RecA and RecB: probing complexes of DNA repair proteins with mitomycin C in live *Escherichia coli* with single-molecule sensitivity

Alex L. Payne-Dwyer[1,2†], Aisha H. Syeda[1,2†], Jack W. Shepherd[1,2], Lewis Frame[3] and Mark. C. Leake[1,2*]

[1] Department of Physics, University of York, York, UK, YO10 5DD;

[2] Department of Biology, University of York, York, UK, YO10 5DD;

[3] School of Natural Sciences, University of York, York, UK, YO10 5DD.

[*]Correspondence: mark.leake@york.ac.uk

[†]These authors contributed equally.

Abstract

The RecA protein and RecBCD complex are key bacterial components for the maintenance and repair of DNA. RecBCD is a helicase-nuclease that uses homologous recombination to resolve double-stranded DNA breaks. It also facilitates coating of single-stranded DNA with RecA to form RecA filaments, a vital step in the double-stranded break DNA repair pathway. However, questions remain about the mechanistic roles of RecA and RecBCD in live cells. Here, we use millisecond super-resolved fluorescence microscopy to pinpoint the spatial localization of fluorescent reporters of RecA or RecB at physiological levels of expression in individual live *Escherichia coli* cells. By introducing the DNA crosslinker mitomycin C, we induce DNA damage and quantify the resulting steady state changes in stoichiometry, cellular protein copy number and molecular mobilities of RecA and RecB. We find that both proteins accumulate in molecular hotspots to effect repair, resulting in RecA stoichiometries equivalent to several hundred molecules that assemble largely in dimeric subunits before DNA damage, but form periodic subunits of approximately 3-4 molecules within mature filaments of several thousand molecules. Unexpectedly, we find that the physiologically predominant forms of RecB are not only rapidly diffusing monomers, but slowly diffusing dimers.

Keywords: recombination; repair; DNA damage; mitomycin C; super-resolution microscopy; single-molecule tracking; Slimfield.

1. Introduction

Accurate duplication of the genome is crucial in all organisms, accomplished by a sophisticated molecular machine known as the replisome [1]. An inability to accurately replicate genetic material can lead to cell death and/or cancers [2,3]. Mitomycin C (MMC) is a naturally occurring antibiotic that can be used to controllably disrupt DNA replication, and thus a valuable reagent in studying DNA repair processes. It is used as a chemotherapeutic in treating several cancers [4] and retinopathies [5], and acts by targeting DNA deoxyguanosine (dG) residues [6], forming intrastrand or interstrand crosslinks [7]. If unrepaired, these structures can interfere with cellular processes such as transcription and replication, potentially leading to genome instability [8]. An encounter between a mitomycin C-induced crosslink and an approaching replisome may result in replisome disassembly and eventually a double strand break (DSB) [9]. RecBCD recognises DSBs in *E. coli* [10], processing the ends to generate 3'-ended single-stranded DNA (ssDNA) as a landing pad for the principal recombination protein, RecA [10]. Recombination of RecA-ssDNA complexes with the homologous DNA restores the replication fork, on which the replisome can be reloaded. The replisome may resume replication if the blocking adduct is repaired [11]. As a complex of individual RecB, RecC and RecD proteins, RecBCD is a versatile helicase-nuclease and underpins two major pathways for homologous DNA recombination, essential for DSB repair [10]. RecBCD activities involve several processes - it recognises and binds DSBs, begins unwinding both DNA strands, and also degrades both [10]. This activity continues unhindered until it encounters an octameric Chi site that induces a shift in enzyme activity to degrade only the 5'-

ended strand [12,13]. This activity shift results in a 3'-ended ssDNA overhang that facilitates RecA loading. A key function of RecA is its ability to form nucleoprotein filaments on exposed ssDNA in response to damage [14]. These filaments can infiltrate an intact duplex and, on finding homology, recombine with the infiltrated duplex [15,16]. The extension of filaments along the cell accelerates this homology search in a non-linear fashion [17]. Following further processing of the resulting structure, primosome proteins establish an intact replisome thereby enabling replication to resume [18]. Recombination proteins, such as RecBCD, need access to replication-transcription conflict sites and collapsed forks, but if RecBCD is missing then double-stranded DNA (dsDNA) is degraded by exonucleases [19,20], possibly resulting from replisome disassembly. However, how RecA stabilizes blocked forks remains an open question.

The nucleoprotein filaments formed by RecA are both a requisite and a hallmark of the cell-wide SOS response [21–24]. The SOS response is a regulatory shift that promotes cell survival in adverse conditions associated with increased rates of interrupted replication and DNA damage [25]. The SOS response to DNA damage induced by antimicrobials plays a major role in the emergence of persister cells [26] and wider antimicrobial tolerance on a population level [27].

Given these far-reaching implications of RecA and RecB activity as studied comprehensively with mutants [12,24,28–32], it is important to establish the number of molecules present in cells, how they are spatially distributed and organized, and how these are affected by antimicrobials such as MMC. Here, we use millisecond super-resolved Slimfield microscopy [33] in live *E. coli* containing genomically-encoded fluorescent fusions RecA-mGFP [34] and RecB-sfGFP [35]. Since RecA fusion constructs retain only partial function, our approach makes use of a merodiploid RecA fusion that expresses from one copy of the native gene and one copy of the *recA4155* fusion construct [34]. This strain rescues approximately wild-type sensitivity with mixed assemblies of the two RecA proteins [34].

We use Slimfield microscopy to visualise the spatial distribution of RecA and RecB fluorescent proteins in individual cells. From these quantitative images, we identify diffraction-limited local intensity maxima (we denote these as *foci* – see Table 1 for a description of technical Slimfield microscopy nomenclature used in this study) to a lateral spatial precision of 40 nm [36]. Slimfield uses ~millisecond sampling that is sufficiently rapid to link the moving foci derived from the same emitter sources over sequential image frames, following appropriate bespoke particle tracking analysis [33,37,38], into *tracks*. Each of these tracks implies the presence of a particle containing one or more associated molecules; typically more than one prior to photobleaching, so more generally, we term each a *molecular assembly*. These tracks reveal the detailed diffusion of labelled RecA and RecB assemblies in the cytoplasm of a living cell. By using the single-molecule sensitivity of Slimfield microscopy, we are able to quantify single-molecule photobleaching steps in fluorescence intensity, to identify the characteristic brightness of a single fluorescent protein [33]. Not only does this calibration apply to the fraction of the fluorescence intensity for each tracked assembly, but also to the GFP fluorescence in the whole, or part, of each cell. We use this to determine the number of GFP-labelled molecules within each tracked assembly (the *stoichiometry*), and the total number of fluorescently-labelled molecules within each cell (the *cellular protein copy number*), or intracellular segment (the *segment protein copy number*). Those fluorescent molecules which contribute to the copy number above the cell's autofluorescent background but are not detected as foci (typically due to high, uniform emitter density and/or excessive mobility) are denoted the *pool*.

Slimfield has some similarities to single particle tracking photoactivation localization microscopy (sptPALM) [39–41], however, our approach is simpler, requiring only constitutively expressed fluorescent reporters such as GFP, and trades off the condition of observing exclusively single molecules in order to measure the stoichiometry of dynamic assemblies far more accurately. This is a deliberate advantage of our technique over other single-molecule microscopy techniques as previously used to count RecB content in cells molecule-by-molecule [35].

Prior to MMC treatment, only point-like assemblies of RecA or RecB are detectable. RecA presents far brighter fluorescence in a cell than RecB, indicating both a typical stoichiometry and a cellular protein copy number that are 2-3 orders of magnitude greater. On treatment with MMC, we observe an increase in the average cellular protein copy number of RecA, but not of RecB, in each cell, with up to 20% of cells devoid of RecB assemblies. MMC induces the formation of RecA assemblies larger than can be captured in single foci, and we interpret these as RecA nucleoprotein filaments, or bundles of filaments [24,30,42–47], typically associated with the SOS response.

Between cellular states of SOS readiness and MMC-induced response, the stoichiometries of RecA assemblies increase, and the diffusion coefficients of assemblies decrease correspondingly. We also discover surprisingly consistent intervals between the stoichiometries of different assemblies in each condition. We interpret the average number of molecules in the intervals (the *periodicity*) as indicative of an oligomeric structural repeat unit that comprises assemblies. The periodicity of RecA assemblies changes from dimeric in character to groups of roughly 3-4 molecules in response to MMC, while the periodicity of RecB assemblies is dimeric, and insensitive to MMC treatment.

Our results shed new light on the relations between structure and function for RecA and RecBCD in mediating repair upon DNA damage.

Metric / Object	**Definition**
Segment	An area of the image defined by a contiguous subset of pixels in a binary mask. This area either corresponds to a whole cell (a *cell mask*), or more typically a region inside the cell (an *intracellular segment*) of high fluorescent intensity. The term "*segment*" refers to an intracellular segment unless otherwise stated.
Cell mask	A segment containing the outline of one cell. These are extracted using a machine learning protocol (Supplementary Methods).
Intracellular segment	A segment inside the cell. These are extracted from the set of foci localized in that cell by rendering a superresolved image, followed by local Otsu thresholding (Materials and Methods 4.3.5), with the intention of isolating RecA objects that resemble nucleoprotein filaments or bundles.
Assembly	A group of labeled molecules physically associated with one another, either directly or indirectly, such that their diffusive movement is strongly correlated, and therefore always detected in the same track.
Focus (foci)	A spot-like local intensity maximum in a single frame, which corresponds to a localized group of labeled molecules (Materials and Methods 4.3.1). Associated properties include centroid location, total intensity, and signal-to-noise ratio.
Track	A set of foci in adjacent frames that are spatially close enough to form a contiguous trajectory (Materials and Methods 4.3.1). Typically associated with a single molecular assembly, or a group of strongly colocalized assemblies.
Diffusion coefficient	Measure of the random microscopic motion of a specific track based on the increase in the mean square displacement of its intensity centroid over time (Materials and Methods 4.3.2).
Characteristic single-molecule brightness	The average sum of pixel values in foci associated with a single fluorescent reporter molecule (e.g. mGFP), under a fixed imaging condition (Materials and Methods 4.3.3). Equivalent to the modal step size in intensity for tracks in the final stage of photobleaching (Figure S1).
Stoichiometry	The number of fluorescently labeled molecules in a specific track. This is estimated by extracting the sequence of foci belonging to that track, then extrapolating the sum of pixel values in each focus backwards along that sequence to get an initial track intensity that is independent of photobleaching (Materials and Methods 4.3.4). The initial track intensity is then divided by the characteristic single-molecule brightness.
Periodicity	The population-averaged number of fluorescently labeled molecules in inferred repeat units within tracked objects. Estimated by averaging the consistent intervals between nearest-neighbor peaks in the population-level stoichiometry distribution (Materials and Methods 4.3.5).
Integrated intensity	The total fluorescence intensity of a segment in pixel counts, normalised by the characteristic single molecule brightness (Materials and Methods 4.3.6-7).
Cellular (or segment) protein copy number	The average number of molecules in a cell (or intracellular segment), as estimated from the increase in integrated intensity above negative control (i.e. subtracting the contribution from autofluorescence (Materials and Methods 4.3.6-7).
Pool	The intracellular fluorescence which is not detected in tracks.
Pool stoichiometry	The number of untracked, labeled molecules within an area of the pool equal to the size of one diffraction-limited focus (Materials and Methods 4.3.6).

Table 1. Definitions of quantitative analysis metrics for Slimfield.

2. Results

2.1 Abundance of RecA, but not RecB, increases on MMC-induced DNA damage

We first optimised MMC treatment conditions so that they did not cause cellular filamentation in wild type cells (Materials and Methods 4.1, Figure S1) but did induce the SOS response [48], since cells would then be sensitised to MMC if the SOS response is blocked [48]. Filamentation and loss of viability was also minimal for the labeled strains (Figure 1), hence we used the same MMC treatment for all strains. Given that SOS induction in these and related strains typically takes <20 min [34,48], the kinetics of initial SOS induction due to MMC will likely reach steady state within the 180 min MMC exposure that we used. In light of the timescale of the initial SOS induction, RecA or RecB dynamics were not in the scope of our study here, but rather the steady state effect of MMC on the distribution and molecular organization of RecA and RecB.

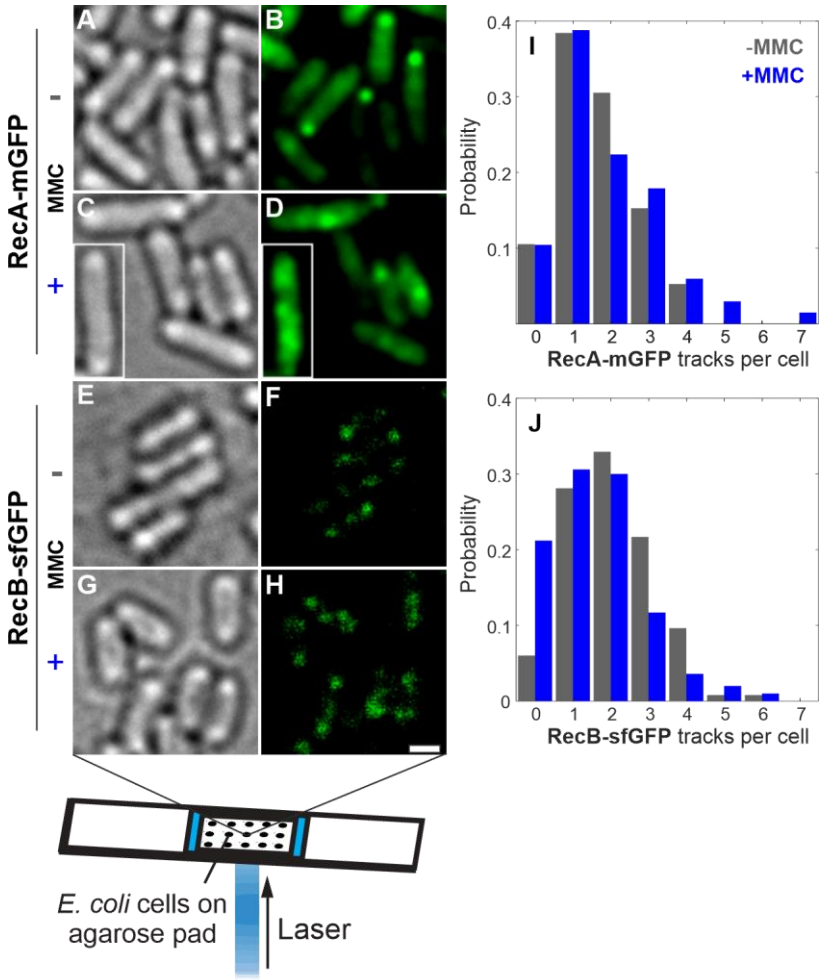


Figure 1. Brightfield and Slimfield (mean average of 3 initial frames) of live *E. coli* in 56-salts minimal media, labeled at RecA-mGFP or RecB-sfGFP before and after MMC treatment. Inset (C,D) is another cell transplanted from the same acquisition outside the cropped field of view at the same scale. Brightness of RecB-GFP Slimfield panels (F,H) scaled 100× *vs*. RecA-mGFP panels (B,D). Scale bar 1 µm. (I,J) Probability distributions for number of tracks detected per cell. Tracks are identified in post-acquisition analysis (Materials and Methods 4.3.1) by first detecting foci as local fluorescent maxima, then linking nearest-neighbour foci in subsequent frames.

We performed brightfield and Slimfield microscopy in each field of view (Materials and Methods 4.2). Binary masks for each cell were extracted independently of any fluorescence signal, using a machine learning segmentation protocol on brightfield images. We then applied these masks to fluorescence images to eliminate extracellular background and facilitate statistics on a cell-by-cell level (Figure S5 and Supplementary Methods). Using home-written automated particle tracking and analysis software ADEMScode [49], we identified fluorescent foci from local intensity maxima in each cell and linked these into tracks (Materials and Methods 4.3.1). We determined the stoichiometry (the number of molecules present) of each track, from the summed pixel intensity values corresponding to the start of each track before any photobleaching has occurred (Materials and Methods 4.3.4), normalized by brightness corresponding to a single molecule of GFP (Materials and Methods 4.3.3).

From the localizations of foci during photobleaching, we also reconstructed superresolved images of fluorescent RecA structures. We extracted binary masks from highly fluorescent regions of interest in these images (denoted *intracellular segments,* or simply *segments* for brevity) using a classical segmentation method (Materials and Methods 4.3.7), which enabled statistics on an intracellular segment level (Figure 3 and Figure S3).

Separately, from the cell masks (or intracellular segments) we also calculated the cellular (or segment) protein copy number (Materials and Methods 4.3.6-7); first we summed the pixel values in each cell or segment area and normalized these by the characteristic brightness of a single GFP to obtain the total intensity within that region, expressed in molecules [50]. Taking the difference from an equivalent area of the control strain that does not express GFP, then yields the cellular or segment protein copy number corrected for any cellular autofluorescence.

Since the RecA-mGFP strain is merodiploid, both the *recA-mgfp* gene fusion construct and the unlabeled endogenous *recA* gene are expressed simultaneously [34]. However, their expression levels are not necessarily identical, nor equivalently inducible by MMC. From previous estimations of the relative *lexA* suppression rates of the relevant *recA* promoters [51], reasonable expectations are that a majority of the RecA present in the cell will be labelled with mGFP, and that RecA-mGFP is 2-3 fold less inducible under the SOS response as endogenous RecA [52]. We estimated the different cellular levels of unlabeled RecA *vs.* RecA-mGFP using Western blotting (Figure S4), which confirmed that the RecA-mGFP was in excess compared to the endogenous protein before and after treatment. Both qPCR and Western blots indicated that both endogenous RecA and RecA-mGFP are inducible by MMC treatment (Materials and Methods 4.4), with the RecA-mGFP indeed about half as inducible (Figure S4). Therefore, the total (i.e. labelled plus unlabelled) amount of RecA protein present, whether as stoichiometry, periodicity or protein copy numbers, is higher than that reported for the RecA-mGFP data directly, by an approximate correction factor of 1.3-fold in the presence of MMC. In the absence of MMC the relative amount of RecA-mGFP to RecA is large enough that the correction factor is effectively 1. As these corrections are indicative, we do not apply them in the early stages of the Results, but present them later only where relevant to interpretations (Results 2.3 and Discussion).

We find that in the absence of MMC, RecA-mGFP has an an approximately uniform distribution in the cytoplasm that is occasionally punctuated by bright fluorescent foci that can be linked into tracks (Figure 1B). The cellular protein copy number of RecA-mGFP increases from 11,400 ± 200 molecules (±SEM) in untreated cells to 19,500 ± 300 molecules in MMC treated cells (Figure S5A). MMC treatment resulted in the subset of these RecA-mGFP molecules that are localized in tracks (*i.e.*, the mean summed stoichiometry of all tracks detected in the whole cell) approximately doubling from 510 ± 30 to 1,080 ± 60 molecules per cell (Table S1). We denote the fluorescently detected, but untracked, molecules of RecA as residing in a *pool*. The pool typically comprises molecules that are sufficiently dim, out-of-focus, and/or rapidly diffusing to evade direct particle-tracking-based detection; here the RecA concentration is exceptionally high such that the stochastic fluctuations corresponding to motion of discrete foci are partly averaged out. During photobleaching, the density of foci decreases, overlap decreases and tracks become more evident. The proportion of RecA-mGFP molecules in tracks is relatively low compared to the pool, but remains representative of the population of assemblies containing RecA-mGFP.

RecB-sfGFP also exhibited fluorescent tracks against a relatively diffuse background, before and after MMC treatment (Figure 1F,H). RecB-sfGFP foci were observed more commonly near the poles of the cell regardless of MMC (Figure 1G-H). Since the RecB-sfGFP fluorescence signal is comparatively small, estimates based on cellular protein copy number must account carefully for autofluorescence due to native components other than GFP. We estimate that the contribution of autofluorescence from the summed pixel intensity values from unlabeled MG1655 parental cells grown and imaged under identical conditions. We find that the mean level of RecB-sfGFP fluorescence was almost three times greater than the cellular autofluorescence (Figure S5B), therefore there is a comparatively large population of the cellular RecB-sfGFP that evades direct particle-tracking-based detection (*c.f.*, slower sampled images from commercial confocal/epifluorescence microscope systems) and thereby comprise a RecB pool.

The cellular protein copy number of RecB-sfGFP does not decrease significantly following MMC-induced DNA damage, comprising 126 ± 11 molecules per cell before treatment and 101 ± 14 molecules following MMC treatment (Figure S5B, Brunner-Munzel (BM) test, n=246, p=0.0216 | NS, not significant at Bonferroni-adjusted $\alpha = 0.01$). However, the mean number of RecB-sfGFP localized into tracks does decrease with MMC; just 13.6 ± 0.5 molecules per cell in all tracks, decreasing to 9.3 ± 0.3 on MMC treatment (BM test, n=246, p<0.001). This is clearly much smaller absolute number of tracked molecules per cell compared to RecA-mGFP, but a similar proportion of the cellular protein copy numbers (ranging from 6-10% in each case). The complementary fractions of the total RecA-mGFP and RecB-sfGFP molecules assigned to their respective pools are thus consistently high (89-95%). In the respective strains, the total concentration of RecB-sfGFP is much lower than that of RecA-mGFP, and this likely indicates the correspondingly more rapid diffusion of RecB-sfGFP species within the pool.

2.2 RecB forms characteristic puncta which are partially lost on MMC exposure

We detected typically 1-3 tracks of RecA-mGFP or RecB-sfGFP in each cell above the local background fluorescence (Figure 1I,J). However, each showed strongly opposing trends in the number of tracks observed upon MMC treatment. While RecA-mGFP showed no significant change in the mean number of tracks on MMC treatment (from 1.66 ± 0.06 to 1.86 ± 0.16 tracks per cell, BM test, n=60, p=0.50 |NS), MMC reduced the population average number of RecB-sfGFP tracks significantly, from 2.06 ± 0.09 to 1.56 ± 0.06 per cell.

If, however, we set aside the fraction of cells with no detected RecB-sfGFP tracks, the change in the mean number of RecB-sfGFP tracks is marginal, from 2.20 ± 0.07 to 1.98 ± 0.07 tracks (BM test, n=234, p=0.006). We see that the cells which continue to harbor RecB tracks are relatively unchanged by MMC, each containing an average of 12.1 ± 0.3 molecules per cell. The unexpected subset of cells that are devoid of RecB-sfGFP tracks increases from 6% to 21% of the population on MMC treatment. These otherwise resemble the other treated cells; rather than filamenting, they retain 92 ± 3% of the population averaged cell length and retain the same pool level of untracked RecB-sfGFP molecules.

The increase in the fraction of cells lacking RecB-sfGFP foci agrees with a model of random, independent survival of assemblies (Figure 1J, the MMC+ condition is consistent with Poisson distribution with same mean; Pearson χ^2 test, dof=6, n=234, p=0.004).

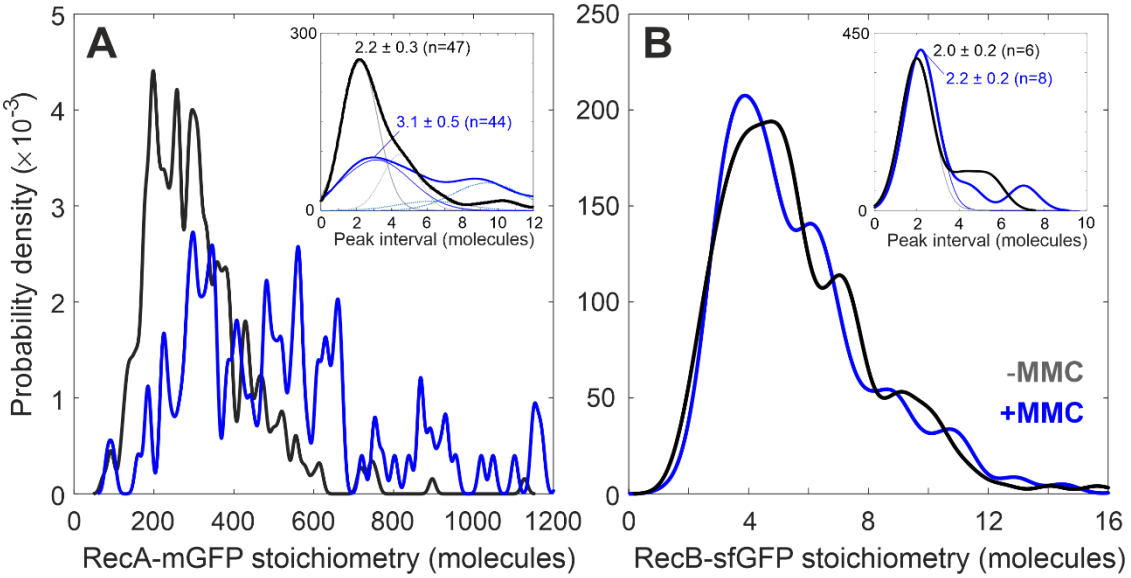


Figure 2. Stoichiometry distributions of detected foci of A) RecA-mGFP and B) RecB-sfGFP with (blue) or without MMC treatment (black), shown as kernel density estimates [53]. The statistics used for MMC- (MMC+) conditions include N=190 (67) whole RecA-mGFP cells containing n=316 (125) tracks, or whole N=249 (307) whole RecB cells containing n=514 (478) tracks within the cell masks. The use of 'probability density' reflects the fact that each distribution is continuous with a total area equal to 1, such that areas under the curve correspond to the probability that the stoichiometry of a given assembly falls within a range. The kernel width (the width for smoothing the discrete stoichiometry of each track) is 0.7 molecules following the known detection sensitivity to single GFP (A, inset and B, both panels), or 8 molecules for clarity (main panel A). Insets are the distributions of intervals between nearest neighbor stoichiometry peaks (solid curves) whose modal position, or periodicity, indicates the number of GFP-labeled molecules in a repeating subunit within molecular assemblies (Table 1). Overlaid are heuristic Gaussian fits that minimize a reduced χ^2 metric, with components of equal width and whose centers are fixed at integer multiples to account for the detected optical overlap of an integer number of subunit repeats of tracked foci. The resulting fits comprise three components for RecA-mGFP with MMC treatment (blue, Pearson's R^2 = 0.979, 5 degrees of freedom (dof)) and two components for RecA-mGFP without MMC (grey, Pearson's R^2 = 0.961, 4 dof). The mode of the peak interval is indicated ±95% confidence interval, alongside the number of contributing peak pairs in the original stoichiometry distribution.

RecA-mGFP foci were approximately two orders of magnitude brighter than those of RecB-sfGFP, corresponding to a greater apparent stoichiometry. A subset of polar assemblies in untreated cells are especially bright (Figure 2B); we defined this subset quantitatively by thresholding at 2× the mean stoichiometry of all assemblies. The mean stoichiometry of these bright assemblies is itself as high as 760 ± 40 molecules (Figure S3A).

On treating with MMC, the RecA-mGFP mean stoichiometry almost doubled from 310 ± 8 to 580 ± 30 molecules per focus, reflecting further local accumulation of RecA-mGFP protein (Figure 2A). We find that most RecA-mGFP molecules comprise an untracked, diffusive pool, in which there are ~30 RecA-mGFP molecules in an area corresponding to that of a typical diffraction-limited focus (which we denote as the *pool stoichiometry*, Table 1). The fact that the relative increase in pool stoichiometry with MMC treatment to ~50 RecA molecules (Figure S5C) is smaller than the fractional increase in the amount of RecA in tracks (Figure 2), indicates that the MMC-driven upregulation of RecA disproportionately affects tracked assemblies. As such, either i) new assemblies are formed which contain much more RecA than those before MMC treatment, or ii) those assemblies that already

contain local concentrations of RecA accumulate more RecA. Under our treatment protocol, these changes do not deplete the reservoir of RecA in the cytoplasm. This observation of localized accumulation of RecA-mGFP is consistent with prior reports of long nucleoprotein filament formation on single stranded DNA [34]. The increased number of RecA tracks we observe upon MMC treatment may therefore indicate greater occurrence of processed ssDNA.

The RecB-sfGFP mean stoichiometry decreases very slightly from 6.6 ± 0.1 to 6.1 ± 0.2 molecules per focus (BM, n=478, $p<10^{-6}$) (Figure 2B). A mean of approximately 6 RecB-sfGFP molecules in each case can be explained if the assembly contains 3 identical subunits whose periodicity is 2 molecules (Figure 2B inset). A pool stoichiometry of ~1 molecule of RecB-sfGFP (Figure S5B,D) suggests that the untracked RecB-sfGFP are likely to be monomers irrespective of MMC treatment (BM test, n=243, p=0.27 | NS). We find that the untracked pool of RecB-sfGFP comprises 90 ± 1% of the total RecB-sfGFP molecules in the cell.

2.3. RecA reorganises into filaments with 3-4-mer subunits in response to MMC

RecA-mGFP and RecB-sfGFP stoichiometry distributions show clear and reproducible peaks (Figure 2A and 2B). One explanation is that each detected fluorescent focus has a diffracted-limited width of ~250 nm that may potentially contain more than one 'subunit' of RecA-mGFP or RecB-sfGFP, bound sufficiently to co-track, such that the measured focus stoichiometry may appear as an integer multiple of that subunit, manifest as periodic peaks on the focus stoichiometry distribution. The expected difference between pairs of values on the stoichiometry distribution is thus either zero or an integer multiple of the periodicity within measurement error. The magnitude of the most likely non-zero pairwise difference value corresponds to the periodicity, with less likely values corresponding to harmonic peaks. Our approach uses a modal estimate of the nearest-neighbor peak intervals (Materials and Methods 4.3.5), and therefore produces a continuous, heuristic estimate for the periodicity. We then compare this periodicity metric to realistic models with integer numbers of molecules. RecA-mGFP tracks have a periodicity of 2.2 ± 0.3 molecules before addition of MMC (Fig 2A inset). This is clearly most consistent with a dimeric subunit of RecA in structures before MMC treatment. After MMC treatment, the most likely interval value is 3.1 ± 0.5 RecA-mGFP molecules, and estimating the additional unlabeled RecA content indicates a likely overall periodicity range of 3-4 RecA molecules (see Discussion and Figure S4).

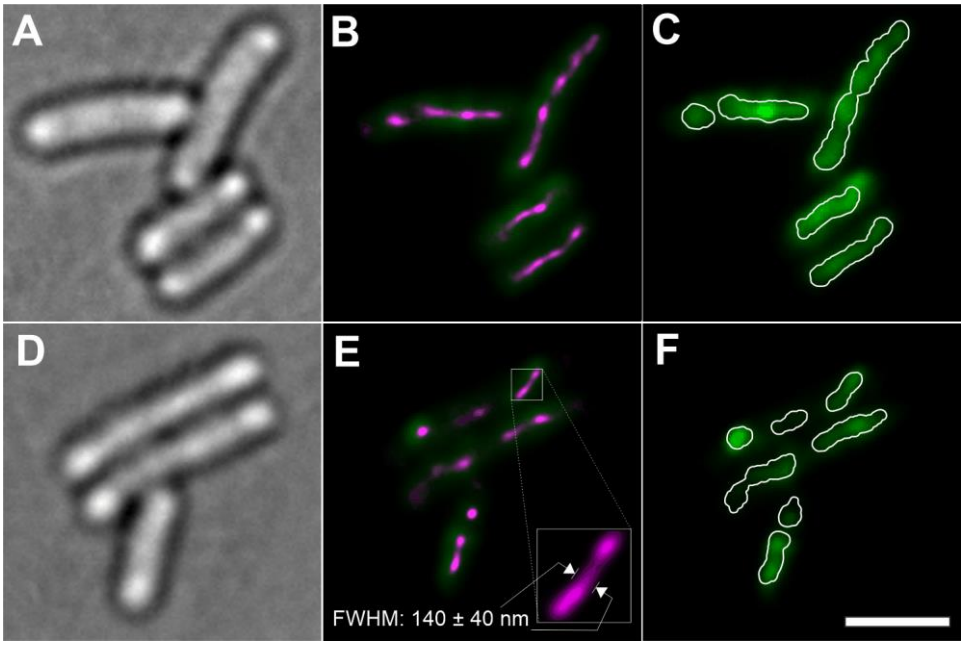


Figure 3. Bundles of RecA-mGFP filaments in MMC treated cells as observed in A,D) brightfield and B,E) The initial Slimfield fluorescent frame (green) overlaid with all super-resolved single-molecule tracks from the acquisition (*ca.* 40 nm spatial precision, with point localizations from foci visualized as a Normalised Gaussian rendering in ThunderSTORM, Materials and Methods 4.3.7), revealing filaments with high spatial precision (magenta); note that the contrast for the green Slimfield channel is set to half to aid the visibility of the superresolution rendering. C,F) Slimfield at full contrast, overlaid with segments derived from each super-resolved bundle by Otsu thresholding and expanding the resulting image masks by the point spread function width of 180 nm, so as to match the diffracted-limited widefield image optical resolution (white overlay); these segments were then used to calculate the segment protein copy number. Scale bar 2 µm.

In MMC-treated cultures we observe strikingly bright, elongated structures (Figures 3B-E). These resemble parallel or intertwined RecA-mGFP nucleoprotein filaments, that we denote as *bundles* following similar observations by others [21,28,34,42,54]. The bundles were identified in a pointillistic manner by overlaying the tracked foci with our measured localization precision of 40 nm. Though it is unclear whether this segmentation is able to distinguish individual filaments or bundles of RecA from one another, the segments reproduce the contiguous morphology of the bright structures at a diffraction-limited optical resolution (Figure 3C,F).

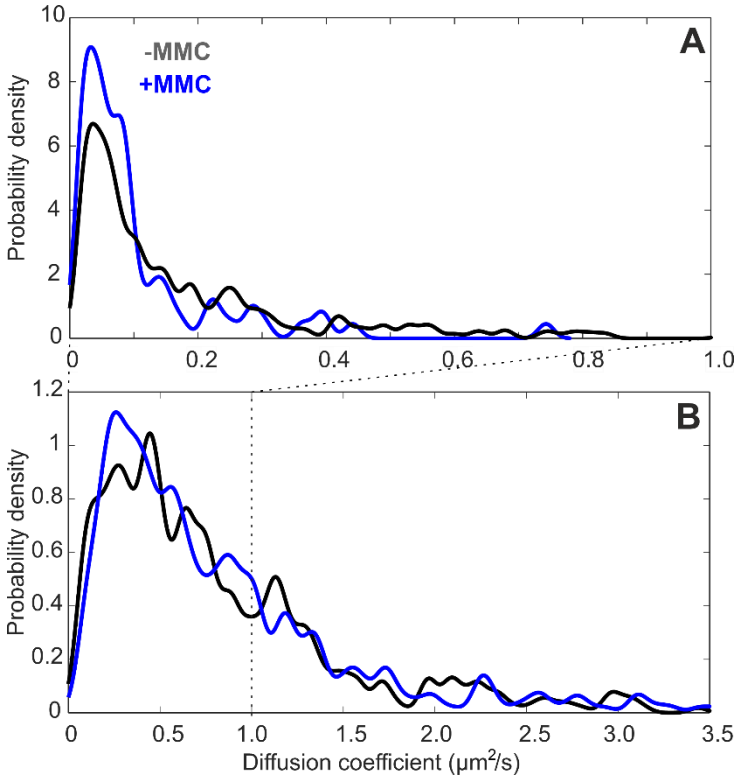
Though single contiguous segments are evident along the full length of some cells (Figure 3C), the mean number of segments is 1.8 ± 0.4 per cell (Figure S3B). We occasionally observed several small segments per cell, in quantitative agreement (Figure S3B) with our expectation that segments occur at random under of a Poisson distribution, albeit conditioned on the presence of at least one segment persistently occurring per cell. Assuming that DNA damage also occurs randomly but under unconditional Poisson statistics, the number of unaffected cells can only be small when there is significantly more than one segment-inducing damage site per cell at any time. It is not clear how many DSBs per cell cycle an *E. coli* culture can sustain without loss of viability, but repeated stalling and collapse of the replisome is common, and cells with single chronic DSBs are known to replicate almost normally within the confines of an elevated SOS response [55]. Under the relatively mild MMC treatment conditions of our study, we observe a comparatively small fraction of the available RecA-mGFP in bundle-associated segments (here the sum of the segment protein copy numbers is <40% of the cell protein copy number, in contrast to 70% [34]). These observations favour an explanation that much of the RecA in elongated MMC-induced structures is bound to a relatively large number of ssDNA nicks as well as a small number of DSBs per cell at any one time.

The high estimated amount of RecA (Figure 2), and the substantial super-resolved breadth of these objects (Figure 3, Figure S3D) above the ~40 nm width of individual filaments [17] suggests that these are bundles comprised of either multiple RecA filaments, and/or multiple windings thereof. We calculate the segment protein copy number within each of these segments in a similar manner to each whole cell. We find the segment protein copy number is 2,800 ± 200 RecA-mGFP molecules (Figure S3A). That means each segment typically includes about three times as much RecA in total than the brightest polar assemblies detected in untreated cells (Figure S3A). Greater than 95% of these segments contain a track whose stoichiometry exceeds twice the mean stoichiometry of untreated cells. The RecA structures observed after MMC treatment cannot therefore be produced solely from the large RecA assemblies prior to MMC treatment, but most likely recruit additional RecA from the cytoplasmic pool. We cannot measure the ratio of RecA to available ssDNA directly, however, the high measured amount of RecA provides some indication that it occurs in high enough excess to form RecA-rich bundles rather than simple nucleoprotein filaments. The binding site density on each helical filament containing ssDNA was found in previous studies to be 1.5 nm per RecA in the presence of ATP [56,57]. As the individual filaments are known to be undersaturated with RecA under physiological conditions [58], one would expect a longer filament per molecule. In contrast, we find that each

349 bundle-associated segment typically measures 900 ± 400 µm (mean ± s.d.) in length, 140 ± 40 µm wide
350 (Figure S3C-E) and no greater than ~0.4 µm deep (based on depth of focus constraints), but contains
351 a quantity of RecA we estimate sufficient to produce >7 µm total length of individual helical filament
352 based on known structures [59]. The longest segments have a more efficient packing density of RecA-
353 mGFP (Figure S3F) which approaches the binding site saturation limit of 1.5 nm / molecule. This link
354 between length and efficiency could result from the functional alignment and elongation of the
355 filament along the cell axis, meaning fewer re-entrant windings of any bundles, and exposure of vacant
356 binding sites to free RecA in cytoplasm.

357 In contrast, the brightest RecA assemblies in untreated cells occur in isolation, and are never elongated
358 but reside within diffraction-limited foci (Figure 1B). Defining these as containing RecA exceeding
359 twice the mean labeled stoichiometry, these occur in 10 ± 3% of untreated cells and have a typical
360 content of 800 ± 100 RecA molecules (Figure S3). This relatively high density is equivalent to >2 µm of
361 filament packing inside a sphere <0.4 µm in diameter. While these assemblies resemble RecA storage
362 bodies, as suggested previously [21], the *recA4155* R28A mutation has been shown to inhibit the
363 formation of true DNA-independent storage bodies [24]. Despite the presence of wild type RecA, it is
364 likely that our observations before MMC treatment indicate DNA-bound RecA bodies that are not
365 filamentous.

366



367

368

369 **Figure 4.** Distributions of instantaneous microscopic diffusion coefficient for tracks of A) RecA-mGFP
370 and B) RecB-sfGFP obtained from Slimfield. Kernel density estimates were generated with a kernel
371 width of 0.008 $\mu m^2/s$ corresponding to the lower bound uncertainty in diffusion coefficient, estimated
372 as the localization precision / $(\text{timestep})^2$. Statistics are as shown for Figure 2.

373

374 The diffusive dynamics of RecA assemblies are also indicative of their state of condensation into
375 filaments. Returning to the tracked foci of RecA-mGFP, we noticed that the mean diffusion coefficient
376 decreases sharply from 0.17 ± 0.02 $\mu m^2/s$ to 0.07 ± 0.01 $\mu m^2/s$ following MMC treatment (Figure 4A).
377 This initially low diffusivity, and the further drop in diffusivity, likely reflect the proportion of RecA

condensed onto ssDNA. MMC induces formation of filaments and these are relatively static on the *ca.* 10 s timescale of the Slimfield acquisition. In contrast, we find that the mean diffusion coefficient of tracked RecB is not significantly affected by MMC treatment (Figure 4B), with untreated and treated values of 0.82 ± 0.03 $\mu m^2/s$ and 0.79 ± 0.03 $\mu m^2/s$ respectively (BM test, n=478, p=0.48 | NS). The diffusivity of RecB-sfGFP in tracks is still lower than expected for a single molecule freely diffusing in bacterial cytoplasm of ~10 $\mu m^2/s$, based on simplistic assumptions of a hydrodynamic diameter of ~10 nm, and contrasts with the large amount of pool RecB-sfGFP that diffuse too quickly to be tracked. This observation hints at the tracked subset of RecB forming larger complexes with other partners not detected here, such as RecC and RecD.

3. Discussion

We used Slimfield to investigate the stoichiometry and spatial location of fluorescently tagged RecA and RecB proteins in live *E. coli* upon treatment with the DNA cross-linking and alkylating agent MMC. RecA and RecB are repair proteins whose involvement in MMC-specific damage repair pathways, as part of the SOS response or otherwise, is unclear. We probed the steady state effect of MMC on RecA and RecB at the minimum inhibitory concentration, which is relevant to sub-lethal antimicrobial exposure. Our results show that the sensitivity and dynamic range of Slimfield is sufficient to quantify counts, either by stepwise photobleaching of multi-molecular complexes or by direct detection of single molecules using millisecond sampling.

RecA assembly formation is not solely correlated with induced DNA damage. Before treatment with MMC, we find that a portion of RecA appears in foci at an average incidence of approximately 2 foci per cell. In 10% of cells, at least one of these foci is especially bright, circular and localized to one of the cell poles. A previous study reports that a minority of cells (4-9%) exhibit spontaneous RecA foci near the poles prior to DSB induction [34]. It has been suggested that wild type RecA foci at the cell membrane might act as nucleation points for later filament formation across DSBs [60], or that these are storage bodies outside the nucleoid [21]. However, the RecA-mGFP strain used here (and in [34]) is a *recA4155* (R28A) genotype which abolishes DNA-independent aggregation of RecA [52,61]. In this strain, we cannot eliminate the possibilities that wild-type RecA forms native storage bodies that are undetected due to exclusion of RecA-GFP, or indeed visible storage structures which do recruit the mutant RecA-mGFP (RecA4155), which would account for the resemblance of detected foci to previous observations of these bodies outside the nucleoid [21,34,61]. In the case where RecA-mGFP cannot participate in storage bodies and can only aggregate in the presence of DNA, there is an alternative explanation for the subset of RecA-mGFP foci we observe, distinct from membrane anchors and storage bodies. These foci do not appear to require RecB for spontaneous assembly [34] indicating that they are independent of DSBs and instead assembled at incidental sites of ssDNA. The foci lie consistently at the periphery of the cell, which indicates they are not likely to be associated with ssDNA within replication forks. These occasionally bright foci may instead simply reflect stochastic ssDNA nicks in a small proportion of cells of an otherwise healthy culture.

Our findings show that the RecA-mGFP copy number increases upon treatment with MMC. We observed a modest increase in the number of tracks, but whose stoichiometry per focus is almost twice those of untreated cultures. This observation of spatially localized RecA and is consistent with significant assembly formation ultimately leading to formation of long nucleoprotein filaments on ssDNA as nucleated from polar locations [34]. These filaments are known to accumulate into bundles as posited by Story et al. [42]. We observed filamentous bundles in MMC treated cultures, possibly due to increased availability of processed ssDNA from DNA damage sites. RecA-assisted homologous recombination and RecA* disassembly occur on a timespan between 15 min [17] and 2 hours [34]. We detect a large increase in RecA stoichiometry (Figure 2) and cellular protein copy number (Figure S5A) and decrease in diffusivity (Figure 4) even after 3 hours' treatment, indicating that RecA bundles continue to form in response to constantly accumulating DNA damage.

Our observation of an about 2 intracellular segments per cell (Figure S3B) is consistent with approximately 2 MMC-induced RecA bundles each extending along opposite halves of a cell (Figure 3F) at any one time in the steady state. This observation may indicate the presence of a double-strand break (DSB) with nearly-bridged loci. However, according to the schemes in previous work [17,34], the development and breakdown of filaments [17] and bundles [34] takes typically <20 min, while for bundles only, recombination is the rate-limiting step, taking up to 90 min [34]. It follows that labeled bundles associated with DSBs would be expected to be bridged for most of their visible lifetime. It is therefore possible that either i) multiple DSBs are present and the segments correspond to different simultaneously bridged DSBs, or that ii) one bridged DSB is present alongside other defects which support RecA filament binding, such as ssDNA nicks.

Intracellular segments were typically aligned along the cell axis (Figure 3B,E) in agreement with the observations of filaments and bundles by other authors [17,34]. While some degree of alignment is expected for all segments much longer than the cell diameter (0.78 ± 0.05 µm), we note that more than half of the detected segments are shorter than this (Figure S3E), which may suggest an alignment mechanism that is not solely due to segement length. Moreover, segments appeared to follow the central axis of the cell, rather than the cell outline (Figure 3B,E), which suggests they fall mostly within the nucleoid rather than residing at the cell membrane, in keeping with the known DNA repair function of the filaments. Filament extension along the cell axis is not predicated on the presence of sister homology [34] but inherently reduces the dimensionality of the search for any homology to one across the cross-section of the cell, independent of cell length or DNA content [17]. Thus, extension vastly accelerates the search time [17]. However, the cause of the extension is unclear. It may reflect simple polymeric elongation under spatial confinement inside the cell, but extension is entropically unfavourable for a flexible polymer. Stiffening and/or thickening of filaments into bundles [34] would therefore faciliate extension. The bundle model in [34] suggested a thickened central backbone flanked by thin filament ends. The bundles observed in our study appear to be thickened with a typical cross-sectional full-width half-maximum of 140 ± 40 nm (Figure S3C) in agreement with previous observation, 160 ± 30 nm [34]. Rather than a monolithic central section, our observations resemble beads on a chain, or a sequence of thick and narrow sections (Figure 3B,E). We find the median width increases rapidly with segment length (Figure S3D), which in this binary framework, suggests the bulk of the increase in bundle length is taken up by the thickened portions and that the thin sections are relatively short. Yet, individual ~40 nm-wide filaments without thickened portions have also been observed previously to extend dynamically on the scale of minutes or less along the length of the cell [17]. We speculate that this suggests an active process of pole-to-pole translocation of thin filament ends (for example, as proposed in [59]), to facilitate the reduced search time.

The observable periodicity of RecA structures could indicate a difference in their macromolecular organization in response to MMC treatment. We observe a change in the periodicity of RecA stoichiometry from ~2 molecules in foci in untreated cells, to a ~3-4-mer within spatially extended filaments following treatment with MMC, after accounting for the unlabelled RecA content per cell with a correction factor of 1.3 ± 0.1 (Results 2.3 and Figure S4). Previous *in vitro* and *in vivo* studies indicate that RecA undergoes linear polymerization in a head-to-tail fashion, with dimeric nucleation points on ssDNA mediated by SSB [62] consistent with our finding of dimeric periodicity prior to treatment. These also provide evidence for stable trimeric, tetrameric, hexameric and the filamentous forms when ssDNA is present [63], consistent with our findings post-treatment. Our snapshot observation of filament stoichiometry cannot shed light directly on models of dynamic nucleation or stepwise growth, as explored in [64–66]. Rather, it explores molecular details of the characteristic protein subunits within the mature filament at steady state. The helical geometry of the filament, with a pitch of 6 RecA molecules per turn, implies that each group of 6 RecA forms a split-ring structure related to the intact hexameric ring of DNA helicases, but distorted axially such that rings each complete a single helical turn around ssDNA [67]. Such ring-shaped hexamers have been identified *in vitro* for both the wild type RecA protein, and the RecA (R28A) mutant [61] that is fused with GFP in our experiment. Even if isolated oligomers were somehow unstable *in vivo*, a polymeric filament could

conceivably still result from a small, periodic barrier to polymerization corresponding to this split-ring distortion. This points to the hypothesis that the fundamental building block of RecA filaments is a factor of 6, if not a hexamer. However, our stoichiometry analysis suggests variability in the total size of assemblies, with our periodicity results indicating a range of 3-4 molecules per subunit. This could reflect trimers which form half-turns in the filament, or perhaps tetramers as an intermediate between preexisting dimers and hexameric rings. Although these data cannot directly establish whether independent oligomers of wild type RecA occur *in vivo* either on DNA or in the cytosol, it is conceivable that assembly and rearrangement of RecA subunits on DNA could generate the canonical ATP-inactive and ATP-active DNA-binding filaments [68,69]. In light of a recent study highlighting the role of RecN in RecA filament formation and activity [59], it is interesting to pose whether RecA assemblies with the dimeric subunit may be devoid of RecN and are ATP-inactive, and if these might then change to a higher oligomeric form upon DNA damage via the involvement of RecN and its associated ATP-activity.

Our measurements confirm that RecA has a very high concentration in the cytosol of live cells. We observe that untreated cultures comprise approximately 11,000 molecules of RecA-mGFP per cell, which increases to 20,000 RecA-mGFP molecules in cells treated with MMC. Of the latter, 28 ± 7% resides in filamentous bundles large enough to be resolved in millisecond widefield fluorescence images. Applying the approximate merodiploid correction factors that we estimated of 1.0 ± 0.1 and 1.3 ± 0.1 respectively (Results 2.3 and Figure S4), the total copy number is approximately 11,400 ± 200 RecA molecules in untreated cells, increasing to 25,300 ± 400 molecules in treated cells. Though less than the 4-5-fold transcriptional increase suggested by qPCR (Figure S4), the more than two-fold increase of total RecA with MMC resembles the increase detected in western blots (Figure S4). While the RecA copy number we estimate in untreated cells exceeds the *ca.* 5,000 molecules reported previously by Lesterlin et al [34], our more direct estimations are of similar order and correlate with previous work indicating 2,900-10,400 molecules, with the high end of this range obtained from cells in EZ rich medium using a ribosome profiling method [70]. Approximately 15,000 RecA molecules per cell in rich medium were reported previously, using semi-quantitative immunoblotting [71]; the same study found that the RecA copy number increased to 100,000 molecules upon MMC treatment. Large discrepancies between studies in the increase in RecA due to MMC treatment are not only due to treatment dose [72] but also arise from differences in *recA* genotype, culture media and growth conditions, as noted by others [21]. In particular, our study uses a minimally inhibitory treatment with MMC (Figure S1).

While the RecA-mGFP protein is not identical to native RecA in its enzymatic activity [32,43,52], the merodiploid strain used in our study also includes a full complement of native RecA from a single allele. This wild-type protein is expected to mitigate the partial loss of sensitivity in RecA-mGFP in any given mixed assembly, as shown on a cellular level by the similar SOS response profile [34] and lack of filamentation under treatment with MMC (Figure 1, Figure S1). The *recA* wild type allele is expressed under control of the native operator, while the *recA-mGFP* allele is expressed under the *recAo1403* operator. In the absence of treatment with MMC, this operator is known to result in an increased transcription rate of the *recA-gfp* allele relative to the wild type *recA* gene under its native promoter by a factor of 2-3, while both alleles are upregulated to the same level under induction of SOS [52]. Using quantitative Western blotting we estimate that prior to MMC treatment, RecA-mGFP is actually present at several tenfold more than the unlabeled protein (Figure S4), and that in the presence of MMC the ratio of RecA-mGFP to RecA is lower, at approximately 3-4 to 1. From these ratios, we derived our approximate correction factors of 1.0 or 1.3-fold for the total amount of RecA protein, in the absence or presence of MMC respectively. While RecA-mGFP is known to label RecA assemblies [34], it cannot form DNA-independent assemblies by itself [52], and is therefore reasonable to conclude that all labelled sites here represent occupied DNA on which wild type RecA and RecA-GFP are interchangeable. Even if the binding partition of wild type RecA were higher, for example reflecting the relative sensitivity (Figure S4A and [52]), the high relative concentration of RecA-mGFP (Figure S4B) would conceivably result in the majority of RecA sites on DNA being occupied by RecA-mGFP.

Lesterlin *et al*. showed that RecA immunostaining of filaments (agnostic as to GFP labelling) correlates with the fluorescent distribution of RecA-GFP [34], proving that dark filaments exclusively of wild-type RecA cannot be present. Though this result could potentially be interpreted that the structure is entirely RecA-GFP and that the highly sensitive wild-type RecA is excluded, this wild-type RecA would have to somehow rescue DNA repair function in the cytoplasm rather than in filaments, which has no known basis. We therefore assume the presence of hybrid filaments. In any case, the effect of excluding wild-type RecA from filaments would simply mitigate our periodicity correction factor toward unity, and narrow our estimate of the periodicity within filaments toward a value of 3 RecA molecules.

Unlike RecA-mGFP, we detected only modest quantities of RecB-sfGFP in untreated cells grown in minimal medium: 13.6 ± 0.5 molecules in tracks, and 126 ± 11 molecules in total per cell based on integrated GFP fluorescence corrected for cellular autofluorescence. Several previous reports also indicate that RecB is very scarce – typically less than 20 molecules per cell [35,73]. One of these studies estimated that there are just 4.9 ± 0.3 RecB molecules per cell using a HaloTag fusion allele labeled with HTL-TMR, and 4.5 ± 0.4 molecules per cell using magnetic activated cell sorting of the same RecB-sfGFP strain that we use here, albeit in M9 medium and restricted to nascent cells for which the average copy number is approximately halved [35]. An earlier mass spectrometry study used intensity based absolute quantification to estimate 9-20 RecB molecules per cell across different stages of growth in M9 minimal media [73]. Ribosome profiling estimated the RecB copy number to be 33-93 molecules per cell in different growth media [70]. However, these techniques are either *ex vivo* or necessitate significantly perturbed intracellular crowding that may conceivably result in potentially non-physiological molecular assemblies.

Comparing the number of RecB in tracks in our present study with the number of RecB in distinct foci per cell reported previously, we find a similar albeit slightly higher estimate, possibly because our approach is based on fluorescent fusions with a high labelling efficiency in unsynchronized cultures, as opposed to selecting nascent cells. However, our measurement of RecB copy number exceeds previous estimates. The large remainder in summed pixel fluorescence intensity may represent two possible contributions. The first is from RecB that diffuses faster than Slimfield can track. The highest diffusion coefficients of tracked RecB assemblies approach 3 $\mu m^2/s$ (95% quantiles, Figure 4). We estimate the limit of measurement as approximately 5 $\mu m^2/s$, though it is conceivable that free monomeric RecB-sfGFP could exceed this, given that it has been estimated to reach diffusion coefficients equivalent to approximately 8 $\mu m^2/s$ in *E. coli* cytosol [74,75]. A second possible source is an increase in net autofluorescence relative to the parental strain when the real RecB are labeled; it is unlikely that could account for the discrepancy, since this would require a 3-fold increase in autofluorescence based on our measurements, and such a drastic increase lacks precedence (for example, upon treatment with MMC at a high level sufficient to induce widespread RecA filamentation, our estimation suggests only an increase in autofluorescence of no more than 20%). Furthermore, the measured rate of photobleaching of the diffuse RecB-sfGFP signal matches that of RecB-sfGFP tracks and is roughly half the rate of the autofluorescent parental cells (Figure S6, Table S2). The implication is that untracked RecB-sfGFP is the major contributor to mean cellular fluorescence, which is then a more accurate reflection of total copy of RecB than simply the number of molecules in tracks.

The cellular protein copy number of RecB does not change significantly with MMC (Table S1, Figure S5B,D), suggesting that there may only be a modest regulatory response to DNA damage. Although MMC is known to induce the SOS response and cell cycle arrest [4,48], *recB* expression is itself not induced directly as part of the SOS response. RecB-sfGFP foci increase neither in number (Figure 1J) nor stoichiometry (Figure 2B), which compares with earlier observations that treatment with MMC under similar concentrations to those used in our study do not significantly change RecB expression [22]. In fact, the number of observed tracks per cell *dropped* considerably after MMC treatment, due to a sharp increase in the proportion of cells in which RecB assemblies were absent, from 6% to 21%. This reduction in RecB assemblies was at odds with our expectation that MMC would eventually

increase the recruitment of RecB in response to damage, if not increase the cellular production of RecB. MMC treatment is known to increase the occurrence of DSBs and thereby drive demand for DSB processing [4] that is typically mediated by RecB. Yet, rather than initiating cellular upregulation of RecB, treatment with MMC acts to partially deplete localized assemblies. Given that DSBs are likely to occur in the majority of cells under our MMC treatment, as indicated by the ubiquitous induction of RecA filaments (Figure 3), the fate of RecB assemblies cannot simply reflect the presence or absence of DSBs. The increase in the fraction of cells lacking RecB-sfGFP tracks is consistent with random, independent survival or breakdown of assemblies (Figure 1J). This result may indicate a situation where pre-existing RecB (hetero)complexes at foci are occasionally disassembled while interacting with sites of MMC-induced DNA damage, such as DSBs. This instability of (presumably heteromeric) RecB assemblies might result from successfully bridged pairs of RecA filaments, however, we did not detect any correlated loss of pairs of RecB foci, as might be expected for recombination events. Notably, the number of tracked foci detected per cell is approximately 2 for both RecA-mGFP and RecB-sfGFP. Future colocalization studies of RecA and RecB assemblies may offer more direct insight into the functional interaction and turnover of these repair proteins in regards to whether the average of 2 be related to the number of replication sites, or perhaps simply reflects a small average number of severe DNA damage sites per cell.

Independent of MMC treatment, we observed a dimeric periodicity for RecB-sfGFP. This suggests that RecBCD heterotrimers occur in pairs *in vivo*. Indeed, earlier *in vitro* studies identified the occurrence of $(RecBCD)_2$ complexes, possibly held together by the nuclease domains of the two RecBCD monomers [12]. However, the authors concluded that the monomeric form is functional while the dimeric form is nonfunctional [13]. Furthermore, crystallization of the RecBCD complex for structural studies contained two RecBCD-DNA complexes in the asymmetric unit [76]. Our observations cannot determine covalent interactions directly between RecB molecules, but their cotracking is very strongly correlated. We can infer two details: first, that the dimeric form of the complex, $(RecBCD)_2$, occurs in live cells, and second, that previous *in vitro* observations of dimers are carried over from their physiological state. Our findings suggest a hypothesis that assemblies with multiple pairs of RecB have a greater activity on DSBs than isolated RecB in the pool. Making the distinction between monomeric RecBCD in tracks and monomeric RecB in the untracked pool suggests that RecB monomers in the pool could potentially act as a reservoir. One may consider the alternative situation, where the monomeric pool are the functional RecB elements and the assemblies are reservoirs that disassemble in response to damage, but this makes less sense, since those monomers would already be in excess. A mean stoichiometry of ~6 molecules indicates that RecB foci may occur as colocalized assemblies that comprise roughly three pairs of RecBCD heterotrimers (Figure 5). It would be interesting to estimate the stoichiometries of RecC and RecD in future studies to understand their association in processing DSBs in greater detail.

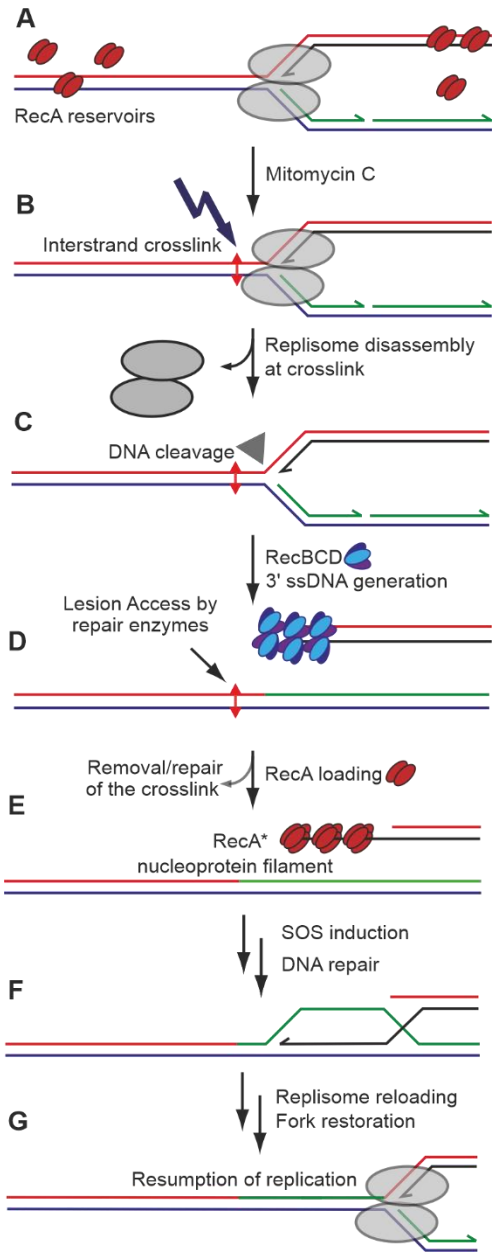


Figure 5. A model of DNA damage caused by treatment with MMC and subsequent repair at the replication site by RecA and RecBCD. A) Intact replication fork; occasional binding of multiple RecA dimers to DNA away from the fork as well as RecA dimers as DNA-free storage bodies in the cytoplasm; B) exposure to MMC and induction of an interstrand crosslink that acts as a barrier to an approaching replication fork; C) replisome dissociates if unable to overcome barrier; dissociated fork is recognised by branched DNA specific endonucleases (filled triangle) that can eventually cause DSBs leading to replication fork collapse; replication fork collapse allows access to repair enzymes to recognise the lesion; D) a newly generated DSB is recognised by RecBCD and processed to generate a 3' single strand end; E) RecA dimers identify the newly generated ssDNA and assemble in groups of 3-4-mers into RecA* filaments; RecA is shown as a short stretch for illustrative purposes but may extend for many thousands of molecules over several hundreds of nm of ssDNA, and these filaments may be twisted and/or grouped into bundles. F) Strand exchange followed by processing of the double strand break, then recombination sufficiently upstream of the lesion and subsequent G) reloading of the replisome. This process allows sufficient time for the repair enzymes to repair the lesion on the template strand,

633 so that replication may resume. For a detailed overview of the possible pathways to fork restoration,
634 refer to [11].

635 While MMC-induced damage constitutes a range of chemical moieties [77], the canonical mechanism
636 of MMC toxicity is of interstrand crosslinks at dG sites [6,7]. The specific repair of interstrand crosslinks
637 (implied in Figure 5) can involve several repair pathways, primarily nucleotide excision repair (NER),
638 which converts the crosslinks into dsDNA breaks [78]. Although NER enzymes such as UvrD typically
639 degrade RecA filaments, NER is involved in the cleavage of damaged replication forks into suitable
640 substrates for downstream processing, including RecA-mediated recombination [79]. Repair of the
641 fork is then completed, for example by PriA-, Rep- and PriC-dependent pathways [11,18,80] on sets of
642 ssDNA and a dsDNA end (Figure 5). The observation of a greater increase in RecA-mGFP copy numbers
643 and foci compared to RecB-sfGFP could indicate a significant proportion of single-strand breaks and
644 single-strand gaps at sites of crosslinks. While two previous studies reported that NER action on
645 crosslinks also produces ssDNA nicks [78,81], we do not know if this applies strictly to MMC-induced
646 NER, as our present work does not pertain to genes that *process* ss-gaps. Future analysis of proteins
647 that process ssDNA breaks may potentially shed light on the relative occurrence of the two types of
648 breaks by MMC and their relation to repair of replication forks.

649 While others have shown that *recB* deletion abolishes UV-induced filaments of RecA [34], we do not
650 know the effect of *recB* deletion and MMC treatment on RecA dynamics. To avoid RecA interference
651 in 'normal' ssDNA processes such as replication, the cell maintains strict control over filament
652 nucleation, based on RecA and associated cofactor concentrations. It is therefore likely that the
653 observed filamentation upon treatment with MMC is dependent on RecBCD, indirectly pointing
654 towards increased occurrences of DSBs in these cells. Alternatively, if RecA nucleation is independent
655 of RecBCD, one might anticipate little change in RecA dynamics upon *recB* deletion. However, further
656 analysis of MMC-dependent RecA stoichiometry and copy number in a strain devoid of RecBCD activity
657 - and with a controlled RecFOR pathway [19] – is needed to differentiate between these models.

658 In conclusion, RecA occurs as assemblies located near poles of wild type cells in a dimeric periodicity
659 consistent with nucleation models. Upon mild treatment with MMC, RecA is upregulated at least two-
660 fold, and assembles into long filamentous bundles on newly generated ssDNA in effectively all cells
661 without exhausting the cytoplasmic reservoir. These mature bundles have a much lower diffusivity,
662 reflecting their aggregation of a few thousand molecules each, with a structural periodicity in the
663 range of 3-4 RecA molecules. The bundles are typically wider than single filaments, but both forms are
664 known to facilitate homology search for homologous base-pairing with an intact duplex. Generation
665 of ssDNA is known to occur at a DSB induced by processing of disassembled forks upon recognition by
666 RecBCD. We observed RecB as a moderately diffusive set of three associated dimers at two locations
667 in the cell, providing further evidence that RecBCD predominantly occurs as pairs of heterotrimers
668 inside the cell at either end of DSBs. Our work implies the existence of a separate, significant reservoir
669 of highly diffusive RecB monomers. Neither of these forms of RecB are upregulated upon MMC
670 exposure, nor do they change their mobility. Accordingly, RecB is not a part of the SOS regulon.
671 Instead, MMC-induced DNA damage impacts the formation – or induces a higher turnover – of these
672 periodic RecB assemblies potentially associated with further DSB repair.

673

674 **4. Materials and Methods**

675 *4.1. Strains, culture and MMC protocol*

676 Three strains of *E. coli* were used in this study without alteration:

677 **Control:**
678 *MG1655*
679 **RecA-mGFP** [34]:
680 *MG1655 rpsL (Str^R,lac+) ygaD1::kan recAo1403 recA4155,4136-gfp901, fhuB::recAwt-cm*
681 **RecB-sfGFP** = MEK706 [35]:
682 *MG1655 recB::sfGFP*

The RecA-mGFP strain used here is the same as in [34]. It is a merodiploid, natively promoted derivative of the SS3085 strain [52,82,83]. It expresses both i) the wild type unlabelled RecA protein from a single, ectopic wild-type *recA* allele, under a wild-type operator, and ii) a labeled mutant protein at the native *recA* site with a mutant operator *recAo1403*, which has a higher transcription rate than wild type [52]. The labelled *recA* is a *recA4155* (R28A) mutant, which complements the wild type recombination function *in vitro* [34,61] but has important differences in self-assembly. Unlike wild type RecA, the *recA4155* forms of RecA cannot alone form assemblies independently of DNA as proven by comparisons *in vitro* [61] and competitive binding studies *in vivo* [52]. The *gfp-901* label is the same as *mut2* (A206T) which corresponds to a monomeric GFP (mGFP) [84]. The notation *recA4136* refers to the insertion of a linker, as well as the *gfp-901* gene, between the penultimate and ultimate stop codons of *recA*. The fusion with mGFP impairs the recombinant sensitivity of the labeled RecA protein; were the strain to include only the fused allele, it would be fully as SOS inducible as wild type MG1655, but only approximately two-thirds as UV resistant, and would be compromised up to *ca.* 10-fold for recombination activity [21,52]. Induction of the SOS response would also take 30 min, or roughly twice as long as wild type (the R28A mutation prevents this from being an additional 2× slower) [52]. The merodiploid strain rescues both these functions and their kinetics: functional RecA filaments labeled with 70% of the total available mutant fusion protein form within just 15 min of DNA damage and reach steady state within 90 min, similar to wild type MG1655 [34].
The RecB-sfGFP fusion was constructed in [35] by wild-type *recB* replacement under plasmid-mediated recombination. N-terminal fusions were shown in [35] to be functional using growth curves and tests of DNA repair, in contrast to C-terminal fusions which may disrupt RecBCD complexation [35].
E. coli strains were grown overnight in 56-salts minimal media at 30°C to mid-log phase in an Innova 44 shaker incubator (New Brunswick). The mid-log phase cultures were concentrated to ~100 cells/ml (OD_{600} ~0.3) and split into two equal fractions. Aliquots were adjusted to either nil (MMC-) or the minimum wt inhibitory concentration of 0.5 µg/ml MMC (MMC+) (Figure S6 and [48]) and incubated at 30°C for a further 3 h (Figure S6). Cells were harvested for microscopy on 1% w/v agarose pads suffused with the same liquid media and imaged within 1 h.

4.2. Slimfield

A custom-built Slimfield microscope was used for single colour, single-molecule-sensitive imaging with a bespoke GFP/mCherry emission channel splitter as described previously [11,37]. The GFP channel was recorded, while the mCherry channel was used only as a negative control. The setup included a high-magnification objective (NA 1.49 Apo TIRF 100× oil immersion, Nikon) and the detector was a Prime95B sCMOS camera (Photometrics) operating in 12-bit gain at 180 Hz and 3 ms exposure/frame, for a total magnification of 53 nm/pixel. The samples were illuminated either in brightfield, or for Slimfield fluorescence in camera-triggered frames by a collimated 488 nm wavelength continuous wave OPSL laser (Coherent, Obis LS) in Gaussian TEM_{00} mode at a power density of 5 kW/cm^2. The number of frames per acquisition was 2,000 for RecA and 300 for RecB strains.

4.3 Quantitative tracking and protein copy number analysis

4.3.1 Identification of Slimfield foci and assignment into tracks

Slimfield image sequences were processed by custom ADEMscode software in MATLAB (Mathworks) [33,80,85–87]. This pipeline identified foci from local maxima in pixel values within individual frames. An iterative Gaussian mask algorithm was used to detect the centroids of foci, using a circular region of interest of radius 5 pixels within a sliding window of 17 pixels. The intensity of each focus was calculated as the sum of the circular region corrected for the average background in the surrounding annular region. The prospective foci were accepted if their intensity was >0.4× the standard deviation in the background region. The nearest neighboring foci in adjacent frames within 8 pixels of each other were assigned to the same track, with a minimum of 4 foci per track. The typical track duration was

limited by diffusion and/or photobleaching to a mean of >13 foci per track over ~75 ms real time, or ~40 ms cumulative exposure (Table S1).

4.3.2. Diffusion coefficient

The centroids of the foci within each track, as generated from the ADEMScode tracking analysis above, were used to calculate displacements over the length of each track in chronological sequence. From these, the mean square displacements (MSDs) of each track were calculated by averaging the square of the displacements across equal lag times, corresponding to all possible intervals between frames up to the length of the track. For each track, the MSDs at the four lowest lag times were linearly interpolated (with a constraint on the fit of passing through a specified intercept on the lag time axis, equal to the square of the measured localization precision of 40 nm divided by the frame interval of 5.7 ms). The initial slope of this fit (and corresponding error) was then divided by a factor of 4 according to the 2D diffusion equation [88] to yield a diffusion coefficient (and error) for that track.

4.3.3 Characteristic single-molecule brightness

The intensity of each focus was estimated by integrating the local pixel values with a local sliding window background subtraction. After photobleaching sufficiently to show single photoactive GFP molecules, the characteristic single-molecule brightness of a single GFP molecule was estimated from the modal brightness of these foci. These were confirmed to be broadly consistent with estimates of the signal per GFP in each dataset were determined from the monomeric intervals in total number of counts due to stepwise photobleaching, as identified by a Chung-Kennedy edge-preserving filter (15 ms window, 50% weighting, Figure S7) [89]. This integrated intensity is characteristic for each fluorescent protein under fixed imaging conditions, although mGFP and sfGFP were found to be indistinguishable in this respect, and hereafter referred to collectively as GFP. To ensure consistent counts per single-molecule probe, analysis was restricted to the uniformly illuminated area lying within half of the $1/e^2$ beamwaist of the excitation laser in the sample plane. The integrated intensity of GFP *in vivo* was found to be within 14% and 9% respective errors in RecA and RecB (88 ± 18 and 177 ± 16 pixel grey values per GFP for the respective gain modes). The combined equivalent is 88 ± 7 photoelectrons per GFP per frame, which is precise enough to unequivocally identify groups or steps of up to 12 GFP molecules.

4.3.4 Stoichiometry

Each track is associated with an assembly that contains a certain number of molecules, or stoichiometry, at the initial point of acquisition. To estimate this stoichiometry for a given track, the intensities of the constituent foci were linearly extrapolated using the first 4 datapoints in the track back to the timepoint of initial laser exposure. This initial intensity of this fit was divided by the characteristic single-molecule brightness signal associated with one fluorescent protein under a fixed excitation-detection protocol. The result is a stoichometry expressed as a number of molecules. The standard error associated with a stoichiometry value of 1 molecule is approximately 0.7 molecules. To avoid undercounting bias due to photobleaching, only tracks in the first 10 frames after laser exposure were considered for stoichiometry estimates.

4.3.5 Periodicity

The distributions of track stoichiometry may show periodic peaks, whose smallest reproducible interval can be interpreted as a physical repeat unit or *periodicity* within assemblies. To calculate periodicity, first the stoichiometries of all tracks within each acquisition were represented as a kernel density distribution. The kernel width used was the empirical standard deviation on the characteristic single molecule brightness of 0.7 molecules [41]. Peaks in this distribution were detected using the MATLAB *findpeaks* function, and the intervals between nearest neighbor peaks were calculated. These sets of nearest neighbor intervals for each acquisition were then aggregated across the relevant population of cells. A second kernel density estimate was calculated over the intervals for a population, with a kernel width of 0.7 molecules multiplied by the square root of the mean stoichiometry, divided by the square root of the number of interpolated intervals. The fundamental value of this interval distribution (corresponding to the center of the leftmost peak in Figure 2 insets)

783 was refined by fitting the curve with a sum of Gaussian terms centerd at multiples of the fundamental
784 value. To accommodate the uncertainty in the single molecule characteristic brightness, the
785 fundamental value of the fit was not constrained to an exact integer value but represents a heuristic
786 model for the periodicity. The number of terms in the fit was set to minimize the reduced χ^2 metric in
787 the fit. This modal value was reported with 95% confidence interval as the periodicity of assemblies in
788 each population. This method of estimating periodicity was verified as independent of the mean
789 stoichiometry using simulated data drawn from noisy Poisson-distributed multiples of an oligomeric
790 ground truth (artificial input value). This analysis reproduced the expectation that the minimum
791 number of tracks required for sufficient peak sampling, and therefore the limit of periodicity
792 detection, scales with the square root of the mean stoichiometry.

793 4.3.6 Cellular protein copy numbers and pool stoichiometry

794 The cellular protein copy numbers as reported in the Results Section 2.1, Table S1, Figure S5 and
795 Discussion correspond to whole cell masks, as identified using the manual annotated machine learning
796 segmentation output from brightfield images (Figure S2 and Supplementary Methods). Integrated
797 intensities of cells (uncorrected cellular protein copy numbers) and pool stoichiometries, were
798 determined not from tracked foci, but directly from the raw image sequences using the CoPro package
799 in ADEMscode software following [37] with the characteristic single-molecule brightness of GFP (as
800 described in 4.3.2), the cell masks, and the camera's dark pixel bias as input. The procedure effectively
801 adds up all of the pixel values within the mask in question in an initial frame, and accounts for the
802 convolution of the 3D cell volume with the widefield point spread function, followed by projection
803 onto a 2D image. To obtain the cellular protein copy number in the labeled strains, and account for
804 the contribution of autofluorescence, we calculated the difference in mean integrated intensity per
805 segment between the labeled and parent strains under the corresponding MMC± condition, adjusted
806 by the ratio of mean segment area. The pool stoichiometry in each cell is a measure of its untracked
807 molecular concentration. It is calculated in CoPro as the cell's integrated intensity, less the mean
808 integrated intensity of the parental cells, less the total stoichiometry of tracked foci in the cell, divided
809 by the area of the cell mask relative to the area within one diffraction limited focus.

810 4.3.7 Super-resolved images and segment protein copy numbers of RecA-mGFP

811 The segment protein copy numbers as reported in Results Section 2.3 and Figure S3, were calculated
812 with ImageJ, using as input the segments corresponding to bundles or foci, instead of whole cells.
813 These segments were obtained starting from the coordinates of localized, tracked foci in Slimfield
814 analysis, from the latter stages of photobleaching below a threshold stoichiometry of 2 molecules;
815 these were imported into ThunderSTORM software [90]. The Visualization module to build a
816 pointillistic super-resolved image at 40 nm lateral spatial precision (as shown in Figures 3B,E) at 5×
817 upscaling (11 nm pixel size), which was then smoothed with a Gaussian filter of 4 pixels' width, and
818 automatically Otsu thresholded to generate a superresolved binary mask. The masks were then
819 expanded by a distance equal to the widefield resolution of 17 pixels (~180 nm) to match the features
820 in the Slimfield images. The integrated intensities were extracted, as in [91], from the sum of
821 fluorescent pixel counts in the Slimfield images (Analyze Particles > Multi-Measure function in ImageJ)
822 less the area multiplied by the camera pixel dark value. To yield segment protein copy numbers (Figure
823 S3A), the resulting integrated intensities were corrected for the relative autofluorescence, by
824 subtracting the integrated intensity of parental cells adjusted by by the ratio of mean segment area.
825 The Multi-Measure output also included the Feret diameter of each segment which was used as an
826 estimate of its end-to-end length (Figure 3D-F).

827 4.3.8 Photobleaching rates

828 Photobleaching rates were estimated by fitting the decrease in background-subtracted cellular
829 protein copy number or mean track stoichiometry over the exposure time using MATLAB *cftool*. The
830 fit consisted of a monoexponential decay to the first 10 frames with variable initial intensity and decay
831 constant, but with a baseline fixed to the average intensity after 50 frames. Fits were then refined to
832 include only data within the initial 1/*e* decay time (Table S2). RecA-mGFP and RecB-sfGFP photobleach

833 decay times were consistently dissimilar at 13 ± 2 and 6 ± 1 frames respectively; sfGFP is typically
834 several-fold less photostable than comparable enhanced GFPs under high intensity illumination [92].

835 4.3.9 Statistical tests

836 We performed multiple statistical comparisons on each set of tracked data (typically ~5: number of
837 tracks, stoichiometry, periodicity, diffusivity, copy number), which we account for using the standard
838 Bonferroni correction; the significance level is adjusted downwards by a factor of the number of
839 comparisons, $\alpha = 0.05/5 = 0.01$).

840

841 *4.4 Gene expression assays*

842 4.4.1 Quantitative PCR (qPCR)

843 Treated and untreated cultures were grown as in section 4.1. Total RNA was then isolated using
844 Monarch Total RNA Miniprep Kit (New England Biolabs). cDNA was synthesised from 350 ng of total
845 RNA from each sample using Superscript IV reverse transcriptase (Invitrogen) according to the
846 manufacturer's instructions using random hexamer primer (ThermoScientific).

847 The cDNA was then subjected to qPCR using Fast SYBR Green Master Mix (ThermoFisher) in a
848 QuantStudio 3 Real-Time PCR System. The *recA* primer pair amplified *recA* cDNA in the wild type and
849 both *recA* and *recA-GFP* mRNA in the labeled strain. *recA-GFP* alone in the labeled strain was amplified
850 using *GFP* primer pair. 16s rRNA was used as a housekeeping control.

851 Data obtained was analysed using the standard curve method [51]. Standard curves were generated
852 from serial dilutions of PCR products with known concentrations derived from genomic DNA. Fold
853 increase in mRNA levels was calculated by dividing the values obtained for treated mRNA with the
854 untreated. Results are shown in Figure S4A.

855

Primer	Sequence 5'-3'	Complementary region
oAS216	GCAGGCACTGGAAATCTGTG	*recA* (forward)
oAS217	GCCGATTTCGCCTTCGATTTC	*recA* (reverse)
oAS220	CTACAAGACACGTGCTGAAGTC	*GFP* (forward)
oAS221	AGTTGTATTCCAATTTGTGTCCAAGAATG	*GFP* (reverse)
oAS23	GTAGAATTCCAGGTGTAGCGGTG	16s rRNA (forward)
oAS24	CATCGTTTACGGCGTGGACTACCAG	16s rRNA (reverse)

856 **Table 2. Primers used for qPCR to quantify mRNA of *recA*, *recA-gfp* and housekeeping gene *rrsA*.**

857

858 4.4.2 Western blots

859 Six samples of normalised *E. coli* cell cultures were prepared as above (section 4.1) in 1 ml aliquots at
860 OD_{600} ~0.2. The cells were isolated using centrifugation at 10,000 × *g* for 2 min in a microfuge to
861 prepare them for SDS-PAGE / immuno-detection. The cell pellets were resuspended in 75 µl of SDS
862 loading buffer and boiled for 5 min at 95°C before application of 15 µl onto a 4-20 % gradient gel. The
863 gel was subsequently transferred to nitrocellulose and the membrane was placed in blocking solution
864 (PBS-T, 5% (w/v) non-fat milk) for 3 h. Primary antibody (anti-RecA) was incubated at 1/500 overnight
865 in blocking solution before the membrane was washed (4 x 5 min) in blocking solution. Secondary
866 antibody (goat anti rabbit-HRP) was incubated at 1/2,000 dilution for 4 h in blocking solution before
867 the membrane was again washed (4 x 5 min) in blocking solution. A final wash in PBS was performed
868 before development using ECL and image acquisition (iBRIGHT). Results are shown in Figure S4B.

869

870 **Author Contributions:**

871 A.P-D, A.S, and M.L. designed the research; A.S. cultured and treated cells; A.P-D performed
872 microscopy and data analysis, visualization and curation, J.S. and L.F. wrote and validated
873 segmentation software; A.P-D and A.S drafted the paper; A.P-D, A.S, J.S. and M.L. edited the paper;

M.L. supervised and administered the project. This research was funded by BBSRC, grant numbers BB/P000746/1 and BB/N006453/1, and EPSRC, grant number EP/T002166/1.

Data Availability Statement: The raw imaging data is available on reasonable from https://doi.org/10.5281/zenodo.6639101; the MATLAB tracking analysis code can be found at https://github.com/alex-payne-dwyer/single-molecule-tools-alpd. The U-Net image segmentation architecture originated from code obtained from the NEUBIAS Academy workshop (http://eubias.org/NEUBIAS/training-schools/neubias-academy-home/).

Acknowledgments:

The authors thank Dr. Christian Lesterlin for the gift of the RecA-mGFP strain and Prof. Meriem El Karoui for the RecB-sfGFP strain. We thank the Biosciences Technology Facility for technical assistance with gene expression assays.

Conflicts of Interest:

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

892 **References**

893 1. Reyes-Lamothe R, Sherratt DJ, Leake MC. 2010 Stoichiometry and architecture of active DNA
894 replication machinery in Escherichia coli. *Science* **328**, 498–501.
895 (doi:10.1126/science.1185757)

896 2. Surova O, Zhivotovsky B. 2013 Various modes of cell death induced by DNA damage. *Oncogene*
897 **32**, 3789–97. (doi:10.1038/onc.2012.556)

898 3. Lambert S, Carr AM. 2013 Impediments to replication fork movement: Stabilisation,
899 reactivation and genome instability. *Chromosoma* **122**, 33–45. (doi:10.1007/s00412-013-0398-
900 9)

901 4. Tomasz M. 1995 Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.* **2**, 575–9.
902 (doi:10.1016/1074-5521(95)90120-5)

903 5. Arranz-Marquez E, Katsanos A, Kozobolis VP, Konstas AGP, Teus MA. 2019 A Critical Overview
904 of the Biological Effects of Mitomycin C Application on the Cornea Following Refractive Surgery.
905 *Adv. Ther.* **36**, 786–797. (doi:10.1007/s12325-019-00905-w)

906 6. Weng M-W, Zheng Y, Jasti VP, Champeil E, Tomasz M, Wang Y, Basu AK, Tang M-S. 2010 Repair
907 of mitomycin C mono-and interstrand cross-linked DNA adducts by UvrABC: a new model.
908 *Nucleic Acids Res.* **38**, 6976–6984. (doi:10.1093/nar/gkq576)

909 7. Bizanek R, McGuinness BF, Nakanishi K, Tomasz M, McGuinness BF, Nakanishi K. 1992 Isolation
910 and Structure of an Intrastrand Cross-Link Adduct of Mitomycin C and DNA. *Biochemistry* **31**,
911 3084–3091. (doi:10.1021/bi00127a008)

912 8. Dronkert MLG, Kanaar R. 2001 Repair of DNA interstrand cross-links. *Mutat. Res.* **486**, 217–
913 247. (doi:10.1016/s0921-8777(01)00092-1)

914 9. Legerski RJ. 2010 Repair of DNA interstrand cross-links during S phase of the mammalian cell
915 cycle. *Environ. Mol. Mutagen.* **51**, 540–551. (doi:10.1002/em.20566)

916 10. Smith GR. 2012 How RecBCD Enzyme and Chi Promote DNA Break Repair and Recombination:
917 a Molecular Biologist's View. *Microbiol. Mol. Biol. Rev.* **76**, 217–28. (doi:10.1128/mmbr.05026-
918 11)

919 11. Syeda AH, Hawkins M, McGlynn P. 2014 Recombination and Replication. *Cold Spring Harb.*
920 *Perspect. Biol.* **6**, 1–14. (doi:10.1101/cshperspect.a016550)

921 12. Amundsen SK, Smith GR. 2019 The RecB helicase-nuclease tether mediates Chi hotspot control
922 of RecBCD enzyme. *Nucleic Acids Res.* **47**, 197–209. (doi:10.1093/nar/gky1132)

923 13. Taylor AF, Smith GR. 1995 Monomeric RecBCD enzyme binds and unwinds DNA. *J. Biol. Chem.*
924 **270**, 24451–8. (doi:10.1074/jbc.270.41.24451)

925 14. Howard-Flanders P, West SC, Stasiak A. 1984 Role of RecA protein spiral filaments in genetic
926 recombination. *Nature* **309**, 215–220. (doi:10.1038/309215a0)

927 15. Kowalczykowski SC, Eggleston AK. 1994 Homologous pairing and DNA strand-exchange
928 proteins. *Annu. Rev. Biochem.* **63**, 991–1043. (doi:10.1146/annurev.bi.63.070194.005015)

929 16. Mazin A V., Kowalczykowski SC. 1999 A novel property of the RecA nucleoprotein filament:
930 Activation of double-stranded DNA for strand exchange in trans. *Genes Dev.* **13**, 2005–16.
931 (doi:10.1101/gad.13.15.2005)

932 17. Wiktor J, Gynnå AH, Leroy P, Larsson J, Coceano G, Testa I, Elf J. 2021 RecA finds homologous
933 DNA by reduced dimensionality search. *Nature* **597**. (doi:10.1038/s41586-021-03877-6)

934 18. Windgassen TA, Wessel SR, Bhattacharyya B, Keck JL. 2018 Mechanisms of bacterial DNA
935 replication restart. *Nucleic Acids Res.* **46**, 504–519. (doi:10.1093/nar/gkx1203)

936 19. Dimude JU, Midgley-Smith SL, Rudolph CJ. 2018 Replication-transcription conflicts trigger
937 extensive DNA degradation in Escherichia coli cells lacking RecBCD. *DNA Repair (Amst).* **70**, 37–
938 48. (doi:10.1016/j.dnarep.2018.08.002)

939 20. Repar J, BriŠki N, Buljubašić M, Zahradka K, Zahradka D. 2013 Exonuclease VII is involved in
940 'reckless' DNA degradation in UV-irradiated Escherichia coli. *Mutat. Res. - Genet. Toxicol.*
941 *Environ. Mutagen.* **750**, 96–104. (doi:10.1016/j.mrgentox.2012.10.005)
942 21. Ghodke H *et al.* 2019 Spatial and temporal organization of RecA in the Escherichia coli DNA-
943 damage response. *Elife* **8**, e42761. (doi:10.7554/eLife.42761)
944 22. Khil PP, Camerini-Otero RD. 2002 Over 1000 genes are involved in the DNA damage response
945 of Escherichia coli. *Mol. Microbiol.* **44**, 89–105. (doi:10.1046/j.1365-2958.2002.02878.x)
946 23. Adikesavan AK, Katsonis P, Marciano DC, Lua R, Herman C, Lichtarge O. 2011 Separation of
947 Recombination and SOS Response in Escherichia coli RecA Suggests LexA Interaction Sites. *Plos*
948 *Genet.* **7**, e1002244. (doi:10.1371/journal.pgen.1002244)
949 24. Gruenig MC, Renzette N, Long E, Chitteni-Pattu S, Inman RB, Cox MM, Sandler SJ. 2008 RecA-
950 mediated SOS induction requires an extended filament conformation but no ATP hydrolysis.
951 *Mol. Microbiol.* **69**, 1165–1179. (doi:10.1111/j.1365-2958.2008.06341.x)
952 25. Butala M *et al.* 2011 Interconversion between bound and free conformations of LexA
953 orchestrates the bacterial SOS response. *Nucleic Acids Res.* **39**. (doi:10.1093/nar/gkr265)
954 26. Podlesek Z, Žgur Bertok D. 2020 The DNA Damage Inducible SOS Response Is a Key Player in
955 the Generation of Bacterial Persister Cells and Population Wide Tolerance. *Front. Microbiol.* **11**.
956 (doi:10.3389/fmicb.2020.01785)
957 27. Memar MY, Yekani M, Celenza G, Poortahmasebi V, Naghili B, Bellio P, Baghi HB. 2020 The
958 central role of the SOS DNA repair system in antibiotics resistance: A new target for a new
959 infectious treatment strategy. *Life Sci.* **262**. (doi:10.1016/j.lfs.2020.118562)
960 28. Rehrauer WM, Kowalczykowski SC. 1996 The DNA binding site(s) of the Escherichia coli RecA
961 protein. *J. Biol. Chem.* **271**, 11996–12002. (doi:10.1074/jbc.271.20.11996)
962 29. Handa N, Amitani I, Gumlaw N, Sandler SJ, Kowalczykowski SC. 2009 Single Molecule Analysis
963 of a Red Fluorescent RecA Protein Reveals a Defect in Nucleoprotein Filament Nucleation That
964 Relates to Its Reduced Biological Functions. *J. Biol. Chem.* **284**, 18664–18673.
965 (doi:10.1074/jbc.M109.004895)
966 30. Ivancic-Bace I, Vlasic I, Salaj-Smic E, Brcic-Kostic K, Ivančić-Baće I, Vlašić I, Salaj-Šmic E, Brčić-
967 Kostić K. 2006 Genetic evidence for the requirement of RecA loading activity in SOS induction
968 after UV irradiation in Escherichia coli. *J. Bacteriol.* **188**, 5024–5032. (doi:10.1128/JB.00130-06)
969 31. Anderson DG, Kowalczykowski SC. 1998 Reconstitution of an SOS response pathway:
970 Derepression of transcription in response to DNA breaks. *Cell* **95**, 975–9. (doi:10.1016/S0092-
971 8674(00)81721-3)
972 32. Adams D *et al.* 2007 Selection of Reca+ Recombinant Cosmids - an Easy Method for Making
973 Reca Strains Temporarily Rec+, Permitting P1-Mediated Transduction in a Reca- Background
974 and Transduction of a Reca- Mutation. *J. Biol. Chem.* **8**, 88–94. (doi:10.1074/jbc.M110.194407)
975 33. Plank M, Wadhams GH, Leake MC. 2009 Millisecond timescale slimfield imaging and
976 automated quantification of single fluorescent protein molecules for use in probing complex
977 biological processes. *Integr. Biol.* **1**, 602–612. (doi:10.1039/b907837a)
978 34. Lesterlin C, Ball G, Schermelleh L, Sherratt DJ. 2014 RecA bundles mediate homology pairing
979 between distant sisters during DNA break repair. *Nature* **506**, 249–253.
980 (doi:10.1038/nature12868)
981 35. Lepore A, Taylor H, Landgraf D, Okumus B, Jaramillo-Riveri S, Mclaren L, Bakshi S, Paulsson J, El
982 Karoui & M. 2019 Quantification of very low-abundant proteins in bacteria using the HaloTag
983 and epi-fluorescence microscopy. *Sci. Rep.* **9**, 7902. (doi:10.1038/s41598-019-44278-0)
984 36. Lenn T, Leake MC. 2012 Experimental approaches for addressing fundamental biological
985 questions in living, functioning cells with single molecule precision. *Open Biol.* **2**, 120090.
986 (doi:10.1098/rsob.120090)

987 37. Wollman AJM, Leake MC. 2015 Millisecond single-molecule localization microscopy combined
988 with convolution analysis and automated image segmentation to determine protein
989 concentrations in complexly structured, functional cells, one cell at a time. *Faraday Discuss.*
990 **184**, 401–424. (doi:10.1039/C5FD00077G)

991 38. Wollman AJM *et al.* 2022 Critical roles for EGFR and EGFR-HER2 clusters in EGF binding of
992 SW620 human carcinoma cells. *J. R. Soc. Interface* **19**, 20220088. (doi:10.1098/rsif.2022.0088)

993 39. Manley S, Gillette JM, Patterson GH, Shroff H, Hess HF, Betzig E, Lippincott-Schwartz J. 2008
994 High-density mapping of single-molecule trajectories with photoactivated localization
995 microscopy. *Nat. Methods* **5**, 155–157. (doi:10.1038/nmeth.1176)

996 40. Stracy M, Kapanidis AN. 2017 Single-molecule and super-resolution imaging of transcription in
997 living bacteria. *Methods* **120**, 103–114. (doi:10.1016/j.ymeth.2017.04.001)

998 41. Jin X *et al.* 2021 Membraneless organelles formed by liquid-liquid phase separation increase
999 bacterial fitness. *Sci. Adv.* **7**, eabh2929. (doi:10.1126/sciadv.abh2929)

1000 42. Story RM, Weber IT, Steitz TA. 1992 The structure of the E. coli recA protein monomer and
1001 polymer. *Nature* **355**, 318–325. (doi:10.1038/355318a0)

1002 43. Long JE, Renzette N, Centore RC, Sandler SJ. 2008 Differential Requirements of Two recA
1003 Mutants for Constitutive SOS Expression in Escherichia coli K-12. *PLoS One* **3**, e4100.
1004 (doi:10.1371/journal.pone.0004100)

1005 44. Lee AM, Singleton SF. 2006 Intersubunit electrostatic complementarity in the RecA
1006 nucleoprotein filament regulates nucleotide substrate specificity and conformational
1007 activation. *Biochemistry* **45**, 4514–4529. (doi:10.1021/bi052296x)

1008 45. Wang Y, Bortner C, Griffith J. 1993 Reca Binding to Bulge-Containing and Mismatch-Containing
1009 Dnas - Certain Single-Base Mismatches Provide Strong Signals for Reca Binding Equal to
1010 Multiple Base Bulges. *J. Biol. Chem.* **268**, 17571–17577.

1011 46. Schlacher K, Cox MM, Woodgate R, Goodman MF. 2006 RecA acts in trans to allow replication
1012 of damaged DNA by DNA polymerase V. *Nature* **442**, 883–887. (doi:10.1038/nature05042)

1013 47. McGrew DA, Knight KL. 2003 Molecular design and functional organization of the RecA protein.
1014 *Crit. Rev. Biochem. Mol. Biol.* **38**, 385–432. (doi:10.1080/10409230390242489)

1015 48. Dapa T, Fleurier S, Bredeche MF, Matic I. 2017 The SOS and RpoS regulons contribute to
1016 bacterial cell robustness to genotoxic stress by synergistically regulating DNA polymerase pol
1017 II. *Genetics* **206**. (doi:10.1534/genetics.116.199471)

1018 49. Miller H, Zhou Z, Wollman AJM, Leake MC. 2015 Superresolution imaging of single DNA
1019 molecules using stochastic photoblinking of minor groove and intercalating dyes. *Methods* **88**,
1020 81–88. (doi:10.1016/j.ymeth.2015.01.010)

1021 50. Leake MC, Chandler JH, Wadhams GH, Bai F, Berry RM, Armitage JP. 2006 Stoichiometry and
1022 turnover in single, functioning membrane protein complexes. *Nature* **443**, 355–358.
1023 (doi:10.1038/nature05135)

1024 51. Peterson KR, Wertman KF, Mount DW, Marinus MG. 1985 Viability of Escherichia coli K-12 DNA
1025 adenine methylase (dam) mutants requires increased expression of specific genes in the SOS
1026 regulon. *MGG Mol. Gen. Genet.* **201**, 14–19. (doi:10.1007/BF00397979)

1027 52. Renzette N, Gumlaw N, Nordman JT, Krieger M, Yeh SP, Long E, Centore R, Boonsombat R,
1028 Sandler SJ. 2005 Localization of RecA in Escherichia coli K-12 using RecA-GFP. *Mol. Microbiol.*
1029 **57**. (doi:10.1111/j.1365-2958.2005.04755.x)

1030 53. Leake MC. 2014 Analytical tools for single-molecule fluorescence imaging in cellulo. *Phys.*
1031 *Chem. Chem. Phys.* **16**, 12635–47. (doi:10.1039/c4cp00219a)

1032 54. Shi W-X, Larson RG. 2007 RecA-ssDNA filaments supercoil in the presence of single-stranded
1033 DNA-binding protein. *Biochem. Biophys. Res. Commun.* **357**, 755–760.
1034 (doi:10.1016/j.bbrc.2007.04.014)

1035 55. Darmon E, Eykelenboom JK, Lopez-Vernaza MA, White MA, Leach DRF. 2014 Repair on the Go:
1036 E. Coli maintains a high proliferation rate while repairing a chronic DNA double-strand break.
1037 *PLoS One* **9**. (doi:10.1371/journal.pone.0110784)

1038 56. Ellouze C, Takahashi M, Wittung P, Mortensen K, Schnarr M, Norden B, Nordén B, Norden B.
1039 1995 Evidence for Elongation of the Helical Pitch of the Reca Filament Upon Atp and Adp
1040 Binding Using Small-Angle Neutron-Scattering. *Eur. J. Biochem.* **233**, 579–583.
1041 (doi:10.1111/j.1432-1033.1995.579_2.x)

1042 57. Yu X, Jacobs SA, West SC, Ogawa T, Egelman EH. 2001 Domain structure and dynamics in the
1043 helical filaments formed by RecA and Rad51 on DNA. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8419–
1044 8424. (doi:10.1073/pnas.111005398)

1045 58. Zhao B *et al.* 2017 ATPase activity tightly regulates RecA nucleofilaments to promote
1046 homologous recombination. *Cell Discov.* **3**, 16053. (doi:10.1038/celldisc.2016.53)

1047 59. Chimthanawala A, Parmar J, Kumar S, Iyer KS, Rao M, Badrinarayanan A. 2021 SMC protein
1048 RecN drives translocation and remodelling of RecA filament for homology search. *BiorXiv* ,
1049 2021.08.16.456443. (doi:10.1101/2021.08.16.456443)

1050 60. Rajendram M *et al.* 2015 Anionic Phospholipids Stabilize RecA Filament Bundles in Escherichia
1051 coli. *Mol. Cell* **60**, 374–384. (doi:10.1016/j.molcel.2015.09.009)

1052 61. Eldin S, Forget AL, Lindenmuth DM, Logan KM, Knight KL. 2000 Mutations in the N-terminal
1053 region of RecA that disrupt the stability of free protein oligomers but not RecA-DNA complexes.
1054 *J. Mol. Biol.* **299**, 91–101. (doi:10.1006/jmbi.2000.3721)

1055 62. Roy R, Kozlov AG, Lohman TM, Ha T. 2009 SSB protein diffusion on single-stranded DNA
1056 stimulates RecA filament formation. *Nature* **461**, 1092–7. (doi:10.1038/nature08442)

1057 63. Norden B, Takahashi. M, Takahashi M, Nordén B. 1994 Structure of RecA-DNA complex and
1058 mechanism of DNA strand exchange reaction in homologous recombination. *Adv. Biophys.* **30**,
1059 1–35. (doi:10.1016/0065-227X(94)90009-4)

1060 64. Galletto R, Amitani I, Baskin RJ, Kowalczykowski SC. 2006 Direct observation of individual RecA
1061 filaments assembling on single DNA molecules. *Nature* **443**, 875–8. (doi:10.1038/nature05197)

1062 65. Hernández-Tamayo R, Steube N, Heimerl T, Hochberg G, Graumann PL. 2022 ATPase activity of
1063 B. subtilis RecA affects the dynamic formation of RecA filaments at DNA double strand breaks.
1064 *BiorXiv* , 2022.02.15.480544. (doi:10.1101/2022.02.15.480544)

1065 66. Danilowicz C, Vietorisz E, Godoy-Carter V, Prévost C, Prentiss M. 2021 Influences of ssDNA-
1066 RecA Filament Length on the Fidelity of Homologous Recombination. *J. Mol. Biol.* **433**, 167143.
1067 (doi:10.1016/J.JMB.2021.167143)

1068 67. Yu X, Egelman EH. 1997 The RecA hexamer is a structural homologue of ring helicases. *Nat.*
1069 *Struct. Biol.* **4**, 101–104. (doi:10.1038/nsb0297-101)

1070 68. Yu X, VanLoock MS, Yang S, Reese JT, Egelman EH. 2004 What is the structure of the RecA-DNA
1071 filament? *Curr. Protein Pept. Sci.* **5**, 73–79. (doi:10.2174/1389203043486883)

1072 69. Chen Z, Yang H, Pavletich NP. 2008 Mechanism of homologous recombination from the RecA-
1073 ssDNA/dsDNA structures. *Nature* **453**, 489–494. (doi:10.1038/nature06971)

1074 70. Li GW, Burkhardt D, Gross C, Weissman JS. 2014 Quantifying absolute protein synthesis rates
1075 reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–35.
1076 (doi:10.1016/j.cell.2014.02.033)

1077 71. Stohl EA, Brockman JP, Burkle KL, Morimatsu K, Kowalczykowski SC, Seifert HS. 2003
1078 Escherichia coli RecX inhibits RecA recombinase and coprotease activities in vitro and in vivo.
1079 *J. Biol. Chem.* **278**, 2278–85. (doi:10.1074/jbc.M210496200)

1080 72. Giacomoni PU. 1982 Induction by mitomycin C of recA protein synthesis in bacteria and
1081 spheroplasts. *J. Biol. Chem.* **257**, 14932–14936. (doi:10.1016/s0021-9258(18)33373-8)

1082 73. Soufi B, Krug K, Harst A, Macek B. 2015 Characterization of the E. coli proteome and its
1083 modifications during growth and ethanol stress. *Front. Microbiol.* **6**, 103.
1084 (doi:10.3389/fmicb.2015.00103)

1085 74. Elowitz MB, Surette MG, Wolf PE, Stock JB, Leibler S. 1999 Protein mobility in the cytoplasm of
1086 Escherichia coli. *J. Bacteriol.* **181**, 197–203. (doi:10.1128/jb.181.1.197-203.1999)

1087 75. Nenninger A, Mastroianni G, Mullineaux CW. 2010 Size dependence of protein diffusion in the
1088 cytoplasm of Escherichia coli. *J. Bacteriol.* **192**, 4535–40. (doi:10.1128/JB.00284-10)

1089 76. Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB. 2004 Crystal
1090 structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* **432**, 187–
1091 93.

1092 77. Burby PE, Simmons LA. 2019 A bacterial DNA repair pathway specific to a natural antibiotic.
1093 *Mol. Microbiol.* **111**, 338–353. (doi:10.1111/mmi.14158)

1094 78. Peng X, Ghosh AK, Van Houten B, Greenberg MM. 2010 Nucleotide excision repair of a DNA
1095 interstrand cross-link produces single- and double-strand breaks. *Biochemistry* **49**.
1096 (doi:10.1021/bi901603h)

1097 79. Wood RD. 2010 Mammalian nucleotide excision repair proteins and interstrand crosslink
1098 repair. *Environ. Mol. Mutagen.* **51**. (doi:10.1002/em.20569)

1099 80. Syeda AH, Wollman AJM, Hargreaves AL, Howard JAL, Brüning J-G, McGlynn P, Leake MC. 2019
1100 Single-molecule live cell imaging of Rep reveals the dynamic interplay between an accessory
1101 replicative helicase and the replisome. *Nucleic Acids Res.* **47**, 6287–6298.
1102 (doi:10.1093/nar/gkz298)

1103 81. Verhoeven EEAA, Kesteren M Van, Moolenaar GF, Visse R, Goosen N, Van Kesteren M,
1104 Moolenaar GF, Visse R, Goosen N. 2000 Catalytic sites for 3' and 5' incision of Escherichia coli
1105 nucleotide excision repair are both located in UvrC. *J. Biol. Chem.* **275**, 5120–3.
1106 (doi:10.1074/jbc.275.7.5120)

1107 82. Massoni SC, Leeson MC, Long JE, Gemme K, Mui A, Sandler SJ. 2012 Factors Limiting SOS
1108 Expression in Log-Phase Cells of Escherichia coli. *J. Bacteriol.* **194**, 5325–5333.
1109 (doi:10.1128/JB.00674-12)

1110 83. Reuter A, Hilpert C, Dedieu-Berne A, Lematre S, Gueguen E, Launay G, Bigot S, Lesterlin C. 2021
1111 Targeted-antibacterial-plasmids (TAPs) combining conjugation and CRISPR/Cas systems
1112 achieve strain-specific antibacterial activity. *Nucleic Acids Res.* **49**, 3584–3598.
1113 (doi:10.1093/nar/gkab126)

1114 84. Renzette N, Sandler SJ. 2008 Requirements for ATP binding and hydrolysis in RecA function in
1115 Escherichia coli. *Mol. Microbiol.* **67**. (doi:10.1111/j.1365-2958.2008.06130.x)

1116 85. Llorente-Garcia I *et al.* 2014 Single-molecule in vivo imaging of bacterial respiratory complexes
1117 indicates delocalized oxidative phosphorylation. *Biochim. Biophys. Acta - Bioenerg.* **1837**, 811–
1118 24. (doi:10.1016/j.bbabio.2014.01.020)

1119 86. Dresser L, Hunter P, Yendybayeva F, Hargreaves AL, Howard JAL, Evans GJO, Leake MC, Quinn
1120 SD. 2021 Amyloid-β oligomerization monitored by single-molecule stepwise photobleaching.
1121 *Methods* **193**, 80–95. (doi:10.1016/j.ymeth.2020.06.007)

1122 87. Shepherd JW, Payne-Dwyer AL, Lee J-E, Syeda A, Leake MC. 2021 Combining single-molecule
1123 super-resolved localization microscopy with fluorescence polarization imaging to study cellular
1124 processes. *JPhys Photonics* **3**. (doi:10.1088/2515-7647/ac015d)

1125 88. Robson A, Burrage K, Leake MC. 2013 Inferring diffusion in single live cells at the single-
1126 molecule level. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120029. (doi:10.1098/rstb.2012.0029)

1127 89. Leake MC, Wilson D, Bullard B, Simmons RM, Bubb MR. 2003 The elasticity of single kettin
1128 molecules using a two-bead laser-tweezers assay. In *FEBS Letters*, pp. 55–60.
1129 (doi:10.1016/S0014-5793(02)03857-7)

90. Ovesný M, Křížek P, Borkovec J, Švindrych Z, Hagen GM. 2014 ThunderSTORM: A comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**, 2389–2390. (doi:10.1093/bioinformatics/btu202)

91. Payne-Dwyer AL, Leake MC. 2022 Single-molecular quantification of flowering control proteins within nuclear condensates in live whole Arabidopsis root. In *Chromosome Architecture: Methods in Molecular Biology* (ed MC Leake), p. 2476. New York, NY: Springer US, Humana. (doi:10.1007/978-1-0716-2221-6)

92. Pédelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. 2006 Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88. (doi:10.1038/nbt1172)