Supporting Online Material for

Stoichiometry and architecture of active DNA replication machinery within *Escherichia coli*

Rodrigo Reyes-Lamothe, David J. Sherratt and Mark C. Leake*

*To whom correspondence should be addressed. E-mail: m.leake1@physics.ox.ac.uk

This PDF file includes:

Supporting Online Material

MATERIALS AND METHODS

Construction and characterization of fluorescent fusion strains.

Genes were replaced by their fluorescent versions as described previously (*S1*). In brief, genes coding for YPet and Kanamycin resistance cassette were amplified from the plasmid pROD10 (carrying 11aa linker preceding *ypet* and followed by *frt* flanked *kan*), using primers with 50 nt overhangs with homologous to the place to be inserted. DNA fragments were gel purified and ~1 μg was used for electroporation of AB1157 cells overexpressing λ-Red proteins from pKD46 (*S2*). Correct insertion of the fragment into the chromosome was assayed by PCR.

In the case of DnaN, where YPet was fused to its N-terminus, a similar strategy was used but using the plasmid pROD44 carrying the same genes in an inverted order (*frt* flanked *kan* and *ypet* followed by 11aa linker). After insertion, *kan* was removed by expressing Flp recombinase to prevent changes in the expression of *dnaN* caused by the promoter of *kan*, leaving an *frt* scar immediately upstream of the initiation codon of *YPet*.

DnaX (γ) was produced by site directed mutagenesis of a plasmid carrying the *dnaX-ypet* fusion followed by a Km^R marker. The original frameshifting sequence (gcaaaaaagagtga) was mutated to a sequence that prevents the frameshift (gcGaaGaagagtga) (*S3*); this allele was used to replace the original copy of *dnaX*. γ-mYPet was expressed from a xylose promoter at the *argE* locus in the *E. coli* chromosome. The fusion gene was first cloned in a plasmid and carried the sequence encoding γ followed by an 8 aa flexible linker and *mYPet* and a linked Cm^R gene.

χ and ψ fused to a degron tag were constructed using a plasmid carrying a sequence coding for a 6 aa linker, myc tag, 2 aa linker and the DAS+4 degron tag (S A G S A A E Q K L I S E E D L S S A A N D E N Y S E N Y A D A S) (*S4*), followed by a Km^R gene. Primers carrying 50 nt sequences identical to the 3' end and downstream region of *holC* and *holD* were used to amplify the fragment from the plasmid and subsequent introduction to the cell by lambda red. Gene fusions were subsequently introduced into a strain carrying a *ΔsspB* and an ectopic copy of *sspB* under the regulation of arabinose promoter (gift from R. Sauer's lab).

To ensure that the cells carrying the gene fusions behave as wild type, their growth and replication parameters were determined. The generation time for wt cells grown at 37°C in minimal medium with glycerol as a carbon source was ~110 minutes (*S1*). The generation time in both minimal medium and rich medium for the fusion strains was similar with less than 10% of change, suggesting that the gene insertions did not affect their growth (table S1). The distribution of DNA content in a population was then compared by FACS flow cytometry (fig. S2B and S10). The strains did show patterns similar to wt in asynchronous and

synchronized cultures (by run-out experiments). Further confirmation on the benign effect that the gene fusions had on cells was found when chromosomal loci were marked by the fluorescent repressor-operator system. Fusion strains showed distributions the *oriC* proximal loci *ori1* similar to those reported earlier for wt (*S5*).

YPet-His Purification

AB1157 cells carrying a pBAD33 derivative expressing YPet-His (*S6*), were grown at 30^{o}C to a A_{600} ~0.5 and then induced with 0.2% arabinose for 4 h. 500 ml of induced culture were centrifuged at 4 krpm/20 min/4^{o}C and resuspended in 10 ml of Storing Buffer (10 mM Tris pH 7.5, 250 mM NaCl, 10% glycerol) and stored at -20^{o}C. Cells suspensions were thawed, imidazole added to get a concentration of 5mM along with a cocktail of protein inhibitors (Roche). Cells were lysed by sonication and debris removed by centrifugation at 20 krpm/30 min/4^{o}C. 3 ml of TALON® resin (Clonetech) in a column was equilibrated with 10 ml of binding buffer (5 mM Tris pH 7.5, 250 mM NaCl, 10 mM imidazole, 10% glycerol). The lysate was then loaded in the column and later washed with 25 ml of binding buffer. The column was further washed with 25 ml of washing buffer (25 mM Tris pH 7.5, 250 mM NaCl, 10 mM imidazole, 10% glycerol) and eluted with elution buffer (25 mM Tris pH 7.5, 100 mM NaCl, 250 mM imidazole, 10% glycerol) collecting 0.5 ml fractions. The aliquots containing the highest concentration of protein were mixed and dialysed for 5 h at room temperature, and then overnight at 4^{o}C in 10mM Tris (pH7.5), 10 mM EDTA. Protein was then injected into a Q-column, previously equilibrated with buffer A (25 mM Tris pH 7.5, 25 mM NaCl, 1 mM DTT), and a gradient between 25 mM and 250 mM NaCl was done using different proportions of buffer A and buffer B (25 mM Tris pH 7.5, 500 mM NaCl) over 20 min (collecting 0.5 ml/ 30 s). Salt was removed as before by dialysis and protein was stored at -20^{o}C. Protein concentrations were determined by a BCA Protein Assay (Pierce) and assessment of the level of purification was by SDS-PAGE.

Determination of cellular Ssb-YPet concentration by western blots

Cells were grown in LB or M9-glycerol to an A_{600} of ~0.5 and ~0.2 respectively, centrifuged at 4 krpm/10 min/4^{o}C, resuspended in 200 µl of Cracking Buffer (50 mM Tris-HCl, pH 6.8, 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, and 10% glycerol) and stored at -80^{o}C until used. To find the relation between A_{600} and number of cells, dilutions of the cultures with known A_{600} were plated on LB agar and the number of c.f.u. determined. Samples were briefly sonicated to reduce viscosity, mixed with SDS loading buffer (250 mM TrisHCl pH 6.8, 10% SDS, 30% glycerol, 0.02% bromophenol blue, 0.5 mM DTT) and PBS, to reach a volume of 10 µl. Aliquots of YPet-His were diluted in PBS and different volumes

were transferred to tubes and mixed with SDS loading buffer. Samples were boiled for 5 min and loaded into a 10% acrylamide SDS-PAGE gel. Proteins were blotted on a nitrocellulose membrane (Amersham HyBond ECL) using Transfer Buffer (190 mM Tris, 88 mM glycine, 20% methanol) and transferred at 400 mA/4^{o}C/3 h. Membrane was blocked with 5% BSA in PBS overnight at 4^{o}C. Labeling was done by using a 1:1000 dilution of rabbit anti-GFP polyclonal antibody (Clontech) and a 1:2000 dilution of goat anti-rabbit IgG-Cy5 conjugated secondary antibody (ECL Plex, Amersham). Detection and quantification of bands was done by using 633/675 filters of FUJI Phosphoimager and programs Image Reader V1.8E and Image Gauge V3.3, respectively.

Standard curves using different amounts of pure YPet showed a lower detection limit of 2 ng of protein and a linear relation with concentrations between 5-50 ng. Extracts obtained from a known number of cells carrying Ssb-YPet grown in M9-Gly or LB were loaded into gels containing a standard curve of YPet. A band of ~45 kDa corresponding to the size expected for Ssb fusion (close to the calculated 46.5 kDa) was observed in blots. Quantification of bands, conversion to number of molecules by comparison with the standard curve and normalization by the number of cells loaded, provided an estimate of number of molecules per cell. The average and standard deviations of three independent experiments using cells grown in both conditions were of $(9.8 \pm 2.6) \times 10^3$ and $(2.9 \pm 1.2) \times 10^3$ molecules per cell for LB and M9-Gly respectively. These numbers correlate with previous estimates ranging from 1000-2000 tetramers per cell (*S7*), and show that the number of Ssb molecules when grown in LB was around two times higher than that of cells in minimal media, corroborating previous observations on the effect of growth conditions (*S8*).

Preparation of cells for microscopy

Cells were grown in M9-glycerol overnight at 37^{o}C with shaking, and sub-cultured in the morning into the same media for 3 h at 30 ^{o}C. Poly-L-lysine (0.1% w/v) was injected into a custom made flow-cell of volume ~5 µl (*9*), inverted and incubated for 2 min at room temperature and then 100 µl minimal media wicked through to wash out unbound poly-L-lysine. 20 µl cells directly from the growth medium were then wicked through and the flow-cell inverted for 20 min to allow cells to bind to the coverslip surface. 100 µl minimal media was then wicked through to wash out unbound cells.

Microscopy

We used a bespoke inverted fluorescence microscope (*S10, S11*) with a 100x Plan Fluor 1.45 NA oil immersion objective (Nikon) and an *xyz* nanometer-precise positioning stage (E-503.00; Physik Instrumente). Brightfield illumination used a fiber-coupled tungsten-halogen

source; laser excitation used a TEM00 plane-polarized continuous-wave 532 nm DPSS laser (Laser2000 UK), filtered (laser-line 532 nm), expanded three times. The primary beam is split in to two independently attenuated paths by a polarizing beam-splitting cube. This generates a separate widefield excitation path which could be used for epifluorescence, and a slimfield illumination mode. For the widefield path the beam is focused onto the back-focal-plane of the objective lens via a dichroic mirror (long-pass 530 nm). The field width was ~30 μm, typical intensity ~280 W cm^{-2}. The separately-shuttered slimfield excitation path (*S12*) directs a collimated laser beam to under-fill the back-aperture of the objective lens (*S13*). The effect of this is to generate a focused Gaussian profile at the level of the sample which is expanded laterally to be larger than a diffraction-limited spot, similar in appearance to the excitation volume generated from previous methods employing defocusing optics to a commercial microscope system (*S14*). To avoid known artifacts in estimating the width from purely analytical formulations (*S13*) we sampled the intensity profile directly using a 20 nm diameter yellow fluorescent bead (Molecular Probes) raster-scanned across the field of view and measuring the total bead intensity at intervals of 10 pixels. We then raster-scanned the same bead in the reverse direction at the same scan speed to correct for bead photobleaching. A 2D radial Gaussian function was then fitted to the corrected intensities. Here we selected the standard deviation width σ_{xy} to be ~3 μm (equivalent to a FWHM of ~7 μm) with intensity ~6.5 kW cm^{-2} so as to encompass a whole *E. coli* bacterial cell. This imaging, which allows quantitative detection of single fluorescent molecules at 3 ms capture rates, permits visualization of fast diffusing proteins, which would appear blurred and hidden by camera noise using slower video-rate microscopy.

Fluorescence emissions generated at the sample plane of the microscope are passed through the dichroic mirror, filtered using the relevant band-pass emission filter in the selected filter set and imaged at ~50 nm per pixel in frame-transfer mode (either 3 ms per frame for slimfield illumination or 40 ms per frame for epifluorescence) by a 128x128-pixel, cooled, back-thinned electron-multiplying charge-coupled-device camera (iXon DV860-BI, Andor Technology, UK). For slimfield, typically a maximum of 100 continuously illuminated frames were taken in each run. For FRAP investigations of the Ssb-YPet strain, the focused beam path was positioned over a fluorescent spot and the shutter opened for 200 ms. Subsequent fluorescence recovery of the bleached area was monitored by acquiring single frame images over a time-series for either a short time scale ([1, 2, 4, 8, 12, 16, 20, 30, 60 s post-bleach) or longer time scale ([1, 2, 4, 8, 16, 32, 64, 128, 256 s post-bleach) regime (fig. S1).

For investigating targeted proteolysis of degron-tagged χψ, or deletion of χ (fig. S3), and for investigating co-localization of γ-YPet and ε-mCherry (fig. S4) cells were visualized

in epifluorescence mode using a 100× objective on a Nikon Eclipse TE2000-U microscope, equipped with a Photometrics Cool-SNAP HQ CCD camera. When using degron fusions, cells were induced for 2 h using 0.5% arabinose to ensure complete degradation, while strains carrying γ-mYPet were induced with 0.1% xylose for 40 min to produce the fluorescent fusion.

Detecting and quantifying fluorescent spots in images

Images were analyzed using a toolkit custom-written in LabVIEW 8.5 (*S9-S11, S15- S17*). Custom pattern recognition software was first applied to locate the perimeter of the cell bodies from the brightfield images. This was used to create an image mask for subsequent fluorescence analysis. This insured that detected spots in fluorescence intensity could be mapped back to specific cells and also permitted subsequent coordinate transformation in reference to the orientation of the cell relative to the camera axes. Then, a frame-average image was complied from 30 consecutive fluorescence image frames (90 ms total integration time). "Hotspots" in intensity of the frame-averaged image were then detected automatically using custom code to locate circular intensity distributions of width 100-350 nm (Danielsson distance map algorithm), comparable to the expected point spread function width. A circular region of interest (ROI) was created around each putative spot of radius 8 pixels. Having defined the position of the ROI on the frame-averaged image, analysis was then performed on the raw data series, one image at a time, letting the outer ROI position remain fixed but allowing an inner circular ROI within this to vary its position to best fit the intensity centroid. The intensity in each ROI was modeled as a radial Gaussian plus a uniform baseline of background noise. The intensity contribution due to the YPet (the "spot intensity") was calculated as follows:

1. We used an inner circular mask for the contribution of the spot of diameter 5 pixels to the ROI centered on the intensity centroid.
2. We convolved intensities within the mask by a two-dimensional radial Gaussian function of fixed width 3 pixels and generated a new estimate for the centroid.
3. We iterated steps 1-2 until convergence (generally less than 10 loops). For 3 ms per image frame in slimfield mode using surface-immobilized YPet as the sample this resulted in an ultimate centroid r.m.s. precision of ~50 nm.
4. We defined the background intensity as the mean intensity within the fixed ROI external to the inner circle mask. The contribution to the background count due to diffusive YPet in the cytoplasm (i.e. that not bound in the replisome complex) was calculated for each track as the initial background intensity per pixel after subtracting the autofluorescence contribution per

pixel measured in the non-YPet parental strain and the instrumental background (total of both combined 1030 ± 15 counts).

5. We defined a preliminary spot intensity $I(t)$ at a time t as the sum of all intensities within the inner circular mask after subtraction of the background from each individual pixel value.

6. A small correction to the spot intensity was applied to account for any non-uniformity in the slimfield excitation volume across the field of view due to its Gaussian shape in the sample plane: we multiplied each spot intensity by the factor $\exp(r^2/2\sigma_{xy}^2)$ where r is the distance from the spot centroid to the center of the slimfield excitation volume in the sample plane and σ_{xy} is the standard deviation width of the excitation field (3 μm). Even for the longest bacterial cells in which the spots were most distant from the center of the slimfield volume the correction factor never exceeded 1.1 (i.e. less than 10% difference between corrected and pre-corrected values).

7. A Gaussian fit was then performed on the spot intensity component optimizing both amplitude and width.

This resulted in a fully automatic method for characterizing fluorescent spots on each separate image frame within the fixed hotspot ROI which could quantify the total pixel intensity minus the background detector noise, the size of the spots and the position of the spot to within typically ~50 nm precision.

Estimating stoichiometry of spots.

Spot intensity data were then collated for each cell strain and the distribution of estimated pixel intensity binned on a histogram. This resulted in multiple distinct peaks separated by a roughly constant spacing. We also performed experiments on purified YPet by immobilizing the protein to the coverslip surface using conjugation via the anti-YPet antibody following an earlier protocol used for GFP (*S18*). We found that the center of the lowest order peak from the live cell imaging was always within ~10% of the unitary photobleaching peak for surface immobilized pure YPet performed *in vitro* (Fig. 1). This was also found to be in reasonable agreement with estimating the unitary peak with a Fourier spectral method which constructed a power spectrum from the periodicity in the intensity trace *(S9)*. This consisted of using an edge-preserving Chung-Kennedy algorithm of two adjacent windows run across the data whose output was the mean from the window possessing the smallest variance. We then calculated all pairwise differences in the filtered intensity *I(t)*:

$$\Delta I_{ij} = I(t_i) - I(t_j)$$

for all data point pairs for which the time $t_i > t_j$. The distribution of these differences was calculated using typically 2,000 bins for each trace and normalized by the number of pairwise differences. We calculated the single-sided power spectrum from each pairwise displacement histogram. Peak detection was automated using commercial code (LabView 8.5, National Instruments, Austin, TX) with the unitary photobleach step peak taken as that detected at the highest spatial frequency in each trace, taking a detection threshold as three standard deviations above the high spatial frequency noise floor (*S9*) equivalent to a probability confidence level of p<0.3%. The corresponding value of spatial frequency was inverted to give the characteristic unitary photobleach step size in terms of intensity counts.

This approach of filtration combined with Fourier spectral analysis offered an improvement over single-stage methods of detecting individual step events and compiling in effect an average number from all of these, since here the algorithm instead detects the underlying periodicity in the levels of intensity. Thus it is relatively insensitive to a false detection output from the filter. Where the combined filtering and Fourier spectral approach has weakness is that one loses the temporal information of when exactly any given step event occurs; for much work on for example molecular motors this would be a significant problem, however for the estimation of the unitary step size in a photobleaching trace this does not matter, hence it is an algorithm of choice.

The traces here were found to be noisier than those of earlier studies which used the Fourier spectral approach, primarily since the exposure time per frame was at least an order of magnitude smaller. This resulted in greater variance of the predicted unitary step size when comparing individual photobleach traces for spot intensity. To improve the reliability for the estimation of the size of the unitary YPet photobleaching step in intensity we modified the method by performing a multiple Gaussian fit to the collated, binned spot intensity data, and taking the center of the lower order peak as the best estimate for the unitary step size of YPet *in vivo* in that particular cell strain, I_{YPet}. The values of I_{YPet} across the different cell strains used in this study agreed to within ~20% in the range 1080-1290 counts. Each individual spot intensity trace was then fitted by a single exponential decay function $I(t)=I_0\exp(-t/t_b)$, where t_b was the optimized photobleach time (typically ~45 ms). The number of YPet molecules associated with each fluorescent spot was then estimated as the initial intensity I_0 divided by the appropriate I_{YPet} value for that cell strain (*S9*).

We then constructed the distribution of stoichiometries using an unbiased kernel density estimation via a Parzen window method (*S19*): here we convolved the stoichiometry dataset with a Gaussian kernel of set width equal to the noise of the measurement (equivalent to typically ~1 YPet molecule in terms of peak-to-peak amplitude of the intensity signal). Strains PolIII, β-YPet, ε-YPet and τ-YPet showed clear indication for a bimodal distribution in stoichiometry such that the center position of one peak was roughly twice that of the other

(Fig. 2). This was consistent with a fluorescent spot encompassing either one or two replication forks; in the latter case if the physical separation of the two forks was less than the diffraction-limited optical resolution of ~250-300 nm, as determined by the point spread function width, then they could not be resolved separately but instead detected as a single spot. For these strains, the stoichiometry distribution was fitted with a 2-Gaussian model constraining the peak position of one to be exactly twice that of the other, but permitting the widths and amplitudes of both to be unconstrained.

The χ-YPet, ψ-YPet and Ssb-YPet strains indicated more than 2 clear peaks. The mean separation of the Ssb peaks was periodic enough to be estimated from the fundamental peak of the power spectrum of the distribution, indicating a periodicity of ~4 molecules (Fig. 2).

The contribution to fluorescence intensity from a dark, immature population of YPet was investigated by completely photobleaching the cell for several seconds in the presence of 50 μg/ml chloramphenicol to suppress protein expression and then measuring any subsequent recovery in fluorescence after ~60 min. We found that the level of recovery of fluorescence was comparable to background noise indicating a negligible immature YPet population.

Estimating the quantity of cytoplasmic YPet not bound in the replisome.

The mean intensity per pixel at point (x_0,y_0,z_0) not within a distinct fluorescent spot, $I_m(x_0,y_0,z_0)$ was modeled as the 3D convolution integral of the point spread function $P(x,y,z)$ of a single YPet molecule with the spatial distribution for number density of YPet in the cell $dN/dV(x,y,z)$ and normalized local excitation intensity $L(x,y,z)$ multiplied by the intensity due to a single YPet molecule I_{YPet}:

$$I_m(x_0,y_0,z_0)\,dA = I_{YPet}\frac{dN}{dV}\otimes P\otimes L = I_{YPet}\iiint_{cell}\frac{dN}{dV}P(x-x_0,y-y_0,z-z_0)L(x,y,z)dxdydz$$

$$= nI_{YPet}\iiint_{cell}P(x-x_0,y-y_0,z-z_0)\exp\left(-\left(\frac{x^2}{2l_x^2}+\frac{y^2}{2l_y^2}+\frac{z^2}{2l_z^2}\right)\right)dxdydz$$

$$\equiv nI_{YPet}S$$

Here, P was measured from a previous study *(11)* using a z-stack of images for 20 nm yellow fluorescent beads (Molecular Probes) immobilized to the coverslip. The limits of the integral are defined by the boundaries of the cell body in x, y and z. The number density of YPet molecules per unit volume assumed freely diffusive in the cytoplasm, n, is approximated as being constant throughout the cell. I_{YPet} is the total intensity due to a single photoactive YPet

molecule excited at the origin (i.e. when L=1). The function L represents the spatial distribution of the slimfield illumination over the cell and is similarly modeled as a 3D Gaussian with $l_x = l_y = \sigma_{xy}$ = 3.0 μm and $l_z \approx 2.5\sigma_{xy}$ = 7.5 μm (*S10*). The pixel area at the sample plane is dA (50nmx50nm). The values of S were estimated by numerical integration using values of x_0 and y_0 over a range ± 2 μm centered on a model cell composed of a cylinder length 2 μm with hemispherical caps of diameter 1 μm (*S11*) (fig. S9). This indicates that, in the central region of the cell image bounded by a rectangle of length 1 μm and width 0.5 μm in which the majority of spots were detected, S = 6.1 ± 0.6 (mean ± SD)

We then estimated the mean pixel intensity not associated with distinct foci for each strain, after subtracting the contribution from cellular autofluorescence (~30 counts per pixel) and instrumental background (~1,000 counts per pixel) assessed by imaging the parental non-YPet cell strain under the same microscopy conditions. For example, for the Ssb-YPet strain we measured I_m = 545 ± 180 counts. Using I_{YPet} = 1140 counts (Fig. 1) this indicates that n = 0.08 ± 0.02 YPet molecules per voxel. We estimated that the average volume of an Ssb-YPet strain cell was ~16,500 voxels. This indicates a mean of ~1,320 ± 420 molecules per cell. However, since the cell length can change by a factor of ~two during the cell cycle the effective associated error for a cell sampled randomly during the cell cycle is ± 600 molecules per cell, which is relevant in light of making comparisons from previous studies using bulk biochemical methods. The range of estimated copy numbers for delocalized proteins in the cytoplasm for the other strains used in this study is roughly 30-600 per cell, indicated in table S1. In general our mean estimates are lower by a factor of ~2 compared to earlier bulk-assay based findings, which is likely to be due to our using minimal M9 growth media instead of LB.

Analysing spot size and shape.

To estimate the mean size and shape of putative single-replisome spots (cells containing two distinct spots), cell images in each strain were first rotated to align all long axes parallel to the camera axis (denoted as the x-axis). Spot images were normalized by pixel intensity, superimposed and a mean for each pixel value in a 16x16 array centered on the spot centroid calculated. Estimates for mean FWHM $\langle\sigma\rangle$ of a symmetrical 2D Gaussian fit and the ratio of the FWHM σ_x/σ_y for the 1D Gaussian fits through axes parallel and perpendicular to the cell long-axis through the centroid were then performed. Averaging across the PolIII, ε, τ, χ and ψ-YPet strains indicated $\langle\sigma\rangle$ = 305 ± 30 nm and σ_x/σ_y = 1.06 ±0.06. Performing a Student t-test at a confidence level of p = 0.05 indicates no statistical difference to either $\langle\sigma\rangle$ or σ_x/σ_y across the mean spot data for these strains. The same procedure applied to surface-

immobilized YPet but omitting the rotation step indicates a FWHM width of ~250 nm (fig. S7).

Mean spot data for strains YPet-βand Ssb-YPet indicated a statistically significant difference to both $<\sigma>$ and σ_x/σ_y manifest as a less circular appearance of the mean spot shape extended parallel to the cell long axis. For the YPet-β strain separate analysis was done on individual, non-averaged spot images. The shape of these spots was found to consist of two populations: Type I spots (27%) were reasonably circular such that $\sigma_x/\sigma_y < 1.2$; population Type II spots (73%) had a more extended appearance parallel to the long axis of the cell such that $\sigma_x/\sigma_y \geq 1.2$ (fig. S8). The long axis intensity profile of Type I spots could be fitted using a single Gaussian of width 305 nm consistent with the mean $<\sigma>$ found previously while the Type II spots could not be fitted adequately with a single Gaussian function, but instead were fitted well by two Gaussians. Each Gaussian here was constrained to a width of 305 nm with the ratio of the heights constrained to 2.0 ± 0.3 but the separation *d* of the respective Gaussian centers allowed to vary freely. Good fits (probability of true fit >90% based on generated χ^2 values) were obtained over a range of *d* of ~50-250 nm. Since β is a dimer (*S20*), and assuming ~6 copies per single-replisome spot (Fig. 2) this indicates that Type I spots consist of three β dimers in the "central" replisome region whereas Type II spots consist of two β dimers in the central replisome region, and one β dimer outside of this zone at a variable distance of the order ~50-250 nm.

Modeling turnover of Ssb-YPet

We modeled the experimental FRAP results on the Ssb-YPet strain (fig. S1) as turnover of the Ssb-YPet tetramer sub-unit binding to and unbinding from ssDNA within a closed reaction-diffusion environment confined to the finite volume of the cell in which total content of Ssb-YPet is in steady-state.

Here we denote:

$S_F(t)$ = Unbound number of Ssb-YPet tetramers free in the cytoplasm at time t ($t \geq 0$) following initial focused laser bleach.

$S_B(t)$ = Bound number of the Ssb-YPet tetramers to ssDNA.

$S_T(t)$ = Total number of Ssb-YPet tetramers in the cell.

$S_B^*(t)$ = Bound number of photoactive Ssb-YPet tetramers to ssDNA.

f = fraction of Ssb-YPet photobleached following initial focused laser bleach.

k_1 = on-rate per Ssb-YPet tetramer for binding to ssDNA.

k_{-1} = off-rate Ssb-YPet tetramer for unbinding from ssDNA in absence of HU.

k_{-2} = off-rate Ssb-YPet tetramer for unbinding from ssDNA in presence of 100 mM HU.

Since Ssb-YPet is in steady-state:

$\frac{\partial S_T}{\partial t} = 0$. Thus, S_T = constant= $S_F + S_B$

We denote the reaction-diffusion equations as:

$$\frac{\partial S_F}{\partial t} = D\nabla^2 S_F \quad (1)$$

$$\frac{\partial S_B}{\partial t} = k_1 S_F - k_{-n} S_B \quad (2)$$

We assume that the presence of 100 mM Hydroxyurea (HU) impairs completely the functionality of PolIII (*S21*) thereby preventing PolIII from displacing bound Ssb as it moves along the DNA. Thus we assume the off-rate will be affected by the presence of HU, but not the binding on-rate of Ssb. Here, *n* is either 1 or 2 depending on the absence or presence respectively of HU. *D* is the effective diffusion coefficient of the Ssb-YPet tetramer in the cytoplasm. However, the typical diffusion time scale τ is set by $\sim L^2/D$ where *L* is the typical length dimension of the cell ~1 μm and *D* for typical small proteins and molecular complexes in the cytoplasm is ~10 $\mu m^2/s$, indicating τ ~10 ms. However, the FRAP data indicate observable turnover over a time scale of ~10-100 s. Thus this is clearly a reaction-limited regime (*S22*) and so Eqn. 1 can be ignored. Also, at equilibrium (for example, before the focused laser bleach):

$$\frac{\partial S_{B,eq}}{\partial t} = 0 \therefore k_1 S_{F,eq} = k_{-n} S_{B,eq} \therefore k_1 = \frac{k_{-n} S_{B,eq}}{S_T - S_{B,eq}} \quad (3)$$

Where $S_{B,eq}$ and $S_{F,eq}$ are the value of S_B and S_F respectively at equilibrium. We assume the binding kinetics of photoactive Ssb-YPet are identical to those of photobleached Ssb-YPet and that the population of bleached and non-bleached are ultimately well-mixed, indicating:

$$S_B^* = S_B(1 - f) \quad (4)$$

Under general non-equilibrium conditions, Eq. 1 can be re-written:

$$\frac{\partial S_B}{\partial t} = k_1(S_T - S_B) - k_{-n} S_B$$

Solving this and substituting in for Eqs. 3 and 4 indicates:

$$S_B^*(t) = S_B(t)(1-f) = \frac{1-f}{k_1+k_{-n}}\left(k_1 S_T - \left(k_1 S_T - (k_1+k_{-n})S_B(0)\right)\exp\left(-(k_1+k_{-n})t\right)\right)$$

$$= (1-f)\left(S_{B,eq} - \left(S_{B,eq} - \frac{S_B^*(0)}{1-f}\right)\exp\left(-(k_1+k_{-n})t\right)\right)$$

$$= S_{B,eq}^*\left(1-\left(1-\frac{S_B^*(0)}{S_{B,eq}^*}\right)\exp\left(-k_{-n}t\left(1+\frac{S_{B,eq}}{S_T-S_{B,eq}}\right)\right)\right)$$

$$= S_{B,eq}^*\left(1-(1-\alpha)\exp\left(\frac{-k_{-n}S_T t}{S_T-S_{B,eq}}\right)\right)$$

Here α is the ratio of the bound photoactive component of Ssb-YPet at zero time (i.e. immediately after the initial focused laser bleach) to the bound photoactive component of Ssb-YPet at equilibrium. Since intensity $I_B(t)$ of the bound Ssb-YPet component is proportional to the number of photoactive Ssb-YPet molecules, we then write:

$$I_B(t) = I_{B,eq}\left(1-(1-\alpha)\exp\left(\frac{-k_{-n}S_T t}{S_T-S_{B,eq}}\right)\right)$$

$$= I_B(\infty)\left(1-(1-\alpha)\exp\left(\frac{-k_{-n}S_T t}{S_T-S_{B,eq}}\right)\right) \qquad (5)$$

The experimental FRAP intensity data from the short and long time scale series were collated, estimating mean and SD values at each time point to generate two separate mean traces for experiments done in the absence (-HU) and the presence (+HU) of HU. The non-FRAP slimfield data indicated a mean Ssb-YPet stoichiometry of ~32 molecules (equivalent to 8 Ssb-YPet tetramers) in the absence of HU (Fig. 2). The mean pre-bleach FRAP intensity for the –HU dataset was divided by 32 to estimate the average intensity per photoactive Ssb-YPet molecule under epifluorescence imaging used for FRAP imaging, indicating ~500 counts per Ssb-YPet molecule. This was consistent with an estimate based on a mean I_{Ypet} value of ~1,000 counts multiplied by a factor of ~40/3 to account for the difference in exposure time per frame and multiplied again by a factor of ~280/6500 to account for the difference in excitation laser intensities between epifluorescence and slimfield. Both the -HU and +HU mean FRAP intensity traces were then normalized using the value of 500 counts per molecule to convert into the equivalent number of photoactive Ssb-YPet molecules. The post-bleach data were then fitted using a function $I(t)=A+B\exp(-t/t_r)$ where t_r is the equivalent 1/e recovery time. This indicated t_r values of 25 ± 8 s and 49 ± 9 s for -HU and +HU datasets respectively. Assuming S_F = ~1320 molecules per cell for the –HU dataset and there are two single-replisome spots per cell of mean stoichiometry S_B of ~32 molecules, this indicates a total Ssb content S_T of ~1384 molecules per cell, which we assume is unchanged upon addition of HU. Addition of HU changes the mean stoichiometry S_B of each single-replisome

spot to ~70 molecules per cell. Substituting in these values indicates k_{-1} = 0.04 ± 0.01 Ssb tetramers per sec, and k_{-2} = 0.02 ± 0.01 Ssb tetramers per sec. Assuming that k_{-1} is equivalent to a spontaneous off-rate plus a forced off-rate due to the motion of the DNA polymerase displacing Ssb as it progresses along the DNA, this indicates that the effective forced off-rate is ~0.02 Ssb tetramers per sec. Assuming a mean of ~8 tetramers per single-replisome spot this indicates a total forced off-rate of ~0.2 Ssb tetramers per second from a single replication fork, or an average effective dwell time of ~5 sec per tetramer.

Although PolIII does not displace Ssb in +HU, other proteins involved in recombination, replication restart and repair may modulate its exchange (*S23*). Therefore, it is likely that the spontaneous off-rate of Ssb from ssDNA may be smaller than what we observe, and correspondingly that mediated from the polymerase may be greater. Future experiments will investigate these possibilities.

Simulating photobleaching and characterizing noise-reduction of steppy data

To test the relative noise reduction in filtering raw photobleach intensity traces, we generated simulated data and ran these through three different candidate filtration algorithms. Simulations were generated using a Monte Carlo method under the assumption of a stochastic Poisson distribution bleach model. Data points were polled at time intervals Δt of 3 ms identical to the experimental protocol. Since a principle conclusion from our data was that there were three polymerases in each active replication fork we initialized each simulation with three photoactive YPet molecules. At each time point each photoactive YPet molecule was polled as to whether or not it would irreversibly photobleach. Assuming a single process Poisson distribution for photoactive lifetime, the probability for this occurring was set at $\Delta t/t_{\text{bleach}}$ where t_{bleach} is the measured bleaching time from the exponential fits to the experimental photobleach intensity traces (equating to the mean photoactive lifetime per individual YPet molecule) estimated at ~45 ms, indicating a photobleaching probability per time step per YPet molecule of ~0.07. This probability was then compared against a pseudo-random-generated number in the range 0-1, and if this number was less than the bleach probability then the YPet molecule was assumed to have irreversibly photobleached in that time step. Each unitary bleaching step was assumed to be 1,100 counts in size as indicated experimentally and pseudo-random Gaussian noise was then added to each time point at a level comparable to that of the experimental data (SD ~500 counts). Each simulation was run up to 150 ms.

Each simulation was then filtered using a Chung-Kennedy (CK) filter (*S24, S25*), a running median (M) filter, or a Savitzky-Golay (SG) filter. CK and M filters were compared using the same window widths, in the range 3-6 data points, the SG filter was set to

polynomial order 3. The results of four typical traces can be seen in fig. S11, showing the effects of increasing window width for the M and CK filters. What can be seen clearly is that the SG filter, although generating significant reductions in noise, fails to follow the sharp transition at the edges due to discrete photobleach events. This is intuitive as it is essentially a running polynomial regression, so will smooth noise fluctuations at a constant underlying intensity level well but fails (often dramatically) at a distinct edge (i.e. a step), which can be seen by eye from the simulated smoothed traces. It is not a de facto edge-preserving filter and is thus a poor choice when dealing with steppy photobleaching data. The CK and M filters are both edge-preserving, but close visual inspection of the filtered traces indicates that the CK filter deviates less from the underlying theoretical noise-free trace than the M filter under our conditions. Measuring the average standard deviation between the noise-free and unfiltered noisy traces for 20 simulated bleaches indicates a value of 526 ± 41 (± SD) counts, consistent with the imposed standard deviation noise level of 500 counts in the simulations. Using a window width identical to that chosen in the experimental protocol of 5 data points, the M filter generated an average standard deviation from the theoretical noise-free trace of 387 ± 78 counts, whereas the CK filter using the same window width gave a comparable value of 320 ± 76 counts, with a ratio of the standard deviations using the two different filters of 1.31 ± 0.17. That is, the standard deviation of the M-filtered noisy traces from the theoretical noise-free traces were on average ~30% higher than those for the CK-filtered noise traces.

Therefore, under our experimental conditions, there is strictly speaking a reduction in true-step detection confidence of the M filter compared to the CK filter. In essence the CK filter is a switching mean filter, so that the forward and backward windows switch in response to their respective levels of data variance which thus preserves the edge but with the output converging on the mean edge-free level. The output of a running median converges similarly to the same expected value of the mean distribution, however it can be shown that the sample variance of the median distribution is actually greater than that of the sample mean distribution by a factor of $\pi/2$ (*S26*) and so the standard deviation of the smoothed output from the theoretical expected value is larger by a factor of ~1.25 for the median compared to the CK filter. This is consistent with our estimation of 1.31 ± 0.17 using simulated data. The confidence of correctly identifying a true step while not identifying a false one decreases with increasing standard deviation of the filtered output, and so there is an improvement in favoring the CK filter over the median, assuming both filters are matched in terms of running window widths. In terms of temporal resolution, there is no intrinsic loss in using the CK filter (*S25*), however there is a caveat in that the width of the running window must be less than the typical dwell time between step events otherwise there is a risk of detecting more than one step event in a single window which produces an erroneous output. Thus, although

the temporal resolution is unaffected there is an upper limit on this window width and thus a limit on the improvement in smoothing the data.

One thing however that is clear in the step-finding community is that there is no single filter/detection algorithm that will suit all cases of steppy data, for example other studies have found conditions in which the median filter appears to perform better than the CK filter (*S27*), but rather there is strong case for using the most reliable algorithm based on simulated realistically noisy data of each particular biological system with a full exploration of parameter space.

SUPPLEMENTARY FIGURES AND TABLES

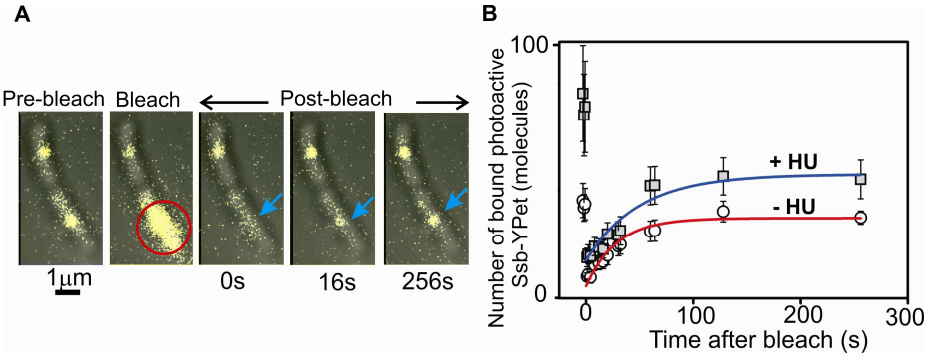


Fig. S1. (A) Brightfield (gray) and epifluorescence (yellow) images overlaid for two Ssb-YPet cells, one of which is subjected to FRAP. Position of laser focal waist indicated (red circle) with subsequent fluorescence recovery of replisome spot (cyan arrow). Minimum-maximum display range for pixel intensity of the bleach image is ~20 times larger than for the other images in the series in order to visualize the extent of the bleach without rendering pixel saturation. **(B)** Mean values of fluorescence intensity of single Ssb-YPet replisome spots during FRAP collating data from both short and long time scale sequences. Results shown in the absence of HU (white circles, *N*=12 cells) or with 100 mM HU added (gray squares, *N*=12 cells) converted into number of bound photoactive Ssb-YPet molecules (Methods), exponential fits (red and blue respectively) indicating 1/e recovery times of 25 ± 8 s (-HU) and 49 ± 9 s (+HU), SEM error bars.

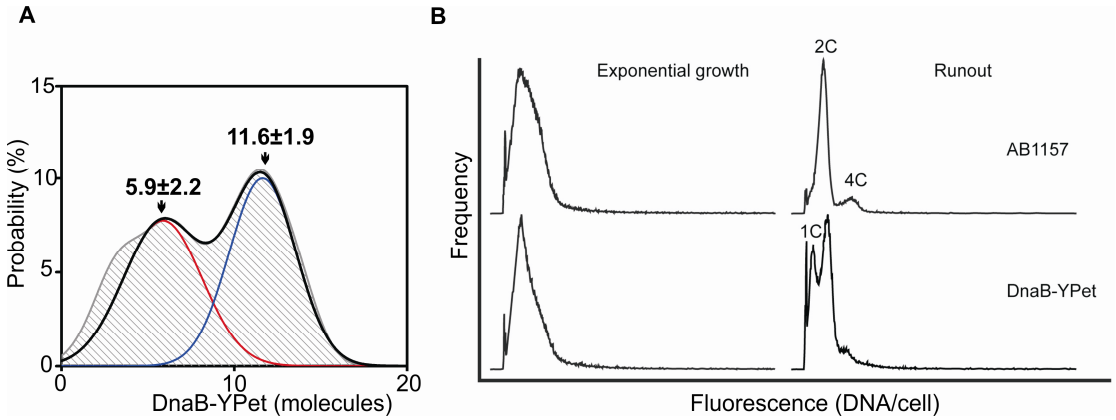


Fig. S2. (A) Stoichiometry distribution for DnaB-YPet based on unbiased kernel density estimation (gray), with 2-Gaussian fit (black) and contributing single Gaussian fit curves (red and blue), mean and SD indicated. **(B)** Equivalent FACS flow cytometry trace compared against parental AB1157; the over-representation of double-hexamers observed in **(A)** is reflected in this trace and may result from abnormal initiation caused by the YPet fusion.

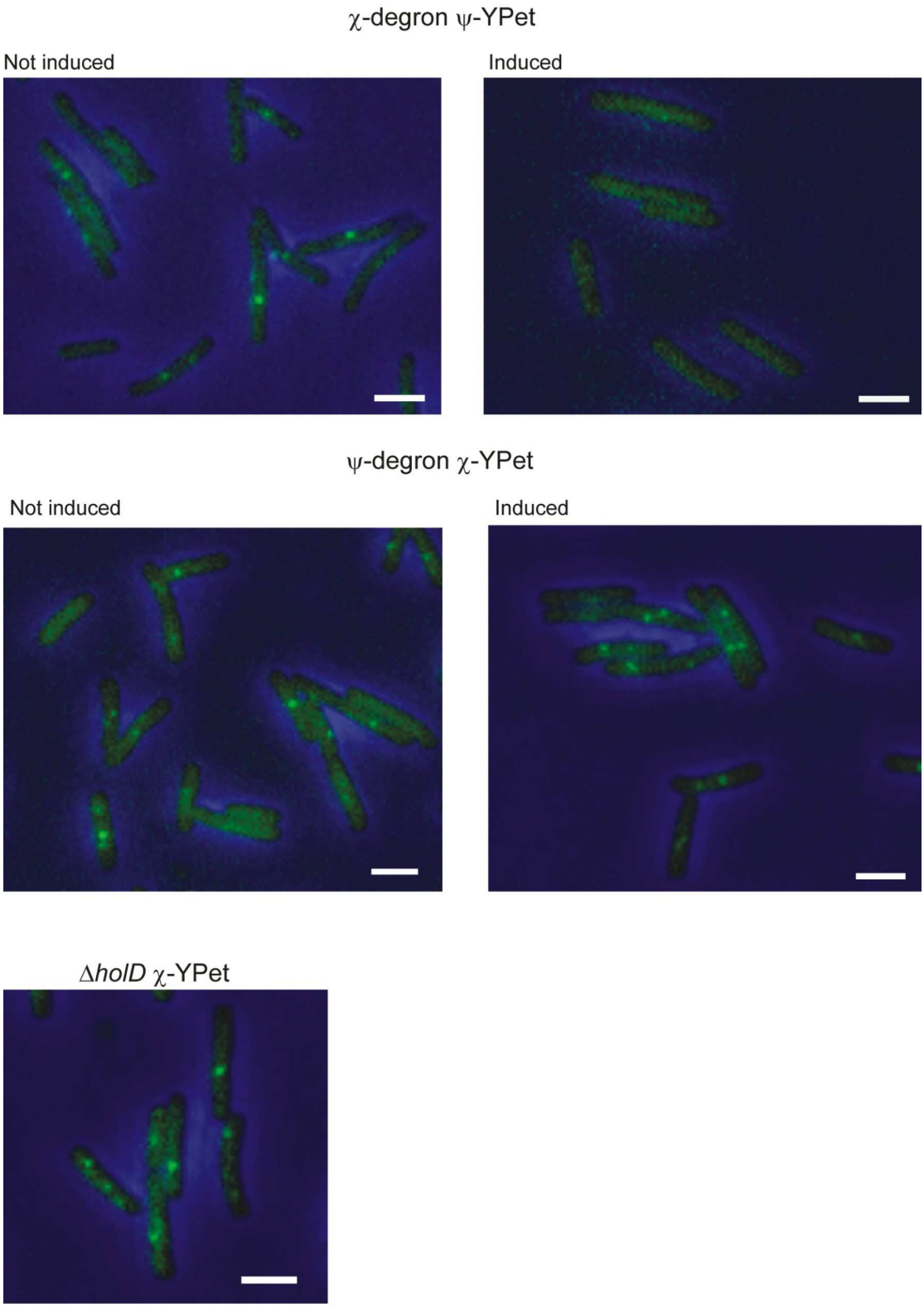
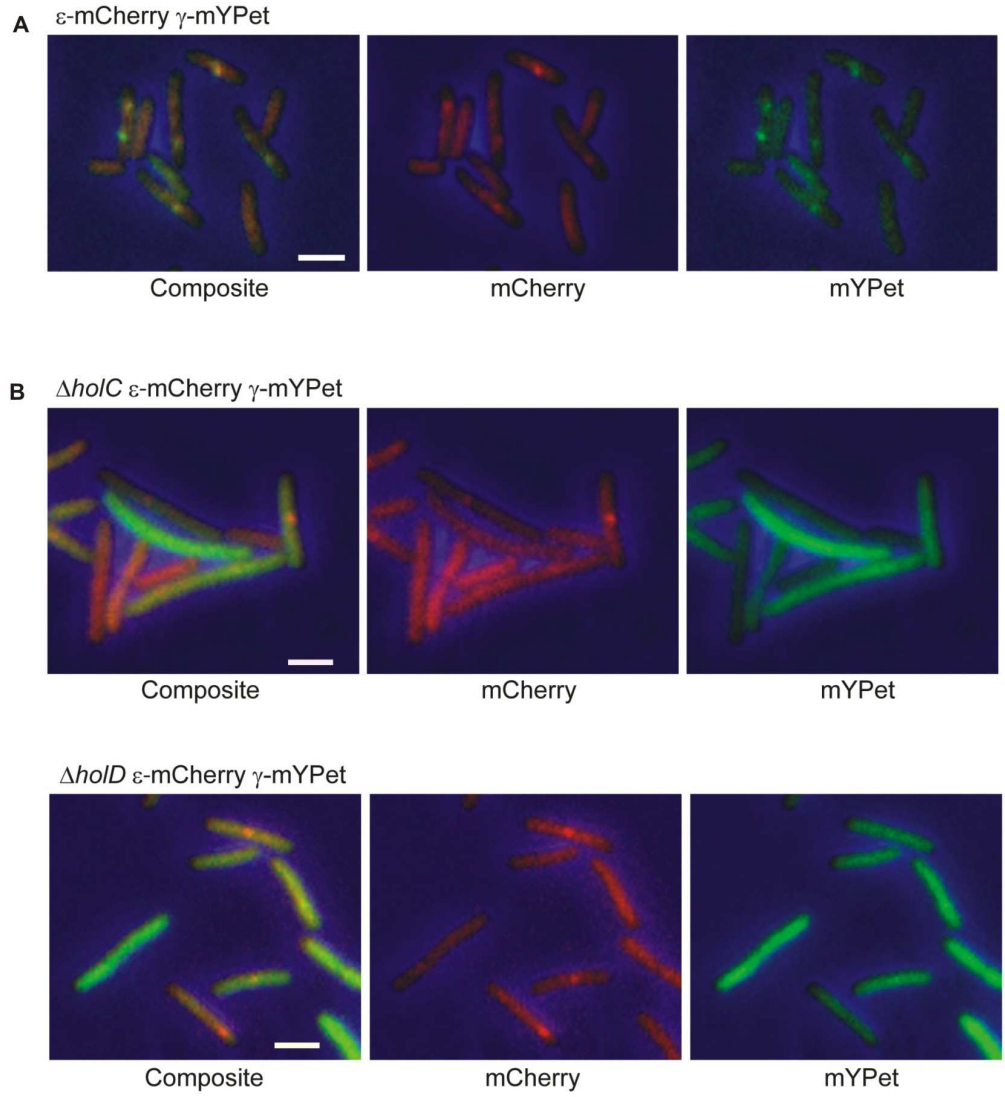


Fig. S3. Formation of foci by ψ-YPet was observed in a strain carrying a χ-degron before but not after degradation was induced. The reciprocal experiment with χ-YPet and a ψ-degron showed no apparent dependency of this protein for ψ since foci persisted even after degradation was induced. The later result was confirmed using a strain carrying a deletion for *holD* (coding for ψ) where χ-YPet foci were still observed. White bar=2 μm.



	Percentage of cells					Co-localization	
# ε Foci	0	1	2	>3	n	(%)	n
wt	40.9	17.3	41.2	0.6	682	91.0	212
Δ***holC***	63.7	9.8	20.6	5.9	471	26.5	189
Δ***holD***	52.4	13.1	33.8	0.7	574	44.0	218

Fig S4. (A) Co-localization of γ-mYPet with ε–mCherry. **(B)** Co-localization of γ-mYPet and ε–mCherry is lost in Δ*holC* and Δ*holD* strains, respectively. The table (bottom) shows the % of cells with the indicated number of ε–mCherry foci for wild type and Δ*holC and* Δ*holD* strains. The Δ*holC* strain grows less well than the Δ*holD* strain, indicating loss of χ is more deleterious than loss of ψ, and shows a higher proportion of non-replicating cells. The % of ε–mCherry foci associated with a γ-mYPet focus is also shown (co-localization). When ε–mCherry foci are associated with a γ-mYPet focus, the latter is almost always of reduced intensity. In other experiments (not shown), targeted proteolysis of a ψ-degron, led to loss of γ-mYPet foci, while in a separate strain, τ-YPet foci remained after ψ degradation.

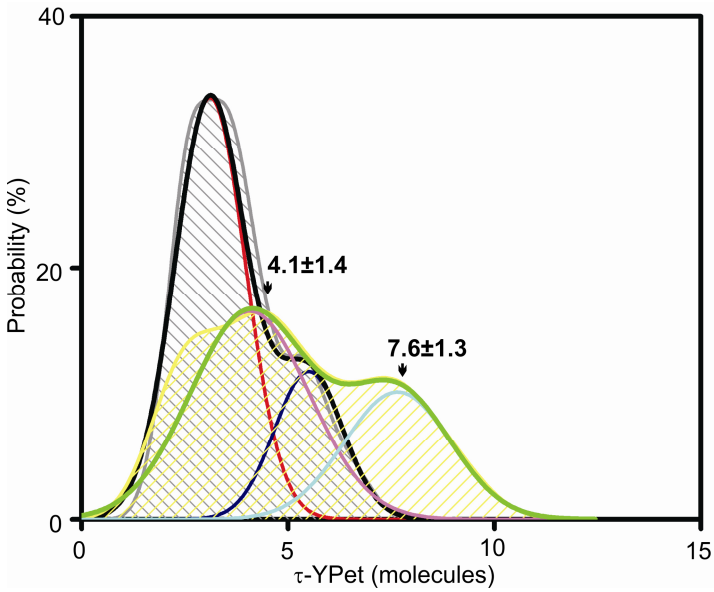


Fig. S5. Stoichiometry distribution for τ-YPet based on unbiased kernel density estimation of the τ-YPet strain expressing the γ component (gray), with 2-Gaussian fit (black) and contributing single Gaussian fit curves (red and blue), plus the τ-YPet strain not expressing γ (yellow), with 2-Gaussian fit (green) and contributing single Gaussian fit curves (magenta and cyan), mean and SD on the non-γ strain indicated.

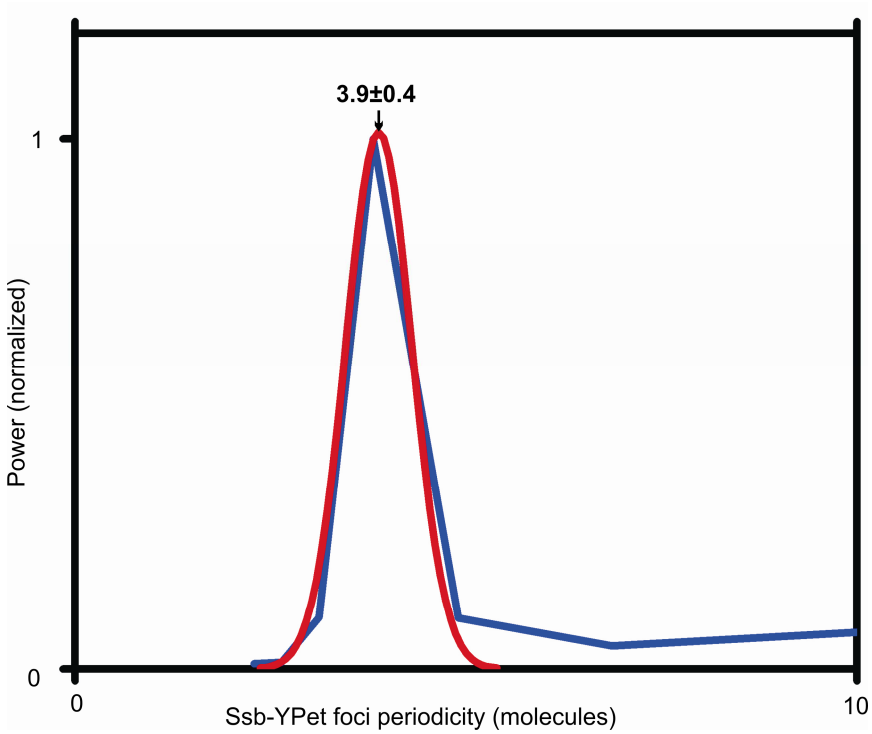


Fig. S6. The power spectrum (blue) for the periodicity in the stoichiometry distribution of the Ssb-YPet strain with Gaussian fit to peak (red), normalized by peak amplitude, mean±SD indicated.

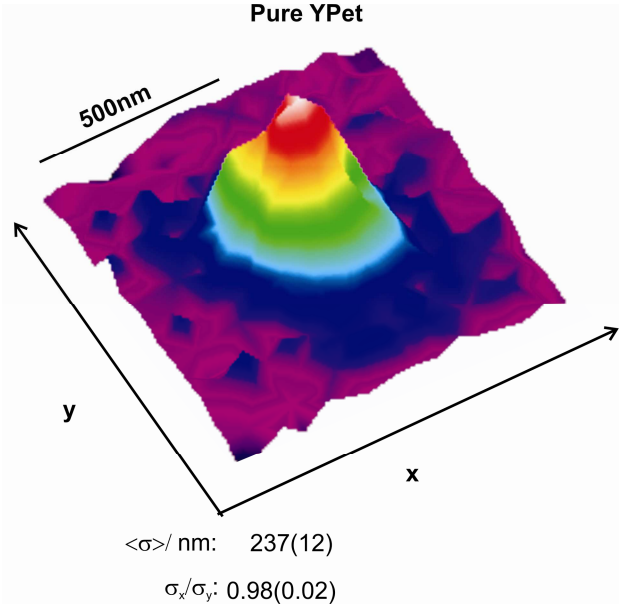


Fig. S7. False-color contour plot indicating the 2D spatial distribution for averaged fluorescent spots corresponding to single surface-immobilized YPet molecules, using *N*=55 separate spots, mean <σ> and σ_x/ σ_y with (SD) indicated.

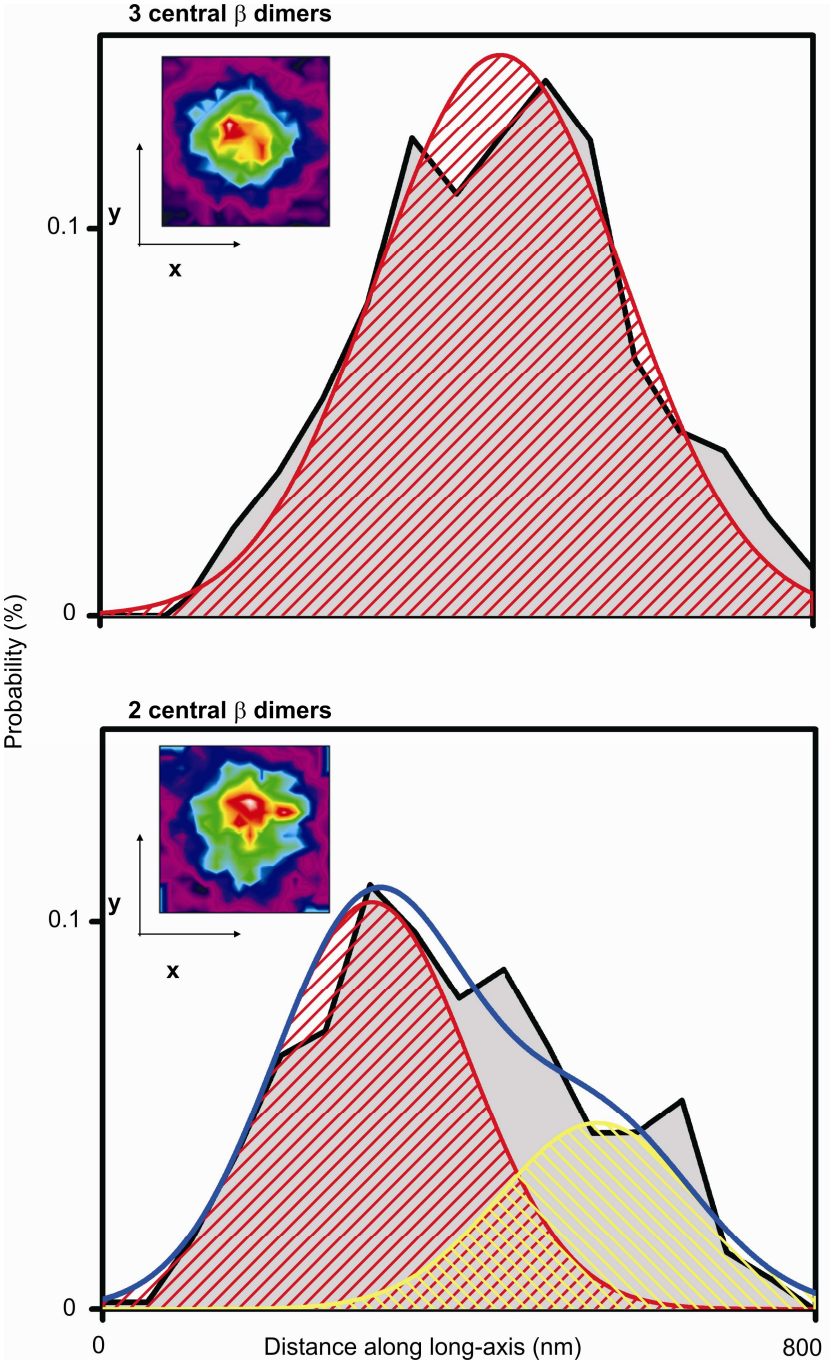


Fig. S8. Intensity profiles (gray) for two non-averaged spots of the β-YPet strain taken parallel to the cell's long-axis through each spot centroid indicating (upper panel) ~three β dimers in the central replisome region of width ~305 nm (red Gaussian fit) and (lower panel) ~two β dimers in the central replisome region (red Gaussian fit) with ~one β dimer displaced ~200 nm from the center (yellow Gaussian fit), combined fit indicated (blue), spot images inset. Of the *N*=64 β spots investigated, 17 (27%) were of the three-central-dimers type and 47 (73%) were of the two-central-dimers type with the non-central dimer displaced in the range ~50-250 nm.

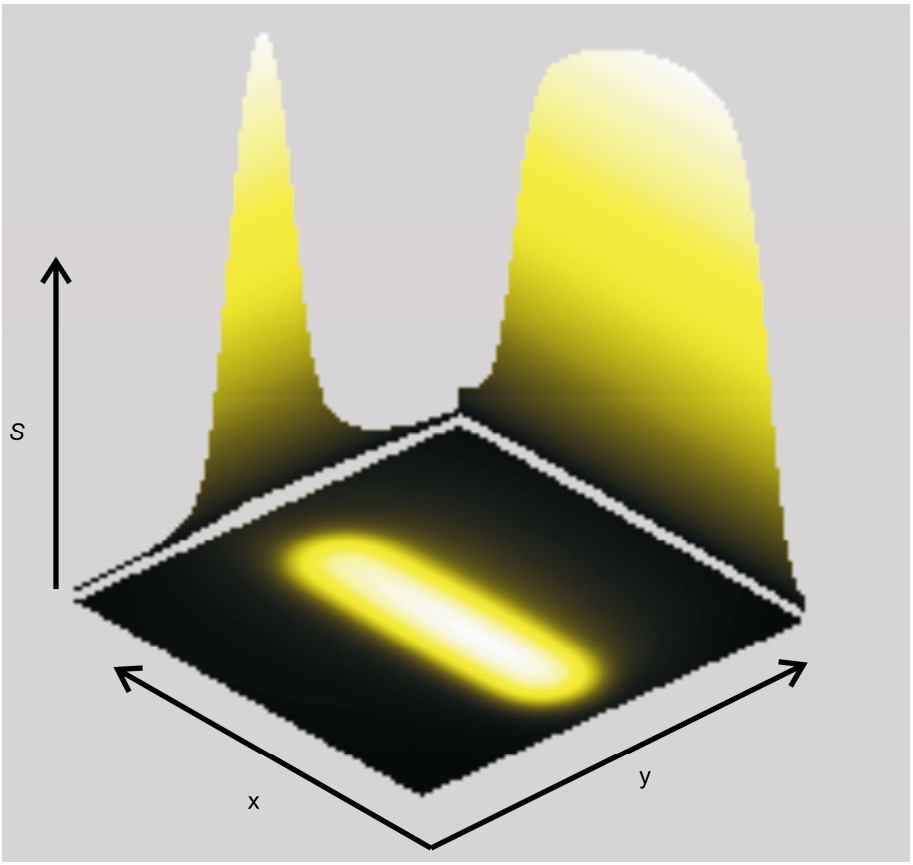


Fig. S9. Projection of the integral *S* on the *xy* plane for the 3D convolution model of cytoplasmic fluorescence intensity of Ssb-YPet in a cell, with profiles parallel and perpendicular to the long axis of the cell indicated.

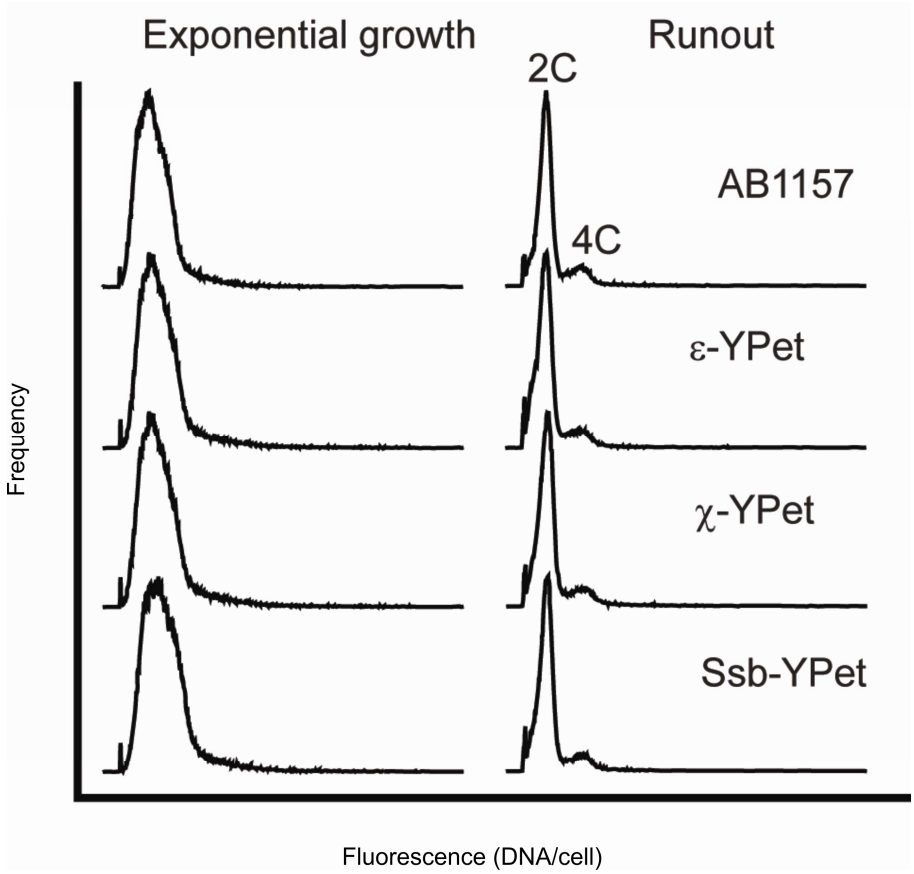


Fig. S10. FACS flow cytometry profiles of exponential cultures grown in M9-Gly at 37ºC before and after replication runout. All fusion strains except that producing DnaB-YPet showed comparable profiles. DnaB-YPet is shown in fig. S2.

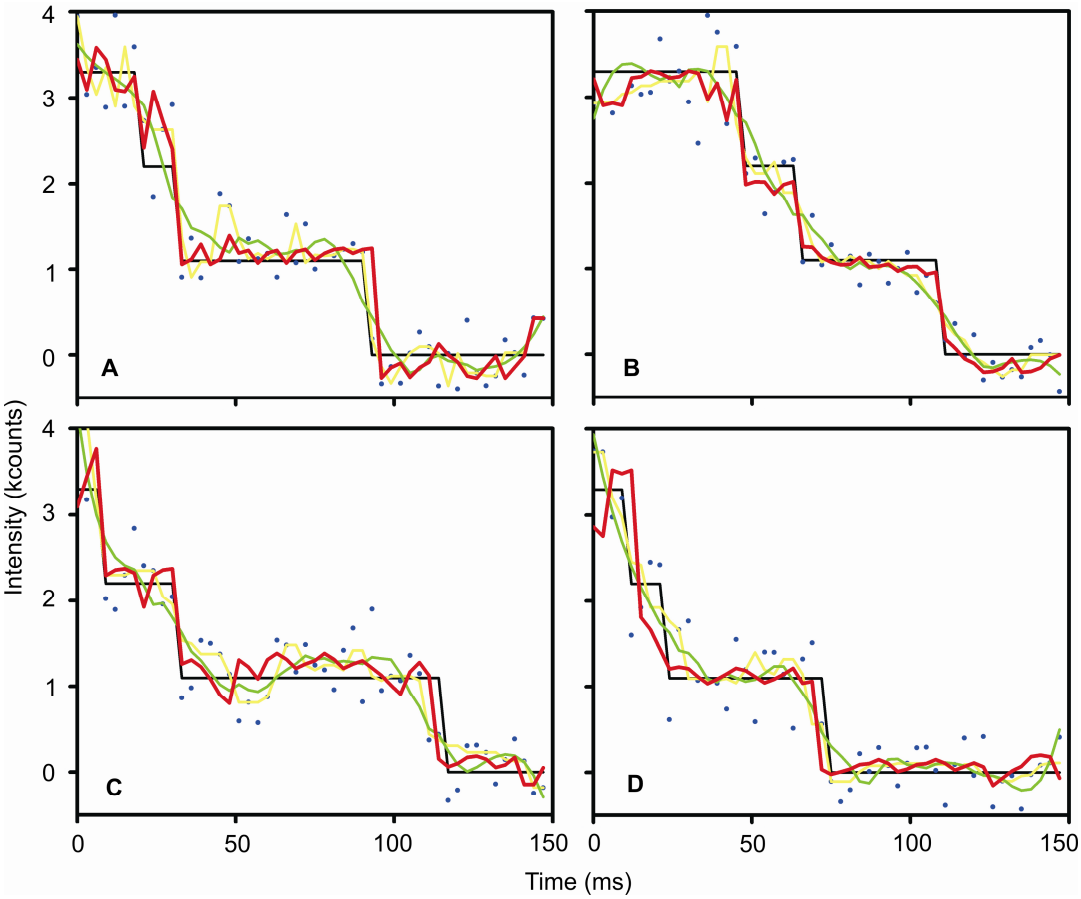


Fig. S11. Simulated photobleach data, showing theoretical noise-free trace (black), theoretical trace plus noise (blue dots), then this filtered using Chung-Kennedy (red), median (yellow) and Savitzky-Golay order 3 (green) filters. Window widths for Chung-Kennedy and median are **(A)** 3, **(B)** 4, **(C)** 5 and **(D)** 6 data points.

Table S1

Protein	YPet fusion/ terminus	Generation time/ min[a]	Replication assay[b]	Mean copies[c] delocalized in cytoplasm[d] /molecules (SD)	Stoichiometry[e] per localized focus/molecules (SD)
DnaB	C	121	FC	300 (210)	6 (2)
α (DnaE)	C	119	FC	80 (50)	3(1)
β (DnaN)	N	119	FC	600 (360)	6 (2)
ε (DnaQ)	C	111	FC	270(160)	3 (1)
τ (DnaX)	C	119	FC	130(90)	3 (1)
τ (*dnaX* $\tilde{\gamma}$)[f]	C	117	ND	180(140)	4 (1)
δ (HolA)	C	117	FC	160(110)	1 (1)
χ (HolC)	C	121	FC	140(160)	4 (2)
ψ (HolD)	C	103	FC	220(130)	4 (2)
Ssb	C	117	FC	1320(420)	32 (12)

[a] Generation time of the wt strain AB1157 in M9-Gly at 37^{o}C was 111 min. In all strains the generation time in LB was 31-32 min.
[b] fig. s2, s10. FC, flow cytometry. ND, not determined
[c] Rounded to nearest 10 molecules
[d] Determined using total fluorescence intensity in cells
[e] Rounded to nearest non-zero integer molecule
[f] Strain produces τ only, unlike the strain above that produces comparable amounts of γ and τ

SUPPLEMENTAL REFERENCES

S1. R. Reyes-Lamothe, C. Possoz, O. Danilova, D. J. Sherratt, *Cell* **133**, 90 (2008).
S2. K. A. Datsenko, B. L. Wanner, *Proc Natl Acad Sci U S A* **97**, 6640 (2000).A.
S3. Blinkova *et al.*, *J Bacteriol* **175**, 6018 (1993).
S4. K. E. McGinness, T. A. Baker, R. T. Sauer, *Mol Cell* **22**, 701 (2006).
S5. X. Wang, C. Possoz, D. J. Sherratt, *Genes Dev* **19**, 2367 (2005).
S6. A. Nguyen, Daugherty PS., *Nat Biotechnol* **23**, 355 (2005).
S7. R. R. Meyer, P. S. Laine, *Microbiol Rev* **54**, 342 (1990).
S8. E. V. Bobst, A. M. Bobst, F. W. Perrino, R. R. Meyer, D. C. Rein, *FEBS Lett* **181**, 133 (1985).
S9. M. C. Leake *et al.*, *Nature* **443**, 355 (2006).
S10. C. J. Lo, M. C. Leake, R. M. Berry, *Biophys J* **90**, 357 (2006).
S11. C. J. Lo, M. C. Leake, T. Pilizota, R. M. Berry, *Biophys J* **93**, 294 (2007).
S12. M. Plank, G. H. Wadhams, M. C. Leake, *Integr. Biol.* **1**, 602 (2009).
S13. S. T. Hess, W. W. Webb, *Biophys J* **83**, 2300 (2002).
S14. T. Schmidt, G. J. Schutz, W. Baumgartner, H. J. Gruber, H. Schindler, *Proc Natl Acad Sci U S A* **93**, 2926 (1996).
S15. M. C. Leake *et al.*, *Proc Natl Acad Sci U S A* **105**, 15376 (2008).
S16. T. Lenn, M. C. Leake, C. W. Mullineaux, *Mol Microbiol* **70**, 1397 (2008).
S17. T. Lenn, M. C. Leake, C. W. Mullineaux, *Biochem Soc Trans* **36**, 1032 (2008).
S18. G. I. Mashanov, D. Tacon, A. E. Knight, M. Peckham, J. E. Molloy,*Methods*. **29**, 142 (2003)
S19. E. Parzen, *Ann. Math. Stat*, **33**, 1065–1076 (1962).
S20. A. Johnson, M. O'Donnell, *Annu Rev Biochem* **74**, 283 (2005).
S21. N. K. Sinha, D. P. Snustad, *J Bacteriol* **112**, 1321 (1972).
S22. B. L. Sprague, R. L. Pego, D. A. Stavreva, J. G. McNally, *Biophys J* **86**, 3473 (2004).
S23. R. D. Shereda, A. G. Kozlov, T. M. Lohman, M. M. Cox, J. L. Keck, *Crit Rev Biochem Mol Biol* **43**, 289 (2008).
S24. S. H. Chung, R. A. Kennedy *J. Neurosci. Methods* **40**, 71 (1991).
S25. D. A. Smith *Phil. Trans. R. Soc. Lond. B* **353**, 1969 (1998).
S26. A. M. Mood, F. A. Graybill, D. C. Boes, *Introduction to the theory of statistics*, 3rd ed. McGraw-Hill International Edition, 257 (1982).
S27. B. C. Carter, M. Vershinin, S. P. Gross, *Biophys. J* **94**, 306 (2008).