Dynamics of bacterial operons during genome-wide stresses

Rahul Jagadeesan[1*], Suchintak Dash[1,†], Cristina S.D. Palma[1,2,†], Ines S.C. Baptista[1,†], Vatsala Chauhan[1], Jarno Mäkelä[3], and Andre S. Ribeiro[1,4*]

[1] Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.

[2] Department of Bioengineering, Rice University, Houston, Texas, USA.

[3] Institute of Biotechnology, Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland.

[4] Faculty of Sciences and Technology, Department of Electrical Engineering and CA3 Group of UNINOVA/CTS, NOVA University of Lisbon, Lisbon, Portugal.

* To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the second, third and fourth authors should be regarded as Joint Second Authors.

Email: andre.sanchesribeiro@tuni.fi

Competing Interest Statement: The authors have no competing interests.

Keywords: Genome-wide stresses; Operon expression dynamics; RNAP fall-offs; Internal Promoters.

SUMMARY

Bacterial gene networks have many operons, each coordinating the kinetics of a few genes in time and strength. In *Escherichia coli*, nearly half of these operons have internal promoters, which allow regulating the downstream genes, independently from upstream genes. We studied their role during genome-wide stresses targeting a few key elements of the transcription machinery, namely, RNAP and gyrase. We show that absolute differences in response strength between genes in the same operon follow a sinusoidal function of the distance between them. This results from the combined effect of elongation fall-offs and internal promoters. The frequency of this function is regulated by the positioning of internal promoters, while its amplitude differs with the perturbation strength. Moreover, we find that positive supercoiling buildup partially explains the RNAP fall-off rates. Finally, we report similar sinusoidal patterns in *E. coli* cells subject to several other stresses, as well as in *Bacillus subtilis, Corynebacterium glutamicum, and Helicobacter pylori*. Overall, our results suggest that the strength, number, and positioning of internal promoters in operons are influenced by a genome-wide need to compensate for RNAP fall-offs, and thus provide distal genes in the operon with similar response strength as genes proximal to the primary promoter.

1. INTRODUCTION

All organisms must coordinate the activities of their genes in time and strength. Commonly in bacteria and archaea species, those sets of genes are placed adjacently in the DNA, in structures named operons [1], where they can be transcribed as single units. For this, upstream from the structural genes is (at least) one strictly regulated promoter [2], controlling their expression. This promoter is both where transcription initiates, as well as where most regulation of gene expression is exerted [3–10].

The dynamic correlation between genes in the same operon [11,12] can benefit the assembly of multi-subunit protein complexes [13,14] and the integration of laterally transferred genetic information [12,15]. It may also assist co-regulating functionally related genes [2,16]. In agreement, genes in the same operon are commonly of the same functional class [16–20].

Although genes in an operon share the same upstream promoter (named here '*primary promoter*'), some genes can nevertheless be differentially expressed from others [21], due to *internal* promoters, placed after the coding region of the first (or more) gene(s) of the operon [22]. For example, the *cmk-rpsaA-ihfB* operon in *Escherichia coli* has a primary promoter, *cmkP*, and several internal promoters [22]. The most downstream of the internal promoters, *ihfBp2*, has preference for σ^{38}, while all others have preference for σ^{70}. Thus, while most genes of *cmk-rpsaA-ihfB* are preferentially expressed during exponential growth, *ihfB* is preferentially expressed during the stationary phase.

Another cause for differential expression between genes in the same operon is premature terminations of transcription elongation, i.e. RNA polymerase (RNAP) fall-offs [23–25]. There are several natural sources of RNAP fall-offs, such as pause-prone sequences, particularly when caused by RNA hairpin loops [23,24]. Arrests can also influence, when ternary complexes remain intact, while further RNA synthesis is blocked [26]. Homopolymer A/T tracts in the DNA [27], other attenuator sequences [28–30], and RNAP collisions (including co-directional) can also enhance premature terminations [31]. Moreover, some operons have adaptive 'premature' rho-dependent fall-off rates [32]. In the *gal* operon, in the presence of galactose, adding glycerol causes low transcriptional polarity, while adding glucose causes higher polarity, by enhancing premature fall-offs rates [33]. Finally, environmental factors, such as temperature shifts, can affect premature dissociation rates in some operons, due to operon-specific mechanisms [35].

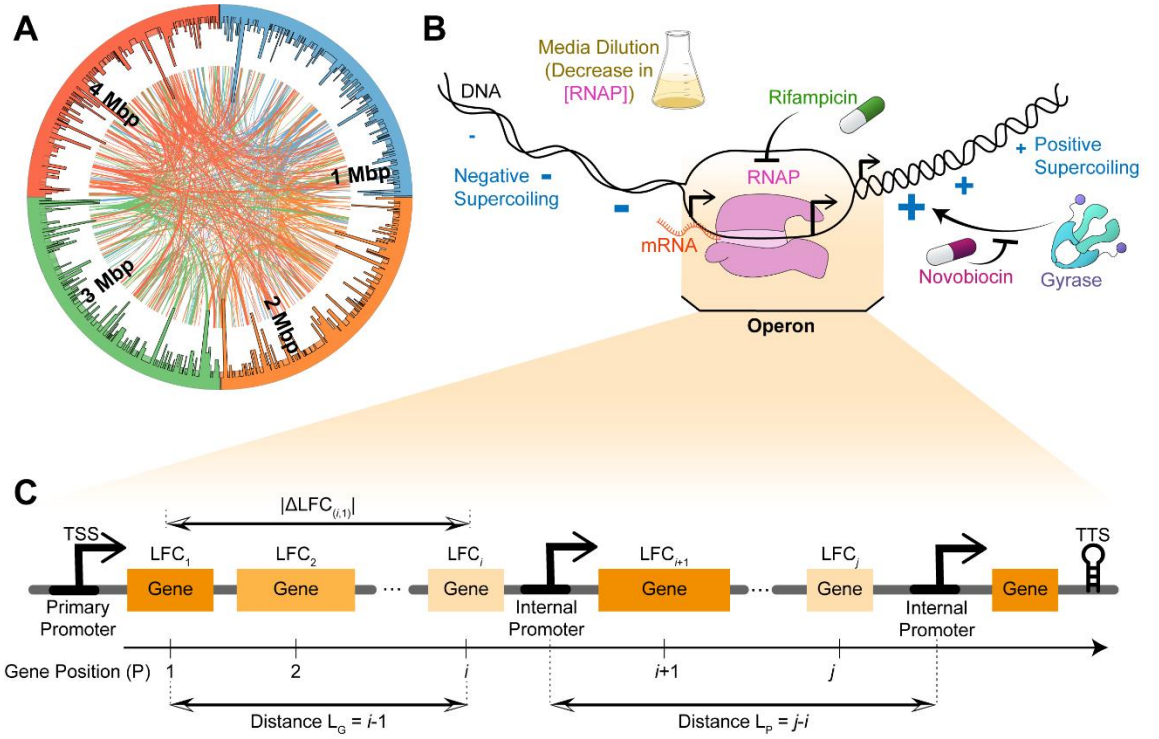

Figure 1: Illustrative summary of our study of the dynamics of operons during genome-wide stresses. **(A)** Illustration of the location of operons in the DNA was constructed using Circos [40]. The external histogram shows the nucleotide length of each operon. Each curved line is a TF interaction between a gene in an operon and another gene. Colors of individual genes are DNA location-based, to facilitate distinguishing interactions. **(B)** Illustration of the three transcription-related, genome-wide stresses studied, specifically caused by media dilution (which reduces RNAP concentration) [40], Rifampicin (which blocks RNAP activity[44]), and Novobiocin (which inhibits gyrase activity) [6,36,40], respectively. Positive and negative

signs illustrate the accumulation of positive and negative supercoiling downstream and upstream the RNAP, respectively. **(C)** Illustration of the features of operons considered, namely, gene positions (*P*) relative to the primary TSS, distance between genes (L_G) and distance between internal TSSs, (L_P). Also shown is a transcription termination site, TTS, present in a few operons. Meanwhile, '|LFC|' stands for the absolute log2 fold change of a single gene response strength to a stress, relative to the optimal condition, and Δ stands for 'difference between'. Finally, the orange gradient coloring the boxes of genes informs about the average responsiveness of the genes within operons to the stresses, as a function of *P*.

Interestingly, during genome-wide stresses, such as media shifts disturbing hundreds of genes at the same time [35,36], the responsiveness from internal promoters may have complex effects on the divergency between upstream and downstream genes. Specifically, the effects likely depend on several factors, including premature RNAP fall-off rates, the location of internal promoters in the operon, and the response strengths of both primary and internal promoters. To study this, we investigated the genome-wide transcription dynamics of the operons of *E. coli* (Figure 1A) during stresses targeting RNAP and gyrase (Figure 1B). Disturbing their activity is known to influence (quasi-synchronously) several hundreds of transcription start sites (TSSs) [36,37] . We analyzed the response strength of each gene, in each operon, as a function of its position (Figure 1C) and as a function of the positioning of internal promoters. In the end, we explore if our findings on the role of internal promoters are general or stress- and/or species-specific, using recently published data on *E. coli* and three other bacteria: *Bacillus subtilis*, *Corynebacterium glutamicum*, and *Helicobacter pylori*.

2. RESULTS

2.1. Internal composition of the operons of *E. coli*.

E. coli has 833 operons (MG1655 strain [6,37,41]), which account for 2708 of the 4724 genes. The nucleotide length of operons and the genes composing them are shown in Supplementary Figures S1A and S1B. Meanwhile, Supplementary Figure S1C shows the number of nucleotides separating genes in the same operon. In general, nearest neighbor genes are closely spaced, as they are separated by ~50 nucleotides on average, regardless of the distance from the primary TSS.

Operons are regulated by a total of 1174 recognized TSS's. However, they only have a total of 174 recognized transcription termination sites (TTSs) [6,37,42]. To investigate this, we studied the distance (in number of nucleotides) between operons and the next gene in the DNA, downstream from the end of the operon. Specifically, we compared when there is versus when there is no TTS in between (red and orange bars in Supplementary Figure S1C, respectively). We find that, for the latter, the average distance is 2× longer. This could be explained by significant RNAP fall-off rates during elongation in non-coding regions of the DNA.

Next, of the 1174 TSSs, 532 (~45%) are internally located, i.e., are downstream from the most upstream gene of the operon, implying that many operons have more than one TSS (Supplementary Figure S1D). Also, the number of internal TSSs is correlated to the nucleotide length of operons (Supplementary Figure S1E). This suggests that internal TSSs could be dynamically pivotal for operons.

Moreover, 422 (51%) operons have known transcription factor (TF) regulation. For comparison, only 483 of the genes not in operons (24%) have TF regulation. Also noteworthy, only 62 operons (~7.7%) are *self-regulated* by a TF (here named "self-TF"). For comparison, 59 (~2.9%) out of 2012 genes not in operons have known self-TF regulation. Overall, operons are more subject to TF regulation than independent genes.

Finally, inside operons, the distance between genes (L_G) and of nucleotides are almost strictly linearly correlated (Supplementary Figure S1F). Thus, for simplicity, from here onwards we identify a gene position (*P*) in an operon by the number of genes from the primary TSS.

2.2 Operons do not influence average single-gene response strengths.

We studied three stresses known to affect hundreds of genes by disrupting transcription. The first stress was caused by Novobiocin, an antibiotic that partially inhibits gyrase [43–47]. Transcription in *E. coli* generates positive supercoiling ahead of the RNAP and negative supercoiling behind it [43,45–48]. While negative supercoiling is removed by a non-energy consuming process, the removal of positive supercoils can take longer, since it requires ATP consumption [49,50]. Moreover, the DNA is organized in discrete, topologically constrained segments. Consequently, positive supercoils buildup (PSB), particularly in more active DNA segments, such as those with highly active operons [51,52].

Novobiocin, by inhibiting gyrase, perturbs DNA supercoiling levels [41,53]. This can influence the TSSs' availability to start transcription [49], RNAP elongation rates [54,55], and the dynamics of open complex formation [49]. Since DNA supercoiling differs between topological domains [56,57], the effect of Novobiocin on operons should differ with their DNA location. We studied both 'strong' (own data, Supplementary Methods 1) as well as 'weak' stress conditions (data from [36]). We classified a stress as 'strong' only when the resulting growth rate differs from the optimal condition (Supplementary Figure S2).

The second stress was caused by introduction of a poorer medium (created by dilutions) (Supplementary Methods 1), known to provoke RNAP concentration to reduce systematically and rapidly, without significantly altering growth rates [56]. All data is from [57] at two time-points: i) following partial changes in RNAP levels and initial gene responses (~120 min after media shift); and ii) following the changes in RNAP levels and changes in TF levels, which subsequently further affect RNA levels (~180 min after media shift, during late exponential growth).

The final stress was caused by Rifampicin (own data, Supplementary Methods 1), an antibiotic that specifically binds RNAP, blocking transcription elongation after the initial 2-3 nucleotides [58]. Rifampin can also affect DNA replication and general stress responses [59,60].

Meanwhile, Supplementary Figure S3 shows the genome-wide distributions of single-gene response strengths (calculated by the log2 fold changes of RNA read counts, *LFC*) to each stress of genes, along with the corresponding mean (μ_{LFC}) and absolute mean ($\mu_{|LFC|}$). Only the latter increases with stress strength. Interestingly, operons do not seem to influence the genes' average, absolute stress responses consistently (with the differences being smaller than the error bars).

2.3 Gene response strengths do not show a consistent trend as a function of their position in the operon.

We next studied the average absolute response strengths of genes in operons, $\mu_{|LFC|}$, as a function of P (Supplementary Figure S4). In some stresses, $\mu_{|LFC|}$ decreases weakly with P, while in others it does not change with P. In all cases, for $P > 10$, the number of data points is smaller than 30, which did not suffice to establish a trend.

From the same data, the mean absolute differences in response strengths between genes in the same operon, $\mu_{|\Delta LFC|}$, change with P (Supplementary Figure S5). In general, genes in position P have response strengths more similar to genes in positions $(P-i)$ and $(P+i)$ than to genes in $(P-(i+1))$ and $(P+(i+1))$, etc., but only for small i (in general, for $i < 4$).

2.4 Differences in response strengths differ non-monotonically with the distance between genes in the same operon.

$\mu_{|\Delta LFC|}$ changes non-monotonically with P (Supplementary Figure S5). To dissect causes, we studied how the distance between genes in the same operon, L_G, affects $\mu_{|\Delta LFC|}$ (Figures 2A to 2C). Surprisingly, in all stresses, $\mu_{|\Delta LFC|}$ can be well fit by a *sine* function of the form:

$$\mu_{|\Delta LFC|}(L_G) = A_0 + A.L_G + B.\sin(2\pi.f.L_G) \qquad (Eq.\ 1)$$

Here, A_0 stands for $\mu_{|\Delta LFC|}(L_G)$ when intercepting the y axis, while A and B are, respectively, the coefficient of the linear term and the amplitude of the best fitting function. Meanwhile, f is the frequency of the wave. Supplementary Table S1 shows the parameter values of each fitted curve. As expected, for all stresses tested, the amplitude increases with the stress strength. More strikingly, the frequency, f, differs little between stresses (ranging only from 0.09 to 0.11, with a mean of 0.10 and a small standard deviation of 0.01). I.e., f appears to be largely independent from both the mechanism affected by the stress, as well as from the stress strength.

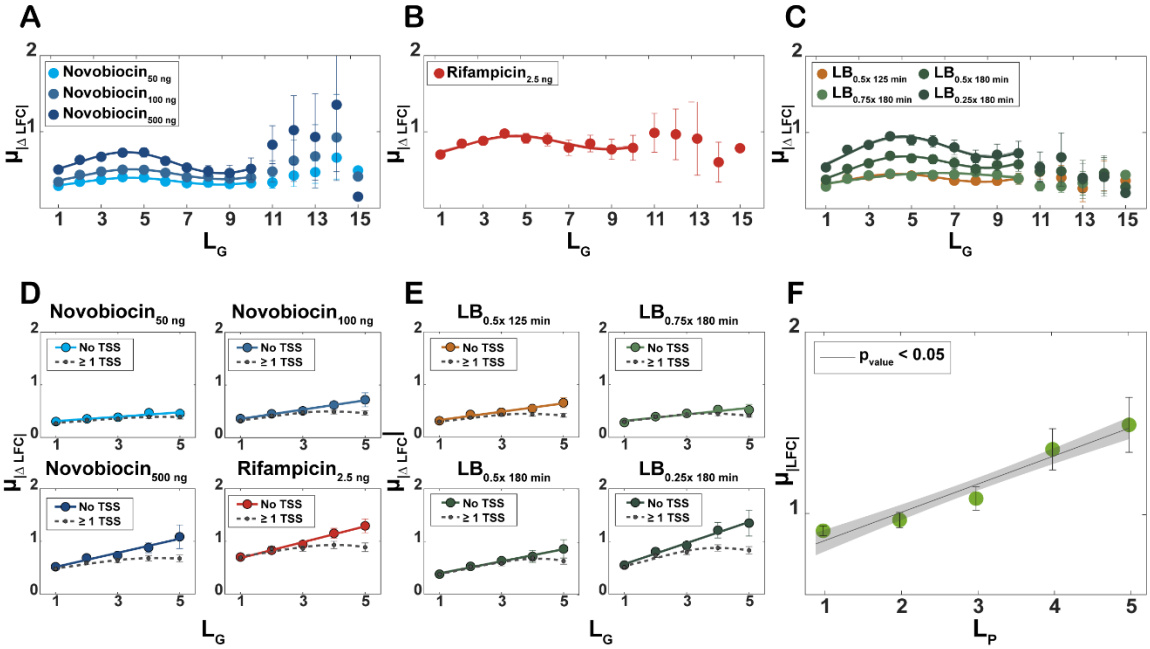

Figure 2: Average absolute differences in the response strengths of genes in the same operon ($\mu_{|\Delta LFC|}$) as a function L_G. Data from all 833 operons of *E. coli*. **(A)** Novobiocin-based stresses. **(B)** Rifampicin-based stresses. **(C)** Shifts in RNAP concentration due to shifts in media richness. From (**A-C**), each data is the $\mu_{|\Delta LFC|}$ of all pairs of genes of the same operon distanced by L_G, and the error bars are the standard error of the mean. Also shown are the best fits of functions of the form of (*Eq. 1*). See Supplementary Table S1, for the best fitting parameter values and R^2 values. **(D, E)** Average absolute difference between the LFCs of pairs of genes in same operon ($\mu_{|\Delta LFC|}$) with and without internal TSS's in between (dotted and solid lines, respectively). Data plotted only for $L_G < 6$ since there are very few pairs of genes without TSSs in between for $L_G > 5$. See Supplementary Table S7, for the best fitting parameter values. **(F)** Average absolute response strength, $\mu_{|LFC|}$, of the first gene immediately downstream of a TSS plotted as a function of L_P (number of genes controlled by that TSS prior to any other downstream TSS, illustrated in Figure 1C). In figures A, B, C, and F, the shadow areas are the 95% confidence intervals (not visible in most cases). In Figures D and E, the dashed lines correspond to the cases (≥ 1 TSS). In (D-F) we also show best linear fits, along with their p values (see Supplementary methods 3).

Interestingly, we observed a similar phenomenon, not previously reported, from measurements of single-cell protein numbers [61]. We considered 444 proteins of cells moved from an overnight culture into a fresh medium [61]. Then, we extracted 207 differences in numbers of proteins of genes in the

same operon ($\mu_{|\Delta Pn|}$). From Supplementary Figure S6 and Supplementary Table S2, $\mu_{|\Delta Pn|}$ follows a similar wave-like pattern as a function of L_G, with an f of ~0.17 Note that, for the same 207 pairs of genes the corresponding f of the sine function fitted to $\mu_{|\Delta LFC|}(L_G)$ has an average of 0.10 and a standard deviation of 0.01 (extracted from all stresses studied).

2.5 Differences in response strengths between genes in the same operon cannot be explained by TF interactions, σ factor preferences, AT-content, or p distance.

We next investigated if TFs influence $\mu_{|\Delta LFC|}$. We plotted $\mu_{|\Delta LFC|}(L_G)$ for genes in the 427 operons that are not subject to TF regulation [60] (Supplementary Figure S7 and Supplementary Table S3). For comparison, we plotted $\mu_{|\Delta LFC|}(L_G)$ for the genes in the 406 operons subject to TF(s) regulation (Supplementary Figure S8 and Supplementary Table S4).

Both operon cohorts can be well fitted by sine functions (average R^2 of 0.85 and all R^2 larger than 0.7) with the frequencies for the same stress between cohorts differing by less than ~10%. The only difference is that genes in operons subject to TF regulation have ~15% higher $\mu_{|\Delta LFC|}$ (in amplitude, because their response strengths are stronger, likely due to the initial responses being "amplified" by the subsequent changes in their regulatory TF numbers, as reported in [61].

We further plotted $\mu_{|\Delta LFC|}(L_G)$ of the 56 operons that express one or more self-regulator TF ("self-TFs") (Supplementary Figure S9 and Supplementary Table S5). Again, sine functions describe well the behavior (average R^2 ~0.9) with frequencies differing by less than 15% from the operons above. Thus, self-TFs do not disturb the phenomenon significantly.

Overall, we conclude that TF interactions do not significantly influence the main parameters of the sine functions that best fit $\mu_{|\Delta LFC|}(L_G)$ (particularly the frequency). This conclusion holds even when the responses of genes directly interacting via TFs exhibit correlation, e.g., following media dilutions (Supplementary Figure S10).

We also failed to find relationships between promoter sequences and their positioning in operons. Specifically, we did not find relationships in the promoter σ factor preferences or AT content [62](Supplementary Methods 4) with their positioning in the operon (Supplementary Figures S11 and S12). Similarly, we found no relationship between the promoter sequence (as measured by p-distances) and the positioning in the operon (Supplementary Methods 5 and Supplementary Figure S13). Consequently, we could not establish a relation between promoter sequences and $\mu_{|\Delta LFC|}(L_G)$.

2.6 The positioning of genes in operons explains $\mu_{|\Delta LFC|}(L_G)$.

To test if $\mu_{|\Delta LFC|}(L_G)$ is related with the placement of genes in operons, we next used null models where the genes' positioning in their operon, or the operon that they belong to, are randomized. We found that, first, when the positioning of each gene in its operon is randomized, $\mu_{|\Delta LFC|}$ no longer changes with L_G (Supplementary Figure S14). Similarly, when randomizing which operon each gene belongs to (while not changing the positions relative to the primary TSS), again $\mu_{|\Delta LFC|}$ does not change with L_G (Supplementary Figure S14). Finally, the same occurs when randomizing both P and the operon each gene belongs to (Supplementary Figure S14).

Moreover, in natural operons, $\mu_{|\Delta LFC|}$ between the most downstream gene in each operon and the first gene downstream the end of the operon is ~2× higher (Supplementary Table S6) than the $\mu_{|\Delta LFC|}$ between nearest neighbor genes in the same operons (Figures 2A-2C). We conclude that,

during genome-wide stresses responses, the low mean values, and the wavelike pattern of $\mu_{|\Delta LFC|}(L_G)$ are both a result of the dynamic correlations between genes in the same operon.

2.7 In the absence of internal TSSs, $\mu_{|\Delta LFC|}$ of genes in the same operon is linearly correlated with L_G.

Next, we investigated which features of the operons can explain how the gene responses relate. First, to dissect the role of premature elongation terminations (RNAP fall-offs), we studied the $\mu_{|\Delta LFC|}$ of pairs of genes *without* internal TSSs in-between. On average, their $\mu_{|\Delta LFC|}$ increases linearly with L_G (Figure 2D and 2E, solid lines). This suggests that elongating RNAPs have relatively homogenous fall-off rates throughout operons during the stresses.

Meanwhile, in pairs of genes with at least one internal TSS in-between, $\mu_{|\Delta LFC|}$ increases less with L_G (Figures 2D and 2E, dotted lines). This suggests that internal TSSs significantly influence the response of genes in downstream positions.

We thus hypothesized that promoters controlling more genes in an operon should have stronger responses, so that all downstream genes are responsive. To test this, we analyzed the response strengths of genes immediately downstream from TSSs and plotted it as a function of the number of downstream genes of that TSS, prior to the occurrence of another TSS. From Figure 2F, as hypothesized, their $\mu_{|LFC|}$ increases linearly with the distance to the closest downstream TSS (L_P).

Given this, we argue that a main role of internal promoters of *E. coli* operons is to counter the effects of RNAP fall-offs, providing stronger response strength to downstream genes more distanced from the primary promoter. To an extent, compared to when they are absent, internal promoters act towards homogenizing the response strengths of genes in the same operon.

2.8 RNAP fall-offs influence differences in response strength of genes in the same operon.

Results in Figure 2 can be explained by a tangible rate of fall-offs of elongating RNAPs (R_f), as this would cause genes more distanced from the closest upstream promoter to have weaker response strengths.

Meanwhile, if tangible fall-off rates exist, there should also exist a significant difference between the numbers of RNA reads from the starting and the ending regions of a gene. Using RNA-seq data, we estimated these average differences for each gene, as a function of its position in the operon (Methods section 4.2). Results in Figure 3A shows that: i) the number of reads between the start and end regions of the gene coding regions differ significantly, suggesting significant rates of elongating RNAP fall-offs; ii) R_f does not differ with P, for $1 < P < 10$, suggesting that events promoting fall-offs do not locate at a specific distance(s) from the primary TSS. As a side note, in the future, it should be of interest to study why $R_f(P)$ is significantly lower for $P = 1$ alone.

We conclude that fall-offs influence differences in response strengths between genes in different positions of their operon, thus playing a key role in how $\mu_{|\Delta LFC|}$ increases with L_G. This conclusion is in line with the degree of presence of TTSs between operons and downstream genes as a function of the distance between them (Supplementary Figure S1C). I.e., if fall-offs are frequent, large distances suffice to uncouple the dynamics of operons and their downstream genes. Meanwhile, lesser spaced downstream genes need a TTS for this purpose.

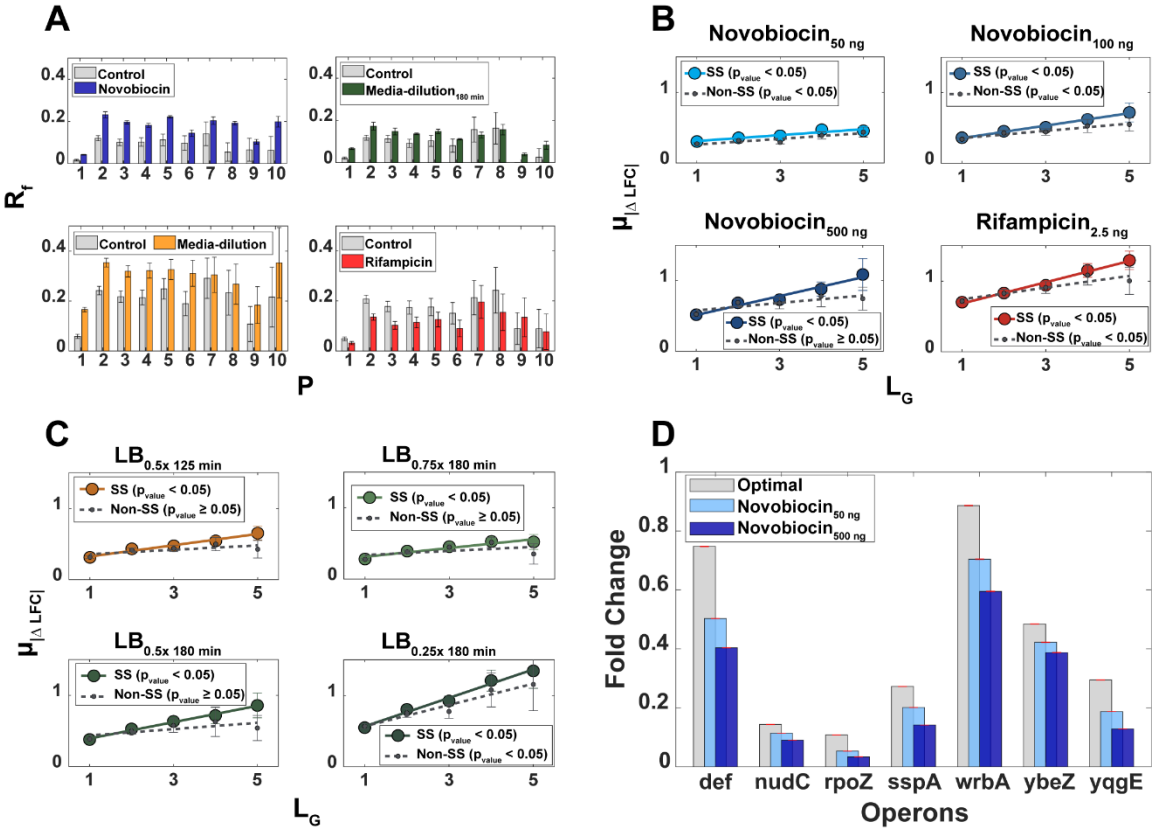

Figure 3: Effects of elongation fall-offs on the responses of operons to the genome wide stresses. (A1-A4) Average differences between the number of RNA reads from the starting region of a gene and the number of RNA reads from the ending region of the same gene, normalized by the number of reads of the starting region. Data is plotted as a function of the position *(P)* of the genes, relative to the primary TSS. The error bars are the standard error of the sum. **(B** and **C)** Average absolute difference between the LFCs of pairs of genes in same operon *($\mu_{|\Delta LFC|}$)* with and without supercoiling sensitive genes in between them. The latter are grey. Data only for $L_G < 6$ since there are very few pairs of genes without TSSs in between for $L_G > 5$. The solid lines correspond to operons without internal TSSs, while the dotted lines correspond to operons without internal TSS but also with supercoiling sensitive genes. For measurements under Novobiocin at 50 ng/µl, the responsiveness of genes with and without high supercoiling sensitivity cannot be visually distinguished (the lack of sufficient data points hampered the use of statistical tests). The p-value of each line is from a statistical test of whether the linear fit differs from a horizontal line, using 'fitlm' in MATLAB. **(D)** Fold changes between the protein levels of pairs of genes in the same operons, as measured by Flow-cytometer using the YFP strain library [61]. The strains are listed in Supplementary Table S19. The red error bars are the standard error of the Fold change.

2.9 No specific short DNA sequences correlates with RNAP fall-off rates.

We investigated if we could relate fall-off rates to specific DNA sequences. First, we searched for 14 DNA sequences (12 nucleotide each) (Supplementary Methods 6) that are known to enhance the propensity for transcriptional pausing [63] since, eventually, these events could also promote RNAP fall-offs (e.g. by enhancing the rates of RNAP collisions). Also, we searched for homopolymetric tracts of A's or T's with 9 nucleotides each, as they have been reported to promote RNAP slippage and fall-offs [64]. Their frequency of occurrences in the DNA of *E. coli* (Operon region) are shown in Supplementary Table S8. We then tested if the frequency of occurrence of these sequences along the DNA coding for the genes correlates with the average values of $R_f(P)$. From Supplementary Figure S15, we found no correlation in either case. Finally, we searched for TSSs along operons in the opposite strand (believed to enhanced RNAP collisions[65,67]), but we did not find them in significant numbers. These results suggest that the falloff rates are homogenous throughout operons.

2.10 Sensitivity to positive supercoiling buildup enhances RNAP fall-off rates.

A past study showed that accumulation of PSB in the DNA can force elongating RNAPs to pause and destabilize bound proteins [66]. Meanwhile, we observed that Novobiocin, which inhibits gyrase, increases R_f for all P (Figure 3A). Moreover, R_f increases with Novobiocin concentration (Supplementary Figure S16).

Thus, we hypothesized that PSB may partially explain premature RNAPs fall-offs (even in optimal conditions). If this holds true, genes in DNA regions more sensitive to PSB should have higher $\mu_{|\Delta LFC|}(L_G)$ in all conditions. To test this, we considered only operons without internal TSSs (as TSSs would interfere with the estimations of fall-off rates). Then, we compared $\mu_{|\Delta LFC|}(L_G)$ of pairs of genes classified as highly sensitive to PSB with other pairs of genes (Supplementary Methods 7). In agreement with our hypothesis, in all stress conditions, $\mu_{|\Delta LFC|}(L_G)$ changes more strongly with L_G for pairs of genes highly sensitive to PSB (Figures 3B and 3C and Supplementary Table S9).

Following this, we measured the protein levels of genes in the same operons while subject to Novobiocin (Supplementary Methods 2). From Figure 3D and Supplementary Figure S17, Novobiocin increased the fold change between the response strength of downstream and upstream genes in the same operon. This suggests that the effects of PSB on RNA numbers propagates to protein levels.

We conclude that PSB is a major cause of premature RNAP fall-offs (likely in all the stresses studied), which generate $\mu_{|\Delta LFC|}(L_G)$. Nevertheless, the $\mu_{|\Delta LFC|}$ of genes not highly sensitive to PSB also changes with L_G (albeit less pronouncedly). Thus, PSB cannot fully explain premature fall-offs. This is expected, since other phenomena (e.g., RNAP collisions) should also cause fall-offs.

2.11 Other stresses cause similar genome-wide responses in E. coli.

If the genome-wide stress response dynamics of operons observed in Figures 2A-2C is due to RNAP fall-offs and internal TSSs, the same responsiveness should be detectable in past data on other stress responses by *E. coli* cells. To test it, we collected publicly available, processed RNA-seq data in NCBI for *E. coli* MG1655 subject to stresses (mutations not included). Specifically, we found data for cells subject to: changes in the microaerobic environment [68], heat-shock [69], hydrostatic pressure [70], chloramphenicol [72], erythromycin [72], and Mercury [71]. Further, we studied cells subject to anaerobic-to-aerobic shift (at 0.5, 1, 2, 5, and 10 min, after the shift) [73].

Analyzing the data while accounting for the operon structure of *E. coli*, we find that, in all cases, $\mu_{|\Delta LFC|}(L_G)$ is well fit by a sine function of the form of *Eq. 1* (Figure 4). Moreover, the mean and standard deviation of the frequency f of the best fitting lines are 0.10 and 0.01, respectively, as in our measurements (Figures 2A-2C and Supplementary Table S10). Meanwhile, the amplitudes differ widely with the nature of perturbation (as in Figures 2A-2C).

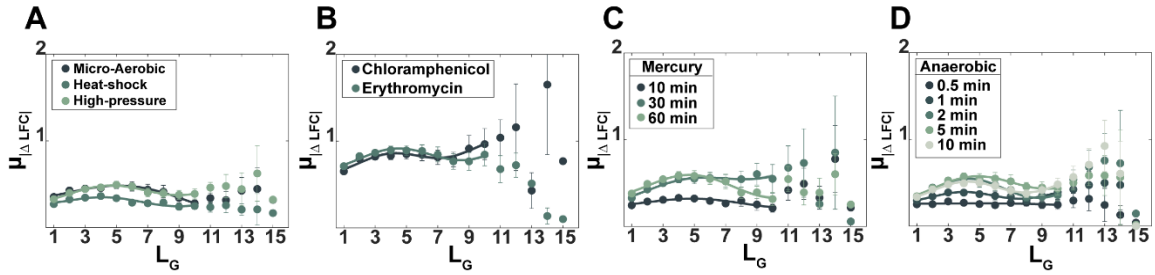

Figure 4: **Average absolute differences in the response strengths of genes in the same operon ($\mu_{|\Delta LFC|}$) as a function of L_G for *E. coli* cells under other stresses**. From **(A)** to **(D),** each data is the $\mu_{|\Delta LFC|}$ of all pairs of genes of the same

operon distanced by L_G genes. The error bars are the standard error of the mean. Also shown are the best fits of functions of the form (*Eq 1*) along with their R^2 values. See Supplementary Table S10, for the best fitting parameter values. In all cases, R^2 is larger than 0.75 (average R^2 of 0.90). Finally, the data on heat-shock responses was obtained by AIR-seq. Data from [68], [69], [70], [72], [72], [71], and [73]. In **(C)** and **(D)**, the time values correspond to when measurements were conducted following the perturbation.

2.12 Other bacteria also exhibit similar genome-wide stress responses.

Many other bacteria have their genome similarly organized in operons, and their sizes (gene numbers) have been reported [71–77] (although numbers and positioning of internal TSSs many times have not). Noteworthy, their operon size distributions (Supplementary Figure S18) cannot be distinguished, statistically, from the size distribution of *E. coli* (Supplementary Table S11). We thus hypothesized that, during genome-wide stress responses, the $\mu_{|\Delta LFC|}$ of genes in operons should differ with L_G as in *E. coli*.

To test this, we collected data on various stress responses of *B. subtilis* [74–80], *C. glutamicum* [81,82] , and *H. pylori* [83]. From Figure 5, in all cases and species, $\mu_{|\Delta LFC|}(L_G)$ is well fit by a sine function of the form (*Eq. 1*). The best fitting frequencies (mean and standard deviation) for each stress and species are shown in Supplementary Table S12-14. While similar, the frequencies differ more between species, than between different stresses in the same species. There are many possible reasons for this, including the differences in arrangement of internal TSSs of each species, different energy needs in their RNAP elongating processes, among other. Meanwhile, as expected, amplitudes differ between species and between stresses Supplementary Table S12-14. Overall, we conclude that the response dynamics of the operons of *E. coli* to genome wide stresses (Figures 2A-2C), is also observed in *B. subtilis*, *C. glutamicum*, and *H. pylori*.

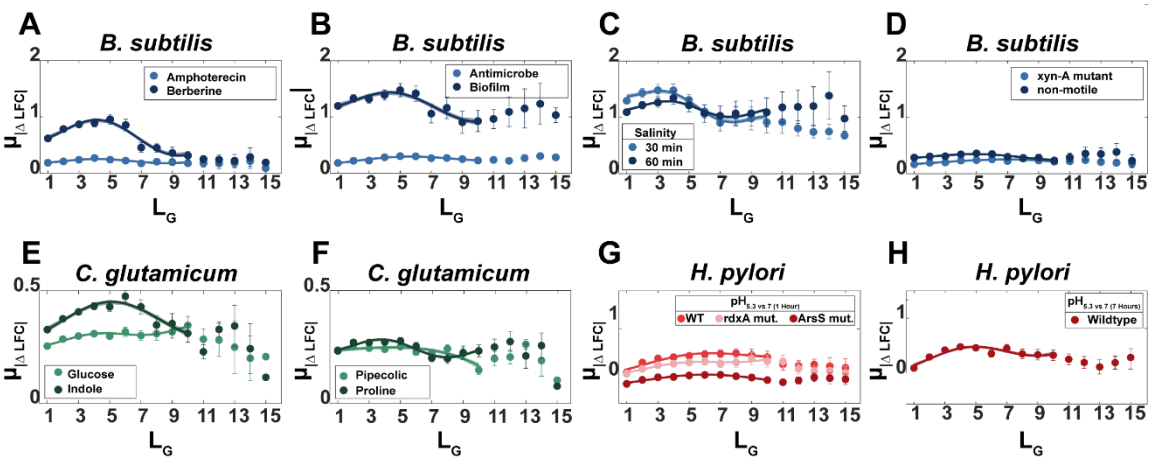

Figure 5: **Average absolute differences in response strengths between genes in the same operon ($\mu_{|\Delta LFC|}$) as a function of L_G in *B. subtilis, C. glutamicum, and H. pylori*.** Each data is the $\mu_{|\Delta LFC|}$ of all pairs of genes of the same operon distanced by L_G genes in between. Error bars are the standard error of the mean. Also shown are the best fits of functions of the form *Eq. 1*. Their R^2 values are shown in Supplementary Table S12-14 along with the best fitting parameter values. Data from [74–80], [81,82] , and [83]. **(A-H)** correspond to data from a specific species and condition, listed on the figure, specifically.

2.13 A stochastic model of operon dynamics can reproduce the results from RNAP fall-offs and internal promoters alone.

We used a stochastic model of transcription in operons to test if the average, genome-wide behavior in Figure 2 can be explained by internal TSSs combined with RNAP elongation fall-offs. Reactions S1-S7 of the model are described in detail in (Supplementary Section 9). Briefly, Reaction S1 accounts for multi-step transcription initiation, starting with RNAP binding to a TSS

(primary or internal) leading to the sequential events closed complex formation, open complex formation, and promoter escape. Meanwhile, Reaction S2 accounts for successful elongation events that produce complete RNAs. Reaction S3 models spontaneous, premature fall-offs of elongating RNAPs. These fall-offs compete directly with the successful elongation events. Finally, Reactions S4-S7 model the various outcomes of potential RNAP fall-offs due to collisions of elongating RNAPs with RNAPs in closed or in open complex formation. Fall-off due to RNAP pausing was not included since we found no correlation between fall-off rates, R_f, and pause sequences (Results section 2.9).

We simulated the model assuming the parameter values in Supplementary Table S15 using the simulator *SGN Sim* [84], whose dynamics is based on the stochastic simulation algorithm [85,86]. For each condition, we performed 1000 numerical simulations, each lasting for 10^6 seconds. From the simulations, we estimated $\mu_{|\Delta RNA|}(L_G)$ between genes in the same operon, which we compared to the empirical data on $\mu_{|\Delta LFC|}(L_G)$.

First, we studied the effects of changing the number of internal TSSs (per gene) of an operon. From Supplementary Figure S19A, in all models, $\mu_{|\Delta RNA|}$ differs with L_G according to a sine-like function, in agreement with the empirical data (Figures 2, 4 and 5). Next, we studied the effects of changing RNAP fall-off rates. Again, in all models, $\mu_{|\Delta RNA|}$ differs with L_G according to a sine-like function (Supplementary Figure S19B). Finally, we tested a model without internal TSSs. In this case, the results (Supplementary Figure S19C) match instead the data in Figures 2D and 2E, as predicted. We conclude that, according to the model, internal TSSs and RNAP fall-off events can explain the data on $\mu_{|\Delta RNA|}(L_G)$ in all conditions for *E. coli*, *B. subtilis*, *C. glutamicum*, and *H. pylori*.

The model also suggests that increasing the number of internal TSSs (while maintaining the gene length of the operon constant) decreases the frequency of $\mu_{|\Delta RNA|}(L_G)$ (Supplementary Figure S19A). To test this prediction, we considered the empirical data on $\mu_{|\Delta LFC|}(L_G)$ solely from operons with more than 3 genes (a total of 423 operons). Then, we separated them into 2 cohorts, one with operons of less than 3 TSSs and the other with the operons of 3 or more TSSs. As predicted, the frequency of the best fitting function of $\mu_{|\Delta LFC|}(L_G)$ of operons with 3 or more TSSs is higher (Supplementary Table S16). Only for condition 'Media richness depletion (0.75× at 180 min)' is this not observed. However, that condition is the weakest media depletion, which results in weaker signals.

Moreover, from Supplementary Figure S19B, the models predict that increasing $k_{fall\text{-}off}$ should increase the amplitude of the best fitting function of $\mu_{|\Delta RNA|}(L_G)$. In agreement, the amplitudes of $\mu_{|\Delta LFC|}(L_G)$ increase with Novobiocin concentration (Supplementary Table S1). Specifically, increasing Novobiocin concentration should increasingly disturb positive supercoiling building. That, in turn, should disturb transcription, likely increasing RNAP fall-off rates [66].

Finally, we used the model to explain why $\mu_{|\Delta LFC|}(L_G)$ can be well-fitted by sine functions. Assuming stable, identical transcription rates at the primary and internal TSSs, the RNA numbers produced by each gene of the operon, when plotted as a function P, can be well fitted by a function of the form: $f(x) = A + B.(C - x) \bmod D$ (Supplementary Figure S20A and Supplementary Table S17). This function is the outcome of RNAP fall-offs, which increasingly weakens the RNA production rates of each gene, along with the effect of the internal promoter, which increases the signal similarly to primary promoter. Next, we converted this data into the state space $|\Delta RNA|$ versus L_G (Supplementary Figure S20B and Supplementary Table S18). In this state space, the data can be well fitted by a sine function instead, supporting the hypothesis that the two factors (RNAP fall-offs and internal promoters) can, alone, explain the sine-like shape of $\mu_{|\Delta LFC|}(L_G)$ in Figures 2A-2C.

2.14 Relationship between the L_P of genes in operons and their biological function.

We investigated if the genes' biological functions may influence their positioning in the operon, using gene ontology [87,88](Supplementary Methods 9). First, we found that, on average, the closer two genes are to each other in the same operon, the more likely it is that they are involved in the same biological processes (Supplementary Figure S21). This conclusion reinforces past findings [17,18,89] using other classification methods.

Next, we studied the mean and standard deviation of L_P of cohorts of genes for specific biological functions (Table 1). We considered gene ontology (GO) terms that are direct descendants (child) of the general GO term GO:0008150 "biological process" [90]. Table 1 shows the 8 of 23 terms considered (rows 1 to 8). Meanwhile, the remaining terms (15 of 23) were not considered, because we found less than 100 genes (genome-wide) with such feature(s). Also shown in Table 1 is gene 'essentiality'. We classified 236 genes as "essential" according to the definition in [91,92,94]. Specifically, i.e., a gene is essential if cells are unable to survive in optimal conditions following its deletion [91,92,94].

From Table 1, there is a significant variability in $\mu(L_P)$ and $STD(L_P)$ between the gene cohorts differing in biological functionality. Moreover, genes associated with the '*regulation of biological processes'* (rows 1, 4, 6, and 8) have small L_P ($\mu(L_P)$ = 0.52). Also, we find that the $\mu(L_P)$ of essential genes equals 0.59, which is close to genes associated with the '*regulation of biological processes*. This may imply that quickly adapting those processes may be a vital goal of *E. coli* cells under the stresses studied. Meanwhile, genes associated with the other functionalities exhibit an average L_P ($\mu(L_P)$ = 0.95), similar to the overall average of all other genes in operons (0.93, row 9) and to the genome-wide average (0.96, row 10). Thus, the distance of genes in operons from their closest upstream promoter relates with their biological functions.

Table 1. L_P of genes in operons for specific biological functionalities. Shown are the average (μ) and the standard deviation (STD) of L_P of cohorts of genes sharing a higher order biological function (based on the gene ontology classification [87,88]. Also shown are the number of genes of each cohort. Gene functions are based on the gene ontology classification in [87,88], except for cohorts 10 to 11, which involve 'all genes' in operons and 'essential genes', which are based on the classification in [91,92,94]. For additional information, see the distribution of operon genes lengths in Supplementary Figure S18 (top left), with a median of 3.

	Functionality	$\mu(L_P)$	$STD(L_P)$	Genes in operons
1	Positive regulation of biological process	0.6	1.22	70
2	Localization	1.09	1.31	449
3	Metabolic process	0.92	1.42	1225
4	Regulation of biological process	0.55	0.99	291
5	Response to stimulus	0.86	1.37	475
6	Biological regulation	0.59	1.00	358
7	Cellular process	0.95	1.38	1598
8	Negative regulation of biological process	0.36	0.63	110
9	All genes not in 1-8	0.93	1.38	597
10	All genes in operons	0.96	1.40	2262
11	Essential Genes	0.59	0.89	222

3. Discussion

Many open questions remain about the general principles governing the formation and evolution of operons. Studies have characterized many individual operons and identified clear phenotypic

advantages of placing those genes under the control of the same promoter(s), due to related biological functions [12,18,19,22,93]. For example, the *rplKAJL-rpoBC* operon codes for proteins RplK, RplA, RplJ, and RplL, which are all 50S ribosomal subunits [95]. The operon further codes for RpoB and RpoC, which are the subunits β and β' of RNAP [22,95]. Lacking any of these proteins in this operon (e.g. if that protein production was controlled by another operon) would probably decrease the efficiency with which *E. coli* forms RNAPs and ribosomes. Such a need for sub-units of complex proteins to be produced at the same time, in the same location, and in similar quantities provides one general explanation for why genes coding for protein subunits are integrated into the same operons. Studies have also shown the benefits of internal promoters in some operons[96]. For example, the *rplKAJL-rpoBC* operon has three promoters, which may allow, in some conditions, to produce the protein pairs (RplK and RplA), (RplJ and RplL), and (RpoB and RpoC) with relatively distinct dynamics [97–99]. This may be advantageous when, e.g., ribosomes and RNAPs should not be produced at the same rate[100]. Meanwhile, decreasing physical stability of mRNAs and increasing transcription elongation failures might limit the growth of operon lengths [101–104]. Due to this, we hypothesized that, in some conditions, some internal promoters may assist as 'signal recovers' between upstream and downstream genes, while diverging their expression levels in other conditions.

Here, we studied how the behavior of genes in operons in the context of genome wide perturbations that, by targeting RNAP and gyrase [36,37], affect both primary as well as internal promoters of most operons, quasi-synchronously. We analyzed this dynamical data while accounting for information on the operon's composition of *E. coli*, which is likely the most dissected bacterial genome. This allowed accessing the degree to which operon lengths and internal promoters can be correlated to the dynamics of RNA production within operons.

We found that differences in response strengths of genes in the same operon are well matched by sine functions of the distance between the genes. These wave patterns emerge from the contribution of internal TSSs, combined with a genome-wide fall-off rate of elongating RNAPs. Meanwhile, in operons lacking internal TSSs (and in transcription units alone), the differences are better matched by lines with positive slope, due to tangible fall-off rates of elongating RNAPs alone. Moreover, we provided evidence that the RNAP fall-off rates can be partially explained by PSB. The findings were reproduced by a simple stochastic model of transcription based on these hypotheses. Finally, from recently published data, we observed the same phenomenon in *E. coli* under various other stresses, and in three other bacterial species (one gram-negative and two gram-positive).

So far, the main biological function associated with internal promoters has been enhancing divergency between upstream and downstream genes, in response to specific signals [33,101], including antibiotics [105]. However, under the genome-wide stresses studied here, they decreased divergency, counteracting the effects of premature RNAP fall-offs. This and the strong correlation between internal promoter response strength and number of downstream genes prior to other internal promoters (Figure 2F) suggest that RNAP fall-off rates may have partially shaped the number, positioning, and strength of internal promoters in operons.

Another way the gene network of *E. coli* may have been shaped by premature RNAP fall-offs is the number and location of TTSs. There are only 174 known TTSs associated with operons in *E. coli*, while operons contain more than 2500 genes [22]. However, a tangible premature RNAP fall-off rate could allow more distanced genes to become uncorrelated, even in the absence of a TTS. In agreement, nearest neighbor genes of different operons that are more distanced are also less likely to have a TTS in between (Supplementary Figure S1C), which implies that more distanced genes have their activity made independent by natural RNAP fall-offs.

The findings reported here could be of use to the Bioindustries, include in Synthetic Biology, and to Biomedicines. Rearranging the positioning of genes and TSSs in operons may allow fine-tuning natural transcriptional programs of stress response. I.e., re-positioning of genes in operons should

allow tuning their response strengths to certain stresses, which could alter certain cellular processes, in order to, for example, minimize resources consumption, and thus improve the yield and sustainability of bioreactors. Similarly, adjustments of gene positioning in operons and the use of internal TSSs could also become useful for tuning the kinetics of synthetic genetic circuits. Finally, if bacteria are placing vital genes (e.g. essential ones) immediately downstream to primary and internal promoters, to ensure their responsiveness to genome-wide stresses, we could potentially use this to search for specific target genes for new antibiotics.

Overall, our results provide new explanations for the naturally evolved lengths of prokaryotic operons, and the location, number, and strength of their internal promoters. The findings may be of use to predict and dissect the same features in the operons of other bacteria.

4. MATERIALS AND METHODS

4.1 Bacterial strains and genome-wide stresses

E. coli (MG1655 strain) is our model organism because, first, its TFN is well mapped [22]. Thus, it is possible to determine the potential the influence of TF interactions. Moreover, sequence-dependent phenomena, e.g., transcriptional pauses, and other elongation-related phenomenon are well known [23,25,63,106,107]. Further, protein numbers follow RNA numbers and there are no known 'strong' mechanisms regulating RNA numbers after transcription [10].

RNA-seq data on genome-wide stresses is partially curated from [36,37]and partially produced for the present study. In all cases, changes in gene expression were measured by RNA-seq by estimating the genes' Log fold changes (LFC) relative to RNA-seq when not adding Novobiocin, not diluting the medias, and not adding Rifampicin, respectively. A total of 4573 unique combination of genes separated by a certain distance was obtained for genes in the same operon (with > 1 genes). For Novobiocin, we have rejected 1015 pairs (unknown differences in LFC). So, we have considered only 3558 pairs of gene, for whose LFC is known and calculated the absolute Delta Fold change between them. Same as Novobiocin, we have rejected 1631 gene pairs for Media-dilution as one we don't know at least one of their LFC, and only 2942 gene pairs were considered (FC of 653 genes is unknown). For Rifampicin we have considered 3897 pairs of genes and rejected 676 pairs.

We also collected data on protein levels by flow-cytometry from cells subject to Novobiocin stresses (50 and 500 µg/mL), respectively (supplementary methods 2). For all stresses, we measured the optical density at 600 nm (OD600nm) to monitor the cell growth, using a Bioteks Synergy HTX Multi-Mode Reader.

4.2 Genome-wide differences in reads between the start and end of genes in operons.

To demonstrate the existence of significant RNAP fall-off rates during elongation, we obtained the read counts of start and stop codons for each gene in each operon using MATLAB 'featurecount' function. Then, we normalized the read counts using Transcripts per million method (TPM) [108].

Next, we conducted a t-test on the distributions of start and stop codons from three biological replicates for each condition, to determine their statistical significance. In cases where the distributions were statistically significant, we investigated whether there was a terminator site located between the start and stop codon. These cases were discarded from further analysis. From the remaining cases, we estimated the fall-off rates, r_f, of genes in the position i of operon j, as follows:

$$r_{f(i,j)} = \frac{reads_{start(i,j)} - reads_{end(i,j)}}{reads_{start(i,j)}} \quad \textbf{(1)}$$

From this, the operon-wide average rate of fall-offs as a function of the position of the genes, R_f, is given by:

$$R_f = \sum_{i=1}^{n} \left(\frac{\left(\sum_{j=1}^{k} r_{f(i,j)} \right)}{k} \right) \quad \textbf{(2)}$$

where k is the number of genes per position.

AUTHOR CONTRIBUTIONS

A.S.R. conceived the study, assisted by R.J. and C.S.D.P. A.S.R supervised the study. R.J. executed most data analysis. S.D. performed most measurements. I.S.C.B. and C.S.D.P. performed supporting data analysis. I.S.C.B. produced the illustrative figures. V.C. assisted in the RNA-seq measurements. J.M. contributed to the data interpretation and suggested additional analysis. R.J. produced the model, supported by C.S.D.P. and A.S.R. A.S.R. and R.J. drafted all documents, supported by I.S.C.B. and S.D., which were revised and commented on by all co-authors. C.S.D.P., S.D., and I.S.C.B. contributed equally and have the right to list their name first.

FUNDING

Work supported by the Jane and Aatos Erkko Foundation [10-10524-38 to A.S.R.] and the Sigrid Jusélius Foundation [A.S.R.]. Work also supported by Suomalainen Tiedeakatemia [to R.J. and S.D.]; Finnish Cultural Foundation [00222452 to I.S.C.B., and 00210818 to C.S.D.P.]; Tampere University Graduate Program [to V.C.], EDUFI [TM-21-11655 to R.J.], and Academy of Finland [349116 to J.M.]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGEMENTS

We would like to acknowledge the assistance of M. Bahrudeen and B. Almeida for suggestions regarding statistical analysis.

CONFLICT OF INTEREST

The authors have no competing interests.

REFERENCES

1. Okuda, S., and Yoshizawa, A.C. (2011). ODB: a database for operon organizations, 2011 update. Nucleic Acids Res. *39*, D552–D555.

2. Jacob, F., and Monod, J. (1961). On the regulation of gene activity. Cold Spring Harb. Symp. Quant. Biol. *26*, 193–211.

3. Browning, D.F., and Busby, S.J.W. (2004). The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. *2*, 57–65.

4. Browning, D.F., and Busby, S.J.W. (2016). Local and global regulation of transcription initiation in bacteria. Nat. Rev. Microbiol. *14*, 638–650.

5. Chamberlin, M.J. (1974). The selectivity of transcription. Annu. Rev. Biochem. *43*, 721–775.

6. Chong, S., Chen, C., Ge, H., and Xie, X.S. (2014). Mechanism of transcriptional bursting in bacteria. Cell *158*, 314–326.

7. McClure, W.R. (1985). Mechanism and Control of Transcription Initiation in Prokaryotes. Annu. Rev. Biochem. *54*, 171–204.

8. McLeod, S.M., and Johnson, R.C. (2001). Control of transcription by nucleoid proteins. Curr. Opin. Microbiol. *4*, 152–159.

9. Ruff, E.F., Record, M.T., Jr, and Artsimovitch, I. (2015). Initial events in bacterial transcription initiation. Biomolecules *5*, 1035–1062.

10. Bernstein, J.A., Khodursky, A.B., Lin, P.-H., Lin-Chao, S., and Cohen, S.N. (2002). Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. U. S. A. *99*, 9697–9702.

11. Swain, P.S. (2004). Efficient attenuation of stochasticity in gene expression through post-transcriptional control. J. Mol. Biol. *344*, 965–976.

12. Rocha, E.P.C. (2008). The organization of the bacterial genome. Annu. Rev. Genet. *42*, 211–233.

13. Pál, C., and Hurst, L.D. (2004). Evidence against the selfish operon theory. Trends Genet. *20*, 232–234.

14. Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998). Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem. Sci. *23*, 324–328.

15. Pál, C., Papp, B., and Lercher, M.J. (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat. Genet. *37*, 1372–1375.

16. Osbourn, A.E., and Field, B. (2009). Operons. Cell. Mol. Life Sci. *66*, 3755–3775.

17. Salgado, H., Moreno-Hagelsieb, G., Smith, T.F., and Collado-Vides, J. (2000). Operons in Escherichia coli: genomic analyses and predictions. Proc. Natl. Acad. Sci. U. S. A. *97*, 6652–6657.

18. de Daruvar, A., Collado-Vides, J., and Valencia, A. (2002). Analysis of the cellular functions of Escherichia coli operons and their conservation in Bacillus subtilis. J. Mol. Evol. *55*, 211–221.

19. Rogozin, I.B., Makarova, K.S., Murvai, J., Czabarka, E., Wolf, Y.I., Tatusov, R.L., Szekely, L.A., and Koonin, E.V. (2002). Connected gene neighborhoods in prokaryotic genomes. Nucleic Acids Res. *30*, 2212–2223.

20. Rogozin, I.B., Sverdlov, A.V., Babenko, V.N., and Koonin, E.V. (2005). Analysis of evolution of exon-intron structure of eukaryotic genes. Brief. Bioinform. *6*, 118–134.

21. Conway, T., Creecy, J.P., Maddox, S.M., Grissom, J.E., Conkle, T.L., Shadid, T.M., Teramoto, J., San Miguel, P., Shimada, T., Ishihama, A., et al. (2014). Unprecedented high-

resolution view of bacterial operon architecture revealed by RNA sequencing. MBio *5*, e01442-14.

22. Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., García-Sotelo, J.S., Alquicira-Hernández, K., Muñiz-Rascado, L.J., Peña-Loredo, P., et al. (2018). RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. Nucleic Acids Res. *47*, D212–D220.

23. Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochem. Soc. Trans. *34*, 1062–1066.

24. Toulokhonov, I., and Landick, R. (2003). The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. Mol. Cell *12*, 1125–1136.

25. Epshtein, V., and Nudler, E. (2003). Cooperation between RNA polymerase molecules in transcription elongation. Science *300*, 801–805.

26. Komissarova, N., and Kashlev, M. (1997). Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. Proc. Natl. Acad. Sci. U. S. A. *94*, 1755–1760.

27. Zhou, Y.N., Lubkowska, L., Hui, M., Court, C., Chen, S., Court, D.L., Strathern, J., Jin, D.J., and Kashlev, M. (2013). Isolation and characterization of RNA polymerase rpoB mutations that alter transcription slippage during elongation in Escherichia coli. J. Biol. Chem. *288*, 2700–2710.

28. Merino, E., and Yanofsky, C. (2005). Transcription attenuation: a highly conserved regulatory strategy used by bacteria. Trends Genet. *21*, 260–264.

29. Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. Nature *289*, 751–758.

30. Naville, M., and Gautheret, D. (2009). Transcription attenuation in bacteria: theme and variations. Brief. Funct. Genomic. Proteomic. *8*, 482–492.

31. Wang, L., Watters, J.W., Ju, X., Lu, G., and Liu, S. (2023). Head-on and co-directional RNA polymerase collisions orchestrate bidirectional transcription termination. Mol. Cell *83*, 1153-1164.e4.

32. Ullmann, A., Joseph, E., and Danchin, A. (1979). Cyclic AMP as a modulator of polarity in polycistronic transcriptional units. Proc. Natl. Acad. Sci. U. S. A. *76*, 3194–3197.

33. Fornwald, J.A., Schmidt, F.J., Adams, C.W., Rosenberg, M., and Brawner, M.E. (1987). Two promoters, one inducible and one constitutive, control transcription of the Streptomyces lividans galactose operon. Proc. Natl. Acad. Sci. U. S. A. *84*, 2130–2134.

34. Nishi, A., and Zabin, I. (1963). Effect of temperature on the differential rate of synthesis of proteins of the lactose operon in E. coli. Biochem. Biophys. Res. Commun. *13*, 320–323.

35. Sanchez-Vazquez, P., Dewey, C.N., Kitten, N., Ross, W., and Gourse, R.L. (2019). Genome-wide effects on *Escherichia coli* transcription from ppGpp binding to its two sites on RNA polymerase. Proc. Natl. Acad. Sci. U. S. A. *116*, 8310–8319.

36. Almeida, B.L.B., M Bahrudeen, M.N., Chauhan, V., Dash, S., Kandavalli, V., Häkkinen, A., Lloyd-Price, J., S D Cristina, P., Baptista, I.S.C., Gupta, A., et al. (2022). The transcription factor network of E. coli steers global responses to shifts in RNAP concentration. Nucleic Acids Res. 10.1093/nar/gkac540.

37. Dash, S., Palma, C.S.D., Baptista, I.S.C., Almeida, B.L.B., Bahrudeen, M.N.M., Chauhan, V., Jagadeesan, R., and Ribeiro, A.S. (2022). Alteration of DNA supercoiling serves as a trigger of short-term cold shock repressed genes of E. coli. Nucleic Acids Res. *50*, 8512–8528.

38. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

39. Lloyd-price, J., Kandavalli, V., Chandraseelan, J.G., Goncalves, N., Oliveira, S.M.D., Häkkinen, A., and Ribeiro, A.S. (2016). Dissecting the stochastic transcription initiation process in live Escherichia coli. DNA Res *23*, 203–214.

40. Kandavalli, V.K., Tran, H., and Ribeiro, A.S. (2016). Effects of σ factor competition are promoter initiation kinetics dependent. Biochim. Biophys. Acta *1859*, 1281–1288.

41. Hartmann, G., Honikel, K.O., Knüsel, F., and Nüesch, J. (1967). The specific inhibition of the DNA-directed RNA synthesis by rifamycin. Biochim. Biophys. Acta *145*, 843–844.

42. Gellert, M., O'Dea, M.H., Itoh, T., and Tomizawa, J. (1976). Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase. Proc. Natl. Acad. Sci. U. S. A. *73*, 4474–4478.

43. Sugino, A., Higgins, N.P., Brown, P.O., Peebles, C.L., and Cozzarelli, N.R. (1978). Energy coupling in DNA gyrase and the mechanism of action of novobiocin. Proc. Natl. Acad. Sci. U. S. A. *75*, 4838–4842.

44. Deng, S., and Stein, R.A. Higgins Organization of supercoil domains and their reorganization by transcription.

45. Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. U. S. A. *84*, 7024–7027.

46. Wu, H.Y., Shyy, S.H., Wang, J.C., and Liu, L.F. (1988). Transcription generates positively and negatively supercoiled domains in the template. Cell *53*, 433–440.

47. Tsao, Y.P., Wu, H.Y., and Liu, L.F. (1989). Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies. Cell *56*, 111–118.

48. Samul, R., and Leng, F. (2007). Transcription-coupled hypernegative supercoiling of plasmid DNA by T7 RNA polymerase in Escherichia coli topoisomerase I-deficient strains. J. Mol. Biol. *374*, 925–935.

49. Ma, J., and Wang, M.D. (2016). DNA supercoiling during transcription. Biophys. Rev. *8*, 75–87.

50. Rovinskiy, N., Agbleke, A.A., Chesnokova, O., Pang, Z., and Higgins, N.P. (2012). Rates of gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome. PLoS Genet. *8*, e1002845.

51. El Hanafi, D., and Bossi, L. (2000). Activation and silencing of leu-500 promoter by transcription-induced DNA supercoiling in the Salmonella chromosome. Mol. Microbiol. *37*, 583–594.

52. Guptasarma, P. (1996). Cooperative relaxation of supercoils and periodic transcriptional initiation within polymerase batteries. Bioessays *18*, 325–332.

53. Mizuuchi, K., O'Dea, M.H., and Gellert, M. (1978). DNA gyrase: subunit structure and ATPase activity of the purified enzyme. Proc. Natl. Acad. Sci. U. S. A. *75*, 5960–5963.

54. Dorman, C.J. (2019). DNA supercoiling and transcription in bacteria: a two-way street. BMC Mol Cell Biol *20*, 26.

55. Revyakin, A., Ebright, R.H., and Strick, T.R. (2004). Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation. Proc. Natl. Acad. Sci. U. S. A. *101*, 4776–4780.

56. Vanderlinden, W., Skoruppa, E., Kolbeck, P.J., Carlon, E., and Lipfert, J. (2022). DNA fluctuations reveal the size and dynamics of topological domains. PNAS Nexus *1*. 10.1093/pnasnexus/pgac268.

57. Postow, L., Hardy, C.D., Arsuaga, J., and Cozzarelli, N.R. (2004). Topological domain structure of the Escherichia coli chromosome. Genes Dev. *18*, 1766–1779.

58. Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S.A. (2001). Structural mechanism for rifampicin inhibition of bacterial rna polymerase. Cell *104*, 901–912.

59. Villain-Guillot, P., Bastide, L., Gualtieri, M., and Leonetti, J.-P. (2007). Progress in targeting bacterial transcription. Drug Discov. Today *12*, 200–208.

60. Wehrli, W. (1983). Rifampin: mechanisms of action and resistance. Rev. Infect. Dis. *5 Suppl 3*, S407-11.

61. Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533–538.

62. Liu, M., Tolstorukov, M., Zhurkin, V., Garges, S., and Adhya, S. (2004). A mutant spacer sequence between -35 and -10 elements makes the Plac promoter hyperactive and cAMP receptor protein-independent. Proc. Natl. Acad. Sci. U. S. A. *101*, 6911–6916.

63. Gabizon, R., Lee, A., Vahedian-Movahed, H., Ebright, R.H., and Bustamante, C.J. (2018). Pause sequences facilitate entry into long-lived paused states by reducing RNA polymerase transcription rates. Nat. Commun. *9*. 10.1038/s41467-018-05344-9.

64. Anikin, M., Molodtsov, V., Temiakov, D., and McAllister, W.T. (2010). Transcript Slippage and Recoding. In Recoding: Expansion of Decoding Rules Enriches Gene Expression Nucleic acids and molecular biology. (Springer New York), pp. 409–432.

65. Sneppen, K., Dodd, I.B., Shearwin, K.E., Palmer, A.C., Schubert, R.A., Callen, B.P., and Egan, J.B. (2005). A mathematical model for transcriptional interference by RNA polymerase traffic in Escherichia coli. J. Mol. Biol. *346*, 399–409.

66. Ma, J., Bai, L., and Wang, M.D. (2013). Transcription under torsion. Science *340*, 1580–1583.

67. Callen, B.P., Shearwin, K.E., and Egan, J.B. (2004). Transcriptional interference between convergent promoters caused by elongation over the promoter. Mol. Cell *14*, 647–656.

68. Liou, G.-G., Chao Kaberdina, A., Wang, W.-S., Kaberdin, V.R., and Lin-Chao, S. (2022). Combined transcriptomic and proteomic profiling of E. coli under microaerobic versus aerobic conditions: The multifaceted roles of noncoding small RNAs and oxygen-dependent sensing in global gene expression control. Int. J. Mol. Sci. *23*, 2570.

69. Meng, L., Guo, Y., Tang, Q., Huang, R., Xie, Y., and Chen, X. (2020). Metabolic RNA labeling for probing RNA dynamics in bacteria. Nucleic Acids Res. *48*, 12566–12576.

70. Guyet, A., Dade-Robertson, M., Wipat, A., Casement, J., Smith, W., Mitrani, H., and Zhang, M. (2018). Mild hydrostatic pressure triggers oxidative responses in Escherichia coli. PLoS One *13*, e0200660.

71. LaVoie, S.P., and Summers, A.O. (2018). Transcriptional responses of Escherichia coli during recovery from inorganic or organic mercury exposure. BMC Genomics *19*, 52.

72. Bie, L., Zhang, M., Wang, J., Fang, M., Li, L., Xu, H., and Wang, M. (2023). Comparative analysis of transcriptomic response of Escherichia coli K-12 MG1655 to nine representative classes of antibiotics. Microbiol. Spectr., e0031723.

73. von Wulffen, J., Ulmer, A., Jäger, G., Sawodny, O., and Feuer, R. (2017). Rapid sampling of Escherichia coli after changing oxygen conditions reveals transcriptional dynamics. Genes (Basel) *8*, 90.

74. Revilla-Guarinos, A., Dürr, F., Popp, P.F., Döring, M., and Mascher, T. (2020). Amphotericin B specifically induces the two-component system LnrJK: Development of a novel whole-cell biosensor for the detection of amphotericin-like polyenes. Front. Microbiol. *11*. 10.3389/fmicb.2020.02022.

75. Avci, F.G., Atas, B., Aksoy, C.S., Kurpejovic, E., Gulsoy Toplan, G., Gurer, C., Guillerminet, M., Orelle, C., Jault, J.-M., and Sariyar Akbulut, B. (2019). Repurposing bioactive aporphine alkaloids as efflux pump inhibitors. Fitoterapia *139*, 104371.

76. Popp, P.F., Benjdia, A., Strahl, H., Berteau, O., and Mascher, T. (2020). The epipeptide YydF intrinsically triggers the cell envelope stress response of Bacillus subtilis and causes severe membrane perturbations. Front. Microbiol. *11*, 151.

77. Nagler, K., Krawczyk, A.O., De Jong, A., Madela, K., Hoffmann, T., Laue, M., Kuipers, O.P., Bremer, E., and Moeller, R. (2016). Identification of differentially expressed genes during Bacillus subtilis spore outgrowth in high-salinity environments using RNA sequencing. Front. Microbiol. *7*, 1564.

78. Futo, M., Opašić, L., Koska, S., Čorak, N., Široki, T., Ravikumar, V., Thorsell, A., Lenuzzi, M., Kifer, D., Domazet-Lošo, M., et al. (2021). Embryo-Like Features in Developing Bacillus subtilis Biofilms. Mol. Biol. Evol. *38*, 31–47.

79. Wang, B., van der Kloet, F., Kes, M.B.M.J., Luirink, J., and Hamoen, L.W. (2022). Transcriptome analyses using regulon, functional category, and operon information with GINtool. bioRxiv. 10.1101/2022.10.24.513545.

80. Syvertsson, S., Wang, B., Staal, J., Gao, Y., Kort, R., and Hamoen, L.W. (2021). Different resource allocation in a Bacillus subtilis population displaying bimodal motility. J. Bacteriol. *203*, e0003721.

81. Walter, T., Veldmann, K.H., Götker, S., Busche, T., Rückert, C., Kashkooli, A.B., Paulus, J., Cankar, K., and Wendisch, V.F. (2020). Physiological response of Corynebacterium glutamicum to indole. Microorganisms *8*, 1945.

82. Pérez-García, F., Brito, L.F., and Wendisch, V.F. (2019). Function of L-pipecolic acid as compatible solute in Corynebacterium glutamicum as basis for its production under hyperosmolar conditions. Front. Microbiol. *10*, 340.

83. Loh, J.T., Shum, M.V., Jossart, S.D.R., Campbell, A.M., Sawhney, N., McDonald, W.H., Scholz, M.B., McClain, M.S., Forsyth, M.H., and Cover, T.L. (2021). Delineation of the pH-responsive regulon controlled by the Helicobacter pylori ArsRS two-component system. Infect. Immun. *89*. 10.1128/IAI.00597-20.

84. Ribeiro, A.S., and Lloyd-Price, J. (2007). SGN Sim, a stochastic genetic networks simulator. Bioinformatics *23*, 777–779.

85. Gillespie, D.T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. *22*, 403–434.

86. Gillespie, D.T. (1977). Exact Stochastic Simulation of Coupled Chemical Reactions. *81*, 2340–2361.

87. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

88. Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. *49*, D325–D334.

89. Riley, M. (1993). Functions of the gene products of Escherichia coli. Microbiol. Rev. *57*, 862–952.

90. EMBL-EBI QuickGO. https://www.ebi.ac.uk/QuickGO/term/GO:0008150.

91. Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., Ote, T., Yamakawa, T., Yamazaki, Y., Mori, H., et al. (2004). Cell size and nucleoid organization of engineered Escherichia coli cells with a reduced genome. Mol. Microbiol. *55*, 137–149.

92. Kato, J.-I., and Hashimoto, M. (2007). Construction of consecutive deletions of the Escherichia coli chromosome. Mol. Syst. Biol. *3*, 132.

93. Okuda, S., Kawashima, S., Kobayashi, K., Ogasawara, N., Kanehisa, M., and Goto, S. (2007). Characterization of relationships between transcriptional units and operon structures in Bacillus subtilis and Escherichia coli. BMC Genomics *8*, 48.

94. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 2006.0008-2006.0008.

95. Steward, K.L., and Linn, T. (1991). In vivo analysis of overlapping transcription units in the rplKAJLrpoBC ribosomal protein—RNA polymerase gene cluster of Escherichia coli. J. Mol. Biol. *218*, 23–31.

96. Igarashi, K., Fujita, N., and Ishihama, A. (1989). Promoter selectivity of*Escherichia coli*RNA polymerase: omega factor is responsible for the ppGpp sensitivity. Nucleic Acids Res. *17*, 8755–8765.

97. Yamamoto, M., and Nomura, M. (1979). Organization of genes for transcription and translation in the rif region of the Escherichia coli chromosome. J. Bacteriol. *137*, 584–594.

98. Ralling, G., and Linn, T. (1984). Relative activities of the transcriptional regulatory sites in the rplKAJLrpoBC gene cluster of Escherichia coli. J. Bacteriol. *158*, 279–285.

99. Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G., and Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. Nat. Methods *3*, 623–628.

100. Chen, M., and Fredrick, K. (2020). RNA polymerase's relationship with the ribosome: Not so physical, most of the time. J. Mol. Biol. *432*, 3981–3986.

101. Wang, Y., Yue, X.-J., Yuan, S.-F., Hong, Y., Hu, W.-F., and Li, Y.-Z. (2021). Internal promoters and their effects on the transcription of operon genes for epothilone production in Myxococcus xanthus. Front. Bioeng. Biotechnol. *9*. 10.3389/fbioe.2021.758561.

102. Li, Y., and Altman, S. (2004). Polarity Effects in the Lactose Operon of Escherichia coli. J. Mol. Biol. *339*, 31–39.

103. Mateus, A., Shah, M., Hevler, J., Kurzawa, N., Bobonis, J., Typas, A., and Savitski, M.M. (2021). Transcriptional and post-transcriptional polar effects in bacterial gene deletion libraries. mSystems *6*. 10.1128/msystems.00813-21.

104. Rho-dependent Transcriptional Polarity in the iZvGMEDA Operon of Wild-type Escherichia coli K12.

105. Munita, J.M., and Arias, C.A. (2016). Mechanisms of antibiotic resistance. Microbiol. Spectr. *4*. 10.1128/microbiolspec.VMBF-0016-2015.

106. Phroskin, S., Rahmouni, R., Mironov, A., and Nudler, A. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science *328*, 504–508.

107. Ma, J., and Wang, M. (2014). Interplay between DNA supercoiling and transcription elongation. Transcription *5*, e28636.

108. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.