Quantitative genomic analysis of RecA protein binding during DNA double-strand break repair reveals RecBCD action in vivo

Charlotte A. Cockram, Milana Filatenkova, Vincent Danos, Meriem El Karoui, David R. F. Leach

To cite this version:

HAL Id: hal-01263631
https://hal.archives-ouvertes.fr/hal-01263631
Submitted on 27 Jan 2016
Quantitative genomic analysis of RecA protein binding during DNA double-strand break repair reveals RecBCD action in vivo

Charlotte A. Cockram*, Milana Filatenkova,b,c, Vincent Danosb,c, Meriem El Karoui,b,c,1 and David R. F. Leach*a

*Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom; bLife Sciences Institute, School of Informatics, University of Edinburgh, Edinburgh EH9 9LE, United Kingdom; and cSynthSys, Centre for Synthetic and Systems Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom

Edited by Gerald R. Smith, Fred Hutchinson Cancer Research Center, Seattle, WA, and accepted by the Editorial Board July 15, 2015 (received for review December 18, 2014)

Understanding molecular mechanisms in the context of living cells requires the development of new methods of in vivo biochemical analysis to complement established in vitro biochemistry. A critically important molecular mechanism is genetic recombination, required for the beneficial reassortment of genetic information and for DNA double-strand break repair (DSBR). Central to recombination is the RecA (Rad51) protein that assembles into a spiral filament on DNA and mediates genetic exchange. Here we have developed a method that combines chromatin immunoprecipitation with next-generation sequencing (ChiP-Seq) and mathematical modeling to quantify RecA protein binding during the active repair of a single DSB in the chromosome of Escherichia coli. We have used quantitative genomic analysis to infer the key in vivo molecular parameters governing RecA loading by the helicase/nuclease RecBCD at recombination hot-spots, known as Chi. Our genomic analysis has also revealed that DSBR at the lacZ locus causes a second RecBCD-mediated DSBR event to occur in the terminus region of the chromosome, over 1 Mb away.

Introduction

DNA double-strand break repair (DSBR) is essential for cell survival and repair-deficient cells are highly sensitive to chromosome breakage. In Escherichia coli, a single unrepaired DNA DSB per replication cycle is lethal, illustrating the critical nature of the repair reaction (1). DSBR in E. coli is mediated by homologous recombination, which relies on the RecA protein to efficiently recognize DNA sequence identity between two molecules. RecA homologs are widely conserved from bacteriophages to mammals, where they are known as the Rad51 proteins (2). The RecA protein plays its central role by binding single-stranded DNA (ssDNA) to form a presynaptic filament that searches for a homologous double-stranded DNA (dsDNA) donor from which to repair. It then catalyzes a strand-exchange reaction to form a joint molecule (3), which is stabilized by the branch migration activities of the RecG and RuvAB proteins (4). The joint molecule is then resolved by cleavage at its four-way Holliday junction by the nuclease activity of RuvABC (5, 6).

RecA binding at the site of a DSB is dependent on the activity of the RecBCD enzyme (Fig. Lf). RecBCD is a helicase-nuclease that binds to dsDNA ends, then separates and unwinds the two DNA strands using the helicase activities of the RecB and RecD subunits (see refs. 7 and 8 for recent reviews). RecD is the faster motor of the two and this consequently results in the formation of a ssDNA loop ahead of RecB (Loop 1 in Fig. Lf) (9). As the enzyme translocates along dsDNA, the 3′-terminated strand is continually passed through the Chi-scanning site thought to be located in the RecC protein (10). When a Chi sequence (the octamer 5′-GCTGGTGG-3′) enters this recognition domain, the RecD motor is disengaged and the 3′ strand continues to be unwound by RecB. Under in vitro conditions, where the concentration of magnesium exceeds that of ATP, the 3′ end (unwound by RecB) is rapidly digested before Chi recognition, whereas the 5′ end (unwound by RecD) is intermittently cleaved (11, 12). After Chi recognition the 3′ end is no longer cleaved but the nuclease domain of RecB continues to degrade the 5′ end as it exits the enzyme (11, 12). Under in vitro conditions where the concentration of ATP exceeds that of magnesium, unwinding takes place but the only site of cleavage detected is ~5 nucleotides 3′ of the Chi sequence (13, 14). Because the RecB motor continues to operate while the RecD motor is disengaged, Loop 1 is converted to a second loop located between the RecB and RecC subunits or to a tail upon release of the Chi sequence from its recognition site. We therefore describe this single-stranded region as Loop/Tail 2 in Fig. Lf. After the whole of Loop 1 is converted to Loop/Tail 2, this second single-stranded region continues to grow as long as the RecB subunit unwinds the dsDNA. The RecBCD enzyme enables RecA protein to load on to Loop/Tail 2 to generate the presynaptic filament necessary to search for homology and initiate strand-exchange (15). Finally, the RecBCD enzyme stops translocation and disassembles as it dissociates from the DNA, releasing a DNA-free RecC subunit (16).

Our understanding of the action of RecBCD and RecA has been the result of more than 40 years of genetic analysis and

Significance

Maintaining genomic integrity is crucial for cell survival. In Escherichia coli, RecA-mediated homologous recombination plays an essential role in the repair of DNA double-strand breaks (DSBs). A greater understanding of the mechanism of homologous recombination requires quantitative analysis of genomic studies in live cells. We have developed a novel method that is able to capture these interactions on a genome-wide scale by combining ChiP-seq and mathematical modeling to interpret the patterns of RecA–DNA interaction during DSB repair (DSBR). This genomic analysis has also revealed unexpected RecA binding in the terminus region of the chromosome, consistent with a second DSBR event (at a distance of 1 Mb) that is indirectly caused by the first DSBR event induced at the lacZ.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. G.R.S. is a guest editor invited by the Editorial Board. Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE71249).

1To whom correspondence may be addressed. Email: Meriem.Elkaroui@ed.ac.uk or D.Leach@ed.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1424269112/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1424269112 PNAS PLUS

PNAS | Published online August 10, 2015 | E4735–E4742

PDMS PLUS
biochemical investigation of these purified proteins in vitro. However, relatively little is known about their activities on the genomic scale. To investigate these reactions in vivo, we have used RecA chromatin immunoprecipitation with next-generation sequencing (ChIP-Seq) in an experimental system that allows us to introduce a single and fully repairable DSB into the chromosome of *E. coli* (1). Because DSBR by the RecBCD complex is a genomic scale. Furthermore, our analysis reveals that DSBR at the terminus region of the chromosome, an unanticipated observation illuminated by the genomic scale of our data.

### Results

**DSB-Dependent RecA Loading to DNA.** We initially investigated the in vivo binding of RecA at the site of a DSB by ChIP and assayed RecA–DNA interactions by quantitative PCR (qPCR). The DSB was generated by SbcCD-mediated cleavage of a 246-bp interrupted DNA palindrome inserted in the lacZ gene (*lacZ::pal246*) (1). In the absence of a DSB, there was no RecA enrichment detected in the 40-kb region surrounding *lacZ* (Fig. 2A). However, following the induction of a DSB there was significant RecA binding detected on both sides of *lacZ* (Fig. 2B). This binding corresponded to the first correctly oriented Chi site on either side of the DSB and spread out over several kilobases of DNA, consistent with the formation of a RecA filament on a single-strand of DNA generated by RecBCD, followed by strand invasion to form a joint molecule. These data suggested that, as expected and consistent with in vitro data, RecA is loaded at the DSB in a Chi-dependent manner (12, 19).

After recognition of the Chi sequence, qPCR analysis detected the binding of RecA to a 30-kb region of DNA surrounding the DSB. Large peaks of RecA enrichment were detected immediately after the first Chi sites on both sides of the DSB, with RecA enrichment decreasing at Chi sites further away from *lacZ* (Fig. 2B). On the origin-proximal side of the break we detected binding of RecA following, not only the first Chi site encountered, but also to subsequent Chi sites. Eighteen-fold RecA enrichment was observed following the Chi site positioned closest to the DSB at a focus on the origin-proximal side and a subsequent peak of 12-fold RecA enrichment was detected at loci positioned ~13 kb origin-proximal to the DSB. This second peak is consistent with the presence of four Chi sites in this region. The origin-distal side of the DSB does not have a second peak, with RecA enrichment plateauing at fivefold enrichment at the sites tested between 7 kb and 13 kb, as expected given the presence of only a single Chi site in this region. The distribution of RecA binding also suggested that the two sides of the DSB might be processed differently, with a higher RecA-enrichment observed at the first Chi position on the origin-proximal side of the break compared with the origin-distal side (Fig. 2B).

**RecA Loading at Synthetic Arrays of Three Chi Sites.** To confirm that RecA was indeed loaded in relation to the recognition of Chi sites by the RecBCD enzyme, we investigated RecA binding in the presence of synthetic arrays of three Chi sites inserted at 3 kb on either side of the DSB. In vitro studies have shown that a single Chi site is recognized by RecBCD with an efficiency of 20–40%, which suggests that following a DSB, a substantial number of RecBCD molecules fail to recognize Chi (20). An efficiency of Chi recognition in vivo similar to that obtained in vitro would explain the observed Chi distribution at *lacZ::pal246*. Previously, arrays of three synthetic Chi sites have been shown to be recognized by RecBCD with an efficiency of 60–80% (21). We reasoned that placing these arrays either side of the DSB would focus a similar proportion of RecA loading closer to *lacZ*. Furthermore, we placed the Chi sites at equal distances (3 kb) on the two sides of the break to increase the symmetry of the reaction. ChIP-qPCR

---

**Fig. 1.** DSBR in *E. coli*. (A and B) Schematic representation of DSB processing by the RecBCD complex. (A) Before Chi recognition, both the RecB and RecD motors progress along the DNA. RecD is the faster motor and as a result a loop of ssDNA (Loop 1) is formed ahead of the slower RecB motor. The 3′ ssDNA strand is scanned for the Chi sequence by the RecC protein. (B) After Chi recognition, RecBCD likely undergoes a conformational change so that only the RecD motor is engaged. The RecA protein is recruited by the RecB nuclease domain and loaded onto the ssDNA loop generated by RecB unwinding to promote RecA nucleoprotein filament formation. In this schematic representation, the Chi site is shown held in its recognition site. However, the Chi site will be released either by disassembly of the RecBCD complex or at some point before this and the second single-stranded region on the 3′ terminating strand will be converted from a loop to a tail. Therefore, this region is denoted Loop/Tail 2. The mathematical model described in SI Appendix does not depend on the ATP/magnesium dependent differential cleavage of DNA strands (7, 8), nor does it depend on the precise time that the 3′ end is released from the complex following Chi recognition. (C) The hairpin endonuclease SbcCD is used to cleave a 246-bp interrupted palindrome inserted in the lacZ gene of the *E. coli* chromosome. Cleavage of this DNA hairpin results in the generation of a site-specific DSB on only one of a pair of replicating sister chromosomes, thus leaving an intact sister chromosome to serve as a template for repair by homologous recombination.
revealed that, in vivo, the triple Chi arrays do indeed stimulate RecA binding closer to the DSB and that binding was enhanced relative to that observed at single endogenous Chi sites (Fig. 2 B and D). Interestingly, the asymmetry in RecA binding to the DNA following a DSB remained, with more RecA bound to the origin-proximal side of the DSB compared with the origin-distal side (Fig. 2D). Furthermore, despite the 38-fold RecA enrichment detected at a locus ~3 kb origin-proximal to the DSB, there was still as much as 15-fold RecA enrichment observed at loci following endogenous Chi sites subsequent to the triple-Chi arrays. This finding confirmed that, like single Chi sites, the triple Chi arrays failed to be recognized in a detectable proportion of the population and that successive Chi sites are required for efficient DSBR.

High-Resolution Analysis of RecA Loading by ChIP-Seq. To quantify RecA binding to DNA in relation to Chi, ChIP was combined with high-throughput sequencing (ChIP-Seq) to provide a genome-wide analysis of RecA–DNA interactions following a DSB (Fig. 3). These experiments were carried out with the arrays of three Chi sites at 3 kb on either side of the DSB site to focus the reaction at equidistant sites on both sides of the break. RecA was loaded, at the site of a DSB in the lacZ gene, in a Chi-dependent manner with approximately twofold more RecA on the origin-proximal side of the DSB compared with the origin-distal side (Fig. 3B). This finding is consistent with the results obtained by ChIP-qPCR (Fig. 2) and suggests that the two DNA ends are not equally competent for Chi recognition. Because previous work has shown that SbcCD generates a two-ended break in a RecB mutant (1), we suggest that approximately half of the DSBs arising at the interrupted palindrome are converted from two-ended to one-ended structures by RecBCD action. This could happen if RecBCD traveling on the origin-distal end catches up with the replication fork and dissociates before recognizing a Chi sequence (Fig. S1).

To further determine the role of subsequent Chi sites during DSBR, we deleted all of the endogenous Chi sites within a 15-kb region on either side of the break, leaving only the triple-Chi arrays positioned 3 kb either side of the lacZ. ChIP-Seq analysis revealed that, although the level of enrichment stimulated by the triple-Chi array remained unchanged, RecA binding was decreased in the 15-kb region where the Chi sites had been deleted (Fig. 3C). Strikingly, this decrease was correlated with an increase in RecA binding caused by Chi sites more than 20 kb away from the DSB (Fig. S2). This finding confirmed that RecBCD enzyme complexes that did not act at the array of three Chi sites progress many kilobases further on the DNA until they do recognize a Chi site.

Inferring the Parameters of RecBCD Activity in Vivo from High-Resolution ChIP-Seq Data Using Mathematical Modeling. We reasoned that the high spatial resolution afforded by ChIP-Seq data could be exploited to reveal quantitative aspects of the molecular behavior of RecBCD-mediated loading of RecA in living cells. To interpret the quantitative information contained in the high-resolution ChIP-Seq data in terms of the parameters of RecBCD action in vivo, we developed a mathematical model of enzyme action. Our mathematical model is based on the known in vitro properties of RecBCD and its crystal structure (7, 8, 22) and is described in detail in the SI Appendix (see also Fig. L4). The model’s main assumptions are the following: Before Chi recognition, the RecBCD complex translocates along the DNA with both the RecD and RecB motors engaged. RecD is the lead motor and the difference in speed of the two motors leads to the accumulation of a ssDNA loop (Loop 1) ahead of RecB that
depends on the motor speed ratio \( V_{\text{BCD}}/V_{\text{RecA}} \) and the distance traveled by the enzyme. Chi recognition is stochastic and we denote \( p_{\text{Chi}} \) the probability of Chi recognition. Upon Chi recognition, the RecD motor is disengaged and Loop 1 is converted to Loop/Tail2, which is extended using the RecB motor and RecA is loaded with equal probability across the ssDNA. We denote \( p_{\text{RecA}} \), the probability of RecBCD dissociating from DNA or stopping RecA loading. Using these assumptions, we calculated the probability of RecA loading at a genomic position in the vicinity of the DSB depending on the position of the DSB and the Chi sites, \( V_{\text{BCD}}, p_{\text{Chi}}, \) and \( p_{\text{RecA}} \).

Using a strain in which we had deleted all of the endogenous Chi sites within a 15-kb region either side of the break, leaving only the artificial Chi arrays positioned 3 kb either side of the \( \lambda \)lacz region, we varied the number of Chi sites in the origin-proximal array from one to six. We then compared the mathematical model prediction to the RecA ChiP-Seq data obtained for strains with double Chi sites and single Chi sites. All Chi sites within a 30-kb region surrounding the DSB were removed in strain (DL4899) lacking the 246-bp palindrome. In A-C red circles represent triple-Chi arrays and green circles single Chi sites, all Chi sites shown are in the active orientation. The raw distributions of hits are depicted in gray. To help visualization, the data were filtered using a loess filter (bandwidth 5,700 nucleotides, span 0.02, red curve). The data shown are from a representative example of two biological replicates. (A) ChiP-Seq analysis of the \( \lambda \)lacz region in the absence of a DSB carried out in strain (DL4899) lacking the 246-bp palindrome. In A–C red circles represent triple-Chi arrays and green circles single Chi sites, all Chi sites shown are in the active orientation.

**Fig. 3.** High-resolution analysis of DSB-dependent RecA loading by ChiP-Seq. (A) ChiP-Seq analysis of the \( \lambda \)lacz region in the absence of a DSB carried out in strain (DL4899) lacking the 246-bp palindrome. In A–C red circles represent triple-Chi arrays and green circles single Chi sites, all Chi sites shown are in the active orientation. The raw distributions of hits are depicted in gray. To help visualization, the data were filtered using a loess filter (bandwidth 5,700 nucleotides, span 0.02, red curve). The data shown are from a representative example of two biological replicates. (B) ChiP-Seq analysis of the \( \lambda \)lacz region undergoing DSBR in a strain (DL4900) carrying the 246-bp palindrome located at 0 kb on the map. (C) ChiP-Seq analysis of the \( \lambda \)lacz region of strain DL5215 demonstrates that removing the endogenous Chi sites within a 30-kb region surrounding the DSB stimulates RecA loading at more distant Chi sites during DSBR. White circles represent endogenous Chi sites that have been deleted.

The system uses the fact that a 246-bp interrupted palindrome is cleaved by the SbcCD enzyme at the site of a DNA hairpin structure formed on the lagging-strand template because one copy of the replicated chromosomal copy was cleaved by the SbcCD enzyme at the site of a DNA hairpin structure formed on the lagging-strand template. The system uses the fact that a 246-bp interrupted palindrome is cleaved by the SbcCD enzyme at the site of a DNA hairpin structure formed on the lagging-strand template. The SbcCD enzyme at the site of a DNA hairpin structure formed on the lagging-strand template because one copy of the replicated chromosomal copy (1). We predict the hairpin to be formed on the lagging-strand template because of its single-stranded nature and that under our growth conditions, repair occurs efficiently, presumably using the uncleaved sister chromosome as a template (4, 17). Using ChIP in \( E. coli \) with antibodies against the RecA protein, we investigated the behavior of this protein as it is engaged in repairing this DSB. We
then combined ChIP with whole-genome sequencing to map these RecA–DNA interactions on a genome-wide scale. As predicted by the in vitro biochemistry of RecBCD enzyme, following a DSB, RecA protein is loaded onto DNA in relation to the Chi sites (5′-GCTGGTGG-3′) surrounding the break, and using a simple mathematical model we were able to infer the probability of Chi recognition in vivo. Although the mathematical model we used is specific to RecBCD, such ability to infer the parameters of enzyme action in vivo from a combination of genomic analysis and mathematical modeling promises to be applicable to other macromolecular reactions, such as the activity of RNA polymerase.

**Inference of the Parameters of RecBCD Action in Vivo.** We have coupled quantitative genomic analysis of RecA binding with mathematical modeling of the RecBCD complex in its loading of

![Fig. 4. High-resolution DSB-dependent RecA loading in the presence of one to six Chi sites. (A–F) The efficiency of Chi recognition in vivo was investigated by placing Chi arrays (red circle on the map) containing either 1, 2, 3, 4, 5, or 6 Chi sites on the origin-proximal side of the DSB and using RecA binding (ChIP-Seq) as a measure of Chi recognition. Green circles represent single endogenous Chi sites that are positioned in the active orientation. The number of hits per 250-bp window normalized to the total number of reads in the analyzed region is depicted in gray. To help visualization, the data were filtered using a loess filter (bandwidth 5,700 nucleotides, span 0.057, red curve). The blue line corresponds to the prediction of our mathematical model that estimates the probability that a nucleotide in the vicinity of a DSB is coated by RecA. The data shown are from a representative example of two biological replicates. (G) The three parameters of the model (RecBCD motors’ speed ratio, its probability of Chi recognition and its processivity of RecA loading) were inferred independently on each dataset using a maximum-likelihood strategy (SI Appendix).](image-url)

<table>
<thead>
<tr>
<th>No. of Chi sites in array</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>pChi</td>
<td>0.43</td>
<td>0.35</td>
<td>0.29</td>
<td>0.23</td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td>Vg/V0</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>Pstop</td>
<td>1.1 x 10^-4</td>
<td>1 x 10^-4</td>
<td>1 x 10^-4</td>
<td>1 x 10^-4</td>
<td>1 x 10^-4</td>
<td>0.84 x 10^-4</td>
</tr>
</tbody>
</table>
recA to infer, based on the assumptions of the mathematical model, the molecular parameters of RecBCD action in live cells. Initial analysis using a mathematical model with one mode of action of RecBCD led to estimates of the probability of Chi recognition and processivity values similar to those that had previously been measured in vitro (19, 20). However, we observed that the inferred ratio of the two motors’ speed ($V_B/V_D = 0.94–0.96$) was significantly higher than the value of 0.6 observed in vitro from studies of mutant enzymes defective for the helicase motors of recB or recD (25) and from evaluation of the average rate of Loop 1 formation relative to total unwinding by the wild-type enzyme (9). We calculated (SI Appendix, Eq. [2]) that a ratio of 0.95 would result in the production of a single-strand loop 0.96) was significantly higher than the value of 0.6 observed in vitro from studies of mutant enzymes defective for the helicase motors of RecB or RecD (25) and from evaluation of the average rate of Loop 1 formation relative to total unwinding by the wild-type enzyme (9). We calculated (SI Appendix, Eq. [2]) that a ratio of 0.95 would result in the production of a single-strand loop.
to recombination. However, there are very likely several different reactions taking place. Our observation of a DSBR event close to dif induced by DSBR at lacZ provides a clear physical demonstration of one such interaction in this region.

Experimental Procedures

Bacterial Strains and Growth. All strains are derivatives of E. coli K12 MG1655 (34) and are listed in Table S1. Cells were grown in M9 minimal media supplemented with 0.2% casamino acids, 0.5% glucose, 5 μM CaCl2, and 1 mM MgSO4 at 37 °C. Mutations were introduced by P1 transduction or plasmid-mediated gene replacement (PMGR) (35–38) using the plasmids described in Table S2. All primers used for cloning and genotyping are detailed in Table S3.

ChIP Sample Preparation. All ChIP experiments were performed with ≈ 5 × 10⁸ cells growing in exponential growth phase (OD600nm, 0.2–0.25). RecA–DNA interactions were chemically cross-linked with formaldehyde (Sigma-Aldrich; final concentration 1%) for 10 min at 22.5 °C. Cross-linking was quenched by the addition of glycine (Sigma-Aldrich; final concentration 0.5%). Cells were collected by centrifugation and washed three times in ice-cold 1x PBS. The pellet was then resuspended in 250 μl ChIP buffer (200 mM Tris HCl (pH 8.0), 500 mM NaCl 4% (v/vol) Triton X. Complete protease inhibitor mixture EDTA-free (Roche). Sonication of cross-linked samples was performed using the Diagenode Bioruptor at 30-s intervals for 10 min at high amplitude. After sonication, 350 μl of ChIP buffer was added to each sample, the samples were mixed by gentle pipetting and 100 μl of each lysate was removed and stored as “input.” Immunoprecipitation was performed overnight at 4 °C using 1/100 anti-RecA antibody (Abcam, ab66797). IP samples were then incubated with Protein G Dynabeads (Life Technologies) for 2 h at room temperature. All samples were washed three times with 1x PBS + 0.02% Tween-20 before resuspending the Protein G dynabeads in 200 μl of TE buffer (10 mM Tris (pH 7.4), 1 mM EDTA) + 1% SDS. Next, 100 μl of TE buffer was added to the input samples and all samples were then incubated at 65 °C for 10 h to reverse formaldehyde cross-links. DNA was isolated using the MinElute PCR purification kit (Qiagen). DNA was eluted in 50 μl of TE buffer using a two-step elution. Samples were stored at –20 °C.

Library Preparation for High-Throughput Sequencing. Input and ChIP samples were processed following New England Biolabs’s protocol from the NEBNext ChIP-Seq library preparation kit. Briefly, 200 ng of input and ChIP-enriched DNA was subject to end repair to fill in ssDNA overhangs, remove 3′ phosphates and 5′ phosphorylate the sheared DNA. Klenow exo- was used to adenylate the 3′ ends of the DNA and NEXTFlex DNA barcodes (Bio Scientific) were ligated using T4 DNA ligase. After each step, the DNA was purified using the Qiagen MinElute PCR purification kit according to the manufacturer’s instructions. After adaptor ligation, the adaptor-modified DNA fragments were enriched by PCR using primers corresponding to the beginning of each adaptor. Finally, agarose gel electrophoresis was used to size select adaptor-ligated DNA with an average size of ∼275 bp. All samples were cleaned up using the QIAquick PCR Purification Kit (Qiagen) before being sequenced on the Illumina HiSeq. 2000. RecA binds to both ssDNA and dsDNA in presynaptic and postsynaptic complexes (39–41). It was previously believed that ssDNA could not be detected by ChIP-Seq. However, several studies have recently shown that this is not the case; ssDNA is rendered double-stranded during the library preparation process through the formation of DNA hairpins that arise as a result of microhomology (42–44). This allows the DNA to be amplified and detected by ChIP-Seq. These findings are consistent with our data, which shows a similar pattern of RecA binding detected using both qPCR and high-throughput sequencing.

Chip-Seq Data Analysis. For ChIP-Seq analysis, 50-bp single-end reads were mapped to the E. coli K12 MG1655 (NC000913.3) (34) genome using Novoalign (Novocraft). Novoalign uses the Needleman-Wunsch algorithm to determine the optimal alignment of reads. Before mapping, the 3′ adapter sequences were removed using fastx_clipper and the data collapsed using fastx_collapser to remove identical sequence reads (hannonlab.cshl.edu/fastx_toolkit/index.html). The preparation of Chip-Seq libraries requires a PCR of the adaptor ligated DNA. This can result in PCR duplication of certain DNA fragments. Removing duplicates mitigates the effects of PCR amplification bias so that regions of the genome don’t appear more enriched than they actually are. The Chip-Seq datasets in this study contained ~4% PCR duplicates and these were discarded. The data were also plotted without removing these duplicates and revealed that the trend in RecA binding was unchanged. Sequences were mapped with default parameters, allowing for a maximum of one mismatch per read (novoalign -f DL4900 IP, fasta -d DL4900.genome.ucsc -f UCSC -f DL4900.Psam). To report reads that have multiple alignment loci, we specified the -r parameter as either -r Random or -r None. In the first case Novoalign chooses a single alignment location at random among all of the alignment results; in the second case, only the reads that map to a single genomic location are aligned (www.novocraft.com). PyReadCounters was used to calculate the overlap between aligned reads and E. coli genomic features (45). The distribution of reads along the E. coli genome was visualized using the integrated Genome Browser (46). Full details of all scripts are available upon request.

Identification of RecA-Binding at the DSBR. Because of the specific mechanism of RecBCD-mediated RecA loading observed around the DSBR, classic peak-calling algorithms such as MACS (47) failed to recapitulate the RecA binding at this site. This is because RecA loading at a DSBR is the result of a complex dynamic process that cannot be described as a simple binding event. This suggests that, as has been observed for other datasets (48), the shape of the peaks may carry important information. In particular, we reasoned that given the high spatial resolution of Chip-Seq data, the position and shape of the peaks observed at Chi sites could give us quantitative information about the mechanism of RecBCD-mediated DSBR repair in vivo. Therefore, we developed a mathematical model of RecBCD-dependent RecA loading to evaluate the probability that a nucleotide in the vicinity of a DSBR is coated by the RecA protein. We then used maximum-likelihood estimation to extract the parameters of this model from the dataset. This mathematical model and the associated data analysis are described in detail in SI Appendix.

qPCR. All real-time qPCR reactions were carried out in 15-μl volumes in the MX3000P qPCR machine (Agilent) using the Brilliant II SYBR Green qPCR master mix (Agilent). The temperature profile for all assays was 95 °C for 10 min followed by 40 cycles of 95 °C for 20 s and 60 °C for 60 s. All reactions were repeated in triplicate and the formation of PCR products of the correct lengths was confirmed by agarose gel electrophoresis. A full list of primers used for qPCR is given in Table S4. Assay performance was checked by standard curve for all assays. Data were exported from the MxPro software to Microsoft Excel for analysis. The melting temperature of the qPCR primers was calculated by the manufacturer (MWG Biotech).

ACKNOWLEDGMENTS. We thank Dr. N. Molina and Dr. G. Sanguinetti for advice on data analysis; Dr. Sander Granneman for advice on Chip-Seq analysis; Dr. Ralph Hector for advice on Chip-Seq library preparation; and Dr. M. White for the construction of pDL4690. This research has been supported by an Medical Research Council studentship and a Medical Research Council Centenary Award (to C.A.C.); a Darwin Trust of Edinburgh postgraduate studentship (to M.F.); Marie Curie Fellowship PIEF-GA-2009-254082 – DRIBAC (to M.E.K.); European Research Council Advanced Grant RULE-320823 (to V.D.); a Marie Curie Fellowship PIOF-GA-2009-254082 – DRIBAC (to M.E.K.); European Research Council Advanced Grant RULE-320823 (to V.D.); and a Medical Research Council programme Grant G0901622 (to D.R.F.L.).


