significant enrichment (more than two standard deviations, 2σ) of RecA–ssDNA in *Ter*C-*Ter*B region on the W-strand, and *Ter*A-*Ter*C region on the C-strand (Figure 1D). For the lagging strands, there are two prominent regions ∼20 kb long in which the levels of RecA–ssDNA exceed SSB–ssDNA by greater than two standard deviations. These are located ∼22 kb to the left of *dif* on the W-strand and ∼20 kb to the right of *dif* on the C-strand (Figure 1D, green arrows). These regions are considered in more detail below. Outside the *Ter* region, the differences in RecA–ssDNA *vs*. SSB–ssDNA are considerably less than 2σ (Figure 1D). We further note that the relatively high level of noise occurring in the leading-strand profiles compared to lagging-strand profiles is a consequence of far fewer numbers of mapped reads in the former (Figure 1D).

Fine-grained analysis of RecA and SSB co-localization on lagging-strand ssDNA

A fine-grained analysis was performed to compare the distributions of RecA–ssDNA and SSB–ssDNA on the lagging-strand over the entire genome. We identified chromosomal hot regions in which binding of RecA and SSB to ssDNA shows maximal enrichment, i.e. regions ≥100 bp with ChIP-seq read coverage falling within the upper tail of a Poisson distribution ($P < 0.01$) (see Materials and Methods). The spatial distributions of RecA and SSB hot regions show striking similarities, which begin to appear on the lagging-strand at both sides of OriC and reaching their highest values ∼100 kb distal from OriC (Figure 1C). The concentrations of hot regions remain essentially flat between OriC and *Ter*l on the W-strand and between OriC and *Ter*J on the C-strand, and gradually decrease approaching *Ter*C, with two minor peaks ∼20 kb located on both sides of *dif* (Figure 1C). Approximately half of the hot RecA–ssDNA and SSB–ssDNA regions are overlapping in cells growing in mid-log phase (Table 1, 49%). The fraction of overlap drops to ∼25% in Δ*mutS,* Δ*uvrD* or *recA730* backgrounds and in stationary phase cells (Table 1). In the absence of RecB, there appears to be a small reduction in overlap (Table 1, 42%). The overlap fraction increases during mid-log phase in Δ*recO* (70%), Δ*recB*Δ*recO* (61%), Δ*recB*Δ*recF*(56%), Δ*radD* (71%) and *ssb-113* (78%) (Table 1). The largest increase occurs in UV-irradiated cells lacking RecO, showing a 91% overlap (Table 1, Δ*recO* 20 min post UV). The strong overlap in RecA, SSB binding to ssDNA at 'hot' genomic regions on lagging-strands over the entire genome suggests that RecA and SSB may be co-recruited to RecBCD-processed 3′-ssDNA tails, while perhaps RecF and RecO facilitate RecA-mediated replacement of SSB in the hot binding regions. We speculate that the elevated overlap of RecA and SSB observed for the Δ*recB*Δ*recO* and Δ*recB*Δ*recF* double mutants provides circumstantial evidence for a yet-to-be identified RecBCD and RecFOR loading pathway that facilitates the overlapping binding of RecA and SSB to the same ssDNA regions genome-wide.

RecA–ssDNA and SSB–ssDNA profiles in *E.coli* mutant backgrounds

To further the analysis of the ssDNA binding profiles for RecA and SSB in a variety of mutant backgrounds presented in Supplementary Figures S2 to S20, the genome-wide spatial distribution of RecA and SSB on ssDNA on the W-strand (Figure 2, blue traces) and C-strand (Figure 2, red traces) is summarized for wild-type and mutant mid-log phase and stationary phase cells (Figure 2). In the following discussion, we combine these data with those in Table 1. The W- and C-strand traces intersect at OriC and *Ter*C, with the intersection points marked by the positions of the leading- and lagging-strands in the two replichores. In MG1655 cells, the replication forks are initiated at both sides of OriC and are fused at *Ter*C, even in cells lacking the Tus termination protein (Figure 2, *Δtus*). The ssDNA profiles for RecA and SSB on the W- and C-strands for wild-type cells are altered in the mutant cells (Figure 2, Supplementary Figures S22 and S23), presumably dependent on each protein's role during replication, recombination, and repair.

We have used single and double deletion mutants in a variety of recombination genes to determine changes in the magnitude and location of RecA and SSB binding to ssDNA (Table 1 and Figure 2). In the absence of RecB (Δ*recB*), the fraction of reads that are ssDNA in the RecA pull downs is reduced by 2.4-fold from 17.8% to 7.4%, and a similar 2-fold reduction occurs in the absence of RecF (Δ*recF*). The 23.4-fold preference favoring RecA binding to lagging-strand over leading-strand DNA is reduced to 8.8-fold in the absence of RecB and to 10.1-fold in the absence of RecF (Table 1). There is a similar reduction in RecA binding and strand bias in Δ*recB*Δ*recF* double mutants, although the two mutations together do not reduce the bias further (Table 1). The retention of RecA–ssDNA binding in Δ*recB* and Δ*recF* mutants again suggests that there is a non-negligible component of RecA binding that does not

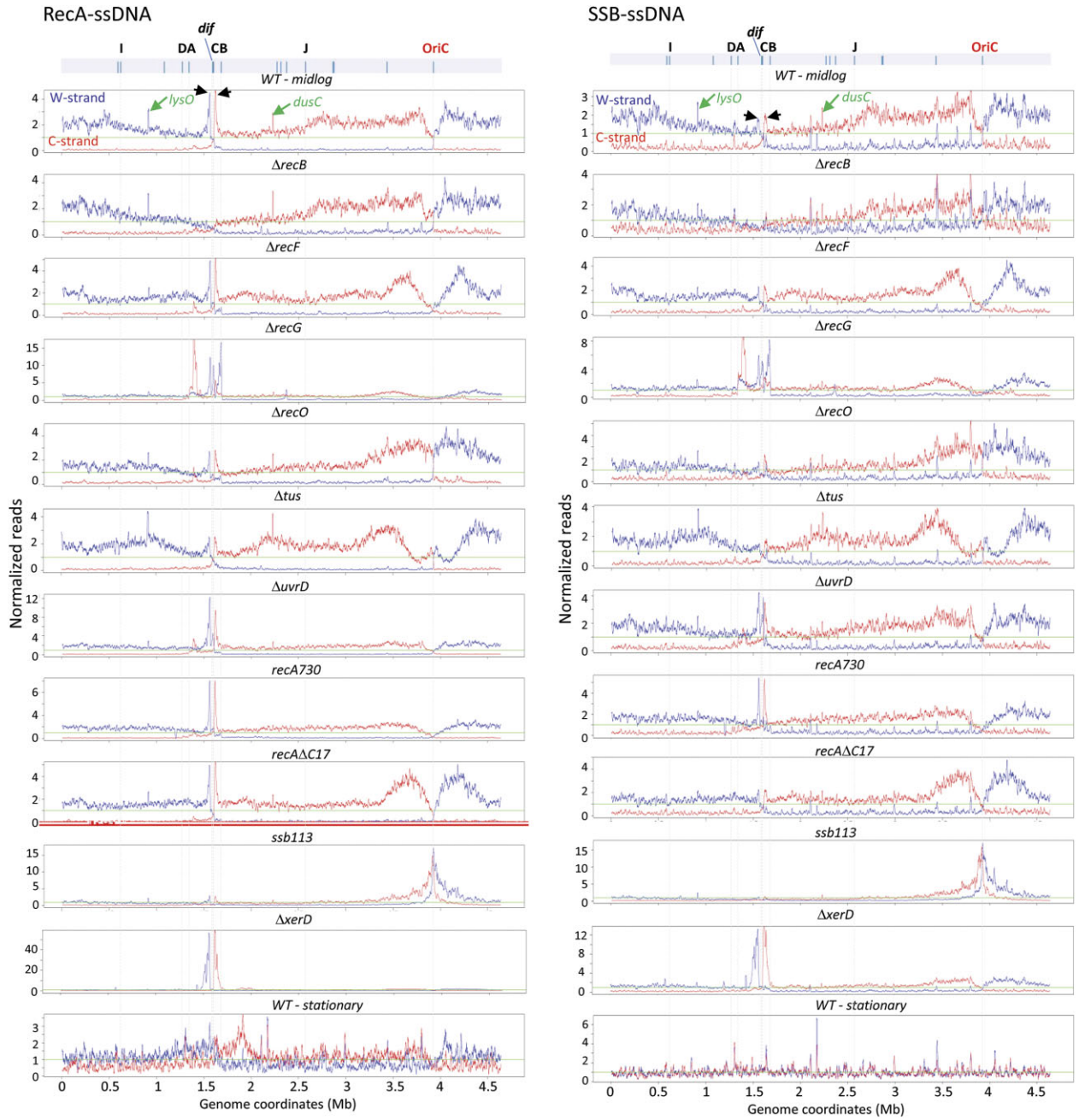


Figure 2. Genomic profiles of RecA–ssDNA and SSB–ssDNA in *wild-type* and selected mutants of *E. coli*. Normalized coverages of RecA–ssDNA and SSB–ssDNA are shown as blue traces for the W-strand, and red traces for the C-strand. Green lines in graphs (value = 1) indicate the genome average coverage. Each data point represents a moving average of RecA–ssDNA and SSB–ssDNA in a 10 kb window. Green arrows indicate peaks on the lagging strands at *lysO* and *dusC* loci. Black arrows denote peaks located on both sides of *dif*. Positions of the replication origin (OriC), *dif*, and 14 *Ter* sites are shown as tick marks at the top. *Ter*A, *Ter*B, *Ter*C, *Ter*D, *Ter*I and *Ter*J are labeled as A, B, C, D, I and J, respectively.

require RecBCD or RecFOR for loading RecA onto ssDNA.

Other mutants that reduce RecA strand bias include Δ*recG*, *recA730* and *ssb-113* (Table 1). Each is a special case for purposes of discussion. The elimination of RecG results in a 2-fold reduction in RecA–ssDNA binding accompanied by the largest reduction (∼4 fold) in lagging-to-leading strand bias observed in this study. Even so, the strand bias remains sizable, with ∼6-fold more RecA bound to the lagging-strand (Table 1). Figure 1C and Supplementary Figures S8, S22 provide an interesting comparison. In wild type cells, replication in the two replichores largely converges on *Ter*C, where a peak of RecA binding is seen in Figure 1C. In the *recG* deletion mutant, there is instead an accumulation of RecA and to some extent SSB binding at *Ter*A and *Ter*B for the C- and W-strands, respectively, indicating a DNA metabolism event at these loci that generates ssDNA and requires RecG to resolve. The effects of RecG

mutants at the sites of replication termination have been explored by Rudolph and colleagues (42,48).

In RecA730 (E38K) cells constitutively induced for the SOS response there is an approximate 2-fold reduction in RecA binding in ssDNA and in strand bias. The RecA E38K protein has enhanced binding to SSB-coated ssDNA and to dsDNA (49,50). The binding to dsDNA may be responsible for the reduction in strand bias.

In *ssb-113* mutant cells there is a strong, ∼6-fold, reduction in RecA strand bias, but a ∼2-fold increase in the fraction of RecA bound in ssDNA – this is the only genetic background among the seventeen mutants studied during exponential growth that exhibited an increase in the fraction of ssDNA reads in RecA pull downs (31.3%) compared to wild-type cells (17.8%) (Table 1). Cells with the SSB-113 protein are temperature sensitive and sensitive to DNA damage (40). Replication is somewhat compromised in these cells, and there is likely to be an increase in ssDNA due to repeated replisome stalling during the early stages of replication, that could generate numerous gaps bound by RecA and SSB in the OriC region (Figure 2, Supplementary Figures S19 and S23). The mutation, a Pro to Ser substitution in the penultimate residue of SSB, affects the interaction of the SSB C-terminus with a range of other DNA repair proteins (51) including the replication machinery (52–54).

In the case of SSB, the binding fraction to ssDNA remains essentially unaffected (∼13%) in the presence of either Δ*recB* and Δ*recF* single- or Δ*recB*ΔrecF double mutants (13 to 15%) (Table 1). There is, however, a small (∼2-fold) reduction in lagging- to leading-strand bias the absence of RecB, which is not seen in Δ*recF* single mutants or Δ*recB*Δ*recF* double mutants (Table 1).

In contrast to RecA binding in ssDNA, fractions of SSB–ssDNA and strand bias are insensitive to the deletion of *recG* (Table 1), although a major shift in the location of major SSB–ssDNA peaks from *Ter*C to *Ter*A and *Ter*B is seen (Supplementary Figure S7 versus Figure 1B) as is the case for RecA. SSB–ssDNA binding and strand bias is also essentially unchanged in *recA730* cells. However, the fraction of SSB–ssDNA is increased by ∼3-fold in *ssb-113* cells along with a ∼2-fold reduction in strand bias. As observed for RecA–ssDNA binding, the *ssb-113* genetic background was the only mutant cell showing greater SSB–ssDNA occupancy (35.1%) compared to wild-type cells (13.0%) (Table 1). It is noteworthy that in *ssb-113* cells, ssDNA occupancy for SSB (35.1%) is essentially the same as that for RecA (31.3%), as is also true for wild-type cells SSB (13.0%) and RecA (17.8%). The increase in ssDNA reads in *ssb-113* cells may again reflect a general increase in genomic ssDNA in this strain.

The profiles presented in Figure 2 and Supplementary Figures S22 and S23 complement the data in Table 1 and allow further analysis of several features of the RecA and SSB binding profiles. We note again the reduction in the fraction of ssDNA reads in the RecA and SSB pull downs in a region of ∼80 kb on either side of OriC. This suggests the presence of what might be termed a RecA/SSB partial exclusion zone near the origin, perhaps reflecting the presence of ssDNA protection by proteins that bind in the region where replication is initiated. This effect is significantly enhanced, and the region expanded when *recF* is deleted, perhaps suggesting that RecF is involved in the generation of RecA/SSB binding substrates specific to the region. Surprisingly, essentially the opposite effect occurs in cells lacking *recO*, where exclusion around the origin largely disappears for both RecA and SSB. It appears that RecO may play a role in creating the exclusion zone. In *ssb-113* cells, the exclusion zone is replaced by a sizable concentration of RecA and SSB binding near the origin that could inhibit replication initiation.

Furthermore, there is enhanced binding of RecA and SSB on ssDNA near the replication terminus. A large RecA–ssDNA peak is found ∼20 kb at both sides of *dif* (Figures 1C and 2, black arrows). There is also a reduction in SSB peaks in several mutant strain profiles. The peaks disappear entirely in a *recB* mutant, which may suggest that for this case, that RecA binding could involve dsDNA break repair. Frequent dsDNA breaks could be occurring at *dif* sites that might arise from a need to resolve dimeric chromosomes during cell division (46,55). Deletion of XerD, which is known to be an essential component of chromosome dimer resolution (37), resulted in a large increase in RecA and SSB peaks on the lagging strands flanking *dif*. These peaks dominate the RecA–ssDNA and SSB–ssDNA profiles (Figure 2 & Supplementary Figure S22). As mentioned earlier, the peaks are spread out when *recG* is absent, which could result from an 'over-replication' in the Ter region (42). The height of the peaks increases in cells lacking UvrD, which plays a role in removing RecA protein from the DNA after it has acted (56).

These results differ substantially in stationary phase cells, where in the absence of normal chromosomal replication, the SSB–ssDNA profile is flat for W- and C-strands, while a preference for RecA–ssDNA binding to lagging-strand persists by ∼1.9-fold (Figure 2, Table 1).

High levels of ssDNA gaps on the leading strands at *Ter*A, *Ter*B, *Ter*C and *Ter*D

We now examine more closely the RecA and SSB ssDNA read peaks associated with the replication terminus. Replication fork arrest has been reported to take place during rapid cell growth in the 'non-permissive' directions mediated by the Tus-*Ter* complex at four *Ter* sites (57), *Ter*A, *Ter*D on one side and *Ter*C and *Ter*B on the other side (Figure 3A, *sketch*). The coverage maps for ssDNA showed high peaks for SSB–ssDNA at *Ter*C and *Ter*B on the W-strand and at *Ter*A and *Ter*D on the C-strand (Figure 1C, Supplementary Figures S1B-S18B). These peaks occur in ∼230 bp regions surrounding each *Ter* site, which correspond to the presence of leading-strand gaps (Figure 3A, B, Supplementary Table S1). However, there are no corresponding ssDNA gaps on the lagging-strands (Figure 3B, Supplementary Table S2). There is also a peak for RecA–ssDNA that is observed only at *Ter*C (Figure 3B, Supplementary Table S1). Each of the SSB and RecA peaks is eliminated in a Δ*tus* strain, which indicates that the ssDNA gap formation is strictly dependent on the presence of Tus protein (Figure 3A, Supplementary Table S1). The SSB–ssDNA and RecA–ssDNA peaks are also observed in stationary cells, suggesting that ssDNA gaps at *Ter* sites persist even after completion of chromosomal replication. The data are consistent

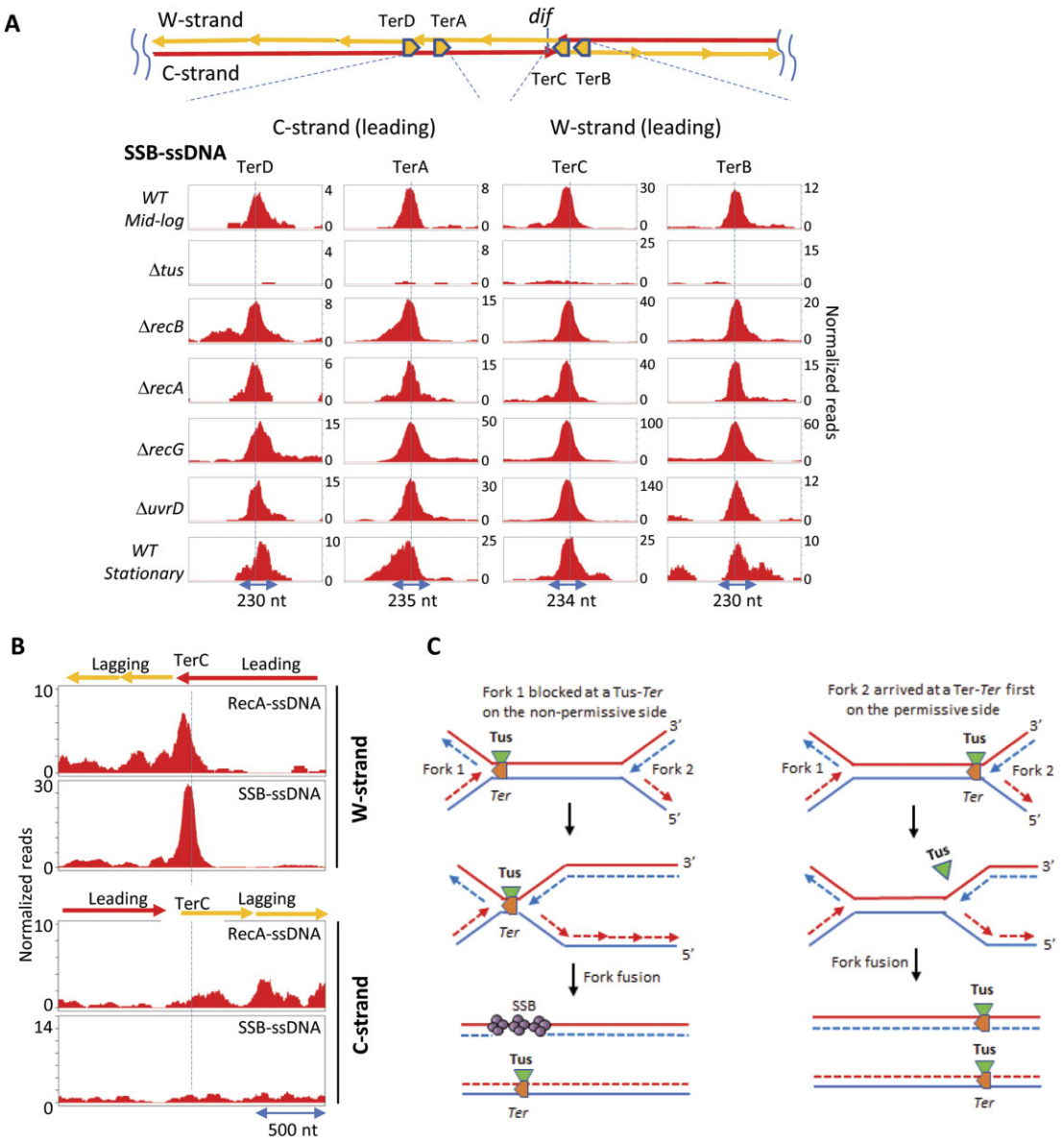


Figure 3. Formation of ssDNA gaps on the leading strands at *Ter*A, *Ter*B, *Ter*C and *Ter*D sites. (**A**) SSB–ssDNA coverage on the leading strands in 1 kb regions around the *Ter* sites. Peaks (∼230 nt) at *Ter*A, *Ter*B, *Ter*C and *Ter*D are evident in wild-type and selected mutants, but not in a *Δtus* strain. Numbers on Y-axis represent normalized mapped reads. (**B**) Zoom-in for a 2 kb region around *Ter*C with mapped RecA–ssDNA and SSB–ssDNA on the W- and C-strands in mid-log phase wild-type cells. Leading and lagging strands are indicated by solid red and orange lines with arrows indicating directions of the replication forks. (**C**) A model depicting formation of ssDNA gap at a *Ter* site when a replication fork is blocked at a Tus-*Ter* complex on the non-permissive side (left panel). Fork fusion at Tus-*Ter* results in a 230 nt gap on the leading strand (left panel). The absence of an ssDNA gap at a *Ter* site when a fork first approaches a Tus-*Ter* complex on the permissive side (left panel). The Tus protein is displaced allowing fork fusion to occur outside the *Ter* site.

with a model of 230 nt gap formation on the leading strand when two replisomes fused at a Tus-*Ter* site (Figure 3C).

If the presence of SSB is used as an approximate measure of the frequency of ssDNA gap formation, the relative frequencies of ssDNA gaps occurring can be estimated by taking the ratios of SSB–ssDNA read coverages at arbitrary sites to the average coverage. For mid-log wild-type cells, the frequency of ssDNA gap formation is highest at *Ter*C (59-fold increase), followed by *Ter*B (20-fold increase), *Ter*A (15-fold increase) and *Ter*D (9-fold increase) (Supplementary Table S1). In the absence of Tus (*Δtus*), a smaller 6.7-fold increase in ssDNA gaps was observed only at *Ter*C (Supplementary Table S1). It is important to note that these ChIP-seq signal enrichments at Ter sites might arise from a small subpopulation of cells containing fused replisomes at the Tus-*Ter* site (55,58–60).

Gap formation for each of the other mutants remained highest at *Ter*C, occurring in a wide range, *ssb-113* (25.4-fold increase), Δ*recG* (113-fold increase), and Δ*uvrD* (394-fold increase). Compared to wild-type cells, ssDNA gaps are increased significantly at each of the four *Ter* sits in cells lacking UvrD, and at *Ter*A, *Ter*D, and *Ter*C in cells

lacking RecG (Supplementary Table S1). Δ*recG* cells also contain SSB–ssDNA peaks at four distant *Ter* sites, *Ter*G and *Ter*J on the W-strand, and at *Ter*E and *Ter*H on the C-strand (Supplementary Figure S8). These distal *Ter* site enrichments were absent in wild-type cells and in the other mutants. High levels of ssDNA gaps observed in the Δ*recG* strain seem likely to be caused by aberrant replication initiation following fork mergers in the terminus region where replication could be blocked from progressing past the Tus-*Ter* sites (42,48,61). An increase in ssDNA gaps at the *Ter* site in Δ*uvrD* cells would be consistent with a reduction in the displacement of Tus from the Ter site. The UvrD helicase has been shown to involve in the removal of Tus protein during chromosome replication (58,62).

Despite high frequencies of ssDNA gap formation at *Ter*A-D sites, significant recruitment of RecA was observed only at *Ter*C (Figure 3B, Supplementary Table S1). The efficiency of RecA recruitment to ssDNA gaps at *Ter*C in wild-type and mutants were calculated as a ratio of RecA–ssDNA enrichment relative to SSB–ssDNA enrichment occurring at *Ter*C. For wildtype, the ratio value is 0.9 (Supplementary Table S1). *Ssb-113* along with Δ*radD*, Δ*recB*Δ*recF,* Δ*recB*Δ*recO*, Δ*recO* and Δ*recG* show the largest effects on RecA recruitment, reducing the RecA/SSB enrichment ratio to 0.1 to 0.3 (Supplementary Table S1).

Visualizing RecBCD-mediated RecA loading at dsDNA break sites

Nucleotide-level maps of RecA–ssDNA allow the visualization of RecA loading on ssDNA at fragile sites containing a high frequency of dsDNA breaks. In mid-log wild-type cells there are two large peaks (Figure 4A, *dif*-Left & *dif*-Right) located on the lagging strands flanking *dif*, which is the site at which sister chromosomes are segregated by XerCD site specific endonuclease and resolvase (63). Xer recombination is required to resolve chromosome dimers occurring in about 15% of rapidly growing cells (64). Recent genomic analyses revealed a high frequency of spontaneous dsDNA breaks occurring at or near *dif* (17,55,65). The first RecA peak (*dif*-Left) is located on the W-strand on the left side of *dif*, and the second peak (*dif*-Right) is located on the C-strand on the right side of *dif*. Each peak extends over a 24 – 26 kb region, beginning at the first Chi sites, which are located ∼20 kb at the left and right sides of *dif* (Figure 4A). Deletion of the XerD recombinase led to a large increase in RecA binding, characterized by a greater than 22-fold enrichment of RecA–ssDNA in the Δ*xerD* strain, compared to about a 3-fold enrichment for wild-type cells (Figure 4A, C, Supplementary Figure S22). These data are consistent with large increase of dsDNA breaks occurring at *dif* in daughter cells following dimer chromosome resolution in the absence of XerD (66).

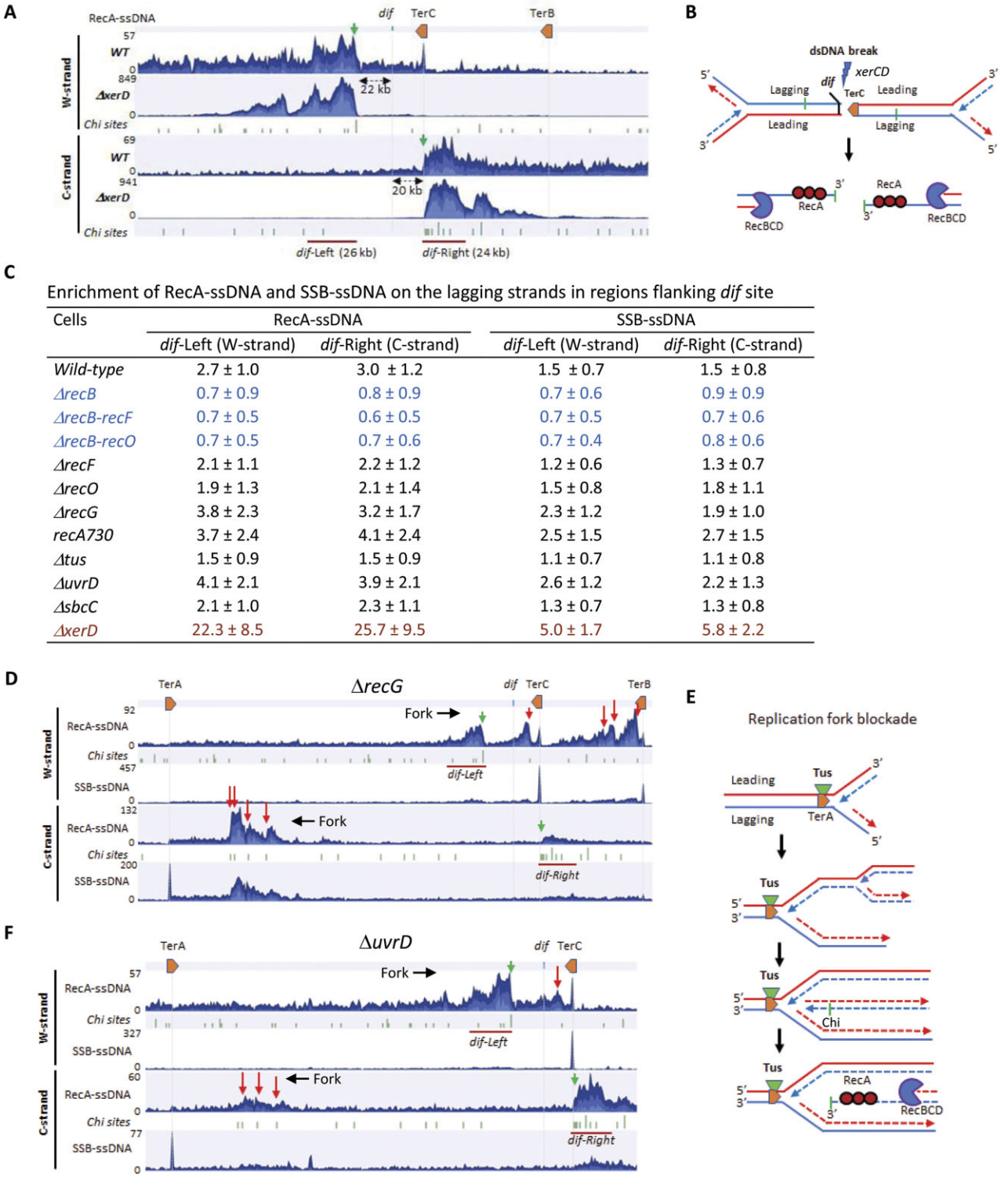
The absence of *dif*-Left & *dif*-Right peaks in Δ*recB,* Δ*recB*Δ*recF,* Δ*recB*Δ*recO* (Figure 4C) suggests an absolute requirement for RecBCD proteins. The location of the *dif*-Left and *dif*-Right peaks is consistent with RecBCD-mediated processing and RecA loading on resected 3′-ssDNA ends that are formed after dsDNA breaks at or near *dif* (Figure 4B). Higher levels of RecA–ssDNA enrichment at *dif*-Left and *dif*-Right in Δ*recG*, Δ*uvrD* and *recA730* backgrounds could reflect higher frequencies of dsDNA breaks at *dif* sites compared to wild-type cells. A somewhat smaller enrichment of SSB–ssDNA was also observed on the lagging-strands at the same *dif* loci, which suggests that SSB may also be recruited to 3′-ssDNA ends that form via resection by RecBCD (Figure 4C).

Helicase deletion mutants Δ*recG* and Δ*uvrD* exhibit extensive DNA over-amplification in the *Ter* region (Figures 2, 4D and F, Supplementary Figures S8, S15). Over-amplified forks, passing through a last Tus-*Ter* complex in the permissive direction would likely encounter a Tus-*Ter* complex located on the non-permissive face. Replication fork blockage by the Tus-*Ter* complex in the non-permissive direction would lead to formation of a single-ended dsDNA break, which can be processed subsequently by RecBCD. RecA could then be loaded onto the nascent ssDNA 3′-end starting at the Chi sites (Figure 4E). Mapping of RecA–ssDNA at the *Ter* region in Δ*recG* cells clearly reveal the presence of peaks that coincide with Chi sites located on the W-strand at left side of *Ter*C and *Ter*B and on the C-strand at the right side of *Ter*A. Similar peaks were observed at *Ter*C and *Ter*A in Δ*uvrD* cells (Figure 4D and 4, red arrows pointing down).

Dynamic changes in RecA and SSB interactions with genomic DNA following UV-irradiation

Dynamic changes of RecA–ssDNA and SSB–ssDNA landscapes after UV-irradiation are revealed by comparing snapshots of profiles taken at 1 min, 5 min, 10 min and 20 min after UV-irradiation (100 J/m^2) with those of non-UV-irradiated cells (Table 1, and Figure 5). There are large changes in RecA and SSB landscapes on gDNA that occur immediately after UV irradiation. For wild-type cells, at 1 min post UV, the fractions of bound ssDNA significantly decreased, from 17.8% to 3.9% for RecA, and from 13% to 8.4% for SSB. At the same time, the ratios of bound ssDNA on lagging/leading strands also decreased from 23.4 to 7.4 for RecA and 6.2 to 2.5 for SSB (Table 1, Figure 5D). The lagging/leading strand ratios of bound ssDNA continued to decrease at longer post UV times reaching values of 1.4–1.5 for RecA and SSB. In contrast, the fractions of bound ssDNA gradually recovered, attaining high levels at 20 min post UV for both RecA (55%) and SSB (71%) (Table 1, Figure 5D).

A reduction in the ratio of RecA and SSB bound to ssDNA on the lagging versus leading strand is consistent with ssDNA gap formation occurring on the leading strand caused by replication fork arrest or lesion-skipping at UV lesions. A similar trend is observed for Δ*recB* and Δ*recO* mutants, although the fractions of RecA-bound ssDNA is significantly lower in 20 min post UV-cells that results presumably by a deficiency in RecBCD- or RecFOR-dependent loading (Table 1, Figure 5D). Along with a large increase in RecA and SSB bound to ssDNA on the leading-strand post UV, there are large peaks of RecA–ssDNA and SSB–ssDNA on W- and C-strands covering 15 kb regions on both sides of OriC (Figure 5A–C, and Supplementary Figure S24). The presence of high concentrations of RecA and SSB bound to ssDNA proximal to OriC suggest that



Cells	RecA-ssDNA		SSB-ssDNA	
	dif-Left (W-strand)	*dif*-Right (C-strand)	*dif*-Left (W-strand)	*dif*-Right (C-strand)
Wild-type	2.7 ± 1.0	3.0 ± 1.2	1.5 ± 0.7	1.5 ± 0.8
ΔrecB	0.7 ± 0.9	0.8 ± 0.9	0.7 ± 0.6	0.9 ± 0.9
ΔrecB-recF	0.7 ± 0.5	0.6 ± 0.5	0.7 ± 0.5	0.7 ± 0.6
ΔrecB-recO	0.7 ± 0.5	0.7 ± 0.6	0.7 ± 0.4	0.8 ± 0.6
ΔrecF	2.1 ± 1.1	2.2 ± 1.2	1.2 ± 0.6	1.3 ± 0.7
ΔrecO	1.9 ± 1.3	2.1 ± 1.4	1.5 ± 0.8	1.8 ± 1.1
ΔrecG	3.8 ± 2.3	3.2 ± 1.7	2.3 ± 1.2	1.9 ± 1.0
recA730	3.7 ± 2.4	4.1 ± 2.4	2.5 ± 1.5	2.7 ± 1.5
Δtus	1.5 ± 0.9	1.5 ± 0.9	1.1 ± 0.7	1.1 ± 0.8
ΔuvrD	4.1 ± 2.1	3.9 ± 2.1	2.6 ± 1.2	2.2 ± 1.3
ΔsbcC	2.1 ± 1.0	2.3 ± 1.1	1.3 ± 0.7	1.3 ± 0.8
ΔxerD	22.3 ± 8.5	25.7 ± 9.5	5.0 ± 1.7	5.8 ± 2.2

Figure 4. Direct visualization of RecA loading on the 3′-end of ssDNA generated by RecBCD in exponentially growing *E. coli*. (**A**) RecA–ssDNA coverage maps at the Ter region in wild-type and Δ*xerD* cells. Two peaks of lagging strand RecA–ssDNA, a 26 kb region on the left side of *dif* (*dif*-Left) and a 24 kb region on the right side of *dif* (*dif*-Right) are indicated. Both RecA–ssDNA peaks start at the first Chi sites (downward green arrows) on the W- and C-strands. (**B**) Schematic representation for the processing of dsDNA ends and RecA loading after a dsDNA break at or near *dif*. End resection by RecBCD generates ssDNA 3′-end on the lagging strands at Chi sites followed by RecA loading. (**C**) RecBCD-dependent enrichment of RecA–ssDNA and SSB–ssDNA on the lagging strands in the *dif*-Left and *dif*-Right regions. (**D**) 'Over-replication' in the *Ter* region in a Δ*recG* mutant generates replication forks that encounter a Tus-*Ter* complex at *Ter*A, *Ter*C and *Ter*B in the non-permissive direction. Peaks of RecA–ssDNA on the W-strand at *Ter*C and *Ter*B, and on the C-strand at *Ter*A starting at Chi sites are indicated by red downward arrows. (**E**) Blockage of a replication fork at a Tus-*Ter* complex in the non-permissive direction creates single-ended dsDNA breaks. RecBCD resects dsDNA ends to generate a ssDNA 3′-end for RecA loading. (**F**) RecA–ssDNA peaks (red downward arrows) representing replication fork blockage in the non-permissive direction at *Ter*A and *Ter*C in a Δ*uvrD* mutant.

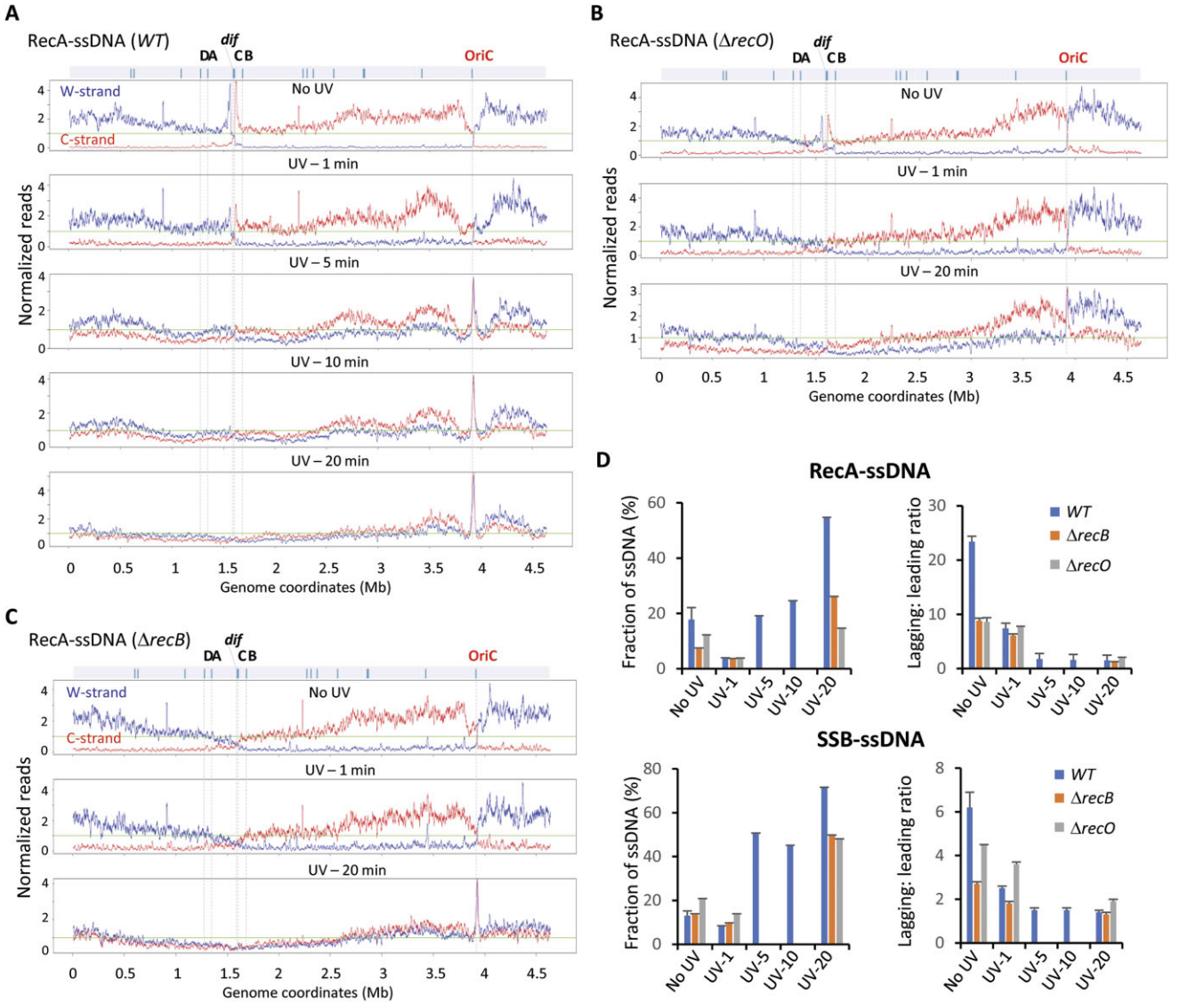
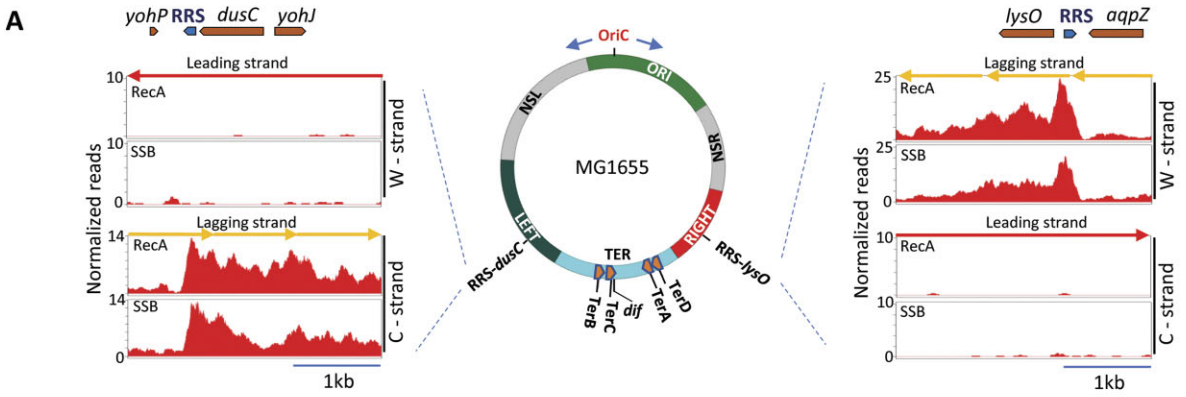


Figure 5. Dynamics of RecA and SSB loading on genomic ssDNA after UV-irradiation of *E. coli*. (A–C) Changes in RecA–ssDNA distribution on W- and C-strands before and after UV-irradiation in *WT* (**A**), Δ*recB* (**B**) and Δ*recO* (**C**). Normalized coverages of RecA–ssDNA are shown as blue traces for the W-strand, and red traces for the C-strand. Green lines (value = 1) refer to the genome average coverage. Each data point represents a moving average of RecA–ssDNA in 10 kb windows. For all graphs, the positions of the replication origin (OriC), *dif* and 14 *Ter* sites are shown as tick marks at the top. *Ter*A, *Ter*B, *Ter*C, *Ter*D, *Ter*I and *Ter*J are labeled as A, B, C, D, I and J, respectively. (**D**) Changes in the fraction of ssDNA (%) bound to RecA and SSB before and after UV-irradiation (UV-1 - 1 min post UV, UV-5 – 5 min post UV, UV-10 – 10 min post UV and UV-20 – 20 min post UV), and the ratios of RecA–ssDNA and SSB–ssDNA on the lagging vs leading strands. Data are shown as the mean ± standard deviation for 2 replichores.

DnaA-dependent replisome loading and firing could still be occurring at OriC, but with very little or no DNA synthesis occurring post UV (67,68).

Replication risk sequence (RRS)

We now return to the two major peaks of RecA and SSB bound to ssDNA on the lagging-strand proximal to genes *lysO* and *dusC* (Figures 1C and 2, green arrows). Both regions are situated symmetrically and ∼650 kbp to either side of *dif*. The peaks of RecA and SSB binding to ssDNA in these regions is substantial, indicating that ssDNA gaps are formed often in these regions. Each contains a nearly perfect copy, inverted, of a 222 bp sequence highly enriched for GC bp (Figure 6A, Supplementary Figure S25). We refer to this sequence as a replication risk sequence (RRS). When this region is expanded, the binding peaks of RecA–ssDNA and SSB–ssDNA are seen as maximal at the end of the repeats, and then extend further by about 1.4–2.5 kb (Figure 6A). These major peaks are also seen in cells lacking *recB*. However, the enrichment is reduced in cells lacking *recF* or *recG* (Figure 6B). These peaks are thus likely to reflect a process at least partly dependent on both the RecF and RecG proteins, most likely some aspect of post-replication gap repair. The repeats may trigger the formation of post-replication gaps at a rate greater than that seen in other parts of the genome. These RecA binding peaks are also reduced in the RecA730 mutant with its enhanced DNA binding proper-

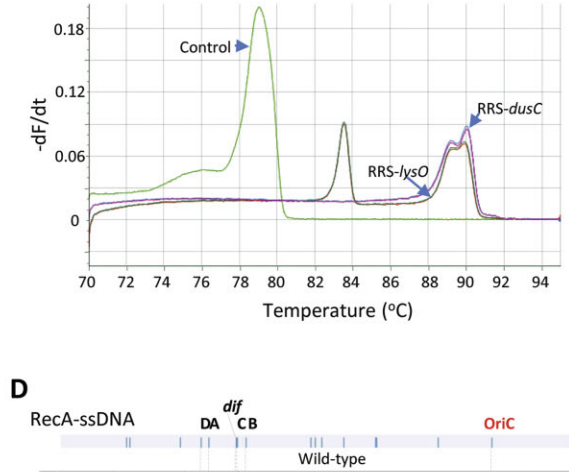


B

Enrichment (fold-increase) of RecA-ssDNA and SSB-ssDNA on the lagging strands at *lysO* and *dusC* loci*

Cells	RecA-ssDNA		SSB-ssDNA	
	lysO	*dusC*	*lysO*	*dusC*
Mid-log				
Wild-type	5.6 ± 2.2	3.5 ± 1.2	5.6 ± 2.1	3.2 ± 1.7
Δmfd	4.8 ± 2.2	3.8 ± 1.7	5.2 ± 2.9	3.7 ± 2.0
ΔmutS	5.3 ± 2.2	3.6 ± 1.7	4.9 ± 2.5	3.8 ± 2.5
ΔradD	6.5 ± 1.7	4.3 ± 1.7	7.8 ± 3.5	4.7 ± 2.2
ΔrecA	–	–	5.5 ± 2.8	3.4 ± 2.3
ΔrecB	5.9 ± 2.5	4.4 ± 2.3	5.4 ± 2.3	3.2 ± 1.8
ΔrecF	3.2 ± 0.9	2.4 ± 1.3	3.7 ± 2.0	2.5 ± 1.3
ΔrecG	3.7 ± 1.6	3.5 ± 1.6	4.7 ± 2.5	3.3 ± 1.7
ΔrecO	4.3 ± 1.9	3.4 ± 1.9	6.7 ± 3.4	3.2 ± 1.7
ΔrecBΔrecF	3.5 ± 0.9	3.3 ± 1.2	4.3 ± 1.7	3.3 ± 1.1
ΔrecBΔrecO	3.9 ± 1.2	3.0 ± 1.3	4.2 ± 2.1	2.8 ± 1.4
ΔsbcB (exo1)	5.6 ± 2.1	4.9 ± 1.9	6.0 ± 2.9	4.2 ± 1.9
ΔsbcC	4.4 ± 1.5	3.1 ± 1.3	5.2 ± 2.4	3.0 ± 1.4
Δtus	6.2 ± 2.7	4.8 ± 1.9	7.0 ± 2.9	4.1 ± 2.2
ΔuvrD	4.9 ± 1.7	2.9 ± 1.8	4.3 ± 2.4	2.6 ± 1.3
recA730	2.8 ± 0.8	2.1 ± 1.0	3.2 ± 1.5	2.3 ± 1.1
recAΔC17	3.7 ± 1.2	2.5 ± 0.9	3.3 ± 1.9	2.4 ± 1.0
ssb-113	5.0 ± 2.8	2.5 ± 1.2	7.3 ± 4.3	3.2 ± 1.8
ΔRRS-lysO	0.9 ± 0.5	4.1 ± 1.3	1.4 ± 0.5	3.2 ± 1.3
ΔRRS-dusC	4.4 ± 1.7	0.9 ± 0.4	5.2 ± 2.1	1.3 ± 0.7
Stationary				
Wild-type	2.1 ± 1.1	1.7 ± 1.6	1.1 ± 0.8	1.4 ± 1.3

* *lysO* site represents 1.4 kb region including RRS sequence and *lysO* gene, *dusC* site represents 2.4 kb region including RRS, *dusC* and *yohJ* genes



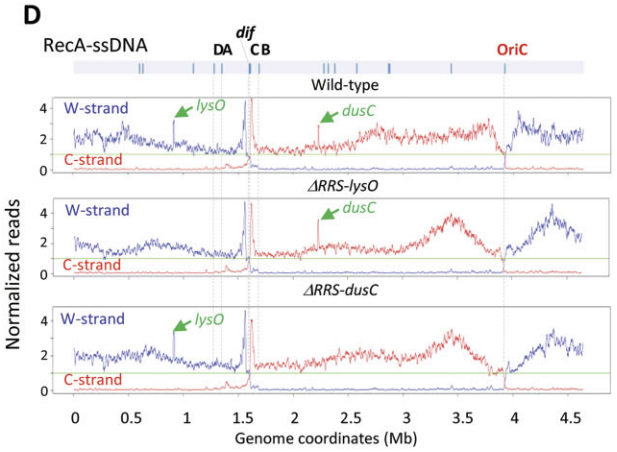


Figure 6. Peaks of RecA–ssDNA and SSB–ssDNA on the lagging strands at *lysO* and *dusC* loci. (**A**) Distribution of mapped RecA–ssDNA and SSB–ssDNA at *lysO* and *dusC* regions. Numbers on the Y-axis represent normalized coverages of mapped reads on W- and C-strands. Sketch in the middle depicts positions of RRS-*lys*O, RRS-*dusC*, *dif*, *Ter* sites, four macrodomains (ORI, LEFT, RIGHT, TER) and two non-structured domains (NSL and NSR). The leading and lagging strands are shown as solid red lines with arrows and orange lines with arrows pointing in the direction of replication. Blue arrow in *lysO* and *dusC* regions indicate positions and orientation of 222 bp imperfect inverted repeat RRS (Replication Risk Sequences). (**B**) Enrichments of RecA–ssDNA and SSB–ssDNA on the lagging strands at *lysO* and *dusC* loci in wild-type and mutant cells. The RecA–ssDNA (or SSB–ssDNA) enrichment levels (mean ± standard deviation) were calculated as fold-increases in the read coverage on the lagging strands at *lysO* and *dusC* loci relative to the average coverages for the entire lagging strands. (**C**) Melting profiles of RRS-*lysO*, RRS-*dusC* and a control sequence. Melting kinetics were monitored by changes in SYBR Green fluorescence as a function of time (-dF/dt) *vs*. temperature. The control represents a sequence having a similar length located 60 bp upstream of RRS-*dusC*. Melting curves for RRS-*lysO* and RRS-*dusC* are shown in duplicate. (**D**) RRS are required for the presence of RecA peaks at *lysO* and *dusC* loci. Wild-type cells contained two RecA–ssDNA peaks, one at *lysO* and one at *dusC* loci (green arrows), whereas a Δ*RRS-lysO* strain contains a single RecA peak at the *dusC* site, and a Δ*RRS-dusC* strain contains a single RecA peak only at the *lysO* site.

ties. Similar but somewhat less prominent peaks are seen in the SSB profiles. The sequence of the GC-rich repeated elements has a great potential for the formation of secondary structure when the region is unwound by DnaB or other helicases. Melting analysis (Figure 6C) showed the presence of two distinct melting peaks at high temperatures (89°C and 90°C) for RRS-*dusC*, and three melting peaks (83.5°C, 89°C and 90°C) for RRS-*lysO*. In contrast, a control sequence having a similar length located 60 bp upstream of RRS-*dusC* has only one dominant melting peak. The presence of multiple melting peaks for RRS sequences indicate that they can readily form stable secondary structures, which may adversely affect DNA replication, especially on the lagging strands. Cloned RRS cannot be sequenced using standard commercial Sanger DNA sequencing protocols and require specialized protocols employing increased heat.

Sequences highly homologous to RRS are widely distributed in enterobacteria. They are generally found in pairs flanking the genomic Ter region, although some species have one or more additional RRS (Supplementary Figure S25). RRS are required for the presence of RecA and SSB peaks at *lysO* and *dusC* loci. Single deletions of each of the RRS in *E. coli* were constructed using the λ Red-based method (69). The ssGap-seq data showed the absence of RecA and SSB peaks at the *lysO* locus in a Δ*RRS-lysO* strain, whereas a Δ*RRS-dusC* strain lacks RecA and SSB peaks at the *dusC* site (Figure 6B, D, Supplementary Figure S26). This indicates that the RRS are in fact responsible for the peaks of enhanced RecA loading onto adjacent ssDNA. To date, we have been unable to construct and confirm a strain lacking both RRS. This may indicate that the presence of at least one RRS is important but additional work is required for confirmation.

DISCUSSION

There are three primary products of this work. First, we describe a new method, ssGap-seq, that permits the mapping of proteins bound to ssDNA on a genomic scale. The other two products represent two categories of results that arose from extensive application of the method. The first of these were patterns that are expected based on past research. Unsurprising or not, the expected patterns provide an important validation of the method. The second category constitutes results that were entirely novel and which are largely inaccessible using other methods.

Among the expected findings are: (a) a strong bias of RecA and SSB on the lagging strand template, as might be expected based on the patterns RecA-mediated repair processes, (b) the profiling that shows a lack of protein loading bias in stationary phase and an increase in RecA and SSB loading after UV irradiation, and (c) the high levels of RecA and SSB loading near the terminus and the large effects of that loading seen in mutant cells lacking RecG function. Although unpredicted, the substantial concentration of RecA and SSB-bound ssDNA near the origin in the *ssb-113* strain is probably to be expected as well. Both double strand break repair (involving either leading or lagging strand replisome encounters with template discontinuities) or post-replication gap repair on the lagging strand template would focus RecA binding on the lagging strand template. Higher levels of DNA damage should trigger higher levels of RecA-mediated DNA repair, thus explaining the observed increases in RecA binding to ssDNA observed. Unscheduled replication initiation occurs at the terminus when RecG activity is missing (42,48,61), likely explaining the increased concentration of RecA and SSB-bound ssDNA near the terminus. SSB-113 is a mutant SSB protein with a substitution of Ser for Pro at the penultimate residue 176. This change in a C-terminal residue renders growth temperature-sensitive and weakens interactions between SSB and a variety of proteins including ExoI (51) and the replication machinery (52–54). The results in Figure 2 and Supplementary Figure S23 show that in *ssb-113* cells, replication stalls repeatedly during its early stages, leaving numerous gaps that are bound by RecA and SSB.

The novel findings are numerous. Notable ones discussed further below include: (1) the presence of an exclusion zone for RecA and SSB-bound gaps near the replication origin, (2) a clear distinction between results obtained with *recF* and *recO* mutant cells, (3) the observation that the lagging strand bias for RecA binding is reduced but not eliminated in mutants lacking both RecB and RecO or RecF, (4) the patterns of RecA distribution observed in the absence of RecG or Tus, (5) the substantial deposition of RecA and SSB at the terminus in mutant cells lacking XerD function, and (6) the discovery of a pair of genomic repeats that we have designated RRS (for replication risk sequence) that trigger substantial deposition of RecA protein in their immediate vicinity. Each of these represents a potential entry point for significant new research paths. We will treat each of these separately and in succession.

(1) *A RecA/SSB exclusion zone on ssDNA near the origin of replication*. At this point, we can only speculate about the exclusion of ssDNA-bound RecA or SSB near the origin. The regulation of replication initiation is highly complex (70) involving DNA sites a considerable distance from *oriC*. The exclusion of RecA and SSB binding to ssDNA is limited to a region encompassing around 100 kbp around the origin and is partial. Understanding its functional significance will require considerable additional work.

(2) *Distinctive results with RecF versus RecO.* The RecA/SSB exclusion zone is expanded in cells lacking RecF and RecG (but nearly eliminated in cells lacking RecO (Figure 2, Supplementary Figure S22). In addition, the deposition of RecA near the RRS sites (described below) is reduced in *recF* strains but not in *recO* strains (Figure 6B). The *recF*, *recO* and *recR* genes have been defined as an epistasis group involved in the loading of RecA onto SSB-coated ssDNA gaps based on many different observations (71–80). However, more recent studies have begun to phenotypically distinguish the RecF and RecO proteins that may reflect some distinct roles in DNA metabolism. In general, the RecF protein is implicated in targeting of gap repair to lesion-containing post-replication gaps and may act via an interaction with replisomal proteins. RecO is needed to load RecA protein there once the gaps are identified. Resistance to particular DNA damaging agents is more dependent on RecF than RecO and vice versa (81–83).

Of the RecFOR proteins, only the RecF protein is toxic to cells when over-expressed (84–86). This represents a key and dramatic distinction between RecO and RecF. There is also a growing literature linking RecF protein to a function at the replisome (82,87–89). The recent ability to visualize single molecules in living cells has demonstrated that RecF and RecO do not colocalize and exhibit very different spatiotemporal behavior, with RecF often co-localizing with the replisome (82). RecG plays a role both in post-replication gap repair (90) and in replication termination (42,48,61). The current results suggests that the repair of post-replication gaps plays a role early in replication that is somehow distinct from its role in later stages of the replication cycle. RecO may play some special role in establishing the exclusion zone.

(3) *The lagging strand bias does not entirely reflect RecA loading by RecFOR or RecBCD.* Combining *recF*/*recO* mutants with Δ*recB* mutants reduces but does not eliminate the lagging strand bias for RecA binding to ssDNA. As the Δ*recF*Δ*recB* and Δ*recO*Δ*recB* double mutants should eliminate the known RecA loading pathways onto SSB-coated ssDNA, the results may indicate the existence of a third RecA loading pathway not yet elucidated. A third pathway has been proposed by Zahradka and colleagues (45).

(4) *Patterns of RecA distribution in the absence of RecG or-Tus; implications for genome stability.* Genomic ssDNA binding profiles of RecA and SSB tend to coincide globally, except for the replication terminal regions (Figure 1D, Supplementary Figures S2C to S4C, S6C to S20C). The terminal regions have been observed to be a source of genetic instability, which is typically associated with dsDNA breaks (46,55,65). As expected, there are higher levels of RecA–ssDNA compared to SSB–ssDNA in this region. The RecA deposition is highly dependent on RecB, indicating RecA recruitment to 3′-ssDNA ends processed by RecBCD. In addition to dsDNA breaks, the presence of prominent SSB peaks bound to the leading strands (Figure 3) indicates the frequent formation of ssDNA gaps in a 230 bp region surrounding *Ter*A, *Ter*B, *Ter*C and *Ter*D. The requirement for Tus protein suggests that gaps at *Ter* sites are likely to form when a Tus-*Ter* complex encounters a first replisome on its non-permissive face, blocking fork progression and forcing fork fusion at the *Ter* site when a second replisome arrives on the permissive face (Figure 3C). Based on the enrichment of SSB–ssDNA (Supplementary Table S1), it appears that most of *Ter* fork fusion events occur at *Ter*C, followed by *Ter*B, *Ter*A and *Ter*D (Supplementary Table S3). The highest frequency of fork fusion at *Ter*C coincides with the strongest paused fork signal measured at *Ter*C in mid-log cells (57). Incomplete replication at a *Ter* site leaving behind an ssDNA gap has been recently attributed to two forks colliding at Tus-*TerB*, based on a biochemical data obtained using a reconstituted plasmid replication system (91). Since ssDNA gaps observed *in vitro* are short ∼15 to 24 nt, which corresponds approximately to the footprint of a Tus-*Ter* complex, they may arise from DNA synthesis blockage by a non-displaced Tus on DNA (91). In contrast, the observed ssDNA gaps (∼230 nt) at *Ter* sites *in vivo* are much larger than the footprint of a Tus-*Ter* complex (92), indicating that replication by both replisomes may have stopped ∼100 nt prior to reaching a Tus-*Ter* site. The shapes of SSB–ssDNA peaks centering around *Ter* sites in wild-type and mutants (Figure 3A), suggest the possibility that Tus protein had been displaced from *Ter* sites to allow binding by SSB. The SSB–ssDNA binding profiles further suggest that fork fusions at *Ter* sites produce ssDNA gaps only on one strand, corresponding to the leading strand in the permissive direction (Figure 3B). It is unclear why a ssDNA gap forms on only one strand at *Ter* sites, but this appears to be a universal feature because leading strand ssDNA gaps are observed at *Ter* sites in all mutant cells (Supplementary Tables S1, S2). The high frequencies of *Ter* ssDNA gaps in stationary cells suggest that gap repair is slow or perhaps absent prior to cell division.

(5) *Patterns of RecA distribution in the absence of XerD; Implications for genome stability.* The deposition of RecA and SSB on ssDNA at sites near the replication terminus in *xerD* mutants is dramatic, by far the most extensive seen in this study (Figure 2, Supplementary Figures S16 and S22). XerD, acting with XerC, is a site-specific recombinase that resolves chromosomal dimers arising via recombination behind the fork during replication (93). Conversion of dimers to monomers is essential for cell survival as the post-replication gap repair via recombination behind the fork, with its potential to create chromosomal dimers, may occur multiple times in every replication cycle (90,94–97). Perhaps the only other process with the capacity to convert the dimers to monomers is RecA-mediated recombination, creating Holliday junctions resolved by RuvC. The presence of so much RecA suggests that, in the absence of XerD, the cells may utilize this secondary albeit less efficient mechanism to resolve chromosomal dimers. Backing up the missing XerCD function with a RecA-mediated recombination would be potentially complex. If RecA-mediated recombination happens in multiple places near the terminus, it might not always yield viable chromosomal products, explaining the growth defects seen in xerCD mutant cells (66). The results here may be providing us with a glimpse of that secondary chromosomal resolution process.

(6) *RRS, a novel genomic repeat element important for replication termination and/or chromosome segregation.* We finish with the RRS, which deserve special mention. The 222 bp RRS sequences are almost perfectly repeated, in inverted fashion at genomic locations located 1.3 Mb apart and symmetrically arranged around *dif*, 650 kb to either side. As seen in Figure 6, these repeats are directly responsible for the greatly enhanced RecA deposition. The pattern of RecA deposition, always on the lagging strand, suggests that a lagging strand replisome is often halted at these repeats, disengaging and leaving behind a substantial gap into which RecA is loaded. The region of RecA loading enhancement is about 2 kb or the length of a long Okazaki fragment. The pattern of deposition, with RecA most concentrated near RRS and trailing off

from there, is consistent with this proposed mechanism of gap creation. These patterns led to the RRS designation, as replication is often compromised at these sites. The two RRS locations are in regions defined previously as being within nucleoid organization LEFT and RIGHT macrodomains, close to the boundaries defined for the TER macrodomain (20). The *matS* sites and the binding sequence for MatP protein involved in TER macrodomain organization and DNA condensation (21) are arrayed between two RRS sites (Supplementary Figure S25B). RRS sequences are widely conserved in enterobacteria in both sequence and genomic positioning. In Kluyvera species, which tend to be about a 25–30% sequence conservation relative to *E. coli*, the RRS sequences are nearly 100% identical. Some species have more than two RRS but always seem to have at least one pair organized as inverted repeats. Their locations often encompass *matS* signature motifs for the TER macrodomain (Supplementary Figure S25), suggesting RRS sequences could play a role in TER macrodomain function in many bacteria. The RecA and SSB deposition in these short DNA segments indicates that the region near RRS is often in a single-stranded state. The RRS repeats are GC-rich and contain substantial regions of potential secondary structure. Successful sequencing of cloned RRS requires specialized Sanger sequencing protocols in which heat is applied. As shown in Figure 6, it is possible to delete one or the other RRS with no apparent effects. However, in spite of much effort, we have to date been unable to construct a mutant strain lacking both RRS, suggesting that retaining at least one is essential. One possibility we have considered is that persistent ssDNA at these locations may create transient swivels that could facilitate some of the steps required to resolve topological issues that accompany replication termination (98). The sites would thereby act as a kind of topological relief valve. These regions would need to be replicated at some point in each cell cycle. The loading of RecA at these sites may facilitate the disruption of secondary structure and the replication of RRS when the swivel is no longer needed. Other roles in facilitating chromosome condensation or other features of Ter macrodomain function are also possibilities that remain to be explored. The RRS appear to play an important role in replication termination and/or chromosome segregation. The discovery of RRS provides an important example of new areas of research opened up by the new ssGAP-Seq method.

The ssGap-seq method

ssGap-seq can be broadly applied. It has the potential to add much to our knowledge of the ssDNA landscape in the *E. coli* genome and the proteins involved in the resolution of single strand gaps. Unsurprisingly, RecA and SSB are often found associated with ssDNA. The new information appears in the details. Going forward, ssGap-seq should provide access to information not previously available. It should enable the characterization of DNA metabolic events involving ssDNA intermediates, especially when used in combination with different genetic backgrounds. With respect to ssDNA gap occupancy, the method is not at all limited to the RecA and SSB proteins. As seen with the RRS sites detected in this study, ssGap-seq also has the potential to define new aspects of bacterial genomic structure and function.

DATA AVAILABILITY

Raw next-generation sequencing data have been deposited to SRA database (SRA Submission ID: SUB12166727, BioProject ID: PRJNA891728).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Institutes of General Medical Sciences [1RM1GM130450 to M.M.C., M.F.G.]; National Institute of Environmental Health Sciences [R35ES028343 to M.F.G.]. Funding for open access charge: NIH [1RM1GM130450].
Conflict of interest statement. None declared.

REFERENCES

1. Shereda,R.D., Kozlov,A.G., Lohman,T.M., Cox,M.M. and Keck,J.L. (2008) SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.*, **43**, 289–318.
2. Rupp,W.D. and Howard-Flanders,P. (1968) Discontinuities in the DNA synthesized in an excision-defective strain of Escherichia coli following ultraviolet irradiation. *J. Mol. Biol.*, **31**, 291–304.
3. Okazaki,R., Okazaki,T., Sakabe,K., Sugimoto,K. and Sugino,A. (1968) Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc. Natl. Acad. Sci. U.S.A.*, **59**, 598–605.
4. Marians,K.J. (2018) Lesion bypass and the reactivation of stalled replication forks. *Annu. Rev. Biochem.*, **87**, 217–238.
5. Langston,L.D. and O'Donnell,M. (2006) DNA replication: keep moving and don't mind the gap. *Mol. Cell*, **23**, 155–160.
6. Krokan,H.E. and Bjoras,M. (2013) Base excision repair. *Cold Spring Harb. Perspect. Biol.*, **5**, a012583.
7. Sancar,A. and Tang,M.S. (1993) Nucleotide excision repair. *Photochem. Photobiol.*, **57**, 905–921.
8. Modrich,P. and Lahue,R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.*, **65**, 101–133.
9. Cox,M.M., Goodman,M.F., Kreuzer,K.N., Sherratt,D.J., Sandler,S.J. and Marians,K.J. (2000) The importance of repairing stalled replication forks. *Nature*, **404**, 37–41.
10. Wong,R.P., Petriukov,K. and Ulrich,H.D. (2021) Daughter-strand gaps in DNA replication - substrates of lesion processing and initiators of distress signalling. *DNA Repair (Amst.)*, **105**, 103163.
11. McCullough,A.K., Dodson,M.L. and Lloyd,R.S. (1999) Initiation of base excision repair: glycosylase mechanisms and structures. *Annu. Rev. Biochem.*, **68**, 255–285.
12. Sancar,A. (1996) DNA excision repair. *Annu. Rev. Biochem.*, **65**, 43–81.
13. Fuchs,R.P. (2016) Tolerance of lesions in *E. coli*: chronological competition between translesion synthesis and damage avoidance. *DNA Repair (Amst.)*, **44**, 51–58.
14. Lovett,S.T. (2017) Template-switching during replication fork repair in bacteria. *DNA Repair (Amst.)*, **56**, 118–128.
15. Kuzminov,A. (1999) Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol. Mol. Biol. Rev.*, **63**, 751–813.
16. Cox,M.M. (2002) The nonmutagenic repair of broken replication forks via recombination. *Mutat. Res.*, **510**, 107–120.

17. Michel,B., Sinha,A.K. and Leach,D.R.F. (2018) Replication fork breakage and restart in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.*, **82**, e00013-18.
18. Anderson,D.G. and Kowalczykowski,S.C. (1997) The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell*, **90**, 77–86.
19. Pham,P., Shao,Y., Cox,M.M. and Goodman,M.F. (2022) Genomic landscape of single-stranded DNA gapped intermediates in Escherichia coli. *Nucleic Acids Res.*, **50**, 937–951.
20. Valens,M., Penaud,S., Rossignol,M., Cornet,F. and Boccard,F. (2004) Macrodomain organization of the Escherichia coli chromosome. *EMBO J.*, **23**, 4330–4341.
21. Mercier,R., Petit,M.A., Schbath,S., Robin,S., El Karoui,M., Boccard,F. and Espeli,O. (2008) The MatP/matS site-specific system organizes the terminus region of the E. coli chromosome into a macrodomain. *Cell*, **135**, 475–485.
22. Kuo,M.H. and Allis,C.D. (1999) In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods*, **19**, 425–433.
23. Cox,M.M. (2001) Historical overview: searching for replication help in all of the rec places. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 8173–8180.
24. Syeda,A.H., Dimude,J.U., Skovgaard,O. and Rudolph,C.J. (2020) Too much of a good thing: how ectopic DNA replication affects bacterial replication dynamics. *Front. Microbiol.*, **11**, 534.
25. Pugh,B.F. and Cox,M.M. (1988) General mechanism for RecA protein binding to duplex DNA. *J. Mol. Biol.*, **203**, 479–493.
26. Clendenning,J.B. and Schurr,J.M. (1994) A model for the binding of E. coli single-strand binding protein to supercoiled DNA. *Biophys. Chem.*, **52**, 227–249.
27. Shan,Q. and Cox,M.M. (1997) RecA filament dynamics during DNA strand exchange reactions. *J. Biol. Chem.*, **272**, 11063–11073.
28. Webb,B.L., Cox,M.M. and Inman,R.B. (1997) Recombinational DNA repair: the RecF and RecR proteins limit the extension of RecA filaments beyond single-strand DNA gaps. *Cell*, **91**, 347–356.
29. Cox,M.M. (2007) Regulation of bacterial RecA protein function. *Crit. Rev. Biochem. Mol. Biol.*, **42**, 41–63.
30. Sweasy,J.B., Witkin,E.M., Sinha,N. and Roegner-Maniscalco,V. (1990) RecA protein of *Escherichia coli* has a third essential role in SOS mutator activity. *J. Bacteriol.*, **172**, 3030–3036.
31. Whitby,M.C., Vincent,S.D. and Lloyd,R.G. (1994) Branch migration of Holliday junctions: identification of RecG protein as a junction specific DNA helicase. *EMBO J.*, **13**, 5220–5228.
32. Kumura,K. and Sekiguchi,M. (1984) Identification of the uvrD gene product of *Escherichia coli* as DNA helicase II and its induction by DNA-damaging agents. *J. Biol. Chem.*, **259**, 1560–1565.
33. Connelly,J.C. and Leach,D.R. (1996) The sbcC and sbcD genes of *Escherichia coli* encode a nuclease involved in palindrome inviability and genetic recombination. *Genes Cells*, **1**, 285–291.
34. Su,S.S. and Modrich,P. (1986) Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 5057–5061.
35. Selby,C.P., Witkin,E.M. and Sancar,A. (1991) Escherichia coli mfd mutant deficient in "mutation frequency decline" lacks strand-specific repair: in vitro complementation with purified coupling factor. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 11574–11578.
36. Chen,S.H., Byrne,R.T., Wood,E.A. and Cox,M.M. (2015) Escherichia coli radD (yejH) gene: a novel function involved in radiation resistance and double-strand break repair. *Mol. Microbiol.*, **95**, 754–768.
37. Blakely,G., May,G., McCulloch,R., Arciszewska,L.K., Burke,M., Lovett,S.T. and Sherratt,D.J. (1993) Two related recombinases are required for site-specific recombination at dif and cer in E. coli K12. *Cell*, **75**, 351–361.
38. Hill,T.M. and Marians,K.J. (1990) Escherichia coli Tus protein acts to arrest the progression of DNA replication forks in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 2481–2485.
39. Lusetti,S.L., Wood,E.A., Fleming,C.D., Modica,M.J., Korth,J., Abbott,L., Dwyer,D.W., Roca,A.I., Inman,R.B. and Cox,M.M. (2003) C-terminal deletions of the Escherichia coli RecA protein. Characterization of in vivo and in vitro effects. *J. Biol. Chem.*, **278**, 16372–16380.
40. Chase,J.W., L'Italien,J.J., Murphy,J.B., Spicer,E.K. and Williams,K.R. (1984) Characterization of the *Escherichia coli* SSB-113 mutant single-stranded DNA-binding protein. Cloning of the gene, DNA and protein sequence analysis, high pressure liquid chromatography peptide mapping, and DNA-binding studies. *J. Biol. Chem.*, **259**, 805–814.
41. Blattner,F.R., Plunkett,G. 3rd, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F. *et al.* (1997) The complete genome sequence of Escherichia coli K-12. *Science*, **277**, 1453–1462.
42. Rudolph,C.J., Upton,A.L., Stockum,A., Nieduszynski,C.A. and Lloyd,R.G. (2013) Avoiding chromosome pathology when replication forks collide. *Nature*, **500**, 608–611.
43. Ivanova,D., Taylor,T., Smith,S.L., Dimude,J.U., Upton,A.L., Mehrjouy,M.M., Skovgaard,O., Sherratt,D.J., Retkute,R. and Rudolph,C.J. (2015) Shaping the landscape of the Escherichia coli chromosome: replication-transcription encounters in cells with an ectopic replication origin. *Nucleic Acids Res.*, **43**, 7865–7877.
44. Seigneur,M., Ehrlich,S.D. and Michel,B. (2000) RuvABC-dependent double-strand breaks in dnaBts mutants require recA. *Mol. Microbiol.*, **38**, 565–574.
45. Buljubasic,M., Hlevnjak,A., Repar,J., Dermic,D., Filic,V., Weber,I., Zahradka,K. and Zahradka,D. (2019) RecBCD-RecFOR-independent pathway of homologous recombination in Escherichia coli. *DNA Repair (Amst.)*, **83**, 102670.
46. Cockram,C.A., Filatenkova,M., Danos,V., El Karoui,M. and Leach,D.R. (2015) Quantitative genomic analysis of RecA protein binding during DNA double-strand break repair reveals RecBCD action in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E4735–E4742.
47. Sinha,A.K., Durand,A., Desfontaines,J.M., Iurchenko,I., Auger,H., Leach,D.R.F., Barre,F.X. and Michel,B. (2017) Division-induced DNA double strand breaks in the chromosome terminus region of *Escherichia coli* lacking RecBCD DNA repair enzyme. *PLos Genet.*, **13**, e1006895.
48. Midgley-Smith,S.L., Dimude,J.U., Taylor,T., Forrester,N.M., Upton,A.L., Lloyd,R.G. and Rudolph,C.J. (2018) Chromosomal over-replication in Escherichia coli recG cells is triggered by replication fork fusion and amplified if replichore symmetry is disturbed. *Nucleic Acids Res.*, **46**, 7701–7715.
49. Lavery,P.E. and Kowalczykowski,S.C. (1992) Biochemical basis of the constitutive repressor cleavage activity of recA730 protein. A comparison to recA441 and recA803 proteins. *J. Biol. Chem.*, **267**, 20648–20658.
50. Henrikus,S.S., Henry,C., McGrath,A.E., Jergic,S., McDonald,J.P., Hellmich,Y., Bruckbauer,S.T., Ritger,M.L., Cherry,M.E., Wood,E.A. *et al.* (2020) Single-molecule live-cell imaging reveals RecB-dependent function of DNA polymerase IV in double strand break repair. *Nucleic Acids Res.*, **48**, 8490–8508.
51. Lu,D. and Keck,J.L. (2008) Structural basis of Escherichia coli single-stranded DNA-binding protein stimulation of exonuclease I. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 9169–9174.
52. Genschel,J., Curth,U. and Urbanke,C. (2000) Interaction of E. coli single-stranded DNA binding protein (SSB) with exonuclease I. The carboxy-terminus of SSB is the recognition site for the nuclease. *Biol. Chem.*, **381**, 183–192.
53. Kelman,Z., Yuzhakov,A., Andjelkovic,J. and O'Donnell,M. (1998) Devoted to the lagging strand-the subunit of DNA polymerase III holoenzyme contacts SSB to promote processive elongation and sliding clamp assembly. *EMBO J.*, **17**, 2436–2449.
54. Yuzhakov,A., Kelman,Z. and O'Donnell,M. (1999) Trading places on DNA–a three-point switch underlies primer handoff from primase to the replicative DNA polymerase. *Cell*, **96**, 153–163.
55. Mei,Q., Fitzgerald,D.M., Liu,J., Xia,J., Pribis,J.P., Zhai,Y., Nehring,R.B., Paiano,J., Li,H., Nussenzweig,A. *et al.* (2021) Two mechanisms of chromosome fragility at replication-termination sites in bacteria. *Sci. Adv.*, **7**, eabe2846.
56. Petrova,V., Chen,S.H., Molzberger,E.T., Tomko,E., Chitteni-Pattu,S., Jia,H., Ordabayev,Y., Lohman,T.M. and Cox,M.M. (2015) Active displacement of RecA filaments by UvrD translocase activity. *Nucleic Acids Res.*, **43**, 4133–4149.
57. Duggin,I.G. and Bell,S.D. (2009) Termination structures in the Escherichia coli chromosome replication fork trap. *J. Mol. Biol.*, **387**, 532–539.
58. Bidnenko,V., Lestini,R. and Michel,B. (2006) The Escherichia coli UvrD helicase is essential for Tus removal during recombination-dependent replication restart from Ter sites. *Mol. Microbiol.*, **62**, 382–396.

59. Kodama,K., Kobayashi,T., Niki,H., Hiraga,S., Oshima,T., Mori,H. and Horiuchi,T. (2002) Amplification of Hot DNA segments in Escherichia coli. *Mol. Microbiol.*, **45**, 1575–1588.
60. Levy,A., Goren,M.G., Yosef,I., Auster,O., Manor,M., Amitai,G., Edgar,R., Qimron,U. and Sorek,R. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, **520**, 505–510.
61. Dimude,J.U., Stockum,A., Midgley-Smith,S.L., Upton,A.L., Foster,H.A., Khan,A., Saunders,N.J., Retkute,R. and Rudolph,C.J. (2015) The consequences of replicating in the wrong orientation: bacterial chromosome duplication without an active replication origin. *Mbio*, **6**, e01294-15.
62. Hiasa,H. and Marians,K.J. (1992) Differential inhibition of the DNA translocation and DNA unwinding activities of DNA helicases by the Escherichia coli Tus protein. *J. Biol. Chem.*, **267**, 11379–11385.
63. Castillo,F., Benmohamed,A. and Szatmari,G. (2017) Xer site specific recombination: double and single recombinase systems. *Front Microbiol*, **8**, 453.
64. Steiner,W.W. and Kuempel,P.L. (1998) Sister chromatid exchange frequencies in *Escherichia coli* analyzed by recombination at the dif resolvase site. *J. Bacteriol.*, **180**, 6269–6275.
65. Sinha,A.K., Possoz,C., Durand,A., Desfontaines,J.M., Barre,F.X., Leach,D.R.F. and Michel,B. (2018) Broken replication forks trigger heritable DNA breaks in the terminus of a circular chromosome. *PLoS Genet.*, **14**, e1007256.
66. Hendricks,E.C., Szerlong,H., Hill,T. and Kuempel,P. (2000) Cell division, guillotining of dimer chromosomes and SOS induction in resolution mutants (dif, xerC and xerD) of Escherichia coli. *Mol. Microbiol.*, **36**, 973–981.
67. Rudolph,C.J., Upton,A.L. and Lloyd,R.G. (2007) Replication fork stalling and cell cycle arrest in UV-irradiated Escherichia coli. *Genes Dev.*, **21**, 668–681.
68. Wendel,B.M., Hollingsworth,S., Courcelle,C.T. and Courcelle,J. (2021) UV-induced DNA damage disrupts the coordination between replication initiation, elongation and completion. *Genes Cells*, **26**, 94–108.
69. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 6640–6645.
70. Katayama,T., Kasho,K. and Kawakami,H. (2017) The DnaA cycle in *Escherichia coli*: activation, function and inactivation of the initiator protein. *Front. Microbiol.*, **8**, 2496.
71. Smith,K.C. and Wang,T.C. (1989) recA-dependent DNA repair processes. *Bioessays*, **10**, 12–16.
72. Clark,A.J. and Sandler,S.J. (1994) Homologous genetic recombination: the pieces begin to fall into place. *Crit. Rev. Microbiol.*, **20**, 125–142.
73. Wang,T.C., Chang,H.Y. and Hung,J.L. (1993) Cosuppression of recF, recR and recO mutations by mutant recA alleles in Escherichia coli cells. *Mutat. Res.*, **294**, 157–166.
74. Lavery,P.E. and Kowalczykowski,S.C. (1988) Biochemical basis of the temperature-inducible constitutive protease activity of the RecA441 protein of Escherichia coli. *J. Mol. Biol.*, **203**, 861–874.
75. Madiraju,M.V., Templin,A. and Clark,A.J. (1988) Properties of a mutant recA-encoded protein reveal a possible role for Escherichia coli recF-encoded protein in genetic recombination. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 6592–6596.
76. Madiraju,M.V., Lavery,P.E., Kowalczykowski,S.C. and Clark,A.J. (1992) Enzymatic properties of the RecA803 protein, a partial suppressor of recF mutations. *Biochemistry*, **31**, 10529–10535.
77. Sawitzke,J.A. and Stahl,F.W. (1992) Phage lambda has an analog of *Escherichia coli* recO, recR and recF genes. *Genetics*, **130**, 7–16.
78. Sawitzke,J.A. and Stahl,F.W. (1994) The phage lambda orf gene encodes a trans-acting factor that suppresses Escherichia coli recO, recR, and recF mutations for recombination of lambda but not of *E. coli*. *J. Bacteriol.*, **176**, 6730–6737.
79. Whitby,M.C. and Lloyd,R.G. (1995) Altered SOS induction associated with mutations in recF, recO and recR. *Mol. Gen. Genet.*, **246**, 174–179.
80. Moreau,P.L. (1988) Overproduction of single-stranded-DNA-binding protein specifically inhibits recombination of UV-irradiated bacteriophage DNA in Escherichia coli. *J. Bacteriol.*, **170**, 2493–2500.
81. Lenhart,J.S., Brandes,E.R., Schroeder,J.W., Sorenson,R.J., Showalter,H.D. and Simmons,L.A. (2014) RecO and RecR are necessary for RecA loading in response to DNA damage and replication fork stress. *J. Bacteriol.*, **196**, 2851–2860.
82. Henrikus,S.S., Henry,C., Ghodke,H., Wood,E.A., Mbele,N., Saxena,R., Basu,U., van Oijen,A.M., Cox,M.M. and Robinson,A. (2019) RecFOR epistasis group: recF and RecO have distinct localizations and functions in *Escherichia coli*. *Nucleic Acids Res.*, **47**, 2946–2965.
83. Odsbu,I. and Skarstad,K. (2014) DNA compaction in the early part of the SOS response is dependent on RecN and RecA. *Microbiology (Reading)*, **160**, 872–882.
84. Sandler,S.J. and Clark,A.J. (1993) Use of high and low level overexpression plasmids to test mutant alleles of the recF gene of Escherichia coli K-12 for partial activity. *Genetics*, **135**, 643–654.
85. Sandler,S.J. (1994) Studies on the mechanism of reduction of UV-inducible sulAp expression by recF overexpression in *Escherichia coli* K-12. *Mol. Gen. Genet.*, **245**, 741–749.
86. Xia,J., Chiu,L.Y., Nehring,R.B., Bravo Nunez,M.A., Mei,Q., Perez,M., Zhai,Y., Fitzgerald,D.M., Pribis,J.P., Wang,Y. *et al.* (2019) Bacteria-to-human protein networks reveal origins of endogenous DNA damage. *Cell*, **176**, 127–143.
87. Sandler,S.J. (1996) Overlapping functions for recF and priA in cell viability and UV-inducible SOS expression are distinguished by dnaC809 in *Escherichia coli* K-12. *Mol. Microbiol.*, **19**, 871–880.
88. Courcelle,J., Carswell-Crumpton,C. and Hanawalt,P.C. (1997) recF and recR are required for the resumption of replication at DNA replication forks in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 3714–3719.
89. Kogoma,T. (1997) Is RecF a DNA replication protein?*Proc. Natl. Acad. Sci. U.S.A.*, **94**, 3483–3484.
90. Romero,Z.J., Chen,S.H., Armstrong,T., Wood,E.A., van Oijen,A., Robinson,A. and Cox,M.M. (2020) Resolving Toxic DNA repair intermediates in every E. coli replication cycle: critical roles for RecG, Uup and RadD. *Nucleic Acids Res.*, **48**, 8445–8460.
91. Jameson,K.H., Rudolph,C.J. and Hawkins,M. (2021) Termination of DNA replication at Tus-ter barriers results in under-replication of template DNA. *J. Biol. Chem.*, **297**, 101409.
92. Mulcair,M.D., Schaeffer,P.M., Oakley,A.J., Cross,H.F., Neylon,C., Hill,T.M. and Dixon,N.E. (2006) A molecular mousetrap determines polarity of termination of DNA replication in E. coli. *Cell*, **125**, 1309–1319.
93. Grainge,I., Bregu,M., Vazquez,M., Sivanathan,V., Ip,S.C. and Sherratt,D.J. (2007) Unlinking chromosome catenanes in vivo by site-specific recombination. *EMBO J.*, **26**, 4228–4238.
94. Romero,Z.J., Armstrong,T.J., Henrikus,S.S., Chen,S.H., Glass,D.J., Ferrazzoli,A.E., Wood,E.A., Chitteni-Pattu,S., van Oijen,A.M., Lovett,S.T. *et al.* (2020) Frequent template switching in postreplication gaps: suppression of deleterious consequences by the Escherichia coli Uup and RadD proteins. *Nucleic Acids Res.*, **48**, 212–230.
95. Jain,K., Wood,E.A. and Cox,M.M. (2021) The *rarA* gene as part of an expanded RecFOR recombination pathway: Negative epistasis and synthetic lethality with *ruvB, recG*, and *recQ*. *PLoS Genet.*, **17**, e1009972.
96. Cooper,D.L., Boyle,D.C. and Lovett,S.T. (2015) Genetic analysis of Escherichia coli RadA: functional motifs and genetic interactions. *Mol. Microbiol.*, **95**, 769–779.
97. Fonville,N.C., Blankschien,M.D., Magner,D.B. and Rosenberg,S.M. (2010) RecQ-dependent death-by-recombination in cells lacking RecG and UvrD. *DNA Repair (Amst.)*, **9**, 403–413.
98. Espeli,O., Mercier,R. and Boccard,F. (2008) DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Mol. Microbiol.*, **68**, 1418–1427.