Assessing *in vivo* the impact of gene context on transcription through DNA supercoiling

Ihab Boulas,[1] Sylvie Rimsky,[1] Olivier Espeli,[1] Ivan Junier,[2,*] and Olivier Rivoire[1,†]
[1] *Center for Interdisciplinary Research in Biology (CIRB),*
Collège de France, CNRS, INSERM, Université PSL, Paris, France
[2] *Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France*

Gene context can have significant impact on gene expression but is currently not integrated in quantitative models of gene regulation despite known biophysical principles and quantitative *in vitro* measurements. Conceptually, the simplest gene context consists of a single gene framed by two topological barriers, known as the twin transcriptional-loop model, which illustrates the interplay between transcription and DNA supercoiling. *In vivo*, DNA supercoiling is additionally modulated by topoisomerases, whose modus operandi remains to be quantified. Here, we bridge the gap between theory and *in vivo* properties by realizing in *Escherichia coli* the twin transcriptional-loop model and by measuring how gene expression varies with promoters and distances to the topological barriers. We find that gene expression depends on the distance to the upstream barrier but not to the downstream barrier, with a promoter-dependent intensity. We rationalize these findings with a first-principle biophysical model of DNA transcription. The model integrates binding, initiation and elongation of RNA polymerases parametrized with available *in vitro* measurements, as well as the action of topoisomerases for which parameters are constrained by our experimental results. By comparing it with the data, our biophysical model supports that TopoI and gyrase must both act specifically, respectively upstream and downstream the gene, and predicts TopoI to be less active than gyrase. It also highlights antagonistic effects of TopoI, which both facilitates elongation and tends to repress initiation. Altogether, our work sets the foundations for a systematic and quantitative description of the impact of gene context on gene regulation.

Gene regulation is most often studied through the lens of transcription factors, leading to its representation as regulatory networks where gene context – the relative location and orientation of genes along DNA – is abstracted away. This simplification has important limitations. It cannot explain, for instance, how reduced bacterial genomes with very few transcription factors generate intricate patterns of gene expression [1–3]. While multiple factors other than transcription factors may be invoked [4–6], the confrontation of transcriptional data with comparative genomics reveals that gene context plays a primary role, at least in bacteria [7]. Accordingly, the expression of a transcription reporter cassette depends strongly on its location along the *E. coli* chromosome [8]. Experimental data also show that a given synthetic regulatory network can behave qualitatively differently in different genetic contexts. [9]. Genome organization is correspondingly found to be evolutionarily more conserved than transcription factor regulation in natural genomes [10]. Yet, how gene context affects gene expression remains poorly understood.

Gene context may impact gene expression in different ways. In bacteria, a simple but pervasive effect is transcriptional read-through, where the absence or the over-riding of terminators cause a downstream co-directional gene to be co-transcribed with an upstream gene [11]. RNA polymerases (RNAPs) may also interact physically, leading to different forms of transcriptional interference [12]. Additionally, the activity of different RNAPs may be coupled indirectly through mechanical perturbation of DNA. Supercoiling, the over- or under-winding of the double helix, is indeed known to affect and to be affected by transcription [13–15]: as an RNAP transcribes, it exerts a mechanical stress on DNA which causes the double helix to be under-wound upstream and over-wound downstream of the gene [16]. This mechanical perturbation can propagate through distances of several kilo-bases [17] to affect neighboring or subsequent initiations [18] and elongations [14] of transcription. Finally, several proteins can impact transcription by modulating supercoiling. These include topoisomerases, which regulate DNA supercoiling [19], as well as nucleoid-associated proteins [20] which may form topological barriers and prevent the diffusion of supercoiling [21, 22].

Conceptually, the simplest situation where gene context can impact gene expression involves a single gene framed by two topological barriers that prevent the diffusion of DNA supercoiling (Fig. 1A). This defines the "twin transcriptional-loop model" introduced thirty five years ago by Liu and Wang to illustrate the interplay between transcription and supercoiling [16], with negative and positive DNA supercoiling generated upstream and, respectively, downstream of an elongating RNAP (Fig. 1A). This model is nowadays at the foundation of all theoretical studies of the impact of gene context on gene expression [23–26]. It is also central to multiple *in vitro* single-molecule experiments that have led to many insights on the translocation of RNAPs along DNA and

---

[*] Electronic address: `ivan.junier@univ-grenoble-alpes.fr`
[†] Electronic address: `olivier.rivoire@college-de-france.fr`

on the activity of topoisomerases [14, 27–29]. As a result, mechanical and topological constraints generated during transcription are well understood at a quantitative level.

The application of the twin transcriptional-loop model to account for *in vivo* phenomena faces, however, two main difficulties. First, our quantitative understanding is limited with respect to the *in vivo* action of topoisomerases and, more particularly in *E. coli*, the relative contribution of DNA gyrase, which removes positive supercoils, and TopoI, which removes negative supercoils. For instance, high-throughput *in vitro* single-molecule assays suggest that the accumulation of positive supercoiling ahead of transcription and its transient release by gyrase produces transcriptional bursts [30] but whether this scenario explains the burst observed *in vivo* depends critically on whether gyrase is limiting *in vivo*, as it has been shown for instance for TopoI [31]. The issue is not only quantitative as the main mode of action of topoisomerases is also not clear: topoisomerases may indeed act either unspecifically or specifically, where specificity may involve DNA motifs [19], DNA mechanical states [32], or interactions with RNAPs [33]. A second difficulty is the diversity of promoter sequences present in genomes, which are well known to differ not only in strength but in their response to DNA supercoiling [34]. These phenotypes cannot currently be predicted accurately from promoter sequences and generally conceal a diversity of underlying physical parameters, including binding, unbinding and initiation rates. As a consequence of these two difficulties, our conceptual and *in vitro* understanding of the interplay between transcription and DNA mechanics cannot presently be applied to a quantitative description of the *in vivo* impact of gene context on gene expression.

Here, we address these difficulties by implementing *in vivo* in *E. coli* different instances of the twin transcriptional-loop model with a single gene insulated from its neighbors (Fig. 1B). We realize this insulation using DNA bridging proteins that we place at varying distances to a range of different promoters (Fig. 1C) and use the data to constrain a first-principle biophysical model of gene transcription where the only free parameters are the mode and intensity of action of topoisomerases. The resulting theoretical model accounts quantitatively for our experimental results and further makes predictions on the mode of action of topoisomerases. Altogether, the combination of our experimental and theoretical models provides a critical missing link between conceptual models, *in vitro* measurements and *in vivo* phenomena, thus paving the way towards a quantitative understanding of the impact of gene contexts on gene expression.

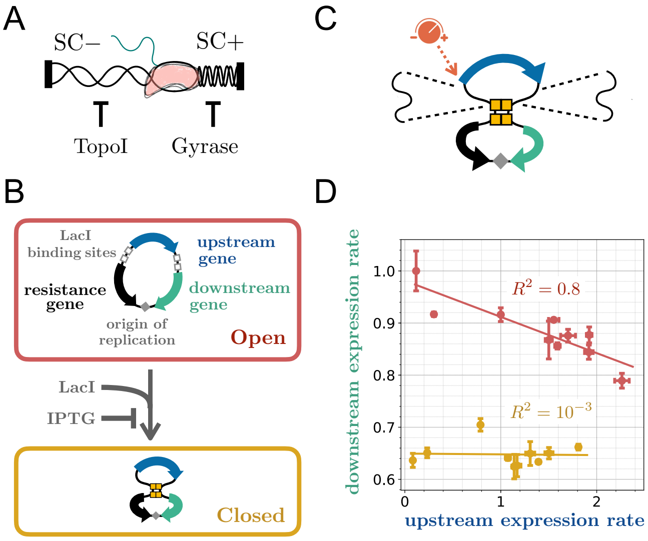

FIG. 1: Conceptual and experimental models – **A.** The conceptual "twin transcriptional-loop model" consists of a single gene delimited by two barriers that prevent the diffusion of supercoiling [16]. A transcribing RNA polymerase generates negative supercoiling upstream and positive supercoiling downstream, which may eventually hinder further transcription due to torsional torques. In *E. coli*, just as in most bacteria, two topoisomerases can resolve these constraints: TopoI, which relaxes negative supercoils, and DNA gyrase, which relaxes positive ones. **B.** We implemented this model on a plasmid with two genes coding for fluorescent proteins, here indicated as upstream and downstream genes, and an antibiotics resistance gene. The upstream gene is flanked by tandems of LacI binding sites: when expressed, LacI forms two loops between which supercoiling cannot diffuse [21, 22], thus insulating the upstream gene. **C.** We built several such systems that differ by the promoter sequence of the upstream gene and the length of the genetic sequences that separate it to the two barriers. **D.** Expression rate of the downstream gene versus expression rate of the upstream gene for given distances but different promoters of the upstream gene (Tab. S1), measured either in the open (in red) or closed (in yellow) system (see Methods for their inference from microplate-reader measurements). Downstream expression rates are normalized by their largest value and upstream expression rates by that of the promoter used downstream when placed upstream. While the expression rates of the downstream and upstream genes are negatively correlated when the system is open, they become uncorrelated when it is closed, consistent with their transcriptional insulation.

I. AN *IN VIVO* TWIN TRANSCRIPTIONAL-LOOP MODEL

To design a genetic system where the transcription of one gene is insulated from the transcription of any other gene, we built on previous *in vitro* results showing how a pair of a tandem of protein binding sites (here *lacO* bound by LacI) can form topologically insulated loops that prevent the propagation of DNA supercoiling from one loop to the other [21, 22]. We introduced such binding sites on a plasmid comprising two co-directional genes separated by a strong terminator in addition to a resistance gene (Fig. 1B). The upstream fluorescent gene is placed in one loop to represent the insulated gene while the downstream fluorescent gene is placed with the resistance gene in the other loop.

The open system displays an interaction between the two fluorescent genes that illustrates the puzzling impact that gene context may have on gene expression: the activity of the downstream gene decreases linearly by up to 20% upon increasing the activity of the upstream gene by changing its promoter sequence (Fig. 1D). The simplest effect, transcriptional read-through, is inconsistent with the data, as it predicts the activity of the downstream gene to increase, not to decrease. Transcriptional interference, while predicting the downstream gene activity to decrease, also predicts that the upstream gene needs to be at least as expressed as the downstream gene to significantly affect it [35] while we observe that considerably weaker promoters have a significant impact on stronger ones (Fig. 1D). Other effects might then be hypothesized as for instance a repression of its initiation due to an excess of positive supercoiling generated by the upstream gene [23]. However, predicting the behavior of this three-gene system requires, first, to understand and quantify the mechanisms at play in the simpler, yet as we shall see already very rich case of a single insulated gene. This single-gene system, which is obtained by closing the loops, is an instance of the twin transcriptional-loop model and we verify that it effectively decouples the expression of the downstream gene from that of the upstream one (Fig. 1D).

II. EXPERIMENTAL RESULTS

A. Downstream versus upstream context

We first study how transcription-induced supercoiling impacts gene expression *in vivo* by changing the upstream or downstream distances between the gene and the topological barriers (Fig. 1C). An increased distance is indeed expected to provide both more DNA to buffer the accumulation of supercoiling and more binding sites for supercoiling-managing topoisomerases (i.e., TopoI and gyrase) to relax this accumulation (Fig. 1A). If the accumulation of supercoiling has an impact on transcription, increasing these distances should therefore modify gene expression levels.

We thus designed different systems where we varied the promoter of the insulated gene (Table S1) and the distance either to the upstream barrier or to the downstream barrier, using promoter-free regions of the λ phage genome. Our measurements were made independent of plasmid copy number (and therefore indirect plasmid size effects) as well as extrinsic factors of variability [36] by normalizing the gene expression rate of the insulated gene with that of a control gene located in the other topologically insulated loop (the "downstream gene" of Fig. 1B), thus defining a relative expression rate (Methods). To assess the sensitivity of gene expression to its downstream (or upstream) context, we compare this relative expression rate in a system with a long distance to the downstream (or upstream) barrier to that in a system with short distances to the two barriers. We call *downstream (or upstream) susceptibility* the ratio of the two rates. We then study these susceptibilities as a function of the promoter strength, which we take to be the relative expression rate measured when both the upstream and downstream distances are short (Methods).

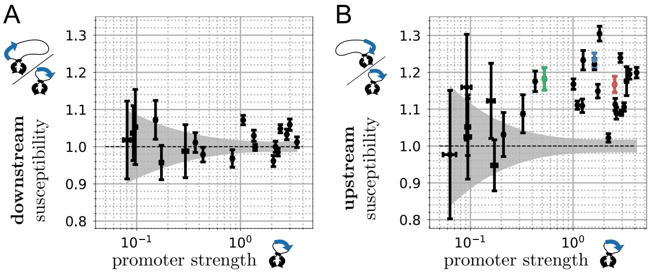

FIG. 2: Susceptibility of gene expression to downstream and upstream contexts – Here we change the context of the insulated gene by introducing a $\sim$ 3 kb sequence either downstream or upstream and consider promoters of varying strengths. **A.** Susceptibility to downstream context versus promoter strength. The downstream susceptibility is defined as the ratio of the expression rate of the insulated gene with a long (3408 bp) distance to the downstream barrier over its expression rate with a short (320 bp) distance. Measurements involving weak promoters are less precise as indicated by the shaded area marking a deviation from unity by less than one standard deviation across replicate measurements (Methods). **B.** Susceptibility to upstream context versus promoter strength. The upstream susceptibility is defined as the ratio of the expression rate of the insulated gene with a long (3205 bp) distance to the upstream barrier over its expression rate with a short ($\sim$ 250 bp) distance. In contrast with downstream susceptibility, it is significantly larger than one for all but one of the strong promoters. The three promoters marked in color are further studied in Fig. 3.

When modifying the downstream distance between the stop codon and the barrier from 320 to 3408 base pairs (bp), we find that gene expression does not vary significantly, irrespectively of the promoter (Fig. 2A). In contrast, when modifying the upstream distance from 250 to 3205 bp between the transcription start site (TSS) and the barrier, we find gene expression to increase by up to 30% (Fig. 2B). For strong promoters (with strength at least 10 times larger than the smallest reported one), this gene expression amplification is statistically significant in all but one case. For weaker promoters, the measurements are less precise and, similarly to the downstream context, we find no evidence of susceptible promoters.

B. Dependence on promoter sequence

The susceptibility to upstream context is not straightforwardly related to the promoter strength: the largest effect is obtained for a promoter whose promoter strength is two-fold smaller than the largest reported one and

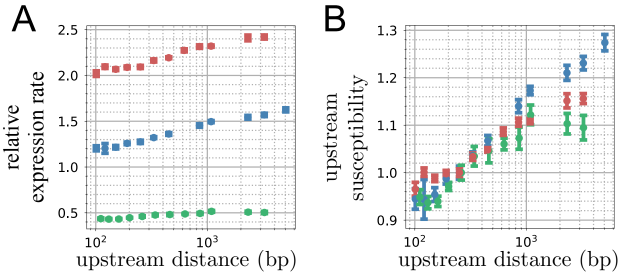

FIG. 3: Dependence on distance to the upstream barrier – **A.** Relative expression rate when varying the distance to the upstream barrier for the three promoters marked in color in Fig. 2B – each point corresponds to an independent measurement. The downstream distance is here 520 bp, while it is 320 bp in Fig. 2B, explaining slight differences of upstream distance at 3205 bp. **B.** Susceptibility to upstream context for the same three promoters. The weak (green) and strong (red) promoters are found to have similar upstream susceptibilities despite having respectively a higher and lower promoter strength than the medium (blue) promoter.

one of the strongest promoters is not susceptible at all (Fig. 2B). Can we rationalize this variability in terms of promoter sequence?

Two factors contribute to promoter strength: binding and initiation. Initiation can be further divided into two steps [37]: the formation of the open complex (OC), which involves a promoter-bound RNAP and a ∼ 12 bp denatured DNA, and promoter escape. The formation of the OC has long been known to be sensitive to supercoiling [38]. Recent work [39] suggests this sensitivity to be primarily modulated by the GC content of the discriminator, a 6 bp long region located immediately downstream of the -10 hexamer. Here, however, we do not observe any significant correlation between the GC content of the discriminator and the upstream susceptibility (Figs. S1-S2).

C. Dependence on upstream distance

Is there an experimental parameter with a systematic impact on upstream susceptibility? Or might the variability of upstream susceptibilities from promoter to promoter conceal an uncontrolled source of variability? An answer is provided by analyzing in more depth how the expression of three specific promoters with different promoter strengths – referred in the sequel as "weak", "medium" and "strong" – depends on the distance to the upstream barrier. In contrast to its intricate dependence on promoter sequence, upstream susceptibility indeed appears to have a simpler, monotonous dependence on the upstream distance (Fig. 3). More specifically, for the three investigated promoters, the susceptibility increases sub-linearly up to distances of the order of 1 kb, beyond which we observe two behaviors with no obvious relationship with promoter strength: on the one hand, the upstream susceptibility of the weak and strong promoters saturate at an amplification of approximately 20% while, on the other hand, that of the medium promoter keeps increasing roughly logarithmically. We also examined the temperature dependence of the medium promoter, finding that its expression levels depend on temperature but not its susceptibility to upstream context (Fig. S3).

III. AN *IN SILICO* TWIN TRANSCRIPTIONAL-LOOP MODEL

Can we build a first-principle biophysical model of the *in vivo* twin transcriptional-loop model that accounts for the different experimental results, namely (i) a susceptibility to upstream context but not to downstream context (Fig. 2), (ii) the dependence of the susceptibility to the distance to the upstream barrier (Fig. 3) and (iii) the non-trivial relationships between the upstream susceptibility and promoter strength (Figs. 2B and 3B)?

A. Modeling transcription

To tackle this problem, we first built a minimal biophysical model of transcription by considering five major stages: RNAP binding to the promoter, formation of the OC, promoter escape, RNAP elongation and transcription termination (Fig. 4). Except termination, which is considered to occur immediately when an RNAP reaches the end of the gene, each of these stages is modeled as a stochastic process with a corresponding rate, with OC formation and RNAP elongation being the only processes sensitive to supercoiling (see below). We further constrain binding to occur only when the promoter is free, i.e., no other RNAP is present within $\ell_{\mathrm{RNAP}} = 30$ bp [40]. As in previous quantitative models [23–26] and consistent with *in vivo* experiments [41], we treat elongating RNAPs as topological barriers and assume DNA supercoiling to relax quickly relative to other time scales [22] so that the supercoiling density is uniform between successive topological barriers (Methods).

B. Known parameters

Elongation involves fixed parameters known from single-molecule measurements (Table I). Namely, RNAP translocation speed has been shown to be a sigmoid function of DNA supercoiling density [42]. Here, we consider a simple binary approximation of this dependence and assume elongating RNAPs to translocate at full speed v_m provided the upstream (downstream) supercoiling densities are above (below) a supercoiling threshold σ_s ($|\sigma_s|$). Below σ_s (above $|\sigma_s|$), RNAPs remain immobile. σ_s thus reflects RNAP stalling as a consequence of the large torque exerted by supercoiled DNA [42] and we take $\sigma_s = -0.062$ (Methods). For the elongation speed, we

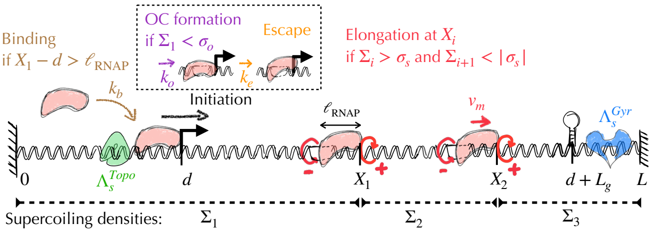

FIG. 4: Schematic representation of our biophysical model of transcription under topological constraints – Transcription includes promoter binding by an RNAP, initiation of elongation which is divided into OC formation and promoter escape, elongation and termination. Elongating RNAPs behave as topological barriers and generate negative supercoils upstream (clockwise red arrows) and positive supercoils downstream (counterclockwise red arrows). The gene is embedded in a domain of length L that is topologically constrained at its extremities. If N RNAPs are elongating (here $N = 2$), $N + 1$ independent topological domains are present whose supercoiling densities are denoted by Σ_i ($i = 1, .., N + 1$). We further indicate the specific action of TopoI (green shape) and gyrase (blue shape) at the extremities of the gene. In addition, TopoI and gyrase may act non-specifically anywhere along the segment.

A. Known parameters		
L_g	gene length	900 bp
L	distance between the 2 barriers	varying
d	distance to upstream barrier	varying
n	number of bp per B-DNA helix	10.5 bp
ℓ_{RNAP}	RNAP exclusion length	30 bp
v_m	elongation speed	$25\,\text{bp.s}^{-1}$
σ_s	SC threshold for elongation	-0.062
$\lambda_{ns}^{\text{Gyr}}$	gyrase non-specific activity	-10^{-4} Lk.bp^{-1}.s^{-1}
B. Expected range of promoter parameters		
k_b	binding rate	$\in [0.01, 10.24]\text{s}^{-1}$
k_o	basal rate for OC formation	$\in [0.01, 10.24]\text{s}^{-1}$
σ_o	SC threshold for OC formation	$\in [-0.05, 0]$
k_e	escape rate	$\in [0.01, 10.24]\text{s}^{-1}$
C. Unknown topoisomerase parameters		
$\lambda_{ns}^{\text{Topo}}$	TopoI non-specific activity	
Λ_s^{Gyr}	Gyr specific downstream activity	
Λ_s^{Topo}	TopoI specific upstream activity	

TABLE I: Model parameters – Our biophysical model involves three types of parameters: **A.** system parameters whose value is either known from the literature or fixed by experimental design; **B.** parameters characterizing each promoter, which are generally unknown and for which we consider a range of values; **C.** unknown parameters related to the *in vivo* activity of topoisomerases, which we estimate using our experimental results. SC stands for supercoiling, OC for open complex, Topo for topoisomerase I and Gyr for gyrase.

take $v_m = 25\,\text{bp.s}^{-1}$, a value reported both in single-molecule *in vitro* experiments [42] and in *E. coli* growing in minimal medium [43], as used in our experiments.

C. Range of promoter-dependent parameters

Binding, OC formation and promoter escape provide a coarse-grained decomposition of the multiple steps of transcription initiation [37, 44]. Kinetic details of each of these stages depend on promoter sequence, but the relationship remains poorly understood [44]. Here, to reflect the diversity of promoter sequences, we consider binding and escape to respectively occur at rates k_b and k_e with values uniformly sampled in the range $[0.01, 10.24]\ \text{s}^{-1}$ [44, 45] (Table I, Methods). For simplicity, we do not explicitly consider promoter unbinding, which we subsume in k_b, nor abortive initiations, which we subsume in k_e.

Next, OC formation involves DNA denaturation. Without DNA supercoiling, kinetics of this denaturation is slow, i.e., the corresponding free energy barrier is high and reflects promoter sequence [44]. DNA supercoiling reduces this barrier, mostly independently of the promoter sequence [46]. Considering that transcription initiation has a sharp sigmoid-like dependence on supercoiling [47] with a negligible rate above a certain threshold, the simplest description of the OC formation is to assume a non-zero rate k_o only if the promoter supercoiling is below a threshold σ_o (Methods). Here, we take k_o in the same range as k_b and k_e ($[0.01, 10.24]\ \text{s}^{-1}$) and σ_o in the range $[-0.05, 0]$ [39].

Finally, we found k_b and k_e to have very similar effects on the results (Fig. S4). For the sake of simplicity, we thus consider in the sequel that promoter escape is immediate once OC is formed ($k_e = \infty$) and discuss only the effect of k_b.

D. Introducing topoisomerase activity

In presence of topological barriers, transcription-generated DNA supercoils may generate strong variations of DNA supercoiling – all the stronger that barriers are closer – which need to be relaxed for transcription to proceed. We thus introduce in our model the stochastic action of TopoI, which removes negative supercoils, and of DNA gyrase, which removes positive ones. We assume that TopoI is active only when the supercoiling is below -0.05, as reported *in vivo* [48], and that gyrase is active only above σ_s to prevent supercoiling from drifting away. Importantly, the *in vivo* modus operandi of topoisomerases remains poorly understood, with distinct scenarios being discussed in the literature (see e.g. [30, 31]). To be comprehensive, we thus consider two non-exclusive scenarios by which each of the two topoisomerases may relax DNA supercoiling.

On the one hand, topoisomerases may act non-specifically at any site (except, to simplify the handling of volume exclusion between DNA enzymes, between two elongating RNAPs). In this case, the corresponding production rates of supercoils, $\lambda_{ns}^{\rm Topo}$ and $\lambda_{ns}^{\rm Gyr}$, are in units of Lk (linking number) per second and per base-pair, meaning that the non-specific activity of topoisomerases depends on the length of the corresponding topological domain. Based on *in vitro* measurement of activity and *in vivo* measurements of the density of active gyrases along DNA [49], we consider $\lambda_{ns}^{\rm Gyr} = -10^{-4}$ Lk.bp^{-1}.s^{-1}. No corresponding measurement is available for $\lambda_{ns}^{\rm Topo}$ and we therefore estimate below an upper bound value using our experimental results.

On the other hand, TopoI and gyrase may act specifically, i.e., at a precise location along the transcription process. In this case, the production rates of supercoils, $\Lambda_s^{\rm Topo}$ and $\Lambda_s^{\rm Gyr}$, are in units of Lk per second, meaning that the specific activities of TopoI and gyrase do not depend on DNA length. Here, in agreement with the reported systematic localization of TopoI at the promoter of genes in various bacteria including *E. coli* [33, 50, 51], we consider the possibility for TopoI to act specifically upstream of transcribing RNAPs. In agreement with gyrase resolving the accumulation of positive supercoiling extremely efficiently [52] and having a biased distribution along bacterial genomes that reflects transcription activity [32, 50], we also consider the possibility for gyrase to act specifically downstream of transcribing RNAPs. As no *in vivo* measurement is available for $\Lambda_s^{\rm Topo}$ and $\Lambda_s^{\rm Gyr}$, we therefore use our experimental results to delineate possible values.

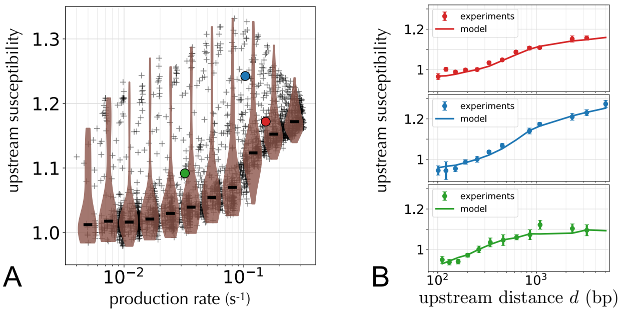

FIG. 5: Upstream susceptibilities in the biophysical model – **A.** Susceptibility to upstream context versus promoter strength for the range of parameters indicated in Table I. Horizontal lines of the violin plots indicate median values. **B.** Upstream susceptibility as a function of the upstream distance obtained in experiments (i.e. results of Fig. 3B) compared to the same quantity obtained in our model for three promoters indicated by colored dots in panel A (see Methods for the values of parameters).

E. Simulations

To implement the transcriptional-loop model, we embed a gene of fixed size $L_g = 900$ bp in a larger domain of size L with the extremities $x = 0$ and $x = L$ defining topological barriers (Methods, Fig. 4). The transcription start site is located at $x = d$ such that the upstream and downstream distances are given by d and $L - L_g$, respectively. Simulations of the transcription process implement the stochastic dynamics of RNAP binding, OC formation, promoter escape, elongation and topoisomerase activities using a discrete-time approach. Transcription rates are measured in a stationary regime by computing the number of transcripts produced per unit of time. Susceptibilities are measured as in experiments by computing the ratio of transcription rates obtained at two different distances (Methods).

IV. MODELING RESULTS

A. Parametrizing topoisomerase activity

First, as the upstream distance increases and in absence of specific activity of TopoI, the non-specific activity of TopoI must increasingly contribute to the upstream susceptibility up to a characteristic distance on the order of $v_m/(n\lambda_{ns}^{\rm Topo})$ where the susceptibility saturates (Fig. S5; $n = 10.5$ is the number of base pairs per DNA helix). The absence of saturation for the medium promoter up to at least $d_{\rm max} = 5$ kb in Fig. 3B thus suggests $\lambda_{ns}^{\rm Topo}$ to be smaller than $v_m/(nd_{\rm max}) \simeq 5.10^{-4}$ Lk.bp^{-1}.s^{-1}. In the following we consider $\lambda_{ns}^{\rm Topo} = 10^{-4}$ Lk.bp^{-1}.s^{-1}, identical to the known value of $-\lambda_{ns}^{\rm Gyr}$.

Second, in absence of any specific activity of either TopoI or gyrase, every stalling of an RNAP would last of the order of 10 to 1000 s – corresponding to $(\lambda_{ns}^{\rm Topo}d)^{-1}$ and $(\lambda_{ns}^{\rm Gyr}(L - L_g))^{-1}$. The production rate of any promoter, including the strongest ones, would therefore be very low, considering $\sim L_g/n \simeq 85$ stalling events. This demonstrates the necessity to consider a specific activity for both TopoI and gyrase, respectively upstream and downstream the gene. We therefore tested a wide range of values of $\Lambda_{ns}^{\rm Topo,Gyr}$ (Fig. S6) and assessed the capacity of the model to reproduce two properties of the dependence of the upstream susceptibility on promoter strength displayed in Figure 2B (where the largest distance is fixed to $d = 3205$ bp): a maximum susceptibility at ~ 1.3 and the susceptibility of the strongest promoters lying between 1.1 and 1.2. The combination $\Lambda_s^{\rm Gyr} = -2.5$ Lk.s^{-1} and $\Lambda_s^{\rm Topo} = 1.4$ Lk.s^{-1} fulfills these requirements. We retain here these values but note that they are not the only one compatible with our results (Fig. S6). More generally, we find that $\Lambda_s^{\rm Gyr}$ should be ≤ -2 Lk.s^{-1} while the corresponding $\Lambda_s^{\rm Topo}$ should lie between 1 and 2 Lk.s^{-1} (Fig. S6). Interestingly, *in vitro* single-molecule experiments have reported a similar value of production rate of supercoils by TopoI, i.e., 1 Lk.s^{-1} [53]. In addition, our inference that $\Lambda_s^{\rm Topo} < |\Lambda_s^{\rm Gyr}|$ and $\Lambda_s^{\rm Topo} < v_m/n \simeq 2.4$ Lk.s^{-1} is consistent with recent *in vivo* results showing that TopoI is limiting for transcription elongation in *E. coli* [31].

B. Capturing promoter variability

Given the above parameters, we can now study how the upstream susceptibility varies both from promoter to promoter and with respect to the distances to the topological barriers. First, we verify the absence of downstream effects (Fig. S7) as a consequence of the downstream barrier only impacting elongation in our model. In the stationary regime, the average time between two transcript productions is indeed independent of the mean elongation time and only reflects initiation times. Moreover, TopoI being less efficient than gyrase ($\Lambda_s^{\mathrm{Topo}} < |\Lambda_s^{\mathrm{Gyr}}|$), elongation and its possible effects on the initiation stage are controlled mainly by the upstream context.

Second, and consistent with Fig. 2B, we verify in Fig. 5A that the upstream susceptibility is not a simple function of promoter strength. More precisely, we obtain an overall shape of the distribution of susceptibilities very similar to experimental results where most of the weakest promoters are not susceptible and most of the strong promoters have a susceptibility above 1.1, with a large variability among strong promoters. The correspondence of the maximal value and variability of the susceptibilities of the strong promoters is expected given that we tuned $\Lambda_s^{\mathrm{Topo}}$ and Λ_s^{Gyr} to capture these features. The correspondence nevertheless extends to the weakest promoters whose insensitivity to upstream context is reproduced without involving any additional fit. Furthermore, we also have the highest susceptibilities occurring for promoter strengths approximately threefold lower than the maximum one. Even more significantly, although we constrained the unknown topoisomerase parameters based on the values of susceptibilities measured at a single upstream distance $d = 3205\,\mathrm{bp}$, our model quantitatively reproduces the full dependence of upstream susceptibility as a function of upstream distance. This is illustrated in Figure 5B where we show how we can find values of k_b, k_o and σ_o for each of the three promoters studied in Figure 3 so as to reproduce the full dependence of their susceptibility as a function of upstream distance. These values are in fact tightly constrained (Fig. S8). For instance, good matches between experimental and theoretical results across all upstream distances as observed in Figure 5B impose to respectively use $\sigma_o = -0.04$ and $\sigma_o = -0.05$ for the weak and medium promoters (Fig. S8). This suggests that our approach may be used to infer promoter parameters.

C. Explaining promoter variability

The dependence of transcript production on the upstream distance reflects antagonist effects that TopoI has on elongation and initiation. TopoI activity is indeed necessary to rescue RNAPs from stalling, and therefore enable elongation, but this activity causes the supercoiling density to jump by finite amounts, which can generate an “excess” of positive supercoiling that inhibits initiation by repressing OC formation (Fig. 6A). This inhibitory effect is prevalent at short upstream distances when TopoI activity induces strong variations of upstream DNA supercoiling density that cause the supercoiling density to be frequently above σ_o, the threshold above which OC formation is prevented. In contrast, variations of TopoI-generated supercoils are dampened by a long upstream distance, with no impact on OC formation (Fig. 6B).

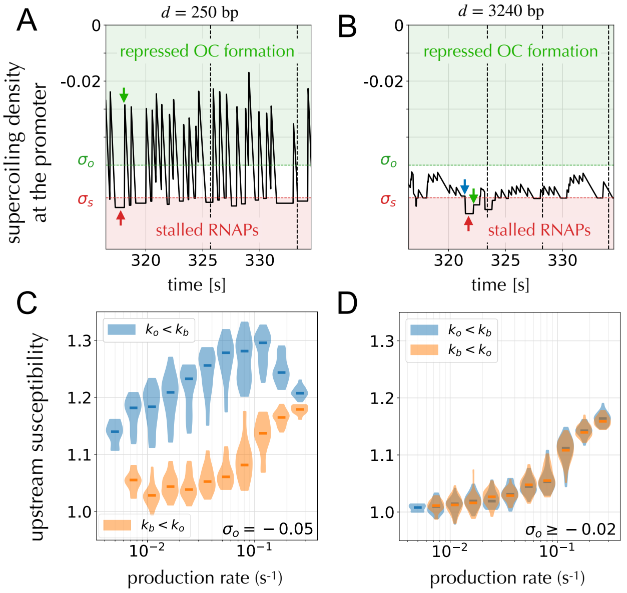

FIG. 6: Simulation results rationalizing upstream susceptibilities – **A.** Values of the DNA supercoiling density at the promoter over a window of time in the stationary regime for short upstream distance ($d = 250\,\mathrm{bp}$) and a promoter corresponding to the medium promoter of Fig. 5B. The large positive jumps are the consequence of TopoI adding one supercoil (green arrow), while decreases are induced by RNAP translocations up to points where DNA supercoiling density is equal to the stalling threshold (σ_s, red dashed line; the red arrow indicates RNAP stalling). When the upstream supercoiling density is above σ_o (green dashed line), OC formation is repressed, preventing new initiations (vertical black dashed lines). **B.** Same as in panel A but for a long upstream distance ($d = 3240\,\mathrm{bp}$), in which case positive supercoils added by TopoI are damped and the upstream supercoiling density remains below σ_o. The blue arrow indicates a non-specific action of gyrase. **C.** Same as in Fig. 5A but considering promoters with $\sigma_o = -0.05$ and distinguishing between those limited by binding ($k_b < k_o$, in orange) and those limited by OC formation ($k_o < k_b$, in blue). **D.** Same as in panel C but considering promoters with $\sigma_o \geq -0.02$, showing no difference between binding-limited promoters and OC formation-limited promoters.

We may also understand how promoters with comparable strength can respond differently to the presence of an upstream barrier by considering two underlying time scales: k_b^{-1}, the time-scale of promoter binding, and k_o^{-1}, the time-scale of OC formation. Indeed, while promoter strength depends roughly symmetrically on k_b and k_o, DNA supercoiling has a direct impact only on OC formation. Promoters with $k_b < k_o$, i.e., limited by OC formation rather than by binding, are therefore expected to be more sensitive to changes of the upstream distance than those with $k_o < k_b$ (Fig. 6C). This effect depends on the value of σ_o, as the lower σ_o is, the more likely it is for the activity of TopoI to prevent OC formation. Here, in agreement with promoter supercoiling densities typically not exceeding -0.02 for an upstream distance $d = 250\,\text{bp}$ (Fig. 6A), differences in upstream susceptibility between promoters with limiting OC formation and those with limiting binding are manifest only when $\sigma_o < -0.02$ (Fig. 6D).

V. DISCUSSION

Gene context is recognized as an important determinant of gene expression with several possible mechanisms at play, including local concentration effects, transcriptional read-through, RNAP interferences and DNA supercoiling. It is generally unknown, however, which mechanism – if any – is prevalent in given *in vivo* conditions. A major impediment has been the absence of data from *in vivo* experiments where the gene context is fully controlled. Here we introduced an insulated genetic system that realizes *in vivo* the simplest case, also known as the twin transcriptional-loop model: a single gene transcribed on a DNA segment delimited by two topological barriers. The study of this minimal system suggests that DNA supercoiling is a prevalent mechanism via which genetic contexts affects expression *in vivo*. It also allows us to assess how DNA supercoiling is handled *in vivo* and how it affects gene expression. We find expression rates to be limited by the presence of an upstream topological barrier but not of a downstream topological barrier. The larger the distance to the upstream distance, the larger the expression rate but the susceptibility of a gene depends non-linearly on the distance and is strongly promoter dependent.

To interpret our experimental results, we developed a first-principle biophysical model of transcription with no free parameter but the mode of action of TopoI and gyrase and the values of promoter parameters. In this model, RNAP elongation generates DNA supercoiling on each side of the elongating RNAP, which in turn affects the elongation of other RNAPs as well as OC formation during initiation. DNA supercoiling on both ends of the gene is then modulated by the action of TopoI and gyrase which we considered to be either non-specific, i.e., scaling with the size of the domain, or specific, i.e., localized at the start or end of the gene. We find the model to account for our experimental data only when TopoI and gyrase are allowed to act specifically. In line with previous works on gyrase [30, 50] and TopoI [33, 50, 51] in various bacteria, including *E. coli*, our results thus demonstrate that these topoisomerases are essential facilitators of transcription. Our analysis further reveals that the production rate of supercoils is lower for TopoI than for gyrase. We also find that RNAPs produce more supercoiling than TopoI can handle. Altogether, our findings therefore show that elongation is mainly controlled by TopoI activity.

While TopoI enables elongation, simulations of our biophysical model reveal an additional antagonistic effect at the core of the large upstream susceptibilities: TopoI represses initiation when the distance to the upstream is too short to dampen changes in supercoiling density. This antagonism is topologically inevitable due to the discrete nature of the supercoils that TopoI adds, which unavoidably translate into discrete increases of the supercoiling density whose size is all the larger that the upstream distance to a topological barrier is short.

Nevertheless, variability in upstream susceptibility is observed among strong promoters which reflects two distinct contributions to promoter strength that are differentially affected by TopoI-induced supercoiling: promoters limited by binding are nearly insensitive to upstream context, while those limited by OC formation are sensitive. Our model indicates that the latter occurs when both the OC formation rate k_o is smaller than the binding rate k_b and when σ_o, the threshold over which OC formation is permitted, is sufficiently close to the RNAP stalling density, σ_s. Predicting the susceptibility of a specific promoter therefore requires the three parameters k_b, k_o and σ_o – to which in practice k_e must be added, which is encompassed in k_b in our model. While a systematic inference of these parameters is beyond the scope of the present work, our study of multiple promoters over a range of parameter values is already highly informative and constrains not only qualitatively but also partly quantitatively the activity of topoisomerases. We thus obtained an upper bound on the non-specific activity rate of TopoI, namely $\lambda_{ns}^{\text{Topo}} < 5.10^{-4}$ Lk.bp^{-1}.s^{-1}, as well as an expected range of values for the specific activity of both gyrA and TopoI, namely $\Lambda_s^{\text{Gyr}} \leq -2$ Lk.bp^{-1}.s^{-1} and $\Lambda_s^{\text{Topo}} \sim 1 - 2$ Lk.bp^{-1}.s^{-1}, respectively. Moreover, reproducing quantitatively the full dependence of the sensitivity of specific promoters on upstream distances as in Fig. 5B strongly constrains the possible values of k_b, k_o and σ_o and hence, provide, a promising road to estimate promoter parameters.

From a genomic perspective, our system purposely defines a limit case where a single gene is fully insulated from other genes. In genomes, no gene is totally insulated from its neighbors but different nucleoid-associated proteins as well as RNAPs themselves may isolate larger groups of genes. In future work, our system could be scaled up to insulate two and more genes and therefore provide valuable information on the consequences of genome organization for gene regulation. In any case, studying the feedback of a single transcribed gene onto itself is a pre-requisite to studies with more intricate gene contexts, as well as a proof-of-concept of their interest.

Additionally, our results are of interest for synthetic biology as they demonstrate a mechanism by which gene expression can be finely controlled. The modulation of gene expression by the distance to an upstream is indeed robust, i.e., independent on the composition of the sequence separating the gene to the topological barriers (Fig. S9), and its simple monotonous dependence is remarkable when contrasted with the complex dependence to promoter sequences (Fig. S2). This is all the more remarkable that the effects are comparable in magnitude to modifying the up-element sub-structure of a promoter (Fig. S10).

Identifying which effects are robust and therefore amenable to an explanation and to experimental control is essential both to the theory and the engineering of biological processes. Counterintuitively, our results suggest that, for transcription, gene context may be more amenable to quantitative explanations and experimental control than promoter sequences despite involving long-range indirect coupling between DNA and RNA polymerases.

VI. METHODS

Strains and plasmids. All measurements were carried out in the *E. coli* MG1655 background. The genetic constructs for the minimal system use the pSC101 origin of replication making it a low copy plasmid and Kanamycin resistant. The upstream gene is the fluorescent gene mCerulean ME, and the downstream the fluorescent gene mVenus ME. Their very similar sequences, comparable folding time and long life times allow for a straightforward comparison of their expression rates. The terminators B0014 and T1 follow mCerulean and mVenus, respectively (SI). For the downstream gene, the Ribosome Binding Site (RBS) is always the same (its sequence can be found in SI) and the promoter is always pR – except for Fig. 1 where it is apFAB61 (see SI). For the upstream gene, the RBS is always apFAB837 and the promoter sequences used can be found in SI. In Fig. 3, the weak, medium and strong promoters are apFAB45, apFAB67 and apFAB70, respectively. Each topological barrier is composed of two tandem lacO biding sites (the plasmid has therefore 4 lacO binding sites in total). The two barriers are also in a tandem orientation with one another. Their sequence differs slightly from that of [21] to avoid unnecessary repeats. Their sequences can be found in SI. The upstream and downstream distances to the barriers were obtained from the PCR of regions of the λ phage genome.

Growth Medium. All of the experiments were carried out in M9 minimal medium using the following recipe: 1X M9 Minimal Salts (from Sigma Aldrich); 0.4% Glucose; 1% Casaminoacids ; 2 mM $MgSO_4$ and 0.1 mM $CaCl_2$. In addition, Kanamycin was added at 50 μg/mL. When IPTG is used, 1mM is added to the media at every step of the protocol below.

Data Acquisition. Glycerol stocks were streaked on resistance agar plates. Single colonies were inoculated in 1 mL of minimal medium with antibiotics, within a 2 mL 96 deepwell plate. Cultures were grown overnight in a thermoblock, at 37° C and 12,000 rpm. Cultures then underwent a 1:500 dilution in 1 mL of minimal medium with antibiotics, within a 2 mL 96 deepwell plate. An outgrowth was run in a thermoblock, at 37° C and 12,000 rpm, for 3h (to an $OD_{600} \approx 0.1$). Cultures then underwent a 1:10,000 dilution and 100 μL of these diluted cultures were aliquoted in a 96 well plates (with black walls and a clear flat bottom). 50 μL of mineral oil was finally added to each well. Time series were acquired in a Tecan Spark Microplate Reader. No Humidity Cassette was used. Temperature was set at 37° C (at least 1h prior to the beginning of the acquisition). Shaking was set on double orbital with amplitude of 3 mm and frequency of 90 rpm. Time points were acquired every 25 minutes, over a total period of $\sim$ 20h. Three quantities were measured at each time point: absorbance (at 600 nm); mCerulean fluorescence (excitation at 430 nm and emission at 475 nm using a manual gain of 90) and mVenus fluorescence (excitation at 510 nm and emission at 550 using a manual gain of 70).

Inference of Gene Expression Rate.

First, the raw temporal data for the optical density and fluorescence is linearly interpolated over 750 points, from the $\sim$ 50 raw data points (using the interp1d module from the SciPy library in Python). The interpolated data is then filtered using a 2nd order polynomial (by a Savitzki-Golay filter using the savgol module from the SciPy library in Python using a window size of 101). The relative differences in gene expression rates are only weakly sensitive to the exact parameters used for the interpolation and filtering.

The gene expression rate (as represented in Fig. 1) is computed as $\alpha_F = (dF_t/dt)/(dN_t/dt)$, where F_t is the fluorescence signal and N_t is the optical density signal (see SI for a justification and a comparison with other approaches). This computation allows for the identification of a region (between the background-dominated early phase and the entrance into stationary phase) during which gene expression rate is stable. This region is identified automatically by minimizing the signal derivative over a temporal region of $\sim$ 1h45. If slower growth (at 29° C instead of 37° C) is used, the duration of the stable signal region extends beyond 5h. Gene expression rate is temporally averaged over this stable region.

The gene expression rate is obtained both for the upstream and downstream genes. They are independently compared in Fig. 1. For Figs. 2-3, the upstream gene expression rate is normalized by that of the downstream gene to remove copy number differences due to changes in plasmid size. This normalization is justified by the independence between the two genes demonstrated in Fig. 1. The relative gene expression rate (as represented in Fig. 2-3) is computed as $\alpha_F^r = (dF_t^{\text{upstream}}/dt)/(dF_t^{\text{downstream}}/dt)$. As for the simple gene expression computation, this computation yields a stable region of gene expression rate which is identified and averaged.

The relative expression rates are obtained for different gene contexts. The strength of a promoter is defined as the relative expression rate measured when both the upstream and downstream distances are short. The susceptibility to the upstream context is the ratio of relative expression rate in the long upstream context over that of in short upstream context, and conversely for the susceptibility to the downstream context. Note, here, that promoter strength should actually be measured in the absence of any context effect. This is nevertheless never the case in practice. We thus checked in our biophysical model that results and conclusions are identical when defining promoter strength from long distances.

For each data point, 4 to 8 replica (constituted of different colonies from a given glycerol stock) were made. Because the inferred gene expression rate comes from a temporal average, we computed the error associated to these replica as the standard error of the mean (their magnitude can be seen in the x and y axes of Fig. 1D for example). The propagation of error when ratios of average gene expression rates are considered (as seen in the y axis of Fig.2A-B) is approximated as

$$\frac{\mu_a \pm \sigma_a}{\mu_b \pm \sigma_b} \approx \frac{\mu_a}{\mu_b} \pm \frac{\mu_a}{\mu_b}\sqrt{(\sigma_a/\mu_a)^2 + (\sigma_b/\mu_b)^2}.$$

where μ stands for mean and σ for standard deviation. In Figs. 2A-B, the grey shadow is obtained by plotting the magnitude of the ratio errors and fitting a 3rd degree polynomial to them. When statistical significance is computed, it is via the independent t-test.

Biophysical model.

The model considers a segment of DNA discretized at the single base level into sites with a topological barrier at $x = 0$, a TSS at $x = d$ and a terminator at $x = d + L_g$ (Fig. 4). The TSS can be occupied by a non-elongating RNAP and sites $X_1, \dots, X_N$ at $d \le X_1 < \dots < X_N < x_p + L_g$ by a varying number N of elongating RNAPs, and we define $X_0 = 0$. Binding can occur if the TSS is free, that is, if $N = 0$ or if $X_1 - d > \ell_{\mathrm{RNAP}}$, with ℓ_{RNAP} the exclusion length of an elongating RNAP. Elongating RNAPs constitute topological barriers and supercoiling relaxes quickly with respect to the time scale of elongation to take an uniform value Σ_i between X_{i-1} and X_i. $\Sigma_i = n\mathrm{Lk}_i/(X_i - X_{i-1}) - 1$, where Lk_i is the corresponding linking number and $n = 10.5$ the number of base pairs per DNA helix. We also define Lk_{N+1}, the linking number of the DNA downstream of the most downstream RNAP and Lk_0 the linking number of the domain when no elongating RNAP is present, and $X_{N+1} = L$ such that $\Sigma_{N+1} = n\mathrm{Lk}_{N+1}/(X_{N+1} - X_N) - 1$.

The dynamics are simulated in discrete time with time unit $\tau_0 = v_m^{-1}$ where v_m is the RNAP translocation speed in bp.s^{-1}. Starting from $N = 0$, a simulation run consisted in performing the following updates, with T the total time of the simulation (in practice, we use a Gillespie algorithm to speed up the simulations):

(1) A new RNAP binds at d with probability $k_b\tau_0\Theta(X_1 - x_p - \ell_{\mathrm{RNAP}})(1 - \delta_p)$ where Θ represents the Heaviside function and $\delta_p = 1$ if the promoter is bound by an RNAP, 0 otherwise.

(2) An RNAP bound at d is considered to form a closed complex with DNA. We then consider the transition to the open complex to occur with probability $k_o\tau_0\Theta(\sigma_o - \Sigma_1)$ where Σ_1 is the supercoiling density behind the last elongating RNAP if there is one ($N > 0$) and $\Sigma_1 < \sigma_o$ otherwise ($N = 0$). (3) The initiation of elongation occurs once the open complex is formed with probability $k_e\tau_0$ (or 1 in the case where $k_e > 1/\tau_0$). The newly elongating RNAP is labeled $i = 1$ and the following updates are made: $N \leftarrow N + 1$, $X_1 \leftarrow d$, $\Sigma_1 \leftarrow \Sigma_1$, $\mathrm{Lk}_1 \leftarrow (1+\Sigma_1)d/n$ and, for $i > 1$, $X_i \leftarrow X_{i-1}$, $\Sigma_i \leftarrow \Sigma_{i-1}$ and $\mathrm{Lk}_i \leftarrow (1 + \Sigma_{i-1})(X_i - X_{i-1})/n$ except Lk_2, which is updated as $\mathrm{Lk}_2 \leftarrow (1 + \Sigma_1)(X_1 - d)/n$.

(4) In the presence of at least one elongating RNAP (i.e., $N \ge 1$), the linking numbers of the upstream and downstream part of the system, Lk_1 and Lk_{N+1}, are updated to account for the actions of TopoI and gyrase. For the non-specific activities, we have $\mathrm{Lk}_1 \leftarrow \mathrm{Lk}_1 + A_T - 2A_G$ and $\mathrm{Lk}_{N+1} \leftarrow \mathrm{Lk}_{N+1} + A'_T - 2A'_G$ where, on the one hand, A_T and A'_T are random variables associated with TopoI activity and drawn from a Poisson distribution with mean $\tau_0\lambda_{ns}^{\mathrm{Topo}}X_1$ (0 if $\Sigma_1 > -0.05$ [48]) and, on the other hand, A_G and A'_G are random variables associated with gyrase activity and drawn from a Poisson distribution with mean $\tau_0\lambda_{ns}^{\mathrm{Gyr}}(L - X_N)/2$ (0 if $\Sigma_{N+1} < \sigma_s$ to prevent supercoiling from drifting away). For the specific activities, we have: $\mathrm{Lk}_1 \leftarrow \mathrm{Lk}_1 + 1$ and $\mathrm{Lk}_{N+1} \leftarrow \mathrm{Lk}_{N+1} - 2$ with respective probabilities $\tau_0\Lambda_s^{\mathrm{Topo}}$ (0 if $\Sigma_1 < -0.05$) and $\tau_0\Lambda_s^{\mathrm{Gyr}}/2$. In the absence of any elongating RNAP, only non-specific activities are considered, and we have $\mathrm{Lk}_0 \leftarrow \mathrm{Lk}_0 + A_T - 2A_G$ with mean of the Poisson random variables A_T and A_G given by $\tau_0\lambda_{ns}^{\mathrm{Topo}}L$ and $\tau_0\lambda_{ns}^{\mathrm{Gyr}}L/2$, respectively.

(5) Each RNAP i, whose order is taken randomly (asynchronous update), moves forwards $X_i \leftarrow X_i + 1$ with probability $\Theta(\Sigma_i - \sigma_s)$ – to avoid artifacts from the discrete nature of the dynamics, we do not consider any exclusion effect between two consecutive elongating RNAPs, supercoiling constrains already preventing RNAPs to pass each other). Following this update, the linking numbers are unchanged but the densities of supercoiling Σ_i and Σ_{i+1} are updated to account for the new distances $X_i - X_{i-1}$ and $X_{i+1} - X_i$.

(6) Any RNAP reaching the terminator at $d + L_g$ is removed ($N \leftarrow N - 1$) and contributes to increase the number of transcripts by one: $M \leftarrow M + 1$.

(7) $T \leftarrow T + \tau_0$.

Using this framework, we estimate the transcription rate $\rho(d)$ of a wide range of promoters (Table I) at various upstream distances d and downstream distance $L - L_g$. This rate is obtained as the number of transcripts M obtained per total time T, $\rho = M/T$. Just as in the experiments, upstream susceptibilities are computed using $\rho(d)/\rho(d = 250\ \mathrm{bp})$. M is taken sufficiently large to get an unbiased estimation of the stationary transcription rate: $M = 10^4$ when testing the full range of promoters

(Fig. 5A) and $M = 10^5$ when analyzing in more detail specific promoters (Fig. 5B).

In Fig. 5A and Fig. 6C-D, we tested 2646 combinations of values of k_b, k_o and σ_o where the frequencies were taken in $\{0.01 \times \sqrt{2}^i\}_{i=0,..,20}$ and σ_o in $\{-0.05, -0.04, -0.03, -0.02, -0.01, 0\}$ ($k_e\tau_0 = 1$ in these figures). In Fig. 5B, the following parameters are used for the weak promoter: $k_b = 0.64.\text{s}^{-1}$, $k_o = 0.04.\text{s}^{-1}$ and $\sigma_o = -0.04$; medium promoter: $k_b = 0.453.\text{s}^{-1}$, $k_o = 0.32.\text{s}^{-1}$ and $\sigma_o = -0.05$; strong promoter: $k_b = 0.32.\text{s}^{-1}$, $k_o = 3.62.\text{s}^{-1}$ and $\sigma_o = -0.04$. $k_e\tau_0 = 1$ for the three promoters (immediate escape).

Stalling torque. The value of the stalling torque $\sigma_s = -0.062$ is based on the relationship $\sigma_s = \Gamma_s/A$ [54], which holds for both super-structured DNA and unstructured DNA. $-\Gamma_s = 18.5\,\text{pN.nm}$ [55] is the RNAP stalling torque and $A = 300\,\text{pN}$ is chosen to be intermediate between the value estimated from single-molecule experiments for super-structured DNA (200 pN) and unstructured DNA (400 pN) [54].

OC formation rate. For the OC formation rate, we consider that the corresponding free energy barrier is reduced by DNA supercoiling independently of the promoter sequence [46]. In this context, the rate can be written as $k_o \exp[-\beta(\Delta G_P + \Delta G_\sigma)]$, where $\Delta G_P > 0$ reflects sequence effect of the promoter and where ΔG_σ reflects mechanical properties of DNA under supercoiling σ. $\beta^{-1} = k_BT$ is the energy unit, with k_B the Boltzmann constant and T the temperature. We can then compare the rate $k_o \exp[-\beta(\Delta G_P)]$ in absence of supercoiling and the rate k_o when ΔG_σ compensates ΔG_P. ΔG_P being in general large with respect to k_BT [44], we have $k_o \exp[-\beta(\Delta G_P)] \ll k_o$. In accord with the sharp dependence of transcription rates as a function of σ [39], we then consider the compensation of ΔG_P by ΔG_σ to occur abruptly at a threshold value σ_o that reflects ΔG_P, i.e., σ_o is promoter dependent. This eventually leads us to use $k_o\Theta(\sigma_o - \sigma)$ for the simplest form of the OC formation rate.

Code availability. A python implementation of the biophysical model is available as a GitHub depository at `https://github.com/ijunier/twinTrans`.

Acknowledgments – We thank Estelle Crozat, Hans Geiselmann, Calin Guet, Thomas Hindré and Bianca Sclavi for fruitful discussions. We acknowledge funding from FRM AJE20160635870, IRIS OCAV Idex ANR-10-IDEX-0001-02 PSL, ANR-18-CE12-0012, LabEx MemoLife and PSL-Qlife.

[1] M. Güell, V. Van Noort, E. Yus, W.-H. Chen, J. Leigh-Bell, K. Michalodimitrakis, T. Yamada, M. Arumugam, T. Doerks, S. Kühner, et al., Science **326**, 1268 (2009).
[2] L. Brinza, F. Calevro, and H. Charles, BMC genomics **14**, 1 (2013).
[3] I. Junier, E. B. Unal, E. Yus, V. Lloréns-Rico, and L. Serrano, Cell systems **2**, 391 (2016).
[4] D. F. Browning and S. J. Busby, Nature Reviews Microbiology **14**, 638 (2016).
[5] S. Miravet-Verde, V. Lloréns-Rico, and L. Serrano, Current Opinion in Microbiology **39**, 89 (2017).
[6] S. Meyer, S. Reverchon, W. Nasser, and G. Muskhelishvili, Current Genetics **64**, 555 (2018).
[7] I. Junier and O. Rivoire, PloS one **11**, e0155740 (2016).
[8] J. A. Bryant, L. E. Sellars, S. J. Busby, and D. J. Lee, Nucleic acids research **42**, 11383 (2014).
[9] A. Nagy-Staron, K. Tomasek, C. C. Carter, E. Sonnleitner, B. Kavčič, T. Paixão, and C. C. Guet, Elife **10**, e65993 (2021).
[10] M. M. Babu, S. A. Teichmann, and L. Aravind, Journal of molecular biology **358**, 614 (2006).
[11] B. Yan, M. Boitano, T. A. Clark, and L. Ettwiller, Nature communications **9**, 1 (2018).
[12] K. E. Shearwin, B. P. Callen, and J. B. Egan, TRENDS in Genetics **21**, 339 (2005).
[13] A. Travers and G. Muskhelishvili, Nature Reviews. Microbiology **3**, 157 (2005).
[14] J. Ma and M. D. Wang, Biophysical Reviews **8**, 75 (2016).
[15] C. J. Dorman, BMC molecular and cell biology **20**, 1 (2019).
[16] L. F. Liu and J. C. Wang, Proceedings of the National Academy of Sciences **84**, 7024 (1987).
[17] L. Moulin, A. R. Rahmouni, and F. Boccard, Molecular microbiology **55**, 601 (2005).
[18] A. Revyakin, R. H. Ebright, and T. R. Strick, Proceedings of the National Academy of Sciences **101**, 4776 (2004).
[19] K. Drlica, Molecular microbiology **6**, 425 (1992).
[20] C. J. Dorman, Biochemical Society Transactions **41**, 542 (2013).
[21] F. Leng, B. Chen, and D. D. Dunlap, Proceedings of the National Academy of Sciences **108**, 19973 (2011).
[22] M. Joyeux and I. Junier, Biophysical Journal **119**, 1215 (2020).
[23] S. Meyer and G. Beslon, PLoS computational biology **10**, e1003785 (2014).
[24] C. Brackley, J. Johnson, A. Bentivoglio, S. Corless, N. Gilbert, G. Gonnella, and D. Marenduzzo, Physical review letters **117**, 018101 (2016).
[25] S. A. Sevier and H. Levine, Physical review letters **118**, 268101 (2017).
[26] S. Tripathi, S. Brahmachari, J. N. Onuchic, and H. Levine, Nucleic Acids Research **50**, 1269 (2022).
[27] L. Bai, T. J. Santangelo, and M. D. Wang, Annu. Rev. Biophys. Biomol. Struct. **35**, 343 (2006).
[28] K. M. Herbert, W. J. Greenleaf, and S. M. Block, Annu. Rev. Biochem. **77**, 149 (2008).
[29] K. C. Neuman and A. Nagy, Nature Methods **5**, 491 (2008).
[30] S. Chong, C. Chen, H. Ge, and X. S. Xie, Cell **158**, 314 (2014).
[31] S. Kim, B. Beltran, I. Irnov, and C. Jacobs-Wagner, Cell **179**, 106 (2019).
[32] D. Sutormin, N. Rubanova, M. Logacheva, D. Ghi-

larov, and K. Severinov, Nucleic Acids Research **47**, 1373 (2019).

[33] D. Sutormin, A. Galivondzhyan, O. Musharova, D. Travin, A. Rusanova, K. Obraztsova, S. Borukhov, and K. Severinov, Nature Communications **13**, 4524 (2022).

[34] B. J. Peter, J. Arsuaga, A. M. Breier, A. B. Khodursky, P. O. Brown, and N. R. Cozzarelli, Genome biology **5**, 1 (2004).

[35] A. C. Palmer, J. B. Egan, and K. E. Shearwin, Transcription **2**, 9 (2011).

[36] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Proceedings of the National Academy of Sciences **99**, 12795 (2002).

[37] K. S. Murakami and S. A. Darst, Current Opinion in Structural Biology **13**, 31 (2003).

[38] P. L. deHaseth and J. D. Helmann, Molecular Microbiology **16**, 817 (1995), URL `https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2958.1995.tb02309.x`.

[39] R. Forquet, M. Pineau, W. Nasser, S. Reverchon, and S. Meyer, Msystems **6**, e00978 (2021).

[40] D. G. Vassylyev, M. N. Vassylyeva, A. Perederina, T. H. Tahirov, and I. Artsimovitch, Nature **448**, 157 (2007).

[41] S. Deng, R. A. Stein, and N. P. Higgins, Proceedings of the National Academy of Sciences of the United States of America **101**, 3398 (2004).

[42] J. Ma, L. Bai, and M. D. Wang, Science **340**, 1580 (2013).

[43] H. Bremer and P. P. Dennis, EcoSal Plus **3** (2008).

[44] E. Ruff, M. Record, and I. Artsimovitch, Biomolecules **5**, 1035 (2015).

[45] W. R. McClure, Annual review of biochemistry **54**, 171 (1985).

[46] A. Revyakin, R. H. Ebright, and T. R. Strick, Proceedings of the National Academy of Sciences **101**, 4776 (2004).

[47] B. El Houdaigui, R. Forquet, T. Hindré, D. Schneider, W. Nasser, S. Reverchon, and S. Meyer, Nucleic Acids Research **47**, 5648 (2019).

[48] E. Zechiedrich, A. B. Khodursky, S. Bachellier, R. Schneider, D. Chen, D. M. Lilley, and N. R. Cozzarelli, Journal of Biological Chemistry **275**, 8103 (2000).

[49] M. Stracy, A. J. Wollman, E. Kaja, J. Gapinski, J.-E. Lee, V. A. Leek, S. J. McKie, L. A. Mitchenall, A. Maxwell, D. J. Sherratt, et al., Nucleic Acids Research **47**, 210 (2019).

[50] W. Ahmed, C. Sala, S. R. Hegde, R. K. Jha, S. T. Cole, and V. Nagaraja, PLOS Genetics **13**, e1006754 (2017), ISSN 1553-7404.

[51] M.-J. Ferrándiz, P. Hernández, and A. G. de la Campa, PLOS Genetics **17**, e1009542 (2021).

[52] R. E. Ashley, A. Dittmore, S. A. McPherson, C. L. Turnbough, K. C. Neuman, and N. Osheroff, Nucleic Acids Research **45**, 9611 (2017).

[53] K. Terekhova, K. H. Gunn, J. F. Marko, and A. Mondragón, Nucleic Acids Research **40**, 10432 (2012).

[54] J. F. Marko, Physical Review E **76**, 021926 (2007).

[55] J. Ma, C. Tan, X. Gao, R. M. Fulbright, J. W. Roberts, and M. D. Wang, Proceedings of the National Academy of Sciences **116**, 2583 (2019).