

Multiscale Spatial Organization of RNA Polymerase in *Escherichia coli*

Ulrike Endesfelder,^{†△} Kieran Finan,^{†¶} Seamus J. Holden,^{‡§△} Peter R. Cook,[¶] Achillefs N. Kapanidis,^{§*} and Mike Heilemann^{†*}

[†]Institute of Physical and Theoretical Chemistry, Johann Wolfgang Goethe-University, Frankfurt, Germany; [‡]Laboratory of Experimental Biophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; [§]Department of Physics, Biological Physics Research Group, Clarendon Laboratory, University of Oxford, Oxford, United Kingdom; and [¶]Sir William Dunn School of Pathology, University of Oxford, Oxford, United Kingdom

ABSTRACT Nucleic acid synthesis is spatially organized in many organisms. In bacteria, however, the spatial distribution of transcription remains obscure, owing largely to the diffraction limit of conventional light microscopy (200–300 nm). Here, we use photoactivated localization microscopy to localize individual molecules of RNA polymerase (RNAP) in *Escherichia coli* with a spatial resolution of ~40 nm. In cells growing rapidly in nutrient-rich media, we find that RNAP is organized in 2–8 bands. The band number scaled directly with cell size (and so with the chromosome number), and bands often contained clusters of >70 tightly packed RNAPs (possibly engaged on one long ribosomal RNA operon of 6000 bp) and clusters of such clusters (perhaps reflecting a structure like the eukaryotic nucleolus where many different ribosomal RNA operons are transcribed). In nutrient-poor media, RNAPs were located in only 1–2 bands; within these bands, a disproportionate number of RNAPs were found in clusters containing ~20–50 RNAPs. Apart from their importance for bacterial transcription, our studies pave the way for molecular-level analysis of several cellular processes at the nanometer scale.

INTRODUCTION

Nucleic acid synthesis in many organisms is highly organized, often performed by assemblies of active polymerases associated with different templates and attached to larger cellular structures. In eukaryotes, for example, active replication forks and proteins are concentrated in distinct factories (1,2). It has also been proposed that analogous transcription factories contain clusters of RNA polymerases (RNAPs) and accessory factors active on multiple templates (3,4). Similarly, active viral polymerases are often immobilized and clustered; for instance, the RNA-dependent RNA polymerases of poliovirus function in large membrane-bound arrays (5), and DNA replication of phage ϕ 29 in *Bacillus subtilis* occurs at the MreB cytoskeleton (6).

It is unclear whether bacterial nucleic acid synthesis is also spatially organized. Thus, in *E. coli*, the active replication machinery is not stably attached to an immobile cellular structure, but instead exhibits constrained diffusion; furthermore, the position of one replication fork is usually independent of the other (7). In the case of transcription, various forms of organization have been proposed. One suggests that transcription occurs at a cytoskeleton, because MreB and RNAP copurify in cell extracts and interact in vitro (8); another finds that the subcellular positioning of the chromosomal loci of membrane proteins is influenced by expression level (9) and one more suggests active RNAPs cluster. Specifically, after growing *Escherichia*

coli in rich media and fixation, green fluorescent protein (GFP)-tagged RNA polymerase appears focally concentrated within the nucleoid; these foci are assumed to be clusters of the highly transcribed ribosomal RNA (*rrn*) operons (10,11), as they disappear when *rrn* transcription is reduced (during the stringent response; (11)) The in vitro aggregation of *E. coli* RNAP into higher-order structures (dimers to octamers) may also reflect bona fide, functionally important, interactions (12). The hypothesis that active RNAPs are immobilized is further supported by the fact that active transcription units interfere with the diffusion of DNA supercoils in vivo (13,14).

Clearly, there is a need to determine accurately and non-invasively the positions of all RNAP molecules in a bacterium. Although immunoelectron microscopy provides sufficient resolution, it requires harsh fixation that can distort nucleoid structure (15) and limit antibody access so that only a fraction of epitopes are labeled (16). Although conventional fluorescence microscopy of GFP-tagged proteins allows positional information to be obtained noninvasively, it has limited spatial resolution (i.e., ~200 nm in the focal plane, and ~500 nm along the optical axis). Here, we use photoactivated localization microscopy (PALM (17,18)) to overcome the diffraction limit. This method relies on the sequential localization of individual molecules to achieve a resolution of 10–40 nm in the focal plane. Using photoactivatable RNAP fusions, which yield a single photoactivation/bleaching cycle per fluorophore, we measure the distributions and numbers of RNAP molecules in cells grown under different conditions. At a single-cell scale, we observe RNAP bands that we attribute to RNAP bound to chromosomal DNA; at the scale within individual bands, we observe clustering that depends on growth media. We

Submitted January 14, 2013, and accepted for publication May 29, 2013.

[△]Ulrike Endesfelder, Kieran Finan, and Seamus J. Holden contributed equally to this work.

*Correspondence: heilemann@chemie.uni-frankfurt.de or kapanidis@physics.ox.ac.uk

Editor: Laura Finzi.

© 2013 by the Biophysical Society
0006-3495/13/07/0172/10 \$2.00

<http://dx.doi.org/10.1016/j.bpj.2013.05.048>



quantify the number and size of clusters and the number of RNAP molecules per cluster, and find small RNAP clusters in a minimal medium, and larger clusters in a rich medium (consistent with a large number of RNAPs transcribing either one or multiple ribosomal RNA operons).

MATERIALS AND METHODS

Genomic manipulation of *E. coli*

Insertions into the *E. coli* genome were created by recombineering as described (19). Briefly, polymerase chain reaction (PCR) fragments encoding 50 bp of homology to the genomic sequence upstream of the insertion site, followed by sequence of the element to be inserted, and then followed by 50 bp homologous to the genomic sequence downstream of the insertion site were prepared using Picomaxx DNA polymerase (Stratagene, Santa Clara, CA), and then purified using Minelute columns (Qiagen, Crawley, UK) followed by isopropanol precipitation. Competent cells were prepared by growing the strain DY330 to an OD_{600} of 0.4–0.6 at 32°C with shaking (200 rpm) followed by 15 min of shaking at 42°C (in a water bath). In a cold room (4°C), cells from 35 mL of culture were pelleted by spinning at $4600 \times g$ for 7 min, washed sequentially with 30 mL and 1 mL of ice-cold 10% glycerol, and resuspended in 200 μ L of 10% glycerol. In an ice-cold electro-oration cuvette (0.1 cm, BioRad, Hemel Hempstead, United Kingdom), 50 μ L of cells were mixed with roughly 1 μ L of 100 ng/ μ L purified PCR product, and electroporated at 1.8 kV. Time constants of successful transformations were always above 5 ms. 1 mL of room temperature lysogeny broth (LB) was immediately added, and the cells were shaken overnight at 32°C before plating on selective media. Colonies were streaked out, and the insertion loci amplified by colony PCR using primers outside the insertion site. The product was then sequenced to ensure proper insertion and absence of mutations in protein-coding sequence. To construct strain KF7-1, the 3' end of the endogenous *rpoC* gene was tagged with a mEos2-AmpR fragment amplified by the primers mEos2rpoCfw and mEos2rpoCrv from the plasmid pRSETamEos2 (Addgene, Cambridge, MA (20)), and transduced into MG1655 using P1 phage transduction. To construct strain KF26, the 3' end of the endogenous *rpoC* gene was tagged with a PAmCherry1-AmpR fragment amplified by the primers mCherry(rpoC)fw and mCherry(rpoC)rv from the plasmid pBADHisB PAm-Cherry1 (21), and transduced into MG1655 using P1 phage transduction (for strains and primers, see Table S2 in the Supporting Material).

P1 phage transduction

Transducing loci from one strain to another was performed using P1 phage (a gift from Dave Sherratt, Oxford University, Oxford, UK). To create P1 lysates, 300 μ L of saturated overnight culture of the donor strain was mixed with 30 μ L 50 mM $CaCl_2$ and 100 μ L P1 lysate (grown on MG1655), and incubated at 37°C for 20 min. The cells were added to 5 mL LB containing 5 mM $CaCl_2$ and shaken at 37°C. After >6 h, 1 mL of chloroform was added to kill any remaining cells, and then removed by spinning at $5000 \times g$ for 10 min and taking the supernatant. Lysates were stored at 4°C.

Transductions were performed by mixing 900 μ L of an overnight culture of the acceptor strain with 100 μ L 50 mM $CaCl_2$, pelleting the cells at $5000 \times g$ for 1 min, removing 900 μ L, and then resuspending the pellet and mixing with 50 μ L of donor phage lysate. Cells were incubated for 20 min at 37°C before the addition of 900 μ L phage buffer (100 mM Na_2HPO_4 , 22 mM KH_2PO_4 , 85 mM NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.001% gelatin) at room temperature, pelleted at $5000 \times g$ for 1 min, resuspended in LB + 5 mM sodium citrate, shaken for 1–3 h at 30°C or 37°C, plated on LB agar + 5 mM sodium citrate containing the appropriate antibiotics, and streaked to single colonies twice on the same media to purify away the phage.

Plasmid construction

Cloning was performed by fusing partially homologous DNA fragments using the In-Fusion recombinase (Clontech, Saint-Germain-en-Laye, France). PCR fragments were purified using a Minelute column (Qiagen, Crawley, UK), and then heated to 80°C for 10 min to remove residual ethanol. pKIE3-1 was constructed by fusing mEos2 (amplified from pRSETa-mEos2 using primers mEos2ampfw and mEos2amprv) with pBAD33 (amplified from pYpet-His using primers pBAD33ampfw and pBAD33amprv).

Western blotting

To isolate total *E. coli* proteins, 50 mL cultures were grown to $OD_{600} = 0.4$, pelleted at $5000 \times g$, washed with 1 mL TE, and resuspended in 600 μ L 5X sodium dodecyl sulfate (SDS) load dye (225 mM Tris-HCl pH 6.8, 5% SDS, 0.25 M dithiothreitol, 0.05% bromophenol blue, 50% glycerol). 1–5 μ L of protein were separated on 7.5% tris-HCl polyacrylamide gels (BioRad) with tris-glycine running buffer (7.55 g/L tris, 47 g/L glycine, 0.25% SDS) using the mini protean system (BioRad, Hemel Hempstead, UK) at 100–200 V. Proteins were transferred to a nitrocellulose membrane using an Iblot system (Invitrogen, Darmstadt, Germany) following the manufacturer's instructions. Membranes were blocked overnight in TBST (Tris-buffered saline + Tween 20; 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Tween 20) with a primary blocking agent, incubated with the primary antibody for 1 h, washed three times with 10 mL TBST, incubated with an antimouse horseradish peroxidase (HRP) (Jackson Immuno Research Laboratories, Inc., West Grove, PA) diluted 1:7500 in 15 mL TBST + 3% bovine serum albumin (BSA) for 1 h, washed 3 times with 10 mL TBST, and once with TBS (Tris-buffered saline; 20 mM Tris-HCl pH 7.5, 150 mM NaCl). Proteins were visualized using the Supersignal West Pico chemiluminescent substrate (Thermo Scientific, Loughborough, United Kingdom) and an LAS4000 CCD camera (Fuji). All primary antibodies were resuspended in 0.5X PBS + 50% glycerol at 1 mg/mL. For detecting the β' subunit of *E. coli* RNAP (mouse monoclonal; Neoclone WP001, Neoclone, Madison, WI), membranes were blocked in 1% low-fat milk, and incubated with 3 mL 1:1000 primary antibody in TBST + 3% BSA. For detection of PAm-Cherry1 (mouse monoclonal; Clontech 632543, Clontech, Saint-Germain-en-Laye, France), membranes were blocked in TBST + 3% low-fat milk, and incubated with 3 mL 1:1000 primary antibody in TBST.

Sample preparation

E. coli were grown with shaking at 200 rpm at 32°C to saturation in overnight cultures containing LB or M9, and were diluted 1/200 into the same media. At $OD_{600} = 0.4$, 1 mL cultures were quickly removed and immediately mixed with a fixation mix, resulting in final concentrations of paraformaldehyde (PFA) and $NaHPO_4$ buffer, pH7.5, of 1% (w/v) and 30 mM, respectively. Cells were fixed for 30 min at room temperature, and then washed 3 times with 1 mL PBS. For immobilization, a Labtek chamber (Nunc, Langensfeld, Germany) was prepared; single chambers were cleaned with 0.5% hydrogen fluoride (HF) for 3 min and then washed with sterile filtered PBS. 0.01% Poly-L-Lysine (Sigma Aldrich, St. Louis, MO) was incubated for 10 min; afterward, the chambers were dried for 10 min. Cell pellets after their last wash were then resuspended in 200 μ L PBS and incubated for 10 min in the prepared Labtek chambers. After washing with 1 mL PBS, the immobilized cells in the chambers were incubated in 1% PFA solution for 5 min before being finally washed 3 times with 1 mL PBS. The samples were stored in PBS at 4°C.

PALM imaging

PALM experiments were performed on a custom-built microscope essentially as described earlier elsewhere (22). A multiline argon-krypton laser

(Innova 70C, Coherent, Santa Clara, CA) and an ultraviolet laser diode emitting at 378 nm (Cube, Coherent) were coupled into a microscope body (IX71, Olympus, Japan) equipped with an oil immersion objective (60× NA 1.45, Olympus, Japan). Fluorescence emission was recorded with an EMCCD camera (Ixon, Andor, Ireland) and appropriate filters (AHF, Tübingen, Germany). Additional optics in the detection path adjusted the image pixel size to 85 nm. PALM imaging was performed at 568 nm excitation for readout (1–4 kW/cm²) and 378 nm for activation of mEos2 and PAmCherry1. Typically, between 5000 and 10,000 imaging frames at 100 ms integration time were recorded. RNAP-PAmCherry1 measurements were made using widefield excitation. Cytosolic PAmCherry1 measurements were performed under HILO excitation (23) to reduce the density of imaged molecules to a level comparable to the RNAP-PAmCherry1 data. The ultraviolet irradiation intensity was increased gradually during a PALM experiment from ~100 μW to 1 mW, and image acquisition was stopped when all fluorescent proteins were converted and detected. PALM images were generated with rapidSTORM (24) using a threshold of 600 photons (above median background noise) and a filter to discard spots with asymmetric shape. Single fluorophores detected in multiple adjacent frames were grouped as one detection event using Kalman filtering (25). The average localization precision was determined experimentally to 18 nm (according to (25)), from which a spatial resolution of ~40 nm was estimated (according to (26)).

Image processing and data representation

All images were contrast-saturated at 1% and rendered using the Gray Look-up-Table in ImageJ (National Institutes of Health, Bethesda, MD). All data were blurred ~2× greater than the average localization precision (~18 nm) for visual clarity. Because RNAP lacks a regular structure (unlike, e.g., microtubules), and is distributed in a random (albeit clustered) fashion, the highly pointillistic nature of raw Thompson blurred images (Fig. S1 A) is difficult to interpret by eye. To facilitate data visualization and manual identification of clusters, we instead Gaussian blurred the data with a 20-nm standard deviation (Fig. S1 B). Note that all quantitative analysis was performed using the list of RNAP localization coordinates, and not the processed images; the latter are presented only for visualization.

Nearest-neighbor distribution analysis of spatial distribution of RNAP

As an initial test for the presence of clustering, the randomness of the data was tested by statistical comparison of the nearest neighbor distribution to that of a simulated data set (Fig. S2 B). An outline of the cell was obtained by performing morphological closing on a binary image of observed localizations and simulated data were generated by distributing N points (equal to the number of observed localizations) randomly within the outline (Fig. S2 B). The nearest-neighbor distributions of each data set were calculated (Fig. S2 B); the residual between the experimental and simulated data set indicates the degree of deviation of the experimental data from an ideal random distribution. This deviation from a random distribution was quantified by calculating the Pearson χ^2 test statistic normalized over all cells in each data set. A normalized χ^2 test statistic close to one indicates that the data are indistinguishable from complete spatial randomness to within the sensitivity of the experiment. Higher values indicate deviation from randomness (i.e., clustering or regular structure).

The presence of chromosomal bands complicated analysis due to the presence of large-scale structure compared to the smaller RNAP clusters that are of high interest in terms of transcription. For this reason, analysis of nearest-neighbor distributions was performed, because it is an inherently short-range metric. Longer range analysis metrics (e.g., Ripley's K function (27)) or the pair correlation coefficient (27) was confounded by the presence of the larger scale structure.

To reduce any effects of the chromosomal bands on the nearest-neighbor analysis, we performed a second set of simulations, which included simu-

lated chromosomal bands. Simulated data were generated by placing a highly blurred image of the experimental data (blurred by a Gaussian with standard deviation of 200 nm) within the cell outline. This blurred image corresponded to a probability distribution that mirrored the large-scale experimental RNAP structure (bands), without containing information on small-scale structure (clustering). Simulated molecular positions were drawn from this distribution. The χ^2 test was repeated using the modified simulation, giving two sets of results No banding and Banding (Fig. S2 C).

Fluorescence recovery after photobleaching (FRAP)

DJ2711 was grown at 32°C overnight in M9 + 0.2% glucose, diluted 1:100 in the same media, and then grown to OD₆₀₀ = 0.1. Agarose pads for imaging were prepared by placing 80 μL of M9 + 0.2% glucose + 3% low melting point agarose between two double strips of autoclave tape (placed on a slide 1 cm apart) and a coverslip. Pads were allowed to set for 5 min before 0.5 μL of cells were sandwiched between the pad and coverslip. Cells were then imaged on an Olympus FV1000 confocal microscope (Olympus, Japan) with a heating stage set to 33°C and an objective heater at 36.5°C. The temperature of the slide was allowed to equilibrate for >1 h, allowing the cells to divide at least once before being imaged. FRAP on the strains MGTRY and KF32 was performed similarly, except that the media contained 0.2% glycerol instead of 0.2% glucose, and arabinose was added to 0.5% after preculture to activate the pBAD promoter. The following imaging settings were used: 501 μm pinhole, 620 V gain, 1% offset, 20 μs/pixel, pixel size = 124 nm × 124 nm, 515 nm laser power = 0.5%. The objective was a UPLSAPO 100X (NA = 1.40). A circular area of 0.6 μm² was photobleached at 100% laser power. The bleaching region was always positioned near the pole of a cell. Cells were always inspected using bright field (using differential interference contrast microscopy) to ensure that no semicompleted septum was present. Images were typically acquired at 200 ms intervals for 40 s. The fluorescence recovery curve, $R(t)$, was created using Fluoview 2.0 (Olympus, Japan) and Excel (Microsoft). First, background was subtracted using an empty region of the image as a reference. Four values were then calculated: I_t^W (the intensity of the whole cell before photobleaching), I_t^P (the intensity of bleaching area before photobleaching), $I_c^W(t)$ (the intensity of the whole cell as a function of time), and $I_c^P(t)$ (the intensity of the bleaching area with time). Photobleaching even a small area also photobleached a substantial fraction of the fluorescent protein in the cell. Therefore, $I_c^P(t)$ was divided by $I_c^W(t)$ for each different time t . This adjustment accounted for loss of fluorescence due to the initial concentrated bleach, as well as subsequent bleaching resulting from imaging. The entire time series of FRAP data was then divided by the normalized initial intensity of the bleached area, I_t^P/I_t^W . This adjustment ensured that the recovery curve would start at, and recover to, a value of 1. In other words, $R(t) = (I_c^P(t)/I_c^W(t))/(I_t^P/I_t^W)$. The $R(t)$ curves from many different cells were then averaged together to produce the FRAP curves.

RESULTS

To monitor the location of RNAP, we tagged one of the large RNAP subunits, β' , with PAmCherry1 by replacing the chromosomal β' gene with a β' -PAmCherry1 C-terminal fusion. As the β' subunit is incorporated into transcriptionally competent RNAP holoenzyme within 2–5 min (28), and as PAmCherry1 fluorophores mature in ~25 min (21), our localizations should then reflect the cellular distribution of core and holoenzyme RNAP. Using Western blots, we established that the FP is incorporated quantitatively; >90% of the β' subunit had a reduced mobility consistent with the added FP tag (Fig. S3 A). Bacteria carrying the

β' -PAmCherry1 gene had the same doubling time as the WT parent in both rich (i.e., LB) and minimal media (i.e., M9 + 0.2% glucose, which we refer to hereafter as M9). (Doubling times in LB were 31.4 ± 0.6 min for WT MG1655 and 31.2 ± 1.6 for the rpoC-PAmCherry1 strain; in M9, they were 102 ± 2 min for MG1655 and 100 ± 1 min for the rpoC-PAmCherry1 strain (Fig. S3 B).) This indicates that β' -PAmCherry1 is fully functional, and that C-terminal β' fusions are well tolerated in *E. coli*, as shown previously (29).

Numbers of RNAP in bacterial cells determined by super-resolution imaging

E. coli expressing RNAP-PAmCherry1 were grown in LB or M9, fixed with formaldehyde, and imaged using PALM. The fluorophore PAmCherry1 was chosen over mEos2 as most photoactivated PAmCherry1 fluorophores photobleached irreversibly in one step, whereas mEos2 exhibited blinking and apparent clustering (see Fig. S2) (30). Fluorophores, which remained active for more than one frame, were grouped into a single localization event (see Methods) (22); then, each event reports on a single photoactivated FP, and allows numbers of molecules of RNAP to be counted. The density of photoactivated fluorophores per field of view was kept low (<2 localizations per frame, Fig. S4 A) to avoid more than one being found within one diffraction-limited area; imaging continued until no more photoactivation occurred (Fig. S4 A). Raw data were post-processed by blurring images according to the experimental localization precision (see Methods).

For cells growing rapidly in LB, we detected an average of 3685 ± 1275 RNAP molecules per cell (Fig. 1, 22 cells); the high standard deviation primarily reflects cell-to-cell heterogeneity (the unsynchronized cells had lengths of 4–11 μm), and not experimental noise (background counts were low, ~ 10 per μm^2 ; Fig. S4 B). For cells growing in M9, we detected 1430 ± 502 RNAPs per cell (24 cells);

this reduction is consistent with the cells being smaller, and numbers scaling with cell volume (Fig. S5 A). The molecular density (in the plane of the two-dimensional (2D) projection) is similar in the two media: 706 ± 119 RNAPs per μm^2 in LB, and 687 ± 76 in M9.

The numbers of RNAPs detected were similar to previous estimates of total cellular RNAP obtained using quantitative Western blotting: MG1655 grown in M9 (at the slightly lower temperature of 30°C) contain 2600 ± 1300 RNAPs/cell (31), and in LB (at 37°C) 2600 ± 300 (32)—values that are within our margin of error. Even so, we believe our values are best seen as lower bounds, as incomplete fluorophore maturation, fixation (Fig. S4 B) and premature photobleaching will probably ensure that some PAmCherry1 molecules go undetected (see Discussion).

RNAP localizes in submicron-sized bands in rich media

In M9, RNAP is spread fairly uniformly throughout the cell (presumably occupying a single nucleoid or—just before cell division—the two nucleoids of future daughters; Fig. 1 B). In contrast, in LB, it is concentrated in 2–8 submicron-sized bands depending on cell size (Fig. 1 A). These bands are regularly spaced but irregularly shaped, and—if approximated as a disk—have a diameter of 500–1000 nm and contain 714 ± 198 detected RNAPs (Fig. S5 B).

For further analysis, we related band number to cell length. In LB, we identified four major groups: short cells with two relatively diffuse bands and no septum, but an area of RNAP exclusion at the mid-cell (Fig. 2 Ai); medium-sized cells with four clear bands and no septum (Fig. 2 Aii); long cells with four broader bands and a septum (Fig. 2 Aiii); and very long cells with 6–8 bands and a septum (Fig. 2 Aiv). Thus, the number of bands scaled with cell length, and so cell-cycle stage (Fig. 2 B). (Long cells with a fully formed septum (group 4; Fig. 2 B) may represent two daughter cells that have yet to separate completely.)

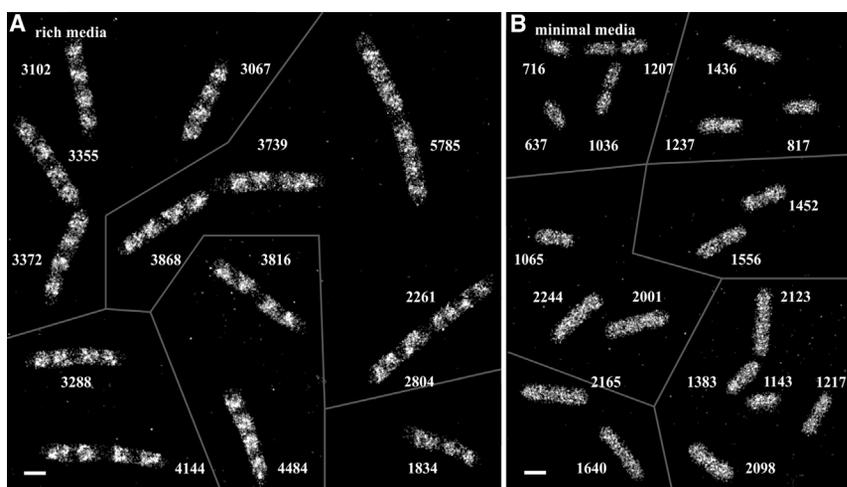


FIGURE 1 Super-resolution map of RNAP distribution in *E. coli*. *E. coli* strain KF26, in which all RNAP β' subunits are tagged with PAmCherry1, was grown at 32°C in (A) LB (a rich media) or (B) M9 (a minimal media) to $\text{OD}_{600} = 0.4$, fixed, and imaged using PALM. Numbers indicate independent molecular localizations per cell. Each panel contains several different images demarcated by gray lines (bars: 1 μm).

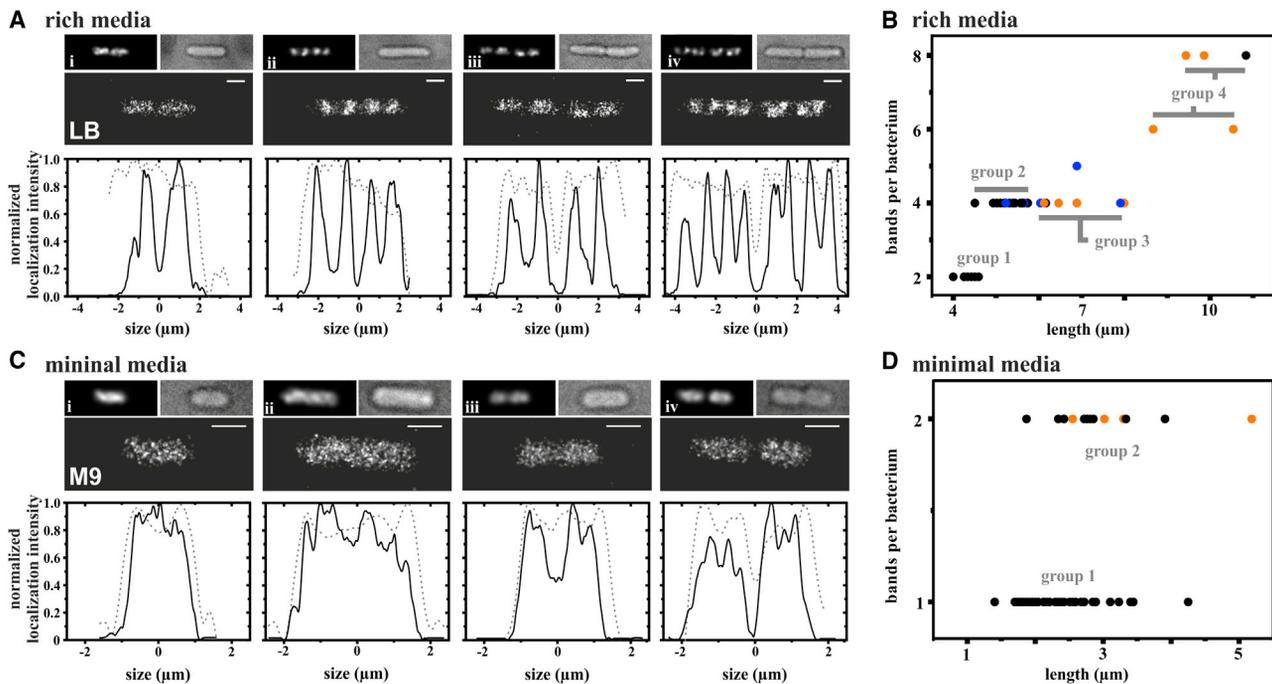


FIGURE 2 RNAP molecules are found in bands. *E. coli* cells expressing β' -PAmCherry1 were grown in LB (A and B) or M9 media (C and D), fixed, and imaged using PALM and bright field microscopy. (A) Three views of four fields are shown (largest images – PALM; top-left image of pair – bright field image; top-right image of pair – reconstruction of a conventional fluorescence image using super-resolution coordinate lists obtained by PALM) above the corresponding longitudinal RNAP localization densities (determined using PALM, solid black line) and intensity profiles (determined using bright field image, dashed gray line) smoothed using a Savitzky-Golay filter. Fields *i–iv* include cells of different length (bar: 500 nm). (B) Number of bands per cell as a function of cell length. Cells lacking a septum are shown in black; cells possessing initiating or fully formed septa are shown in blue and orange, respectively. (C and D) Analysis as in A and B for cells grown in minimal medium.

Clustering of RNAP at the nanoscale

Next, we analyzed the clustering of PAmCherry1-RNAP within the bands seen in LB (bright spots in Figs. 1 A, Fig. 3 Aiii). Superficially, subband clusters had variable shapes and sizes (diameters of 50–300 nm). Notably, few such clusters are visible in M9 (Fig. 3 Aii). To characterize these clusters objectively and quantitatively, we used the DBSCAN algorithm (density-based spatial clustering of applications with noise (33);) which detects clusters based on the local density of points within a search radius (see Methods). To confirm that DBSCAN reports on cluster density and shape, we analyzed simulated images of cells with or without clusters like those seen experimentally; DBSCAN extracted clustering features reliably (Fig. S6). We also applied DBSCAN to cells grown in LB expressing cytosolic PAmCherry1 as a negative control, using cells with localization densities similar to PAmCherry1-RNAP (M9: 687 ± 76 localizations/ μm^2 ; LB: 706 ± 119 localizations/ μm^2 ; PAmCherry1 control in LB: 619 ± 204 localizations/ μm^2 ; see Methods and Table S1).

Cells expressing free PAmCherry1 showed only a small amount of clustering (Fig. 3 Ai, bottom; Fig. S7): single molecules appeared either unclustered (black crosses in Fig. 3 Ai), or in small clusters (i.e., clusters with 5–20 localizations; set of colored circles in Fig. 3 Ai) that could result

from random coincidence of unclustered molecules. This distribution was also reflected in the cluster-frequency histogram (Fig. 3 Bi), which showed a high-amplitude peak of small clusters with 5–10 localizations per cluster (due to random coincidence) followed by a rapid decay to zero amplitude at larger cluster size. In contrast, RNAP in M9 showed several mid-size clusters with up to 86 localizations (Fig. 3 Aii), and the corresponding histogram (Fig. 3 Bii; Fig. S7) confirmed the presence of clusters with ~ 35 localizations (hereafter C35 clusters), which were absent in the negative control (free PAmCherry1; black line); notably, there were few larger clusters with >100 localizations. The C35 clusters are unlikely to be due to density variation between the RNAP in M9 and the control, because the mean molecular densities (above) are similar. Nevertheless, their low density means both that some could arise from random molecular coincidence, and that some might go undetected; as a result, these measurements give an upper bound to the size of RNAP clusters in M9.

Clustering was more apparent in cells grown in LB, where several large clusters with >100 localizations were observed (e.g., see cell with clusters with up to 372 localizations; Fig. 3 Aiii; Fig. S7). The cluster-frequency histogram (Fig. 3 Biii) shows two new populations absent in the negative control or M9 cells: i), clusters with a mean of ~ 70

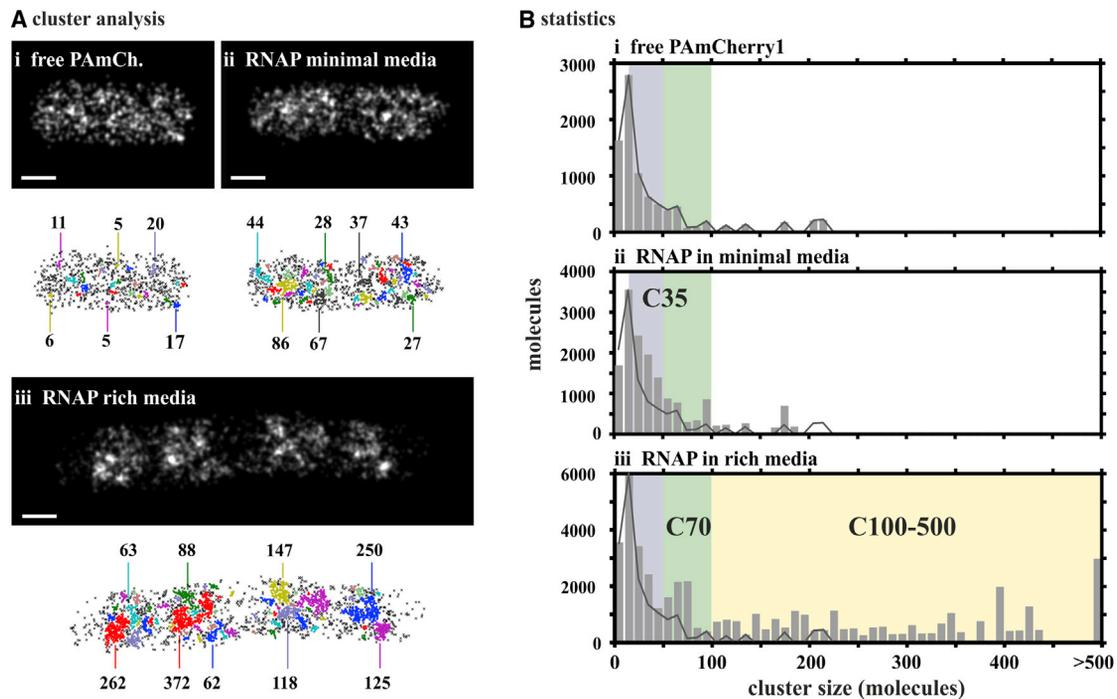


FIGURE 3 Clustering analysis. *E. coli* expressing free PAmCherry1 (i) or β' -PAmCherry1 (ii, iii) were grown in rich LB (i, iii) or minimal M9 media (ii), fixed, and imaged using PALM, and the resulting localization lists subjected to DBSCAN cluster analysis. (A) PALM images are shown above localizations plotted as crosses (crosses in one cluster share the same randomly chosen color). Bars: 500 nm. (B) Sizes of clusters observed in cells in panel A. For ease of comparison, the solid gray line shows the frequency distribution for free PAmCherry1 (i), normalized to the maximum frequency in each panel. Blue, green, and yellow backgrounds identify regions containing clustered molecules.

localizations (hereafter C70 clusters) with a radius of 82 ± 15 nm (the number of C70 and larger clusters scales with cell length; Fig. S8 A), and ii), an extended, slowly decaying, tail of clusters with 100–500 localizations (hereafter C100+ clusters) that probably arise from the overlap of smaller clusters. Such overlap may result trivially from the 2D projection and limitations of DBSCAN (e.g., C372 and C250 in Fig. 3 Aiii appear asymmetric and elongated, and are probably categorized as such large structures simply because of the fortuitous presence of high densities in the connections between smaller clusters), but it may also reflect an underlying physical interaction between smaller clusters.

About 35% RNAP in LB are present in either C70 or C100+ clusters. If we assume that all C100+ clusters arise from overlap of C70 clusters (see Discussion), each chromosomal band would then contain ~ 6 C70 clusters. We note that this estimate is independent of detection efficiency and background. A nearest-neighbor distribution analysis further supports the finding that RNAP-PAmCherry1 is significantly more clustered in LB than cytosolic PAmCherry1 (Methods, Fig. S2). Reducing transcription—either by inducing the stringent response with 1 mg/ml L-serine hydroxamate for 1 h, or addition of 50 $\mu\text{g/ml}$ rifampicin for 120 min—also prevents the formation of large RNAP clusters (Fig. S9).

FRAP analysis shows that most RNAPs are bound to DNA

We also sought to determine what fraction of the RNAPs corresponded to engaged RNAPs. This question has been previously answered using FRAP: in *E. coli* grown in rich media, $>50\%$ of RNAP are immobile for >5 s, and so are likely engaged (10). Yet the fraction of RNAP that are active in cells grown in minimal media remained unknown. To estimate this number, we performed FRAP using M9-grown cells that expressed a bright and photostable RNAP-yellow fluorescence protein (YFP) fusion (Fig. 4). Cells were grown for at least one cell division on the coverslip, cells with signs of septation excluded, and one pole of a cell photobleached. Because cell size in minimal media (2–3 μm length, 0.6 μm diameter) is not much larger than the width of the focused laser beam used (~ 0.3 μm), ~ 30 – 40% of the total RNAP-YFP is bleached (see Fig. 4 A). Subsequently, fluorescence recovery (i.e., redistribution) was complete, but very slow with only a small fast component (see brown lines in Figs. 4, B and C). Specifically, the bleached area recovered $\sim 20\%$ of its final fluorescence during the first 5 s, $\sim 50\%$ in 25 s, and $>90\%$ after 100 s. In contrast, the recovery of an YFP-fusion of the tetracycline repressor (TetR; a sequence-specific transcription factor that also interacts with DNA nonspecifically) was much

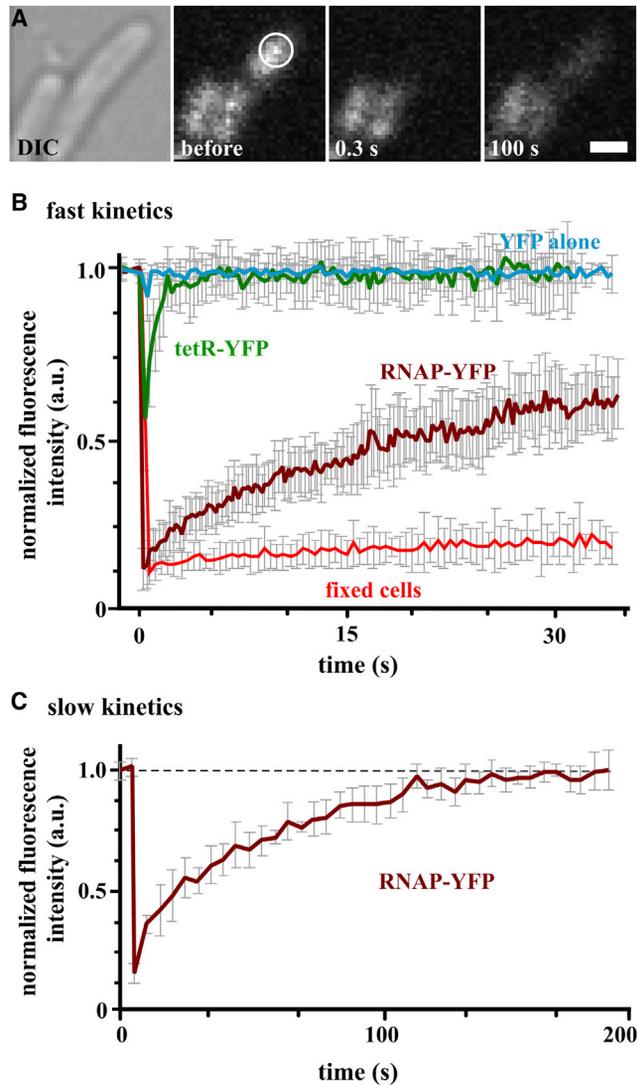


FIGURE 4 FRAP analysis of RNAP mobility. Live cells expressing either the β' -subunit of RNAP fused to YFP, the tet-repressor fused to YFP, or YFP alone were allowed to divide once on an M9-agarose pad at 32°C before the YFP molecules in an area of interest were photobleached, and the rate at which unbleached molecules migrated into the area measured using a confocal microscope. (A) Differential contrast and fluorescence images obtained at different times during a typical FRAP experiment using cells expressing β' -YFP (bar: 1 μm). The area of interest is indicated by a white circle. (B and C) Fast and slow recovery of fluorescence signal in the area of interest normalized relative to that in the whole cell (\pm SD; $n \geq 10$ cells). Fixed cells contained β' -YFP immobilized using formaldehyde before photobleaching.

faster, recovering 50% of the final signal in 0.3 s and >90% in 1.5 s (see *green line* in Fig. 4 B). As expected, both RNAP-YFP and TetR-YFP recover much more slowly than free YFP (>90% recovery in 0.3 s; see *blue line* in Fig. 4 B).

The slow recovery of RNAP indicates that most molecules in the unbleached region are not reaching the bleached area quickly, most likely due to stable association with the chromosome. Intriguingly, the ~100 s needed for full recov-

ery is similar to the time required to transcribe a typical transcription unit. If we assume that initiation takes 1–60 s, and RNAPs are active on transcription units of 0.5–4 kbp (34), while transcribing at a rate of 25–50 bp/s (35) (36), RNAPs should then be immobilized for 11–220 s. The rapidly exchanging fraction of 20% probably corresponds to free cytoplasmic RNAP, as well as RNAP bound nonspecifically and transiently to nucleoid DNA. Our FRAP results in minimal media, combined with published work in rich media (10), establish that 50–80% of RNAP is bound stably to DNA at any moment, and this fraction is likely to be transcriptionally active.

DISCUSSION

Using PALM with a resolution an order of magnitude higher than conventional fluorescence microscopy, we generated detailed maps of RNAP in fixed *E. coli* and determined the minimum number of molecules present in different sub-cellular regions. The possibility that RNAP redistributes during fixation seems unlikely, as several studies using conventional microscopy reveal a similar banded distribution in live cells (10,11). Important issues when imaging any multi-subunit protein are whether the fluorophore seen is present as an individual subunit or in a fully assembled complex, and how the timescale for complex assembly compares to the maturation time of individual FPs. For PAmCherry1, the maturation time of 23 min (21) is longer than the 2–5 min taken to assemble β' into RNAP (28), so we mainly localize complexes at the various phases of the transcription cycle (see also (10), where enhanced green fluorescence protein was used to tag β'). We expect that fixation, fast photobleaching, and the long maturation time of PAmCherry1 probably led us to miss some RNAPs, and speculate that the use of a brighter, more photostable FP with shorter maturation time and better resistance to fixation will increase detection efficiency. Nevertheless, the agreement between our RNAP counts with those previously obtained using quantitative Western blotting (albeit with slightly different temperatures and media) suggests that our detection efficiency is not unreasonably low.

Most RNAP molecules stably associate with DNA in vivo

FRAP shows that cells growing in M9 contain fast- and slow-exchanging fractions of ~20% and ~80%, respectively. This is consistent with results indicating that ~25% RNAP is bound transiently (and so nonspecifically) either as the holoenzyme or core polymerase, and the rest more tightly (which includes 50% that is elongating plus ~25% at the promoter and/or paused (37)). It has been suggested that the low concentration of free RNAP limits transcription (28), and our results are consistent with this (but do not prove it). An analogous study (10) of RNAP using a

moderately rich medium (EZRDM) found that ~50% recovered in <5 s and 50% more slowly (i.e., up to 30 s); the slight differences seen probably reflect different growth rates, nucleoid compactness, number of genes expressed, and number of RNAPs/cell (31).

RNAP occupancy of chromosomal bands in LB

In LB, RNAP is confined to submicron-sized bands (Fig. 1), as reported by Bakshi et al. in moderately rich media (38). Such bands were not seen by Cabrera et al. (11) in their early diffraction-limited study of β' -GFP (10), although they probably underlie the transcription foci that were seen. Because the number and location of bands changes during the cell cycle, we speculate that each band represents a single chromosome loaded with RNAPs.

If each band does, indeed, correspond to a segregating genome, there is good reason to believe that the bands formed by RNAPs might be narrower than the underlying chromosomes. Thus, within each segregating genome the genomic distance of a locus from the *ori* will dictate its position along the cellular axis, with proximity to the *ori* corresponding to proximity to the poles (39). Because—in LB—most active RNAPs are transcribing *oriC*-proximal rRNA operons (40), the *ori*-proximal (i.e., pole-proximal) half of the chromosome should have a higher concentration of RNAPs than the other half. This would lead to successive bands of high and low RNAP intensity, as we see. However, this is only one plausible explanation for the distribution, and further studies will be required to determine the cause of banding.

RNAP occupies two types of large clusters in LB cells: single and multiple *rrn* operons?

We also analyzed RNAP clustering; this is challenging due to the presence of a pool of exchangeable polymerase (25–50%), and the artificial merging of some clusters due to 2D projections of three-dimensional objects. Nevertheless, C70 and C100+ clusters are clearly seen in LB, and some C35 clusters in M9 (Fig. 3) (Fig. 4).

Several lines of evidence suggest the C70 clusters in LB correspond to individual *rrn* operons. They contain at least 70 RNAPs, close to the 80–90 observed by electron microscopy on one single *rrn* operon (41). Moreover, there are ~6 C70 clusters per chromosome (when overlapping clusters are included), which is also close to the expected number of *rrn* operons (i.e., 7–21 (35)). Our observation that reducing *rrn* transcription (by inducing the stringent response, which redistributes RNAPs to other genes) abolishes the C70 clusters is consistent with this. Finally, because ~70% of active RNAPs are closely packed on the *rrn* operons (35), one might expect to see foci representing those operons. (The other 30% are found on non-*rrn* operons, each loaded with one—or few—RNAPs (42).) How-

ever, ultimately, we can only speculate as to what underlies the C70 clusters, and further experiments will be required to answer this question definitively.

If C70 clusters do, indeed, represent isolated *rrn* operons, this allows us to estimate their compaction and size in the cell. Although the 6000-bp operon appears as a linear 2- μ m structure in electron micrographs of Miller spreads, our data suggest that it is folded into a sphere of ~160 nm diameter. Modeling each RNAP as a 16-nm sphere (43), and including the volume occupied by 6000 bp B-DNA, ~8% of the C70 volume would be occupied by RNAP and DNA (Fig. 5). Although this compaction ratio is high, it still leaves enough space for the nascent RNA and associated processing machinery.

In LB, a substantial fraction of RNAPs are found in C100+ clusters, each containing 150–800 molecules; these are likely to correspond to multiple *rrn* operons, with the largest containing 4–7. Operons may appear clustered for trivial reasons, as discussed previously. Alternatively, they may truly lie close together, perhaps in a subcellular structure analogous to the eukaryotic nucleolus—which contains the many different molecular machines required both to make rRNAs (often encoded on different chromosomes) and assemble a mature ribosome (44). Although the existence of a bacterial nucleolus had been suggested (11), additional and direct evidence for the clustering of *rrn* operons will be necessary for definitive conclusions. However, our results do exclude models that require all *rrn* operons to localize within such a structure, because we do observe isolated C70 clusters.

Small RNAP clusters in M9

In M9, we observe C35 clusters that contain ~35 RNAPs. In such slowly growing cells, there are now only ~5 RNAPs per *rrn* operon (as most RNAPs are found on protein-coding genes (35)), and many cells in the population contain two origins (35) and 11 *rrn* operons (four lie close to the origin); then, these C35 clusters could also represent nucleoli. However, some other unknown structure might underpin their organization, as equivalent C35 clusters are found in LB.

Extending the study to living cells

We have seen that various factors combine to make accurate counting and mapping of all RNAPs in a fixed cell difficult, including the poor resolution of light microscopy, the shortcomings of existing FPs (slow maturation (21), fast photobleaching, and poor signal/noise ratios (10)), and difficulties in analyzing clustering in crowded images. Extending such studies to living cells introduces at least two more problems. First, a live cell contains many different populations of RNAPs with varying mobilities (freely diffusing, nonspecifically bound, cycling at the promoter, elongating, paused, and terminating) and it is difficult to simultaneously

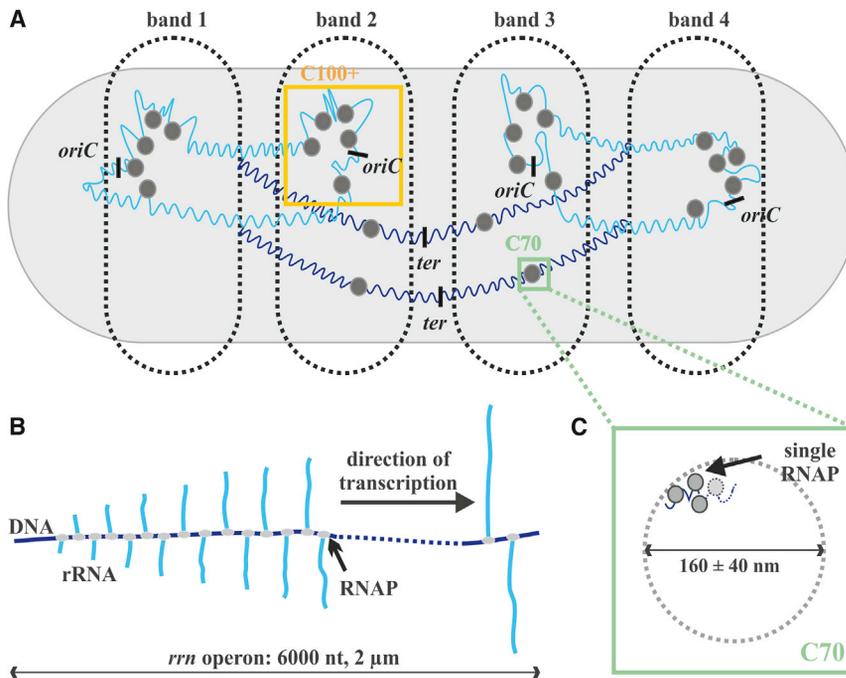


FIGURE 5 Model for the spatial organization of RNAP during growth in rich media. (A) RNAPs are found in C70 clusters (representing single *rrn* operons) and larger C100+ clusters (multiple *rrn* operons). (B) Schematic of an electron micrograph showing a surface-deposited single *rrn* operon of 6000 nucleotides ($\sim 2 \mu\text{m}$) transcribed by 70–80 RNAPs (41). (C) Magnified region in (A) depicting a single C70 cluster, interpreted as a single *rrn* operon in which >70 active RNAPs are tightly packed to occupy (along with the DNA) $\sim 8\%$ of the overall cluster volume.

localize all to within a few nanometers using current techniques. Second, maintaining the same environmental conditions during imaging is challenging (5), especially when depositing rapidly growing cells onto a microscope slide probably induces the stringent response and immediate RNAP redistribution. Such reasons probably underlie why recent studies in living *E. coli* found only 516 and 658 molecules, respectively, of the β (*rpoB*) and ω (*rpoZ*) subunits (45)—values several-fold lower than reported in the literature, and no RNAP clustering (38)—probably because image acquisition took several minutes (when a cell might grow $\sim 200 \text{ nm}$ as an RNAP transcribes several operons). However, we hope our quantitative analysis of the fine structures found in cells fixed after growth at different rates will lay the foundations for higher resolution analyses of all RNAPs in live cells.

SUPPORTING MATERIAL

Supporting methods, two tables, nine figures, and reference (46) are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(13\)00635-8](http://www.biophysj.org/biophysj/supplemental/S0006-3495(13)00635-8).

M.H., U.E., and K.F. are grateful for funding by the Bundesministerium für Bildung und Forschung (BMBF), grant nr. 0315262. K.F. was also supported by the E.P. Abraham Trust, a Clarendon Fund award from the University of Oxford, and an Overseas Research Student Award from the UK government. P.R.C. is E.P. Abraham Professor of Cell Biology and a Professorial Fellow of Lincoln College. A.N.K. and S.J.H. were supported by the European Commission Seventh Framework Program (grant FP7/2007-2013 HEALTH-F4-2008-201418), the Biotechnology and Biological Research Council (grant BB/H01795X/1), and the European Research Council (starter grant 261227). S.J.H. was also supported by a Marie Curie Intra-European Fellowship (grant 297918). We thank Rodrigo Reyes and David

Sherratt (Oxford University) for their gift of the P1 phage, Vladislav Verkhusha (Albert Einstein College of Medicine) for the PAmCherry1 gene, Don Court (NCI Center for Cancer Research) for the DY330 strain, and Ding Jin (NCI Center for Cancer Research) for strain DJ2711.

The authors declare they have no conflict of interest.

REFERENCES

- Cseresnyes, Z., U. Schwarz, and C. M. Green. 2009. Analysis of replication factories in human cells by super-resolution light microscopy. *BMC Cell Biol.* 10:88.
- Hozák, P., A. B. Hassan, ..., P. R. Cook. 1993. Visualization of replication factories attached to nucleoskeleton. *Cell.* 73:361–373.
- Cook, P. R. 1999. The organization of replication and transcription. *Science.* 284:1790–1795.
- Jackson, D. A., F. J. Iborra, ..., P. R. Cook. 1998. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol. Biol. Cell.* 9:1523–1536.
- Lyle, J. M., E. Bullitt, ..., K. Kirkegaard. 2002. Visualization and functional analysis of RNA-dependent RNA polymerase lattices. *Science.* 296:2218–2222.
- Muñoz-Espín, D., R. Daniel, ..., M. Salas. 2009. The actin-like MreB cytoskeleton organizes viral DNA replication in bacteria. *Proc. Natl. Acad. Sci. USA.* 106:13347–13352.
- Rust, M. J., M. Bates, and X. Zhuang. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods.* 3:793–795.
- Kruse, T., B. Blagoev, ..., K. Gerdes. 2006. Actin homolog MreB and RNA polymerase interact and are both required for chromosome segregation in *Escherichia coli*. *Genes Dev.* 20:113–124.
- Malkusch, S., U. Endesfelder, ..., M. Heilemann. 2012. Coordinate-based colocalization analysis of single-molecule localization microscopy data. *Histochem. Cell Biol.* 137:1–10.
- Bratton, B. P., R. A. Mooney, and J. C. Weisshaar. 2011. Spatial distribution and diffusive motion of RNA polymerase in live *Escherichia coli*. *J. Bacteriol.* 193:5138–5146.

11. Cabrera, J. E., and D. J. Jin. 2003. The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. *Mol. Microbiol.* 50:1493–1505.
12. Shaner, S. L., D. M. Piatt, ..., M. T. Record, Jr. 1982. Aggregation equilibria of *Escherichia coli* RNA polymerase: evidence for anion-linked conformational transitions in the promoters of core and holoenzyme. *Biochemistry.* 21:5539–5551.
13. Deng, S., R. A. Stein, and N. P. Higgins. 2004. Transcription-induced barriers to supercoil diffusion in the *Salmonella typhimurium* chromosome. *Proc. Natl. Acad. Sci. USA.* 101:3398–3403.
14. Marenduzzo, D., I. Faro-Trindade, and P. R. Cook. 2007. What are the molecular ties that maintain genomic loops? *Trends Genet.* 23:126–133.
15. Eltsov, M., and B. Zuber. 2006. Transmission electron microscopy of the bacterial nucleoid. *J. Struct. Biol.* 156:246–254.
16. Dürrenberger, M., M. A. Bjornsti, ..., E. Kellenberger. 1988. Intracellular location of the histonelike protein HU in *Escherichia coli*. *J. Bacteriol.* 170:4757–4768.
17. Betzig, E., G. H. Patterson, ..., H. F. Hess. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. *Science.* 313:1642–1645.
18. Greenfield, D., A. L. McEvoy, ..., J. Liphardt. 2009. Self-organization of the *Escherichia coli* chemotaxis network imaged with super-resolution light microscopy. *PLoS Biol.* 7:e1000137.
19. Sharan, S. K., L. C. Thomason, ..., D. L. Court. 2009. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* 4:206–223.
20. McKinney, S. A., C. S. Murphy, ..., L. L. Looger. 2009. A bright and photostable photoconvertible fluorescent protein. *Nat. Methods.* 6:131–133.
21. Subach, F. V., G. H. Patterson, ..., V. V. Verkhusha. 2009. Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nat. Methods.* 6:153–159.
22. Endesfelder, U., S. Malkusch, ..., M. Heilemann. 2011. Chemically induced photoswitching of fluorescent probes—a general concept for super-resolution microscopy. *Molecules.* 16:3106–3118.
23. Tokunaga, M., N. Imamoto, and K. Sakata-Sogawa. 2008. Highly inclined thin illumination enables clear single-molecule imaging in cells (vol 5, pg 159, 2008). *Nat. Methods.* 5:159–161.
24. Wolter, S., M. Schüttelz, ..., M. Sauer. 2010. Real-time computation of subdiffraction-resolution fluorescence images. *J. Microsc.* 237:12–22.
25. Lando, D., U. Endesfelder, ..., E. D. Laue. 2012. Quantitative single-molecule microscopy reveals that CENP-ACnp1 deposition occurs during G2 in fission yeast. *Open Biol.* 2:120078.
26. Bates, M., B. Huang, ..., X. Zhuang. 2007. Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Science.* 317:1749–1753.
27. Stoyan, D. P., and A. Penttinen. 2000. Recent applications of point process methods in forestry statistics. *Stat. Sci.* 15:61–78.
28. Shepherd, N., P. Dennis, and H. Bremer. 2001. Cytoplasmic RNA polymerase in *Escherichia coli*. *J. Bacteriol.* 183:2527–2534.
29. Cabrera, J. E., and D. J. Jin. 2003. Construction, purification, and characterization of *Escherichia coli* RNA polymerases tagged with different fluorescent proteins. *Methods Enzymol.* 370(Pt C):3–10.
30. Annibale, P., M. Scarselli, ..., A. Radenovic. 2010. Photoactivatable fluorescent protein mEos2 displays repeated photoactivation after a long-lived dark state in the red photoconverted form. *J. Phys. Chem. Lett.* 1:1506–1510.
31. Grigorova, I. L., N. J. Phleger, ..., C. A. Gross. 2006. Insights into transcriptional regulation and σ competition from an equilibrium model of RNA polymerase binding to DNA. *Proc. Natl. Acad. Sci. USA.* 103:5332–5337.
32. Piper, S. E., J. E. Mitchell, ..., S. J. W. Busby. 2009. A global view of *Escherichia coli* Rsd protein and its interactions. *Mol. Biosyst.* 5:1943–1947.
33. Ester, M., H.-P. Kriegel, ..., X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. E. Simoudis, J. Han, and U. M. Fayyad, editors. AAAI-Press. 226–231.
34. Cho, B. K., K. Zengler, ..., B. O. Palsson. 2009. The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotechnol.* 27:1043–1049.
35. Bremer, H., and P. Dennis. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. F. C. Neidhardt, editor. American Society for Microbiology, Washington DC. 1553–1569.
36. Golding, I., J. Paulsson, ..., E. C. Cox. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell.* 123:1025–1036.
37. Record, M. T. R., W. S. Craig, ..., P. J. Schlax. 1996. *E. coli* RNA polymerase ($E\sigma^{70}$), promoters, and the kinetics of the steps of transcription initiation. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. F. C. Neidhardt, editor. American Society for Microbiology, Washington DC. 792–821.
38. Bakshi, S., A. Siryaporn, ..., J. C. Weisshaar. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol. Microbiol.* 85:21–38.
39. Sherratt, D. J. 2003. Bacterial chromosome dynamics. *Science.* 301:780–785.
40. Nomura, M. 1999. Engineering of bacterial ribosomes: replacement of all seven *Escherichia coli* rRNA operons by a single plasmid-encoded operon. *Proc. Natl. Acad. Sci. USA.* 96:1820–1822.
41. Miller, Jr., O. L., B. A. Hamkalo, and C. A. Thomas, Jr. 1970. Visualization of bacterial genes in action. *Science.* 169:392–395.
42. Hayashi, A., D. Q. Ding, ..., Y. Hiraoka. 2009. Localization of gene products using a chromosomally tagged GFP-fusion library in the fission yeast *Schizosaccharomyces pombe*. *Genes Cells.* 14:217–225.
43. Murakami, K. S., S. Masuda, and S. A. Darst. 2002. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science.* 296:1280–1284.
44. Boisvert, F. M., S. van Koningsbruggen, ..., A. I. Lamond. 2007. The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* 8:574–585.
45. Taniguchi, Y., P. J. Choi, ..., X. S. Xie. 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science.* 329:533–538.
46. Verheust, C., and D. R. Helinski. 2007. The incC korB region of RK2 repositions a mini-RK2 replicon in *Escherichia coli*. *Plasmid.* 58:195–204.

Title: **Multi-scale spatial organization of RNA polymerase in *Escherichia coli***

Running title: RNAP clustering in *E. coli*

Authors: Ulrike Endesfelder^{1,*}, Kieran Finan^{1,4,*}, Seamus J. Holden^{2,3,*}, Peter R. Cook⁴, Achillefs N. Kapanidis³ and Mike Heilemann¹

Addresses: ¹ Institute of Physical and Theoretical Chemistry
Johann Wolfgang Goethe-University
Max-von-Laue-Str. 7, 60438 Frankfurt, Germany

² Laboratory of Experimental Biophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

³ Department of Physics, Biological Physics Research Group,
Clarendon Laboratory, University of Oxford,
Parks Road, Oxford, OX1 3PU, UK

⁴ Sir William Dunn School of Pathology,
University of Oxford,
South Parks Road, Oxford, OX1 3RE, UK

*contributed equally

SUPPORTING MATERIAL

SUPPORTING METHODS

Choice of photoactivatable fluorescent protein. Some photoactivatable fluorescent proteins (FP) exhibit multiple switching cycles depending on redox properties (1, 2). This behavior is undesired as we intended to take advantage of the stoichiometric ratio between FP labels and RNAPs for quantification. We thus evaluated two FPs (PAmCherry1 and mEos2) by expressing each photoactivatable FP separately in *E. coli*. Looking for the presence or absence of a random spatial distribution, frequent multiple switching cycles should lead to apparent clustering, even for a free and monomeric photoactivatable FP. After FP expression, cell fixation, single-molecule imaging and high-precision localization, we generated super-resolved images and analyzed the spatial distribution of the localizations by comparing the experimental nearest-neighbor distribution with that expected for spatial randomness (Fig. S2). We found that mEos2 showed large deviations from spatial randomness (i.e., apparent clustering) caused by multiple switching cycles, whereas PAmCherry1 was randomly distributed (Fig. S2); this indicates that PAmCherry1 bleaches irreversibly after activation and does not dimerize or multimerize. We therefore chose PAmCherry1 for labeling RNAP.

Clustering analysis. To analyze quantitatively the distribution of cluster sizes and shapes, we performed clustering analysis using the DBSCAN algorithm (short for *density-based spatial clustering of applications with noise*, (3)). DBSCAN clustering is based on the density distribution of points in a dataset. Points with greater than *MinPts* neighbors within a radius ϵ are “density-connected” to their neighbors; regions of adjacent density-connected points are considered to be clustered. Points which are not density connected to any other points are considered to be unclustered “noise” points. *MinPts* was set to 4 as recommended (3).

Cluster analysis algorithms inevitably require the selection of a parameter that determines the sensitivity of the algorithm; in the case of DBSCAN, this critical parameter is ϵ . The quantity $\text{MinPts}/(\pi \cdot \epsilon^2)$ corresponds approximately to the minimum density for adjacent molecules to be considered a cluster. The value of ϵ must be sufficiently large that signal is not completely overwhelmed by noise; here this would result in almost the whole cell being classified as a single cluster. If ϵ is too small, real clusters are missed out, and incorrectly classified as background. The correct choice of ϵ is at a threshold density just large enough to minimize the amount of background particles erroneously grouped into clusters.

Because the mean densities of all three datasets (PAmCherry filtered, M9, LB, see below) were very similar, it was possible to meaningfully use the same value of ϵ for each dataset. This allowed us to make a choice of ϵ based on the PAmCherry control (assumed not to contain any real clusters), determining the maximum value of ϵ which would still yield only a small number of (false) clusters in

that dataset. We found that $\epsilon=30$ nm gave the best result in this regard; this corresponds to a minimum density of ~ 1400 molecules $/\mu\text{m}^2$, twice the mean density observed in the control sample. We emphasize that for clustering analysis using DBSCAN, we highly recommend use of an unclustered experimental control for accurate determination of ϵ .

To confirm that DBSCAN could accurately extract cluster information given the experimentally observed density of background molecules, we simulated cellular distributions containing various types of clusters (i.e., we distributed the localizations randomly or into clusters, while retaining nucleoid size and localization number, and a background density equal to that of the control). We found that using an ϵ value of 30 nm (Fig S6) we were able to accurately detect and quantify simulated clusters over a range including the experimentally detected clusters. Moreover, this same analysis detected only a few small clusters when applied to simulated data sets containing randomly-distributed localizations (Fig. S6).

We also performed the analysis on simulated data for a range of ϵ values (*data not shown*) which confirmed that $\epsilon=30$ nm was optimal for our data; larger values of ϵ increased the number of false clusters, smaller values of ϵ reduced the number of correctly detected clusters.

We investigated whether the variation in density in the LB dataset due to observed banding affected the accuracy of the cluster analysis. We performed simulations of bacteria containing clusters and background as above, but now distributed within simulated bands. The simulated bands were created using highly blurred images of the experimental data (using a Gaussian blurring filter of radius 400 nm). The presence of the bands within the simulated data did not significantly increase the number of erroneously detected clusters (Fig. S6B), nor did it affect accuracy in analysis of simulated clusters (Fig. S6C-D).

Cluster size analysis was carried out by fitting a 1D Gaussian model to peaks identified within the cluster-size histograms. The data was fitted by manually excluding all points outside the region of the peak, followed by ordinary least-squares minimization. The mean number of molecules per cluster was determined from the center of the fitted model, and the uncertainty in this value determined from the standard deviation of that model. The radius of each cluster was calculated as the radius of gyration of all points within the cluster. The mean and standard deviation of the cluster radius was calculated for all clusters within the region of an identified peak. The percentage of molecules in clusters was calculated from the sum of all molecules in clusters with a size greater than the lower edge of the identified peak; no upper bound on cluster size was used in this estimate, in order to include the contributions of unresolved overlapping clusters. The mean number of clusters per cell was calculated from the total number of molecules, the number of cells, and the percentage of clustered molecules; the error on the number of clusters per cell was calculated by error propagation from the uncertainty on the mean number of clusters per cell.

The density of PAmCherry1-RNAP in the M9 RNAP and LB RNAP datasets was similar, with 687 ± 76 molecules/ μm^2 and 705 ± 119 molecules/ μm^2 respectively. However, the cytosolic PAmCherry1 control showed a higher density and much greater spread in density between different cells (1053 ± 485 molecules/ μm^2). To compare the clustering analysis of the PAmCherry1-RNAP and cytosolic PAmCherry1 datasets, cells in the cytosolic PAmCherry1 dataset with a density higher than 1050 molecules/ μm^2 were excluded from clustering analysis. This filtered cytosolic PAmCherry1 dataset had a density of 619 ± 204 molecules/ μm^2 , similar to the PAmCherry1-RNAP data.

Table S1 summarises the observed densities for each dataset. Observed clusters in the LB RNAP dataset are unlikely to arise for the small variations in density compared to the M9 RNAP and filtered cytosolic PAmCherry datasets. Induction of the stringent response, or addition of rifampicin both largely abolished clustering (Fig. S9), despite introducing decreased or increased density respectively (Table S1), further showing that observed clustering was not affected by density variation between the different samples.

SUPPORTING TABLES

Table S1. Observed PAmCherry1 density under the different experimental conditions

Experimental condition	Density ($\mu \pm$ s. d.), mol/ μm^2
Cytosolic PAmCherry	1053 ± 485
Cytosolic PAmCherry, <i>post-hoc</i> filtered	619 ± 204
RNAP-PAmCherry, LB media	705 ± 119
RNAP-PAmCherry, M9 media	687 ± 76
RNAP-PAmCherry, LB media, stringent response	430 ± 95
RNAP-PAmCherry, LB media, rifampicin	650 ± 136

Table S2. Strain list and primers

Strain	Genotype	Source
MG1655	Wild-type E. coli K-12	Jin Laboratory
KF26	MG1655 <i>rpoC-pamcherry1 Amp^R</i>	This study
KF7-1	MG1655 <i>rpoC-meos2 Amp^R</i>	This study
DJ2711	MG1655 <i>rpoC-yfp Amp^R</i>	(8)
KF31	MG1655 pKIE12-1	This study
KF32	MG1655 pBAD33-Ypet-His	This study
KF33	MG1655 pKIE3-1	This study
KF34	MG1655 pBAD\HisB-PAmCherry1	This study
MGTRY	MG1655 tetR-YFP under the pBAD promoter	(9)
Primers	Sequence (5'-3')	
mCherry(rpoC)fw	CCAGCCTGGCAGAACTGCTGAACGCAGGTCTGGGCGGTTCTGATAAC GAGCTCGAGATAATGGTGAGCAAGGGCGAGGAGGATAAC	
mCherry(rpoC)rv	CCCCCATAAAAAAACC CGCCGAAGCGGGTTTTTACGTTATTTGCGG ATTAGTCTGACGCTCAGTGGAAC	
mEos2rpoCfw	CCAGCCTGGCAGAACTGCTGAACGCAGGTCTGGGCGGTTCTGATAAC GAGCTCGAGATAATGAGTGCGATTAAGCCAGACATGAAG	
mEos2rpoCrv	CCCCCATAAAAAAACC CGCCGAAGCGGGTTTTTACGTTATTTGCGG ATTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCACG	
mEos2ampfw	GGCCAGGAGTGAAACGATGAGTGCGATTAAGCCA GAC	
mEos2amprv	CCGGCCACCTTGGCCTTACTCGAGAGATCTTCGTCTGGCATTGTCAG GC	
pBAD33ampfw	TAAGGCCAAGGTGGCCGGTAC	
pBAD33amprv	CGTTTCACTCCTGGCCTTCGTGGCCG	
1stcherryfw	GCCAGGAGTGAAACGATGGTGAGCAAGGGCGAGGAGGATAAC	
1stcherryrv	GCCAGACGCGCTACCCTTGTACAGCTCGTCCATGCCG	
2ndcherryfw	GGTAGCGCGTCTGGCATGGTGAGCAAGGGCGAGGAGGATAAC	
2ndcherryrv	GTGGTGGTGGTGGTGTACTTGTACAGCTCGTCCATGCCG	
pBAD33ampfw2	CACCACCACCACCACCTAAGGC	

SUPPORTING FIGURES S1-S8

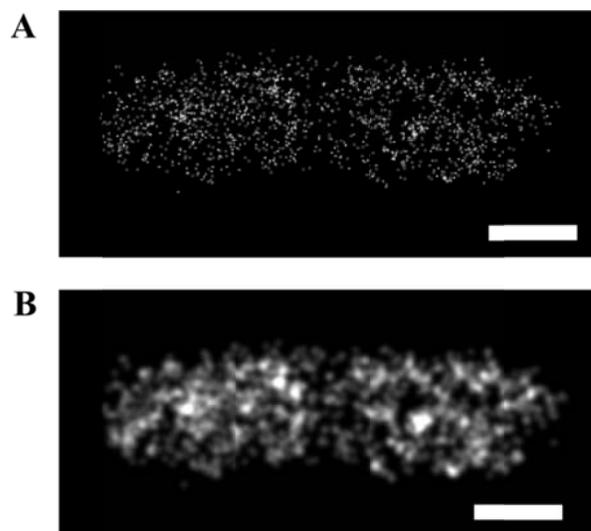


Fig. S1. Graphical representation of the results of PALM imaging.

A. PALM image of RNAP-PAmCherry1 in minimal media, rendered using Thompson blurring (blurring for each localization represented as a 2D Gaussian with standard deviation reflecting the localization precision). Note that, due to the unstructured nature of the data, Thompson-blurred images are difficult to interpret visually.

B. PALM image of RNAP-PAmCherry1 data shown in A, rendered by blurring all localizations with a 2D Gaussian with standard deviation of 20 nm. The increased blurring of the image facilitates visual analysis of the data. Note that all quantitative analysis was carried out on the list of RNAP localization coordinates, not the processed images, which are presented only for the purposes of visualization. Scale bar: 500 nm.

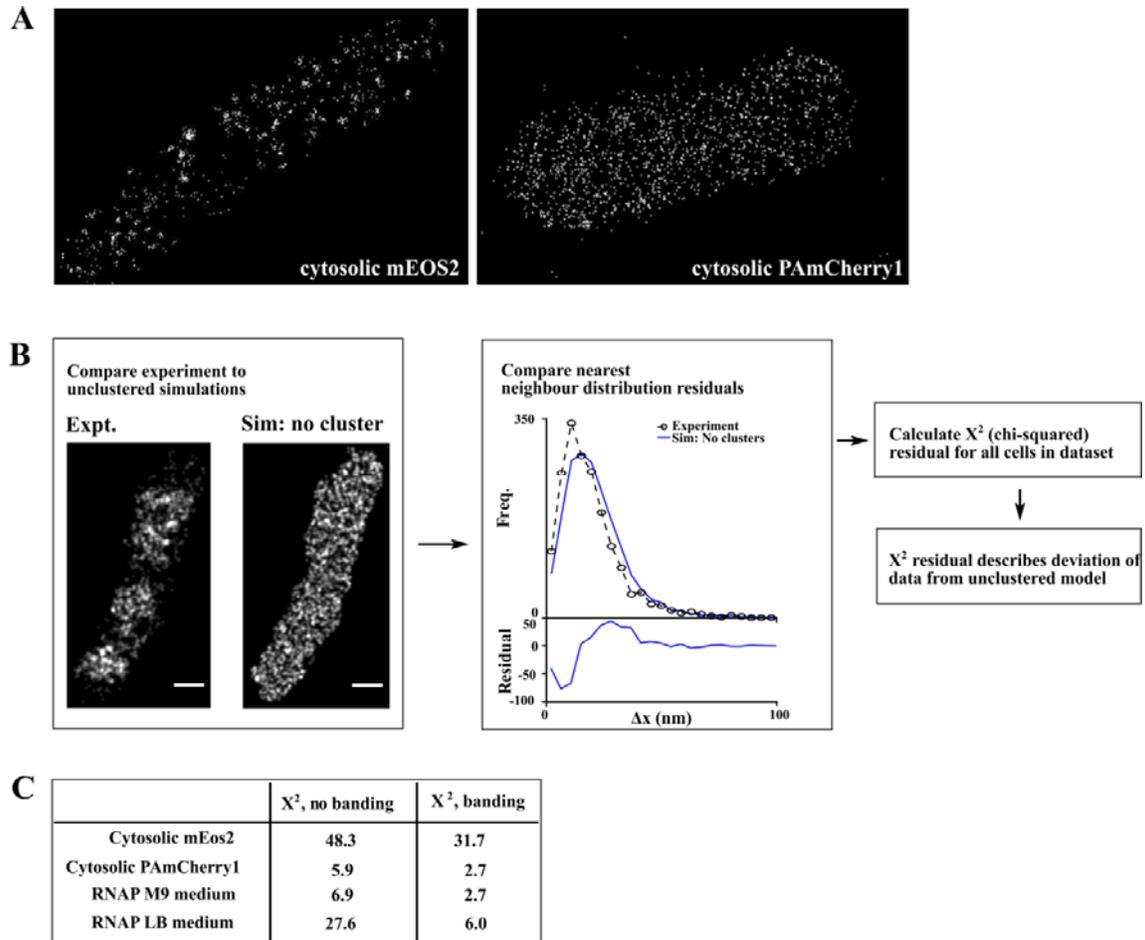


Fig. S2. χ^2 testing for spatial randomness. Bars: 500 nm.

A. Super-resolution images of *E.coli* expressing free mEos2 (left) or PAmCherry1 (right). Cells were grown in LB, fixed, and imaged using our standard PALM protocol. mEos2 localizations show clustering due to chromophore blinking.

B. An illustration of our χ^2 nearest-neighbour-based clustering analysis. Experimentally-acquired localizations in single cells (left image in left panel) were randomly redistributed within cell boundaries to produce a simulated random distribution (right image in left panel). The nearest-neighbour distributions of both data sets were then calculated and compared using a χ^2 analysis (right panels). The χ^2 values increase as the experimental distribution of localizations deviates from randomness.

C. χ^2 values for all data sets compared to uniform distributions either without considering banding, or with banding being considered. In both cases, cytosolic PAmCherry1 and RNAP (tagged with PAmCherry1; cells grown in M9 medium) exhibit similar low χ^2 values, indicating mostly non-clustered distributions, whereas RNAP (tagged with PAmCherry1; cells grown in LB medium) exhibits higher χ^2 values, indicating clustering. Cytosolic mEos2 also exhibits high χ^2 values, reflecting chromophore blinking.

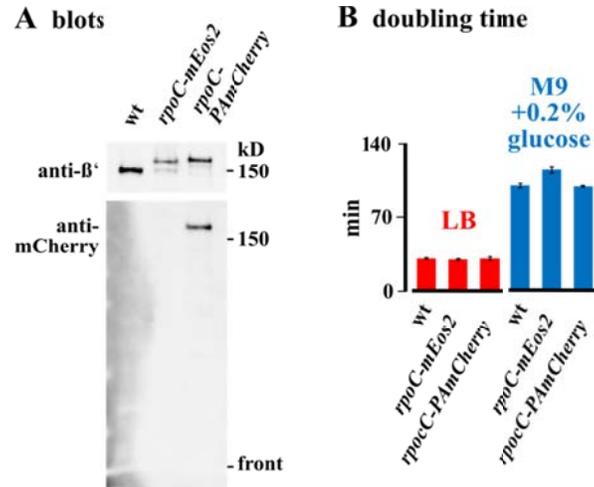


Fig. S3. Validation of tagged β' variants.

A. Western blot of whole cell lysates. The strains MG1655 (wild-type; wt), KF7-1 (rpoC-mEos2) and KF26 (rpoC-PAmCherry1) were grown to $OD_{600}=0.4$, pelleted and lysed. The denatured protein was then separated using SDS PAGE, transferred to a membrane, and subjected to western blotting. Probing with anti- β' antibody (upper panel) reveals that wild-type β' migrates with an apparent molecular weight of ~ 155 kDa, while $>90\%$ of β' -PAmCherry1 migrates slower, consistent with the presence of the tag (which itself is ~ 30 kDa). Probing with anti-mCherry antibody shows that $>90\%$ of the fluorescent protein is attached to β' .

B. Tagging β' with PAmCherry1 has no effect on growth rate. The three strains in (A) were grown in rich media (LB; red bars) or minimal media (M9 + 0.2% glucose; blue bars) with shaking at 32°C . Growth rate was determined by measuring OD_{600} increase during logarithmic phase ($OD_{600} < 0.4$). The presence of the PAmCherry1 tag on β' causes no change in cell doubling time in either media (i.e., $<1\%$), but tagging β' with mEos2 causes a mild growth defect in minimal media. Error bars represent standard deviations ($n=3$).

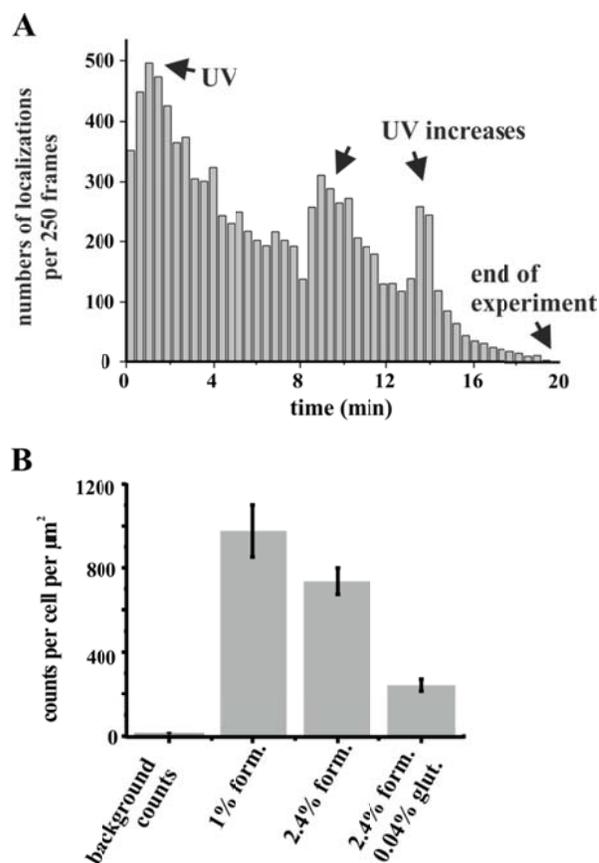


Fig. S4. Single-molecule localizations over time in a PALM experiment, and the effects of fixation.

A. Single-molecule localizations. The number of single-molecule localizations is plotted as a function of acquisition time for a representative image stack. At the beginning of the experiment, the number of single-molecule localizations decreases as the density of photoswitchable fluorescent proteins decreases. The increase of the intensity of UV light used for activation (black arrows) leads to a higher rate of photoactivation. At the end of the experiment, no more fluorescent proteins are detected, which reflects a stoichiometric read-out.

B. The number of single-molecule detections depends on fixation conditions. *E. coli* cells expressing PAmCherry1-tagged variants of RNAP (KF26) were grown in LB media at 32°C, fixed using different fixation mixtures based on paraformaldehyde (PFA; columns 2-4) and glutaraldehyde (column 4), and imaged with PALM. Fixed wild-type MG1655 cells (expressing no fluorescent proteins) produce a very low number of false positive localizations (column 1) due to autofluorescence. Higher concentrations of PFA as well as the addition of glutaraldehyde led to a decrease in the number of detected single molecules.

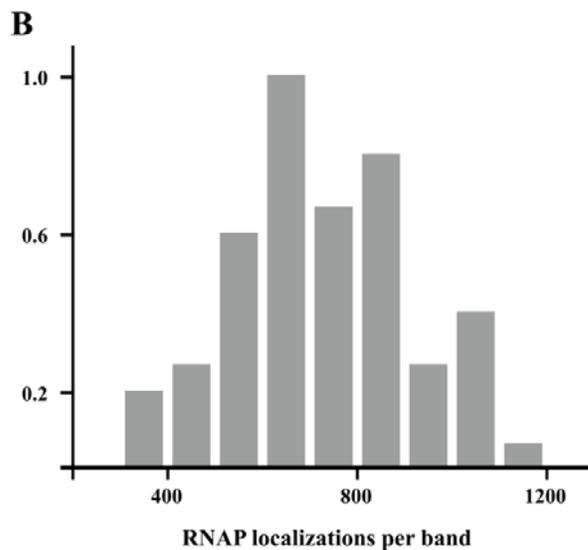
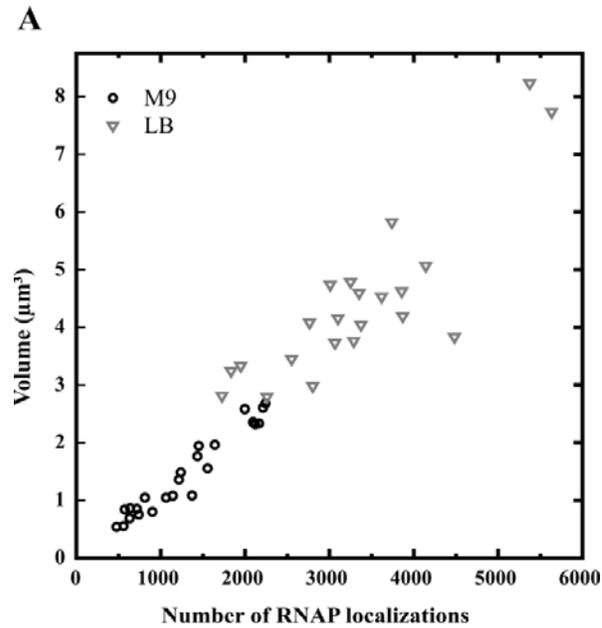


Fig. S5. The number of RNAP localizations depends on cell volume.

A. The number of RNAP localizations per cell scales with cell volume. *E. coli* expressing β' -PAmCherry1 were grown in rich (LB) or minimal (M9) media to mid-log phase, fixed, and imaged with PALM. For each cell (represented by a single point in the plot), the number of detected single RNAP localizations is plotted against cell volume (black circles – M9; grey triangles – LB).

B. Number of RNAP localizations per band. The number of localizations reflects the number of independent localizations in RNAP bands (see Fig. 1A) seen in PALM images of cells expressing β' -PAmCherry1 in LB (cells were grown to mid-log phase). The average number of RNAPs per band was 714 ± 198 (s.d.).

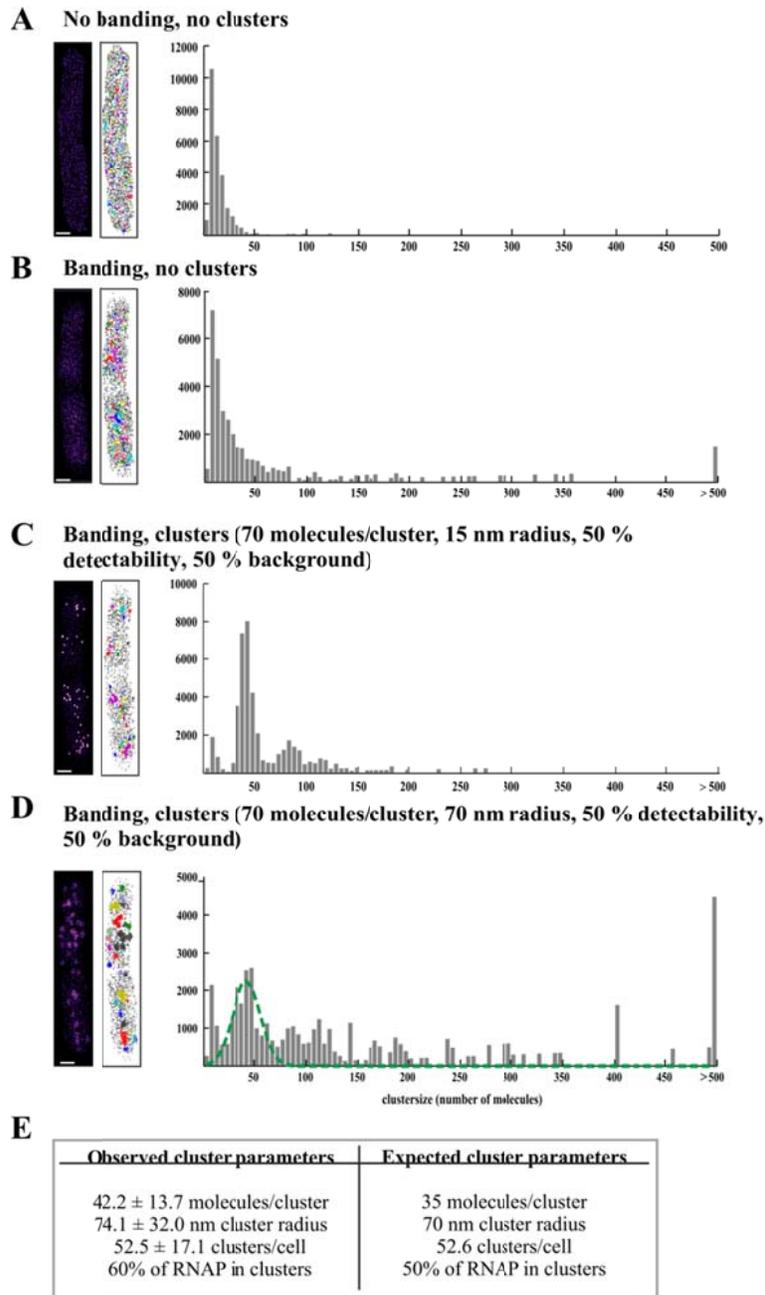


Fig. S6. DBSCAN clustering analysis on simulated data sets. Molecules per cluster found by DBSCAN on simulated data sets. The rightmost points on the histogram x-axes are total counts for $n \geq 500$ (so the large peak represents all additional large clusters outside range plotted).

A. Exemplary simulated data sets for different parameter settings and the corresponding DBSCAN analysis result.

B. In the absence of clustering or chromosomal banding, the cluster histogram shows a random coincidence peak which rapidly tails off.

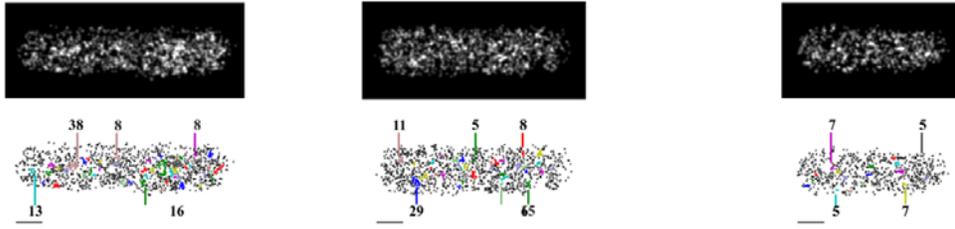
C. Simulated banding increases the extent of the tail to a non-negligible degree. However, random coincidence still tails off rapidly, with only isolated peaks at higher cluster size.

D. In highly clustered data, a clear primary peak is observed (at 35 molecules per cluster due to 50% detectability), followed by multiple secondary peaks due to overlapping unresolved clusters.

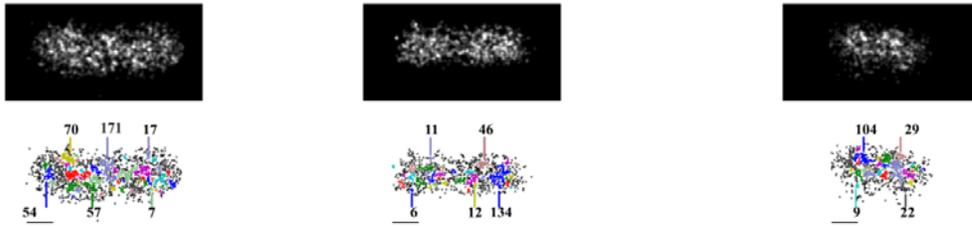
E. Even for large, lower density clusters, cluster properties can still be accurately distinguished. A clear primary peak centered at 42.2 ± 13.7 molecules per cluster is observed, close to the expected value of 35 molecules per cluster. The multiple secondary peaks due to overlapping clusters in this case are not clearly resolved, but instead form an extended, large amplitude tail to the data. This large amplitude tail is not observed in unclustered simulated data.

F. Comparison between observed and expected simulated cluster parameters. The observed cluster radius is defined by the radius of gyration for the cluster.

cytosolic PAmCherry1



M9



LB

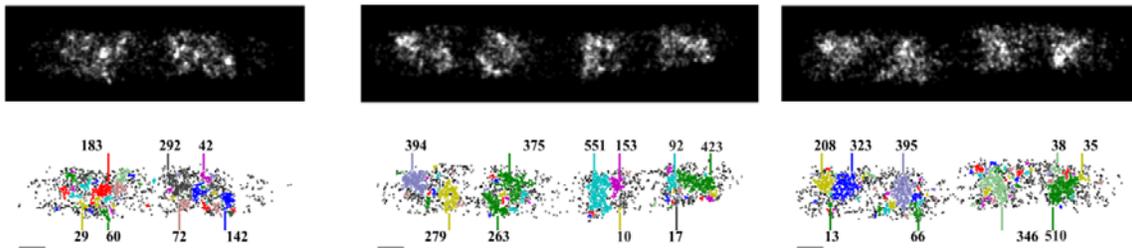


Fig. S7. Additional examples of cells analyzed using DBSCAN. Examples of super-resolved images of bacteria and corresponding DBSCAN analysis, showing clustered (colored circles) and unclustered molecules (black crosses). Numbered labels indicate number of molecules per cluster. For style, see Fig. 4. Scale bar: 500 nm.

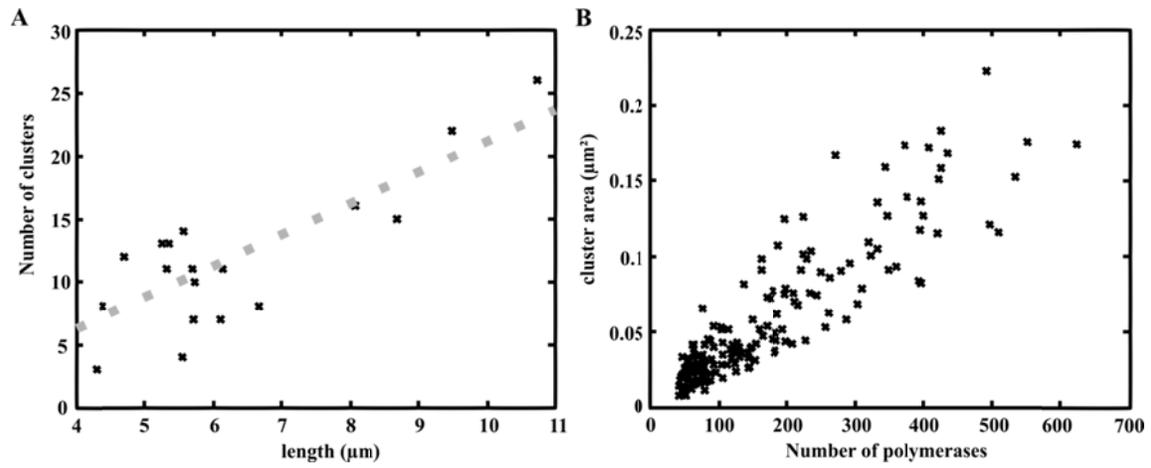


Fig. S8. The number of C70 clusters scales with cell length (A), and the number of RNAP molecules with cluster area (B).

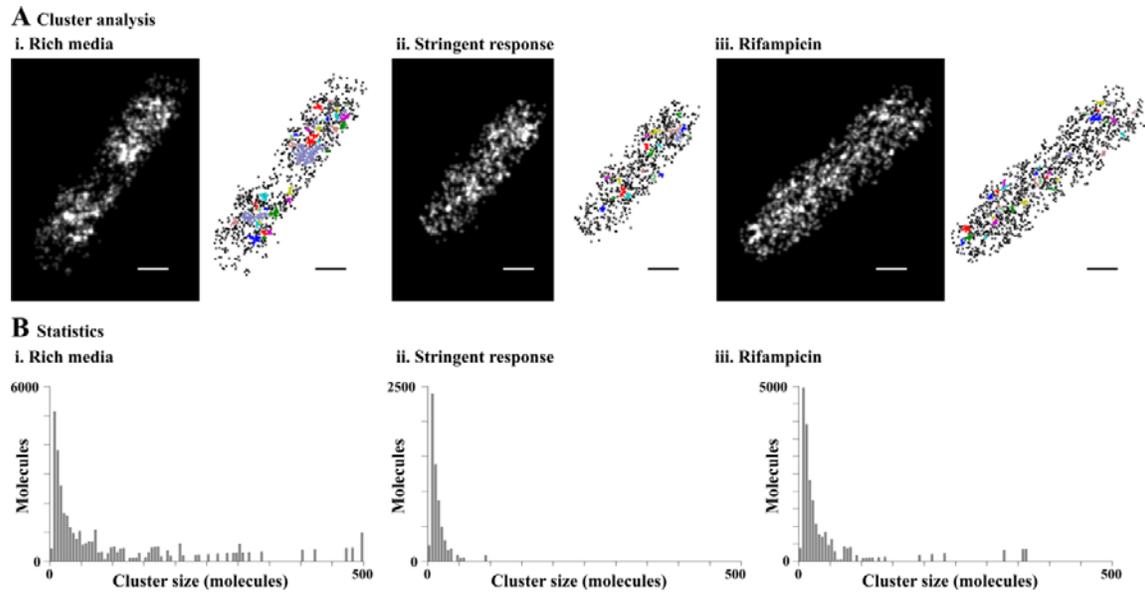


Fig. S9. The stringent response and rifampicin treatment prevent formation of large RNAP clusters. RNAP-PAmCherry1 was used in **(i)** LB (rich media) – as in Fig. 1A, **(ii)** rich media with the stringent response induced by 1 mg/ml L-serine hydroxamate for 1h, and in **(iii)** rich media containing 50 μ g/ml rifampicin for 120 min.

A. Examples of super-resolved images of bacteria and corresponding DBSCAN analysis. Scale bar, 500 nm.

B. Results of DBSCAN clustering analysis.

SUPPORTING REFERENCES

1. Annibale, P., M. Scarselli, A. Kodiyan, and A. Radenovic. 2010. Photoactivatable Fluorescent Protein mEos2 Displays Repeated Photoactivation after a Long-Lived Dark State in the Red Photoconverted Form. *The Journal of Physical Chemistry Letters* 1:1506-1510.
2. Endesfelder, U., S. Malkusch, B. Flottmann, J. Mondry, P. Liguzinski, P. J. Vermeer, and M. Heilemann. 2011. Chemically induced photoswitching of fluorescent probes--a general concept for super-resolution microscopy. *Molecules* 16:3106-3118.
3. Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. E. Simoudis, J. Han, and U. M. Fayyad, editors. AAAI-Press. 226-231.
4. Grigorova, I. L., N. J. Phleger, V. K. Mutalik, and C. A. Gross. 2006. Insights into transcriptional regulation and σ competition from an equilibrium model of RNA polymerase binding to DNA. *Proceedings of the National Academy of Sciences* 103:5332-5337.
5. Ishihama, A. 2000. Functional modulation of Escherichia coli RNA polymerase. *Annual Review of Microbiology* 54:499-518.
6. Bremer, H., and P. Dennis. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*. F. C. Neidhardt, editor. American Society for Microbiology, Washington D.C. 1553-1569.
7. Piper, S. E., J. E. Mitchell, D. J. Lee, and S. J. W. Busby. 2009. A global view of Escherichia coli Rsd protein and its interactions. *Molecular BioSystems* 5:1943-1947.
8. Cabrera, J. E., and D. J. Jin. 2003. The distribution of RNA polymerase in Escherichia coli is dynamic and sensitive to environmental cues. *Molecular Microbiology* 50:1493-1505.
9. Verheust, C., and D. R. Helinski. 2007. The incC korB region of RK2 repositions a mini-RK2 replicon in Escherichia coli. *Plasmid* 58:195-204.