Dissecting the in vivo dynamics of transcription locking due to positive supercoiling buildup

Cristina S.D. Palma, Vinodh Kandavalli, Mohamed N.M. Bahrudeen, Marco Minoia, Vatsala Chauhan, Suchintak Dash, Andre S. Ribeiro

Laboratory of Biosystem Dynamics, BioMediTech, Faculty of Medicine and Health Technology, Tampere University, 33101 Tampere, Finland

ARTICLE INFO
Keywords:
Single-RNA production dynamics
Positive supercoiling buildup
LineWeaver-Burk plots
Transcription locking kinetics

ABSTRACT
Positive supercoiling buildup (PSB) is a pervasive phenomenon in the transcriptional programs of Escherichia coli. After finding a range of Gyrase concentrations where the inverse of the transcription rate of a chromosome-integrated gene changes linearly with the inverse of Gyrase concentration, we apply a LineWeaver-Burk plot to dissect the expected in vivo transcription rate in absence of PSB. We validate the estimation by time-lapse microscopy of single-RNA production kinetics of the same gene when single-copy plasmid-borne, shown to be impervious to Gyrase inhibition. Next, we estimate the fraction of time in locked states and number of transcription events prior to locking, which we validate by measurements under Gyrase inhibition. Replacing the gene of interest by one with slower transcription rate decreases the fraction of time in locked states due to PSB. Finally, we combine data from both constructs to infer a range of possible transcription initiation locking kinetics in a chromosomal location, obtainable by tuning the transcription rate. We validate with measurements of transcription activity at different induction levels. This strategy for dissecting transcription initiation locking kinetics due to PSB can contribute to resolve the transcriptional programs of E. coli and in the engineering of synthetic genetic circuits.

1. Introduction

Transcription in Escherichia coli generates positive supercoiling ahead of the RNAP and negative supercoiling behind it ([11,46,51]; [57,95,99]). Discrete, topologically constrained segments along the DNA cause this process to generate local supercoiling buildup [31,33,41,70,77]. Evidence suggests that this torsional stress can affect gene activity [2,96].

E. coli has (at least) two proteins to resolve torsional stress. Namely, Gyrase removes positive supercoils [9,15,46] while Topoisomerase I removes negative supercoils [9,15,22,35,46,90,96]. Interestingly, in normal conditions, Topoisomerase I removes the negative supercoils at sufficient speed for R loops to not emerge, which is essential for cell survival [16]. This is made possible by the existence of a direct physical interaction between the RNAP and Topoisomerase I, allowing the latter to remove the negative supercoils, as soon as they form [7]. Contrarily to this, the removal of positive supercoils is not as efficient (being an ATP-dependent reaction likely contributes to this [77]), in the sense that positive supercoiling buildup (PSB) is commonly observed, particularly in highly active operons [17,28]. In support, measurements have shown that Topoisomerase I can relax plasmid DNA ~6 times faster than Topoisomerase IV [97], which has the same catalytic rate as Gyrase [84].

As positive supercoils accumulate, elongation slows down and, eventually, there are transient halts in transcription initiation [9,73]. These halts in initiation tangibly decrease RNA production rates and increase transcriptional noise [9,55,59]. Thus, dissection of the in vivo kinetics of transcription locking due to PSB is needed in order to dissect the transcriptional programs of E. coli.

A strategy was recently introduced for dissecting the in vivo kinetics of rate-limiting steps of active transcription initiation from in vivo measurements of individual RNA production events at different RNA polymerase (RNAP) concentrations [48]. It uses a Lineweaver–Burk plot [44] to infer the time-length of events prior and after commitment to open complex formation [58] from measurements of in vitro transcription rates at different RNAP concentrations ([RNAP]) [48]. This is possible due to the independence of the kinetics of the open complex formation from [RNAP], and because there is a range of values of [RNAP] for which the inverse of RNA production rate changes linearly with the inverse of [RNAP] [48].

https://doi.org/10.1016/j.bbagen.2020.194515

Received 14 June 2019; Received in revised form 7 February 2020; Accepted 20 February 2020
Available online 27 February 2020

1874-9399/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses BY-NC-ND/4.0/).
Similarly, chromosomal RNA production rates (particularly of genes in highly transcribed operons) are expected to differ with Gyrase concentration, due to the existence of discrete topological constraints [9,69,70]. Thus, it should be possible to, from in vivo RNA production rates at different Gyrase concentrations, infer the kinetics of in vivo transcription locking due to PSB. For this, it must hold true that there is a range of conditions for which the inverse of the RNA production rate changes linearly with the inverse of Gyrase concentration.

Here we verify this hypothesis and then use this strategy to dissect the contribution of transcription initiation locking due to PSB on the kinetics of RNA production of a chromosomally-integrated gene. We validate the estimation by time-lapse microscopy of single-RNA production kinetics of the same gene when single-copy plasmid-borne, shown to be impervious to Gyrase inhibition. Based on this, we estimate the fraction of time in locked states and the number of transcription events prior to locking, which we validate by measurements of RNA production under the inhibition of Gyrase activity by the addition of Novobiocin (see Section 2.3). Replacing our gene by a gene with a slower transcription rate, we show that changes in the basal transcription rate (expected rate of RNA production in the absence of effects from PSB) affect the contribution of locking due to PSB on effective transcription rates (measured rate of RNA production). Finally, we infer a range of possible transcription initiation locking kinetics in a chromosomal location, obtainable by tuning the basal transcription rate, and validate this inference using measurements of transcription activity at different induction levels.

2. Materials and methods

2.1. Strains and plasmids

We engineered two strains from E. coli BW25993 (lacIq hsdR514 Δ(rhaBAD)M15 Δ(rhaBAD)M16) [10]. In one strain, the target gene \( \text{P}_{\text{lacO3O1}}-\text{mCherry-MS2-BS} \) is integrated into a single-copy F-plasmid (~11 kbp), \( \text{pBELOBAC11} \) (target plasmid). This plasmid is not known to form long-lasting bonds to the cell membrane and is originally responsible for the expression of transient DNA-binding proteins [27,32,60]. In the other strain this plasmid is absent and the same target gene, \( \text{P}_{\text{lacO3O1}}-\text{mCherry-MS2-BS} \), is integrated into the lac locus of the genome using Red/ET recombination (Gene Bridges, Heidelberg, Germany) (Supplementary Figs. S1A and S1B). We found no significant differences in the growth rates of the two strains and the original strain, \( \text{P}_{\text{lacO3O1}}, \) inducible by IPTG, was engineered from the E. coli native \( \text{P}_{\text{lac}} \) by removing the O2 repressor binding site downstream of the transcription start site [62]. Thus, strong topological barriers are not expected to form when fully induced [21]. Also, both strains were transformed with the medium-copy reporter plasmid pZA25-GFP [61] (kind gift from Orna Amster-Choder, Hebrew University of Jerusalem, Israel), coding for the reporter MS2-GFP under the control of the \( \text{P}_{\text{BAD}} \) promoter. The strain with the target gene in a single-copy F-plasmid also contains the native Lac promoter in the chromosome. Thus, it has 4 Lac binding sites. The strain with the chromosome-integrated target gene only has 2 Lac binding sites, as the original Lac promoter was replaced by the target promoter, LacO2O1. However, as both strains overexpress LacI [10], effects of this difference are expected to be negligible.

Both target genes (plasmid and chromosome constructs) code for an RNA with an array of bindings sites (BS) for the modified viral coat protein MS2-GFP [25,67,68]. Due to the multiple BS in the target RNA and the strong binding affinity of each site [25], MS2-GFP tagged RNAs appear as bright spots soon after produced (Supplementary Figs. S2B and S2D). Their maximum fluorescence is reached rapidly (< 1 min) and have long half-lives (Supplementary Section I).

For overexpressing Gyrase, we constructed a plasmid (pZc11 \( \text{P}_{\text{ramm}} \) gyrAB-sfGFP, with ampicillin resistance) with the gyrA and gyrB genes under the control of a Rhamnose promoter. These genes were arranged in a polycistronic manner, using their (identical) ribosome-binding site to maintain the physiological stoichiometry of the two subunits. We amplified the sfGFP using the primers: Forward: 5′ CATATGGACAAAG GAGAAGACCTT T′; Reverse: 5′ CCGGCTTTTGTGACCTCATCCA TGC 3′ with restriction enzymes and cloned it after the gyrAB genes by digestion followed by ligation (Supplementary Fig. S3). We also constructed a plasmid without sfGFP, by digesting with the restriction enzymes NdeI and Nael, followed by ligation, which was transformed to E. coli BW25993 with the \( \text{P}_{\text{lacO3O1}}-\text{mCherry-MS2-BS} \) integrated into a single-copy F-plasmid [27] and to E. coli BW25993 with the \( \text{P}_{\text{lacO3O1}}-\text{mCherry-MS2-BS} \) integrated in the chromosome. Finally, in another strain, we replaced the chromosome-integrated \( \text{P}_{\text{lacO3O1}} \) by the native Lac promoter, followed by the same array of binding sites for MS2-GFP.

To access the intracellular levels of Gyrase A proteins, we used a strain with a gyrA gene endogenously tagged with the YFP coding sequence [85]. From the glycerol stock (~80 °C), cells were streaked on the LB agar plates and incubated at 37 °C overnight. From the plate, a single colony was picked, inoculated in an LB medium supplemented with the antibiotics, and incubated at 30 °C overnight with shaking at 250 RPM. Next, cells were diluted into fresh LB medium to an OD of 0.03 (Optical Density, 600 nm; Ultrospec 10, Amersham biosciences, UK) and grown at 37 °C with 250 RPM until it reaches to the mid-exponential phase (OD ~0.4-0.5).

2.2. Nucleoid visualization by DAPI staining

DAPI (4′,6-diamidino-2-phenylindole) stains nucleoids specifically with little or no cytoplasmic labelling. Gyrase induced and un-induced cells were grown at 37 °C and fixed with 3.7% formaldehyde in phosphate buffered saline (PBS, pH 7.4) for 30 min at room temperature, followed by washing with PBS to remove excess formaldehyde. The pellets were suspended in PBS, and DAPI (2 µg/ml) was added to the suspension. After incubating for 20 min in the dark, cells were centrifuged and washed twice with PBS to remove excess DAPI. Cells were then re-suspended in PBS and 3 µl of these cells were placed on a 1% agarose gel pad for microscopy.

2.3. Growth conditions and induction of the reporter and target gene

From a -80 °C glycerol stock, cells were placed in LB medium agar plates with 34 µg/ml Chloramphenicol and 35 µg/ml Kanamycin (Sigma-Aldrich, USA) and incubated overnight at 37 °C (Innova® 40 incubator, New Brunswick Scientific, USA). Cells were cultured in LB medium from single colonies on LB agar plates with the appropriate concentration of antibiotics and incubated overnight at optimal temperature at 250 rpm with aeration. These cultures were diluted to an optical density (OD\(_{600}\)) of 0.05 in fresh M9 medium, with a culture volume of 20 ml supplemented with the appropriate antibiotics and 0.4% of Glycerol (Sigma-Aldrich, USA), and incubated for 3 h with a 250 rpm agitation until an OD\(_{600}\) of ~ 0.3. Next, to induce MS2-GFP expression, 0.4% of L-Arabinose (Sigma-Aldrich, USA) was added and cells were incubated for another 45 min for sufficient MS2-GFP to accumulate for detecting target RNAs [26]. Next, the target gene was induced by IPTG (Sigma-Aldrich, USA) and cells were incubated for 1 h, prior to image acquisition or RT-PCR. To obtain induction curves of target genes (under the control of \( \text{P}_{\text{lacO3O1}} \) and \( \text{P}_{\text{lac}} \)), 0, 50, 100, 250, 500 and 1000 µM IPTG was added (Supplementary Fig. S4A). Unless stated otherwise, the target genes are always fully induced by 1000 µM IPTG.

We also performed measurements when inactivating and when overexpressing Gyrase. To inactive gyrase, we follow the protocol above, but when MS2-GFP expression is induced, we further added Novobiocin (100 µg/ml) [22]. Since all strains used here contain the gene \( \text{acrA} \), Novobiocin at this concentration is not expected to affect cell division rate [50]. We verified this by measuring growth rates by OD\(_{600}\) for varying Novobiocin concentration (0, 25, 50, 75, 100, 200, 300, 400 µg/ml) [50]. In another experiment, we grew cells at an OD\(_{600}\) of 0.5 in fresh M9 medium, with a culture volume of 20 ml supplemented with the appropriate antibiotics and 0.4% of Glycerol (Sigma-Aldrich, USA), and incubated for 3 h with a 250 rpm agitation until an OD\(_{600}\) of ~ 0.3. Next, to induce MS2-GFP expression, 0.4% of L-Arabinose (Sigma-Aldrich, USA) was added and cells were incubated for another 45 min for sufficient MS2-GFP to accumulate for detecting target RNAs [26]. Next, the target gene was induced by IPTG (Sigma-Aldrich, USA) and cells were incubated for 1 h, prior to image acquisition or RT-PCR. To obtain induction curves of target genes (under the control of \( \text{P}_{\text{lacO3O1}} \) and \( \text{P}_{\text{lac}} \)), 0, 50, 100, 250, 500 and 1000 µM IPTG was added (Supplementary Fig. S4A).
400, 500 ng/μl). The measurements show that growth rates to not differ significantly at 100 μg/ml or lower (Supplementary Fig. S5A). We further verified that Novobiocin does not affect morphology at these concentrations (see Section 3.6).

To overexpress Gyrase, we use Rhamnose (see previous Section [93]. We follow the protocol above but, when inducing MS2-GFP expression, we also added Rhamnose. When applicable, Gyrase and RNAP concentrations were measured 1 h after adding Rhamnose. Gyrase overexpression did not affect bacteria growth (Supplementary Fig. SSB) nor morphology (see Section 3.2).

2.4. RT-PCR

One hour after inducing the target gene, cells were fixed by RNAProtect bacteria reagent (Qigian, Germany), followed by enzymatic lysis with Tris-EDTA Lysozyme (15 mg/ml) buffer (pH 8.3). From the lysates, the RNA content was isolated using RNeasy purification kit (Qiagen) as per the manufacturer instructions. The RNA was then separated by electrophoresis using 1% agarose gel stained with SYBR® Safe DNA Gel Stain (Thermo Scientific, USA). The RNA was intact, with clear bands for the 16S and 23S ribosomal RNA. The RNA yield (~2 μg/μl) and absorbance ratios A260/A280 nm and A260/A230 nm were measured by a Nanovue Plus Spectrophotometer (GE Healthcare Life Sciences, USA). The ratio (2.0–2.1) indicates highly purified RNA. To remove DNA contamination, samples were treated with DNasel (Thermo Scientific, USA) as per the manufacturer instructions. The cDNA was synthesized from 1 μg of RNA using iScript Reverse Transcription Supermix (Biorad, USA) as per the manufacturer instructions. cDNA samples (10 ng/μl) were mixed with qPCR purification master mix with iQ SYBR Green supermix (Biorad, USA) with primers (200 nM) for target and reference genes. 16S rRNA was used as reference. Primers set for target mRNA (mCherry) and reference (16S rRNA) genes were: mCherry (Forward: 5’ CACCTTAAAGCCACAGGC 3’, Reverse: 5’ TGGTGTATCTCGTGGTG 3’), 16S rRNA (Forward: 5’ CTCAGCTCTGTGGTGGAAA 3’, Reverse: 5’ GGACCGCTGGCAACTAAAG 3’). To determine fold changes in mRNA Gyrase, cells were grown in M9 media supplemented with different Rhamnose concentrations. For the Gyrase mRNA (GyrA) gene the primer set was: Forward: 5’ GGATTATGGCGATGCGGTTCAT 3’, Reverse: 5’ CTGGGACAGATTCGTGGCGGT 3’. For mRNA sfGFP the primer set used was: Forward: 5’ GGAAAATACTCTCGTTCGCTGGC 3’, Reverse: 5’ CATACAACTTTGGCATGGCAC 3’. Experiments were performed by a Biorad MiniOpticon Real Time PCR System (Biorad, USA). The thermal cycling protocol was 40 cycles of 95 °C for 10 s, 52 °C for 30 s, and 72 °C for 30 s, with the fluorescence being read after each cycle. For each condition, we performed 3 biological replicates. qPCR efficiencies of these reactions were > 95%. No-RT and no-template controls were used to crosscheck non-specific signals and contamination. Cq values from the CFX ManagerTM Software were used to calculate fold changes in the target gene (normalized to the reference gene) and standard error, using Livak’s 2−ΔΔCT method (40). RT-qPCR results are presented in Table S1.

2.5. Flow cytometry

To measure single cell Gyrase-GFP expression levels, cells were grown as described in Section 2.1. Upon reaching mid exponential phase, cells were diluted 1:1000 into 1 ml PBS vortexed for 10 s and 50,000 cells were tested in each run. Data was collected by an ACEA NovoCyte Flow Cytometer (ACEA Biosciences Inc., San Diego USA) using a blue laser (488 nm) for excitation and the fluorescein isothiocyanate detection channel (FITC) (530/30 nm filter) for emission, at a flow rate of 14 μl/min and a core diameter of 7.7 μM. A PMT voltage of 417 was used for FITC. To avoid background signal from particles smaller than bacteria, the detection threshold was set to 5000 in FSC-H analyses. We set the fraction of the cells used in the analysis (a) to 0.95, to remove any undesired data points from debris, cell doubles etc. Reducing a further did not change the results.

2.6. Western blot

Cells were grown as above until reaching an OD600 of 0.6. Pelleted cells were lysed with B-PER bacterial protein extraction reagent (Thermo scientific) and proteins were extracted. Protein samples were diluted with 4 × laemml sample loading buffer and boiled for 5 min at 95 °C. 30 μg of proteins were loaded on the 4–20% TGX stain free pre cast gel (Biorad) and separated by electrophoresis. Proteins were then transferred to PVDF membrane using Trans-Blot Turbo transfer system (Biorad). The membrane was blocked with 5% non-fat milk and incubated with primary RpoC antibody 1:2000 dilutions (Biologend) overnight at 4 °C and followed by HRP-secondary antibodies 1:5000 dilutions (Sigma Aldrich) for 1 h at room temperature. For band detection, the membrane was treated with a chemiluminescence reagent (Biorad). Images were acquired by the Chemidoc XRS system (Biorad). Band quantification was done using Image lab software (version 5.2.1). For each condition, we performed 3 biological replicates.

2.7. Microscopy and image analysis

Cells were grown as above, and pelleted and re-suspended in ~100 μl of the remaining media. Prior to imaging, cells were placed on a 2% agarose gel pad of M9 medium and kept in between the microscope slide and a coverslip. Cells were visualized by a Nikon Eclipse (Ti-E) inverted microscope with a 100 × Apo TIRF (1.49 NA, oil) objective. Confocal images were taken by a C2+ (Nikon) confocal laser-scanning system with a pinhole size of 1.2 AU. In confocal images, the size of a pixel corresponds to 0.062 μm using a scan area resolution of 2048 × 2048 pixels. MS2-GFP-RNA spots and GyrA-YFP regions were visualized by a 488 nm laser and a 514/30 emission filter, while DAPI-stained nucleoids were visualized by a 405 nm laser and a 447/60 emission filter.

Phase contrast images were taken by an external phase contrast system and DS-Fi2 CCD camera (Nikon). Image sizes were 2560 × 1920 pixels, each pixel corresponding to 0.048 μm. Phase contrast and confocal images were taken simultaneously by Nis-Elements software (Nikon).

From phase contrast images, we segmented cells with the software iCellFusion [78] (Supplementary Figs. S2A and S2C). Errors were manually corrected. Next, phase-contrast and corresponding fluorescence images were aligned by the software CellAging [29]. We used CellAging to detect RNA-MS2-GFP fluorescence spots (Supplementary Figs. S2B and S2D) and assess the intensity of each spot. From these, integer-valued RNA numbers were calculated for each spot (Supplementary Section I).

Nucleoid(s) segmentation was performed as in [63], using a 2D Gaussian approximation, followed by manual corrections. Cells whose size is smaller than 500 pixels were excluded from the analysis since, in general, they were not real cells (e.g. only half of the cell appeared in the image). Also removed were cells larger than 1000 pixels, as they were abnormally elongated. In general, this led to removing < 5% of the cell population.

The segmentation of the intracellular regions with significant Gyra-YFP was done using a tailored software, SCIP [56]. Errors were manually corrected. To remove measurement noise, we applied a 2D Gaussian filter to each region [94].

2.8. Models and simulations

We use stochastic models of gene expression to test if arrests during elongation, caused by PSB, disturb significantly the mean RNA production rate (within realistic intervals of parameter values).

These models are at single-cell, and single-molecule level. Specifically, two models are simulated. One is the Single-Nucleotide
model’ (SN Model), since elongation is modelled at the single nucleotide level (Supplementary section II). The other is the “Minimal model”, as it is designed from the former, but lacks elongation at the nucleotide (Fig. 1, Section 3.1).

The length of each simulation is $10^5\ s$, found to be long enough for not underestimating the mean length of the time intervals between consecutive RNA production events (which would bias the data with right-censoring) [30]. The simulations have a reading time of $1\ \text{s}^{-1}$. The results shown in the Results section are obtained from 100 runs per condition, as this number suffices to obtain consistent results. Finally, the initial components at the start of simulations are 1 promoter (where the transcription start site is located), 1 Gyrase, and 28 RNAPs. In addition, the SN model has also 4058 nucleotides, in the state “unoccupied” (Supplementary Section II) along which elongation will occur.

The models are implemented in the simulator SGNSim [74] and their dynamics follow the Stochastic Simulation Algorithm [23, 24]. In short, the stochastic nature of their dynamics arises from the generation of two random numbers at each step. As described in [24], one of these random numbers determines which is the next reaction (which differs with the propensity of each reaction at that moment), while the other random number determines when the event will occur (which depends on the total propensity when considering all possible reactions combined). SGNSim makes use of ‘Mersenne Twister’ to produce these random variables at each step [57].

### 3. Results

#### 3.1. Expected effects of changing gyrase concentration on the dynamics of transcription

We started by designing a stochastic model of transcription at the single nucleotide level (here named ‘SN model’), described in detail in Supplementary section II and shown in Supplementary Table S2. The model is based on a model proposed in [75] and later used in [53, 72], to which we add positive supercoiling buildup/removal. The reactions composing the model should not be interpreted as elementary transitions. Instead, they represent the rates of the rate-limiting steps of the various events. The model dynamics and simulations are described in Materials and Methods, Section 2.8, while its assumption of homogeneous mixing of RNAP and Gyrase is validated in Supplementary Section III.3.

The model consists of the following events (Table S2): transcription starts when an RNAP finds the promoter (reaction S2.1) and unwinds the DNA for reading and escapes the promoter (reaction S2.5). After this, stepwise transcription elongation is initiated, accounting for realistic RNAP footprint in the DNA template, transcriptional pausing, arrests, editing, premature terminations, pyrophosphorolysis and collisions between RNA polymerases [53].

In addition, the model accounts for the phenomenon of production of positive supercoils during elongation [46, 51, 95], in reaction S2.6 in Supplementary Table S2. As these supercoils accumulate, they enhance the propensity for RNAP arrest (reaction S2.11) [20, 51] and transcription initiation locking (reaction S2.2) [9]. The removal of positive supercoils by Gyrase [84] is also modelled explicitly (reaction S2.17). Finally, the model accounts for the potential accumulation of positive supercoils due to transcriptional activity of neighbor genes (reaction S2.4).

This model does not include RNA degradation, as we measured RNA numbers by MS2-GFP tagging, which prevents degradation for a few hours (Supplementary Section I and Figs. S1 and S2) [26, 86], thus avoiding this source of noise. Further, for the purposes of this work, we are only interested in RNA production rates, which do not depend on degradation.

Results in Supplementary Sections III.1 and III.2 show that this SN model mimics the effects of PSB on the kinetics of stepwise transcription elongation. Namely, from Supplementary Fig. S6A, one finds that the mean elongation time increases as Gyrase numbers are decreased. Meanwhile, from Supplementary Fig. S6B, one finds that this slowdown of stepwise elongation does not affect the mean rate of RNA production, within the range of parameter values tested, which is expected.

Note that in this model, for the realistic range of parameter values considered, once the system reaches steady state (near constant number of RNAPs on the DNA strand), the mean RNA production rate depends only on the rate with which RNAPs initiate new elongation events (reaction S2.1 in Table S2) and on the rate of abortions of elongation (reaction S2.14 in Table S2), with the latter being near negligible (~4% per transcription initiation event). Only in the unlikely scenario of excessive accumulation of RNAPs in the DNA template that would jam the promoter region, would events in elongation affect the mean RNA production rate.

Therefore, for purposes of estimating the effects of changing Gyrase
concentration on the mean RNA production rate, we instead use a minimal model, where elongation is not explicitly represented. Fig. 1 shows the minimal model that, as the SN model, also has a stochastic dynamics in accordance to the SSA (Materials and Methods, Section 2.8). In detail, reaction 1 models RNA production by an active promoter, Pro, and its propensity differs with the basal transcription rate, k1, and with RNA polymerase numbers. Meanwhile, reaction 2, which models transcription locking, is in all identical to reaction S2.2 of the SN model and, thus, its propensity differs with the number of positive supercoils (Supplementary Section II).

Positive supercoils can be generated via reactions 4 and 5 (also as in the SN model) [1,9,51,70,83,84,89]. The propensity of reaction 4 depends on the basal transcription rate, k1, of the gene of interest, as described in [17] and in agreement with results from anchored plasmids [9,87] as well as with results reported here. The parameter λ in reaction 4 accounts for the length of the gene of interest (RNAP will take longer to transcribe a longer gene, during which time positive supercoils are produced). As for reaction 5, responsible for the accumulation of PSB due to the transcription activity in the topological domain of the gene of interest, its kinetics differs with the neighboring activity, which can be tuned by kG and RNAP numbers (Fig. 1).

Once locked, a promoter can become unlocked via reaction 3. The unlocking kinetics can be tuned by the rate constant kunlocked. Because the propensity for locking changes linearly with the number of positive supercoils, the propensity for reaction 3 is kept independent from this number. Else, the overall time spent in locked states would change quadratically with the inverse of Gyrase numbers, and not linearly (Fig. 3 provides empirical support for the assumption that this relationship is linear within realistic ranges of parameter values). Finally, reaction 6 represents the removal of positive supercoils by Gyrase. As this takes place, the propensity for reaction 2 decreases, thereby accounting for the expected decrease in the effects of PSB with increasing Gyrase numbers [9].

To verify that the minimal model constitutes a valid approximation of the SN model, we performed simulations for various Gyrase numbers. Visibly, from Supplementary Fig. S6E, the minimal model matches the mean rate of RNA production of the SN model (and the empirical data) as a function of Gyrase numbers. This is expected since, as noted, all its parameter values are the same as in the SN model, except for reaction 4 in Fig. 1, since this reaction needs to account for the number of nucleotides of the gene of interest (which are modelled explicitly in the SN model). This adjustment is done by having the rate of SC+ production of the gene of interest dependent on its nucleotide length (with λ equaling a tenth of its number of nucleotides, as this is approximately the expected number of SC+ produced during one elongation event [84]).

We then derived an analytical solution of the minimal model, for the inverse of the mean rate of RNA production (r−1) as a function of Gyrase (inset of Fig. 1). Here, ractive, is the mean time between consecutive RNA production events of an unlocked/active promoter, which equals the inverse of kG × RNAP (with RNAP being the number of RNA polymerases). Meanwhile, r−1 is the inverse of the sum of ractive and rlocked with the latter being the mean time spent in locked states (equation in the larger inset in Fig. 1). From this solution, we find that increasing [G] decreases rlocked [9,46], which increases r. In detail, r−1 is expected to change linearly with [G]−1 (large inset, Fig. 1). If this holds true, from measurements of r and [G], it should be possible to extrapolate ractive since ractive should equal r−1 for infinite [G]. Further, from ractive and r, it should be possible to estimate rlocked. Finally, note that while k1 does not affect the mean time for Gyrase to release the gene from a locked state, it does affect the rate of occurrence of locked states.

Interestingly, many plasmids only have weak, transient topological barriers (such as short-term protein-DNA complexes [42]). In particular, aside from when they are anchored to the membrane [3,11,49,71] or have many tandem copies of a DNA-binding site [42], no long-term PSB is expected, since positive and negative supercoils diffuse in opposite directions and annihilate one another [42] (unlike in the chromosome that has topological barriers). As such, it should be possible to simulate the dynamics of plasmid-borne genes using the model in Fig. 1, by setting klock to null, causing rlocked to be null. Consequently, r−1 of a model plasmid-borne gene should equal ractive of the same model gene, when chromosome-integrated. Further, if this holds true, then a plasmid-borne gene can be used as a proxy for the same gene when chromosome-integrated when unaffected by PSB.

3.2. Changing intracellular concentration of gyrase

Above, we hypothesized that r−1 should be linear with respect to [G]−1 within a given range of Gyrase concentrations (see Fig. 1 and Supplementary Section VI). If true, one should observe a line on a Lineweaver–Burk plot [44] of r−1 against [G]−1, from which one can extrapolate ractive. From ractive and r, one can then estimate rlocked.

To test this hypothesis, it is necessary to measure r in cells differing in [G]. For this, we inserted a plasmid carrying a copy of the gyraseA and gyraseB genes under the control of the Rhamnose promoter (pZe11 P_rham gyrAB-sfGFP, Section 2.1 and Supplementary Fig. S4). The region coding for sfGFP allows measuring mRNA coding for Gyrase and the corresponding protein levels produced solely by the plasmid.

We subjected cells to different Rhamnose concentrations until finding a range for which the production rate of the mRNA coding for Gyrase increases linearly with Rhamnose concentration. For this, we performed qPCR using the region of the RNA from the plasmid that is under the control of the Rhamnose promoter (pZe11 P_rham byrAB, Materials and Methods). We further added sfGFP, also under the control of the Rhamnose promoter (pZe11 P_rham gyrAB-sfGFP, Section 2.1 and Supplementary Fig. S4). The region coding for sfGFP allows measuring mRNA coding for Gyrase and the corresponding protein levels produced solely by the plasmid.

We subjected cells to different Rhamnose concentrations until finding a range for which the production rate of the mRNA coding for Gyrase increases linearly with Rhamnose concentration. For this, we performed qPCR using the region of the RNA from the plasmid that is absent in the native RNA coding for Gyrase (i.e. the region coding for sfGFP). In Fig. 2A we find a linear relationship between mRNA fold changes (measured by qPCR) and Rhamnose (0, 0.1, 0.2 and 0.4%). In particular, small deviations from linearity were rejected (p-value > 0.5, see Fig. 2 legend for details).

Next, we verified that cell growth rates were not disturbed in this
range (Supplementary Fig. S5B). Further, to test if morphology was affected, we measured cell areas in the control condition (165 cells analyzed) and when subject to 0.4% Rhamnose (182 cells analyzed). The cell area was obtained from phase contrast images, using the software iCellFusion (Section 2.7). We performed a 2-sample Kolmogorov-Smirnov test and found that, at the significance level of 0.05, the two distributions cannot be distinguished (p-value of 0.5).

Finally, we verified that, within this range of conditions, the mean relative Gyrase expression level changed linearly with the RNA production rate of the plasmid coding for Gyrase (Fig. 2B), as measured by Flow-cytometry (Section 2.5). In particular, small deviations from linearity were rejected (p-value of 0.15, see Fig. 2 legend for details). We thus conclude that the fold change in Gyrase-GFP protein levels correspond to the fold change in the mRNA coding for Gyrase-GFP (Supplementary Section VII and Table S3 show the parameters of the calibration line and procedure).

We expect the quantitative relationship between mRNA and protein numbers of the plasmid-borne Gyrase to be the same as in the native Gyrase mRNA and proteins, since we used the native ribosome binding site in the plasmid construct. Thus, we measured by qPCR the fold change of the mRNA produced by both the native and the plasmid-borne Gyrase genes and used the line in Fig. 2B as a calibration line, to estimate the fold change with Rhamnose in Gyrase protein levels (Supplementary Section VII and Table S3).

Finally, we considered that Gyrase overexpression could change the proteome and, eventually, change cellular functioning (e.g. in 1–2 h). To mitigate effects from this eventuality (to avoid unknown changes in the processes represented in Fig. 1), subsequent measurements were conducted 1 h after inducing Gyrase overexpression (Materials and Methods). Given this and the above, we expect that, for 0.2% or lower Rhamnose concentrations (Fig. 2A), changes in RNA production rate in this time window are largely due to changes in concentrations of the components of the reactions in Fig. 1.

3.3. Transcription rate of a chromosome-integrated gene under the control of PLacO3O1 in the absence of positive supercoiling buildup

Data in [85] indicates that the expression rate of (at least) three of the RNAP sub-units are, in normal conditions, approximately double the average expression rate of E. coli genes. Since several highly expressed genes are supercoiling sensitive [17], it is tangible that Gyrase overexpression may affect [RNAP], which according to the model (Fig. 1), could affect the transcription rate (r) of our gene of interest (Fig. 1). Thus, we first assessed for potential fast changes in [RNAP] when overexpressing Gyrase.

For this, we used the same plasmid as above, with the gyrA and gyrB genes controlled by a Rhamnose promoter, to overexpress Gyrase (but having removed sfGFP, so as to not affect RNA counting or Gyrase functioning, see Materials and Methods). Next, we measured [RNAP] at the different Rhamnose concentrations (0%, 0.1% and 0.2%) by measuring the RpoC protein by Western blot, 1 h after inducing Gyrase overexpression. From Fig. 3A and B, at the same OD600, the [RNAP] differs by 12% between the two extreme conditions. This difference was found to be statistically significant by a 1-sample 2-tailed t-test, with the null hypothesis that the increase is 12% (p-value of 0.42). In addition, we performed a 2-sample, 2-tailed t-test with the null hypothesis that there is no difference between the conditions, which was rejected (p-value of 0.0008). This is expected to partially explain changes in r due to Gyrase overexpression and, thus, needs to be accounted for when quantifying the direct effects of changing [G] (Supplementary Section VIII).

In addition, it is tangible that overexpression of Gyrase could affect the negative supercoiling state of the chromosome, e.g. by introducing negative supercoils [6,47,80]. This, in turn, could affect DNA supercoiling density and its folding and compaction [36,92], which could alter transcription rates by affecting the time-lengths of open complex formations [52].

Unfortunately, we cannot measure directly the in vivo kinetics of open complex formation at a given Gyrase concentration, as this would require measuring the in vivo transcription in cells with differing RNAP concentrations [48,81], which would also affect the intracellular Gyrase concentration. Therefore, instead, we estimated indirectly if Gyrase overexpression (between 0% and 0.2% Rhamnose) suffices to alter significantly the chromosome folding and compaction. For this, we assessed if the nucleoid area (with area being a proxy for compaction strength) is altered by Gyrase overexpression, using DAPI staining and image analysis (Sections 2.2 and 2.7).

The mean and standard deviation of the nucleoid area, when and when not overexpressing Gyrase, are shown in Table S4. We performed a 2-sample student t-test for the null hypothesis that the two data sets of absolute nucleoid area come from the same distribution. The test did not reject the null-hypothesis (p-value > 0.01). We thus conclude that, in the range of Gyrase overexpression levels used here, the nucleoid size was not significantly affected. As such, we do not expect the indirect effects of Gyrase overexpression on DNA supercoiling density to significantly affect the kinetics of open complex formation.

Given this, we again used the plasmid with the gyrA and gyrB genes controlled by a Rhamnose promoter (without sfGFP) to study the effects of Gyrase overexpression on transcription initiation locking due to PSB of a chromosome-integrated gene under the control of PLacO3O1. This promoter was used as its dynamics has been previously studied when plasmid-borne, including using single-RNA MS2-GFP tagging [34,54,64,66,81].

We first measured the absolute mean r−1 in the control condition (Materials and Methods, Section 2.3) by microscopy measurements of integer valued RNA numbers in single cells at different time moments (Supplementary Section IX). The absolute mean r−1 in the control condition was found to equal 1476 s, with a standard error of 145 s.

Using the value of r−1 in the control condition, we scaled the relative qPCR values to obtain the values of r−1 in conditions where Gyrase is overexpressed. Results from qPCR are shown in Supplementary Table S1. In these, the gradually increasing expression of Gyrase did not affect significantly the expression of the 16S rRNA gene. This is expected, since 16S rRNA is a stable component of ribosomes and, thus, should not change significantly between conditions when growth rates are not affected significantly [88] (Supplementary Fig. S5B). Namely, even if the small changes between conditions were considered significant, there is no monotonic change with increasing Rhamnose concentration. As such, 16S rRNA is used as the reference gene.

Next, from Supplementary Table S1 and the microscopy data in the control condition, we obtained absolute rates of RNA production in each condition (black circles in Fig. 3C). Finally, we fitted a line by weighted total least squares (WTLS) [39] (black line in Fig. 3C) to estimate ractive (where the line intersects the Y-axis), when not accounting for changes in [RNAP]. We performed a likelihood ratio test between the best linear fit and fits of higher order polynomials which showed that the linear model best fits the data (p-value > 0.9).

Next, from the [RNAP] in each condition (Fig. 3B) and the model fitting (Supplementary Section IV), we estimated the effects of changes in [RNAP] (Supplementary Section VIII and Supplementary Table S5). Supplementary Figs. S7A and S7B show the Z surfaces of the best fitting models and empirical results. From the R2 values (legend of Supplementary Fig. S7), one finds that the model well-fits the empirical data. We thus used this model to estimate the weight of the changes in [RNAP] (Fig. 3B) on r−1 and then quantified the changes in r−1 due to changes in [G] alone. Results are shown in the blue circles in Fig. 3C. Next, we fitted a line (blue line in Fig. 3C) to these data points using WTLS, from which we estimated r−1 for infinite [G] (i.e. ractive) to be 749 ± 247 s.

Finally, we determined that if the small deviations from linearity are statistically significant by performing a likelihood ratio test between
3.4. Transcription kinetics of \( P_{\text{LacO3O1}} \) when single-copy plasmid-borne

To validate the estimation of \( t_{\text{active}} \), we integrated the same gene under the control of \( P_{\text{LacO3O1}} \) into a single-copy plasmid (Materials and Methods). We expect this to reduce the effects of PSB on the activity of the gene of interest to a minimum. I.e., the value of \( r^{-1} \) of the single copy plasmid-borne gene should approximate the estimated \( t_{\text{active}} \) of the chromosome-integrated gene. If this holds true, adding Novobiocin, which inhibits Gyrase activity \([9,15,22,91]\), should not disturb significantly its activity.

To test this, we performed time-lapse microscopy measurements of RNA numbers in cells subject to 100 μg/ml Novobiocin (Materials and Methods). Images were taken every 15 min, starting 30 min after introducing 1 mM IPTG in the media to ensure full induction of the target gene \([86]\). We also performed measurements where Novobiocin was not added.

From Supplementary Fig. S4B, the RNA production rate of the plasmid-borne gene is not affected by the addition of Novobiocin, as expected if PSB is absent. Meanwhile, in the absence of Novobiocin, we observe the same behavior but higher \( r \), which is consistent with the cells subject to Novobiocin having lesser number of active RNAAP and/or \( G \) factors \([8,13,18,76]\), etc. Further, both behaviors are significantly different from cells with the chromosome-integrated construct subject to Novobiocin, where a clear blocking of the RNA production is observed shortly after adding Novobiocin (Fig. 4, blue line). We conclude that the gene in the single-copy plasmid is not directly affected by Novobiocin, suggesting that it is impervious to the effects from PSB.

In support, according to the model (Fig. 1), for equal mean RNA production rate, the kinetics of RNA production from a gene unaffected by PSB (such as when on a single-copy plasmid) should be less noisy than otherwise (e.g. when chromosome-integrated) \([64]\). Lesser noise should reduce cell-to-cell variability in RNA numbers. To test this, we compared the squared coefficient of variation of RNA numbers in single cells, \( CV^2(\text{RNA}) \), in conditions where the two constructs exhibit the same mean RNA numbers per cell (50 μM IPTG for the plasmid-borne gene and 1000 μM IPTG for the chromosome-integrated gene, Supplementary Fig. S4A). The \( CV^2(\text{RNA}) \) in cells with the chromosome-integrated construct is found to be much higher than in cells with the single-copy plasmid-borne gene (3.18 and 1.58, respectively), in agreement with the model prediction, even though the plasmid-borne gene is being partially affected by LacI repression, which adds variability in RNA numbers \([48]\).

Finally, we verified that the RNA production rate of the single-copy plasmid construct equals the inverse of \( t_{\text{active}} \) of the chromosome construct. For this, we performed microscopy measurements of the integer valued RNA numbers in cells with the plasmid construct and estimated \( r^{-1} \) to be 775 ± 50 s (dark grey triangle in Fig. 3C, for the control condition). This result cannot be distinguished, in a statistical sense,
from the estimate of $t_{\text{active}}$ for the chromosome construct assuming infinite [G] (749 ± 247 s) (in agreement with the model predictions, Fig. 1).

In this regard, in Supplementary Section X, we estimated the minimum PSB effects that would be detectable, provided the same transcription activity (inset in Fig. 4). RNAs were detected by MS2-GFP. For each time point, new cells were taken from the original culture. On average, 200 cells were used per condition. Error bars represent the standard error of the mean. Finally, the inset shows the number of RNA production events per cell relative to the total number of RNAs produced during the measurement time. Data collected at the single RNA level, from time-lapse microscopy measurements with images taken once per minute. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In this regard, note that estimation of $t_{\text{active}}$ for the chromosome construct assuming infinite [G] (749 ± 247 s) (in agreement with the model predictions, Fig. 1).

In this regard, note that estimation of $t_{\text{active}}$ for the chromosome construct assuming infinite [G] (749 ± 247 s) (in agreement with the model predictions, Fig. 1).

Further, we also performed qPCR measurements of the plasmid construct when subject to the same levels of Gyrase overexpression. Results are shown in Fig. 3C. Next, we fitted a line (dark grey line in Fig. 3C) to the data points. From this, we find that the change in RNA production rate of the plasmid gene with gyrase overexpression is ~5 times weaker than in the chromosome-integrated construct. Next, using WTLS [39] we tested if the small deviations from linearity are statistically significant. The test did not reject the linear model (p-value > 0.8). Subsequently, as before, we corrected the data points to account for the changes in RNAP concentrations. Results in Fig. 3 (light grey line) show that the corrected line is nearly horizontal and, as expected, cannot be distinguished from a horizontal line, in a statistically significant, using the same test as above. We conclude that the RNA production kinetics from the plasmid construct is nearly non-responsive to Gyrase overexpression.

In this regard, note that estimation of $t_{\text{active}}$ for the chromosome construct assuming infinite [G] (749 ± 247 s) (in agreement with the model predictions, Fig. 1).

compared to the control condition, accounts for 41% of the decrease in $r^{-1}$, with the remaining 59% being due to increased [G].

3.5. Mean time spent in locked states and average number of transcription events between consecutive locking events

Since Fig. 3 shows that $r^{-1}$ changes linearly with [G]$^{-1}$, we used the Lineweaver-Burk equation [44] to estimate the mean time spent in locked states, $t_{\text{locked}}$, as follows:

$$t_{\text{locked}} = \frac{[G]_2 (\tau^2 - \tau^1)}{([G]_1 - [G]_2)}$$  \hspace{1cm} (1)

From (1), given the control and the condition where relative [G]$^{-1}$ is 0.33 (0% and 0.2% Rhamnose, respectively), we infer $t_{\text{locked}}$ to be 735 s, with a standard error of the mean (SEM) of 341 s (obtained by the Delta Method [5]). Using the other pair of conditions (0% and 0.1% Rhamnose) we obtain the same result, in a statistical sense. As the mean time interval between transcription events is 1476 s, we estimate transcription initiation locking due to PSB to account for ~50% of this interval.

Meanwhile, to estimate the mean number of transcription events between consecutive locked states, $N$, consider that, according to the model:

$$N = \frac{t_{\text{escape}}}{t_{\text{locked}}}$$  \hspace{1cm} (2)

To solve for $N$, we used the value of $t_{\text{locked}}$ obtained above, and $t_{\text{escape}}$ obtained from measurements in [9], which reported that the average DNA binding time of Gyrase is ~333 s while the unbind time is ~103 s [9]. Since Gyrase is expected to resolve multiple positive supercoils during this time [1,84], we assumed that the sum of these times (~1333 s) is an upper bound of the time for a locked gene to escape PSB (i.e. $t_{\text{escape}}$). Introducing the estimated values of $t_{\text{locked}}$ and $t_{\text{escape}}$ in eq. 2, we find that $N$ equals ~1.8 ± 0.84.

3.6. Kinetics of transcription initiation locking in the presence of a gyrase inhibitor

To validate the above estimations, we performed time series measurements at the single-RNA level in cells carrying the chromosome-integrated $P_{\text{LacO3O1}}$ subject to Novobiocin, a Gyrase inhibitor [22]. Assuming that, when Novobiocin first enters the cytoplasm, $P_{\text{LacO3O1}}$ activity is not subject to PSB, then the mean number of RNAs produced until transcription ceases should correspond to the mean number of transcription events between consecutive locking events. As it is not likely that the gene of interest is absent of effects from PSB in all cells, the empirical result should correspond to a lower bound estimate. Interestingly, from the same experiment, it should also be possible to measure $t_{\text{active}}$ (Fig. 1) from the time for RNA production to cease in all cells.

First, we tested whether Novobiocin, at the concentrations used here, affects cell morphology. For this, as above, we measured cell areas in the control condition (165 cells analyzed) and when subject to 100 ng /ml Novobiocin (180 cells analyzed), and then performed a 2-sample Kolmogorov-Smirnov test. We found that, at the significance level of 0.05, the two distributions cannot be distinguished (p-value of 0.13).

Next, we measured integer-valued number of RNAs in individual cells over time, every 15 min, 45 min after inducing the target gene (with IPTG) and adding Novobiocin (Gyrase inhibitor), so as to account for the mean time taken by cells to intake IPTG [64,86] and because only at this moment did we observe any tangible reduction in transcription activity (inset in Fig. 4). RNAs were detected by MS2-GFP tagging, preventing RNA degradation (Materials and Methods). We also performed a control experiment, where Novobiocin was not introduced.

Results in Fig. 4 show that when and only when adding Novobiocin,
the RNA production ceases. In the presence of Novobiocin, on average we observed 0.8 ± 0.11 RNAs per cell after 15 min. Considering mean cell division times (Fig. S5), we estimated the mean number of RNAs produced per cell for 15 min to be ~1.04 ± 0.14. This agrees (statistically) with the above estimation of $N$ (~1.8 ± 0.84). It also agrees with past estimations that, in live cells, transcription initiation locking can occur after less than 5 transcription events [9].

We also extracted the time for transcription events to cease after introducing Novobiocin. For this, we performed additional time-lapse microscopy (1 min interval between images). The number of RNAs produced in individual cells during the observation time were obtained as in (66) and verified by visual inspection. Results in the inset of Fig. 4 show that transcription activity started to be reduced at minute 1 and that no RNA was produced after 10 min, which can be used as a lower bound for $\tau_{active}$ (see above). This agrees with the previous estimation of $\tau_{active}$ (~12 ± 4 min) from Fig. 3.

3.7. Effects of PSB differ with the basal transcription rate

Previous works reported evidence that a gene’s activity affects its own PSB when the gene is on a circular template tethered to a surface [9,87]. We hypothesized that the same occurs on a chromosome-integrated gene, due to discrete topological constraints. This follows from the reasoning that, if the expected time interval between consecutive transcription events becomes longer, while [G] is kept constant, there is more time for Gyrase to resolve transcription initiation locking due to PSB in between transcription events. The model in Fig. 1 accounts for this, as the responsiveness of $r^{-1}$ to changes in [G] should decrease with $k_1$. To test this, we replaced $PLac\text{O}_{3}\text{O}_{1}$ by a native Lac promoter ($PLac$). We chose this promoter because it has similar sequence and known factors.

To an extent, the interpretation of the empirical data relies on the base transcription rate ($r$) of the integrated gene, due to discrete topological constraints. This follows from mutually affect one another [4,14,37,38].

To ensure the model predicted the mean time to lock the promoter due to changes in $r^{-1}$, following Gyrase overexpression, is due to changes in RNAP concentrations, while not perturbing cell growth or morphology. Given this, we developed and validated a method that uses a LineWeaver-Burk plot to dissect, from single-cell, single-RNA data, key kinetic parameters of transcription initiation locking due to PSB. Namely, we dissected the rate of occurrence of these locks and their weight on the effective RNA production rate. Next, we compared with a promoter at the same chromosomal location and similar in structure and regulation but differing in strength. From this, we inferred a range of potential kinetics of transcription initiation locking in a given topological domain that can be achieved by tuning the basal transcriptional rate of the gene of interest. Relevantly, the method was sensitive to detect PSB effects causing a minimum of 1.6 fold changes in transcription rates. Further, we estimate that simple enhancements (e.g. increasing the number of data points used for the LineWeaver-Burk plot from 3 to 10) reduced this to 1.2 fold changes. Other improvements (e.g. higher precision in data collection) should further enhance the sensitivity, which should suffice to, e.g., dissect the effects of interference between closely spaced promoters (Supplementary Section X).

To an extent, the interpretation of the empirical data relies on the models and, thus, is necessary to assess their reliability, i.e. the robustness of their predictions. In this regard, we observed that, first, the models accurately estimated how much of the change in $r^{-1}$, following Gyrase overexpression, is due to changes in RNAP numbers (Fig. 3). In detail, the dynamics of the chromosome integrated gene, when corrected for RNAP changes (with this correction relying on the model), only differs from the plasmid dynamics by 3.5% (not statistically significant) while, prior to considering the model, it differed by 28%. Second, the model predicted the mean time to lock the promoter due to PSB ($\tau_{active}$ ~12 min) from qPCR and population level microscopy data. This estimation was validated by direct measurements using time-lapse microscopy data at the single cell level (one image per minute). In detail, estimated and real data differed solely by ~15% that, when accounting for the measurement error, is also not statistically significant (Figs. 3 and 4). Further, the model accurately predicted (in a statistical sense) how much the basal transcription rate affects the
fraction of time spent by the promoter in locked state (Fig. 6). Finally, the estimations of $k_1$ and $k_{\text{unlock}}$ using the models agree with past estimations (respectively in [9,48]).

Overall, the results suggest that the weight of PSB on the effective RNA production rate of a gene depends not only on the mean activity of the DNA loop that the gene belongs to, but also on the basal transcription activity of the observed gene. This dependence was found to be sufficiently strong to require the introduction of this phenomenon in the model, if one is to predict the effects of changing Gyrase levels on the dynamics of transcription (reaction 4 in Fig. 1). This is because the fraction of time spent in locked states depends not only on the rate of accumulation of positive supercoils, but also on how much time Gyrase have to resolve enough supercoils (in between consecutive transcription events) to avoid reaching a supercoiling density that suffices for promoter locking.

Given that increasing the basal transcription rates enhances the influence of PSB on the effective transcription rate, we hypothesize that, at least in some genes, increasing the basal transcription rate may come at the cost of increased transcriptional noise due to PSB, even if lowering the noise from basal transcription dynamics. We thus expect that the relationship between basal transcription rate and PSB needs to be directly accounted for in models of prokaryotic gene expression. As such, when reducing the SN model (Supplementary Table S2) to a minimal model (Fig. 1), one of the critical components kept from the SN model was reaction 4 (Fig. 1), as it is responsible for the production positive supercoils at a rate that differs with the basal transcription rate of the gene interest.

This was required even though, similar to past models [4,9], there is also reaction 5, which introduces positive supercoils from ‘external’ sources, at a rate that differs with the average transcriptional activity of all genes in the same DNA loop (or topological domain) [17,43] and DNA replication [84]. Interestingly, the existence of this dependency suggests that it should be possible to, to some extent, regulate the robustness of chromosome-integrated synthetic circuits to PSB, by tuning its own transcription rates, as well as placing it in a topological domain with desired mean activity.

In this regard, since increasing the basal transcription rate enhances the effects of PSB, is there an effective upper limit on the transcription rate? If so, this could potentially explain (at least partially) why some genes exist in multi-copy form. Such form would allow crossing this limit, while also supporting more stable expression levels.

Meanwhile, the combination of the results from two different constructs suggest that it may be possible to map a state space of transcription initiation locking of the topological domains of E. coli. However, since domain barriers are not likely to be at fixed sites [45,70,92], it may be necessary to set constructs in various regions of the DNA and measure not only the mean, but also the variability of the propensity for transcription locking as a function of DNA location. Using several constructs, differing in features (e.g. in regulatory mechanisms), should allow accounting for changes in parameters, other than than the basal transcription rate. Namely, while here we mapped a 1-dimensional space by tuning the basal transcription rate, changing other variables would facilitate mapping a multi-dimensional state-space of transcription initiation locking kinetics. We expect such mapping to be of use in dissecting global transcription programs of E. coli, as well as for implementing chromosome-integrated synthetic circuits with predictable kinetics.

Our methodology may also assist in quantifying effects of environmental shifts (e.g. temperature) on the kinetics of transcription initiation locking. One could then explore whether E. coli uses this
phenomenon to adapt to fluctuating environments. This hypothesis is supported by recent observations [64] that cold-shock genes have atypical supercoiling-sensitivity (for unknown reasons). I.e., genes with long-term responses to cold-shocks appear to be impervious to supercoiling, while genes with short-term responses have more-than-expected-by-chance sensitivity to supercoiling. Our methodology may assist in dissecting the responsible mechanisms, e.g. by measuring $\tau_{\text{locked}}$ and $\tau_{\text{active}}$ of these genes following mutations, etc.

We expect our methodology to be compatible with other techniques. E.g., it is potentially valuable to combine it with measurements of local DNA supercoiling density, such as trimethylpsoralen intercalation [40], to quantify the relationship between this density and the effects of PSB on transcription. Similarly, it may be valuable to combine it with the method in [48] to dissect the kinetics of rate limiting steps of active transcription initiation. For chromosome-integrated genes, we expect that only by using both methods will be possible to estimate the times spent prior to open complex formation, since models suggest that this state of activity will differ with the kinetics of promoter locking due to PSB [59], due to the expected competition between the formations of closed complexes and locked states.

Further, our methodology should be applicable using other techniques, such as RNA FISH (Fluorescence in situ hybridization) [79] and RNA aptamer-fluorogen systems [12,19,65,82,98].

Finally, our results derived from a first attempt at dissecting the in vivo dynamics of locking of transcription initiation using a Lineweaver-Burk plot. Many questions remain unanswered and require further study. It may turn out that fluctuations in Gyrase concentration have non-uniform effects at the genome-wide level, due to the dependency on basal transcription rates and mean rates of topological domains. Potentially, this could be used by cells as means to activate specific gene cohorts (e.g. of genes sharing the same topological domain), involved in responsive transcriptional programs. It could also be used to change the state of small genetic circuits responsible for triggering response programs to fluctuations in supercoiling density (e.g. fluctuations in supercoiling densities may alter the stable state of a, e.g., genetic switch with genes in different topological domains). If this holds true, the ‘optimal’ level of Gyrase may differ with the environment and/or internal cell state, depending on whether a given gene cohort (supercoiling density dependent) should be active or not.

In conclusion, the methods and results here presented are expected to support near-future research on the role of Gyrase on the global dynamics of gene regulatory networks.

Funding

This work was supported by the Finnish Academy of Science and Letters [to C.P.]; Pirkanmaa Regional Fund [to V.K.]; Tampere University Graduate Program (Finland) [to V.C. and M.B.]; EDUFI Fellowship [TM-19-11105 to S.D.]; Academy of Finland [295027 to A.S.R.]; and Jane and Aatos Erkko Foundation [610536 to A.S.R.]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author's statement

C.P. and A.S.R. conceived the study. C.P. performed data analysis. C.P., M.B., and A.S.R. performed modelling. V.K.K., M.M., V.C., and S.D. performed measurements. C.P., V.K., and A.S.R. drafted the manuscript. The authors declare no competing interests.

Transparency document

The Transparency document associated this article can be found, in online version.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagrm.2020.194515.
DNA supercoiling is a fundamental regulatory principle that modulates DNA topology and gene expression, as highlighted in "C.J. Dorman, M.J. Dorman. DNA supercoiling is a fundamental regulatory principle." 

Interplay between RNA polymerases on DNA is a critical signal regulating the basal expression of the lac operon in Escherichia coli, as discussed in "D. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions." 


