BBA - Gene Regulatory Mechanisms 1865 (2022) 194812

Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms

journal homepage: www.elsevier.com/locate/bbagrm

Sequence-dependent model of genes with dual σ factor preference

Ines S.C. Baptista [a,1], Vinodh Kandavalli [a,b,1], Vatsala Chauhan [a], Mohamed N.M. Bahrudeen [a], Bilena L.B. Almeida [a], Cristina S.D. Palma [a], Suchintak Dash [a], Andre S. Ribeiro [a,c,*]

[a] Laboratory of Biosystem Dynamics, Faculty of Medicine and Health Technology, Tampere University, Tampere 33520, Finland
[b] Department of Cell and Molecular Biology, Uppsala University, Uppsala 752 37, Sweden
[c] Center of Technology and Systems (CTS-Uninova), NOVA University of Lisbon, 2829-516 Monte de Caparica, Portugal

ARTICLE INFO

Keywords:
Dual σ factor preference
Sequence-dependent gene expression model
RNA-seq
Flow cytometry
E. coli

ABSTRACT

Escherichia coli uses σ factors to quickly control large gene cohorts during stress conditions. While most of its genes respond to a single σ factor, approximately 5% of them have dual σ factor preference. The most common are those responsive to both σ^{70}, which controls housekeeping genes, and σ^{38}, which activates genes during stationary growth and stresses. Using RNA-seq and flow-cytometry measurements, we show that 'σ^{70+38} genes' are nearly as upregulated in stationary growth as 'σ^{38} genes'. Moreover, we find a clear quantitative relationship between their promoter sequence and their response strength to changes in σ^{38} levels. We then propose and validate a sequence dependent model of σ^{70+38} genes, with dual sensitivity to σ^{38} and σ^{70}, that is applicable in the exponential and stationary growth phases, as well in the transient period in between. We further propose a general model, applicable to other stresses and σ factor combinations. Given this, promoters controlling σ^{70+38} genes (and variants) could become important building blocks of synthetic circuits with predictable, sequence-dependent sensitivity to transitions between the exponential and stationary growth phases.

1. Introduction

In *E. coli*, genes are expressed by RNA polymerases (RNAP) core enzymes, which have 5 subunits ($\alpha_2\beta\beta'\omega$). When bound to a σ factor, they become able to recognize specific promoters and, from there, synthesize RNA [1]. Transcription is regulated mostly at the promoter regions, which typically harbor transcription factor (TF) binding sites (TFBS) and other regulatory sequence motifs [2–7]. This regulation is essential for cellular adaptability to both internal as well as external conditions [8,9].

Since σ factors are needed for transcription and because cells are able to regulate their numbers, they are themselves a regulatory mechanism of gene expression [4,10–13]. *E. coli* has seven different σ factors [14]. During exponential growth in optimal conditions, RNAP mostly transcribes genes with preference for σ^{70} (i.e. genes whose promoters are more likely to be recognized by σ^{70} than by other σ factors [15]). These genes are mostly responsible for basic cell functions [15]. Other σ factors are only present under specific stresses [16].

As an example, when the environment becomes depleted of components required for cell growth, *E. coli* usually switch to stationary growth, largely triggered by the appearance of RpoS (σ^{38}), which activates ~10% of the genome [17,18], leading to key phenotypic modifications [18–24]. Meanwhile, since the concentration of RNAP core enzymes remains relatively constant [25] and since the RNAP numbers are lower than σ factor numbers, this will force σ factors to compete for RNAP's [10,12,14,20,26,27]. Consequently, when σ^{38} numbers increase, the genes unresponsive to σ^{38}, if previously active, will be indirectly negatively regulated [10,26,28].

For the σ factor regulatory system to be efficient, promoters need to have a high affinity to one σ factor (strong binding interaction) while also having little to no affinity to other σ factors. In agreement, only a small fraction of promoters can recognize more than one σ factor [6,29,30]. Of these, the most common (84%) are the promoters responsive to both σ^{70} as well as σ^{38} [31] (Supplementary Table S3 and Section 2.4.1). Following the workflow in Fig. 1, we investigated how the promoter sequences recognized by both σ^{70} and σ^{38} relate to the dynamics of the genes that they control (named 'σ^{70+38} genes') prior, during, and following the transition to stationary growth. A graphical representation of the sequence conservation (sequence logo) of the nucleotides between the positions −41 to −1 relative to the transcription start site of the promoters considered is shown in Fig. 2.

* Corresponding author at: Arvo Ylpön katu 34, 33520 Tampere, Finland.
E-mail address: andre.sanchesribeiro@tuni.fi (A.S. Ribeiro).
[1] Equal contributions.

https://doi.org/10.1016/j.bbagrm.2022.194812
Received 2 December 2021; Received in revised form 8 March 2022; Accepted 16 March 2022
Available online 22 March 2022

2. Materials and methods

2.1. Strains and media

E. coli strains and plasmids are listed in Supplementary Tables S1 and S2. We used YFP strains from the genetic stock center (CGSC) of Yale University, U.S.A [33] and, as support, a low-copy plasmid fusion library of fluorescent (GFP) reporter strains to track promoter activity [34]. In both, the fluorescent proteins under the control of the promoters of our genes of interest have been shown to be a good proxy for the native protein levels. For simplicity, we refer to these fluorescence levels as 'protein levels'.

We used a RL1314 strain (rpoC::GFP) generously provided by Robert Landick (University of Wisconsin-Madison), to measure RNA Polymerase levels [35]. Their *rpoC* gene codes for β' sub-unit endogenously tagged with GFP. Since *rpoC* codes for the β' subunit, a limiting factor in the assembly of RNAP holoenzyme [36,37], its numbers serve as a good proxy for RNAP numbers. For simplicity, [RNAP] refers to the concentration of both RNAP core and holoenzymes in cells. Finally, we used a MG*mCherry* (*rpoS*::mCherry) strain to measure RpoS levels (kind gift from James Locke [19]). Their *rpoS* gene codes for σ^{38} endogenously tagged with mCherry. Finally, we used wild-type K12 MG1655 strain for control.

We used M9 medium (1xM9 Salts, 2 mM MgSO4, 0.1 mM CaCl2; 5xM9 Salts 34 g/L Na2HPO4, 15 g/L KH2PO4, 2.5 g/L NaCl, 5 g/L NH4Cl) supplemented with 0.2% Casamino acids and 0.4% glucose, and Luria-Bertani (LB) medium with 10 g peptone, 10 g NaCl, and 5 g yeast extract in 1000 mL distilled water. We used the antibiotics kanamycin and chloramphenicol from Sigma Aldrich, U.S.A.

2.2. Growth rate and growth phase

Growth rates were measured by spectrophotometry (BioTek Synergy HTX Multi-Mode Microplate Reader). From a glycerol stock (−80 °C), cells were streaked on LB agar plates (2%) and incubated at 37 °C, overnight. A single colony was picked, inoculated in LB medium with antibiotics (Section 2.1), and incubated at 30 °C overnight with shaking. Overnight cultures were further diluted into fresh medium to an optical density of 600 nm ($O.D._{600}$) of 0.01 and incubated for growth by shaking at 250 rpm at 37 °C. OD_{600} was recorded every 20 min for 800 min. Cells were extracted at 150 min and 700 min after inoculation into fresh medium to represent cells in exponential and stationary growth phases, respectively (Fig. 3A).

2.3. Gene expression measurements

To measure gene expression, we performed flow cytometry using an ACEA NovoCyte Flow Cytometer equipped with yellow (561 nm) and blue lasers (488 nm) and controlled by Novo Express V1.50 (Supplementary Section S1.1). Meanwhile, to measure σ^{38} *over time*, we used spectrophotometry (Supplementary Section S1.2). To convert protein expression levels to concentration levels, we used cell areas (as proxy for cell volume) which were obtained by microscopy and image analysis (Supplementary Section S1.3). Finally, we measured changes in RNA levels between growth conditions by RNA-seq (Supplementary Section S1.4).

2.4. Promoter sequences

2.4.1. σ factor preference

Supplementary Table S3 informs on the genes' σ factor preference. From Regulon DB v10.5 (August 14, 2019), we obtained lists of all transcription units (TUs), promoters (including σ factor preferences), and genes of *E. coli* [32]. Recently, we compared our lists with information obtained July 1, 2021 and found no changes that would affect the conclusions. TUs only differed by ~1%, promoters by ~0.5%, and genes by ~1%.

From the 3548 TUs (gene(s) transcribed from a single promoter), we extracted 2179 with known promoters, containing 2713 genes in total. To minimize interference to the classification of σ factor preferences arising from transcription by multiple promoters, we narrowed our list to 1824 genes transcribed by only one promoter and, of those, to the 1328 genes with known σ factor preference. From those, 1242 have a preference for only one σ factor, including 931 with a preference for σ^{70}, and 93 with a preference for σ^{38}. Conversely, 76 genes have promoters with a preference for two σ factors. Out of these, 64 are transcribed by only one promoter with a preference for σ^{70} and σ^{38}.

2.4.2. Sequence logo

Promoter sequence logos were created using WebLogo [38]. In each position (from −41 to −1), we counted how many times a nucleotide is present in all promoters considered. Then, we stacked all nucleotides (A, C, T, G) on top of each other, and sorted from the least found one in the bottom to the most present one on the top. For each position, we quantified its 'bit', as the difference between the maximum information possible (entropy given the 4 nucleotides) and the information considering the variability of the nucleotides (sum of the entropy for each of

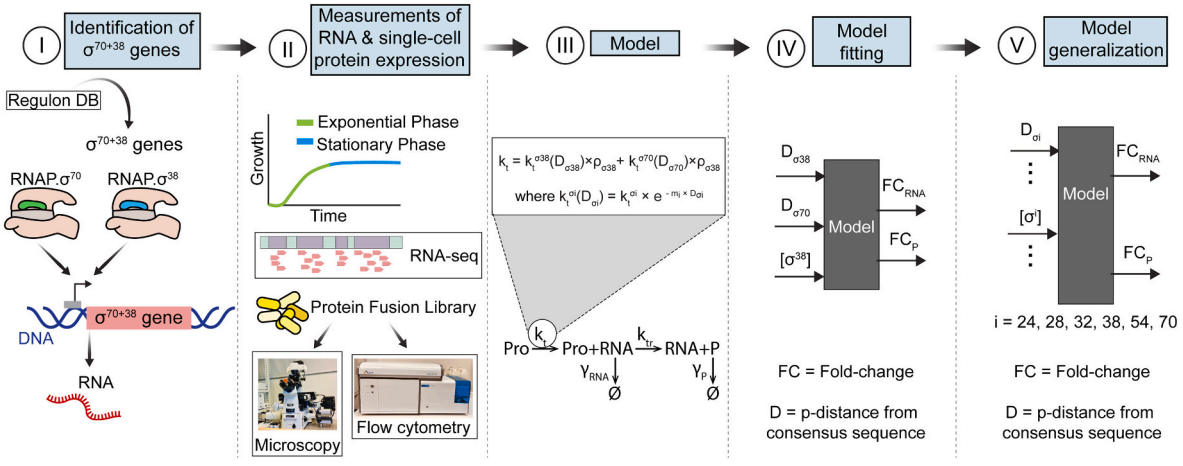

Fig. 1. Workflow. (I) From Regulon DB [32], we identified genes controlled by single promoters with preference for both σ^{70} and σ^{38} ('σ^{70+38} genes'). (II) Next, we measured RNA and single-cell protein levels of σ^{70+38} genes in the exponential and stationary growth phases. (III) Then, we proposed an empirically based model of gene expression fold changes of σ^{70+38} genes in RNA (FC_{RNA}) and protein numbers (FC_P) between growth phases. (IV) Afterwards, we tuned the model to fit how the promoter sequence affects the response to σ^{38}. (V) Finally, we generalized the model to be applicable to genes with preference(s) for any set of σ factors.

the 4 nucleotides) in that position (observed entropy): Bit $= \log_2 4 - \sum_{n=1}^{4} f_n \times \log_2(f_n)$. From this, the height of each letter is set to be proportional to its frequency of occurrence. Finally, the height of each position was normalized, so as to equal its corresponding amount of information (with the more conserved positions having more bits) [39].

2.4.3. p-Distance

The p-distance D of a promoter [40] is the fraction of its nucleotides between positions −41 to −1 (assuming that transcription start site starts at position +1) that differ from the consensus (most common) nucleotide in that position of a cohort of genes (here, genes with preference for a given σ factor). We extracted the consensus sequences related to each σ from RegulonDB [32]. To measure the D of promoters with preference for both σ^{38} and σ^{70}, we calculated the p-distances $D_{\sigma 38}$ and $D_{\sigma 70}$, and then an overall p-distance, Δ, defined as: $\Delta = D_{\sigma 38} - D_{\sigma 70}$.

2.5. Statistical analysis

We used two-sample Kolmogorov-Smirnov tests (KS tests) to compare distributions. Also, we used the Distribution Fitter app (MATLAB) to perform Gaussian fittings, the MATLAB's curve toolbox to fit curves, linear regression models to fit linear correlations, and least squares fitting [41] to fit Hill functions (Supplementary Section S1.6). To fit and validate surfaces, we used cross-validation resampling [42]). Finally, we used Fisher's exact tests to find overrepresentations in gene ontology (Supplementary Section S1.6). To compare if the linear fits of two different groups differ significantly, we extracted the fits' slope and intercept values and performed an ANOVA F-tests (of the null-hypothesis that the values are the same for both groups) for each estimated value.

3. Results

3.1. RNA fold changes when shifting to stationary growth

To study the dynamics of σ^{70+38} genes, we first identified when, after placing cells in fresh media, they transition from exponential and stationary growth (Section 2.2). We used both wild-type (WT, control) cells and a MGmCherry strain [19] carrying fluorescently tagged σ^{38}. From Fig. 3A–C, cells are in the mid-exponential growth phase 150 min after moved to fresh medium, while their σ^{38} levels are low. Meanwhile, at 500 min onwards, cells are in stationary growth, while their σ^{38} levels are high. Conversely, the RNAP concentration was only ~7% higher during stationary growth (Supplementary Fig. S3C–D).

We then performed RNA-seq, at 150 min and 500 min, and calculated the $\log_2$ fold changes in RNA levels between those moments (LFC_{RNA}) (Section 2.3). In Fig. 3D, we show the fitting of Gaussian functions to the LFC_{RNA} distributions of σ^{70+38} genes, σ^{70} genes, and σ^{38} genes, as well as all other genes. In general, σ^{70+38} genes were nearly as upregulated as σ^{38} genes. On the other hand, σ^{70} genes were weakly downregulated, likely due to the expected indirect negative regulation [10,12,14,20,26,27], i.e. the decrease in their expression levels because of the lower concentration of RNAP core enzymes. Finally, most other genes (~2737 out of the 4308 genes) were relatively unresponsive.

3.2. Propagation of shifts in RNA levels to protein levels

Next, we measured single-cell protein levels (Section 2.3) of 9 of the 64 σ^{70+38} genes (only 15 of the 64 are in the YFP library (Section 2.1) and 6 have too weak signals). These 9 genes, according to their LFC_{RNA}, should cover most of the range of response strengths of the σ^{70+38} genes as measured by RNA-seq (Supplementary Section S1.5).

From the protein levels, we extracted the mean protein levels, μ_P, and the corresponding squared coefficient of variation, CV^2_P, during exponential and stationary growth, respectively. We then calculated the $\log_2$ fold changes in μ_P, LFC_P, when shifting from exponential to stationary growth.

Confronting the changes due to the shift in the growth phase in the protein levels and in RNA levels (Fig. 4A), we found that they linearly positively correlate ($R^2 = 0.62$, F-test p-value <0.05) implying that the changes in RNA levels propagated to the protein levels.

Notably, most proteins measured (7 of 9) exhibited $LFC_P < 0$, while only 3 of 9 of the corresponding RNAs had $LFC_{RNA} < 0$. Given this discrepancy, we performed additional measurements (in the exponential and stationary growth phases) of other single-cell protein numbers controlled by σ^{70+38} promoters but using instead a GFP Promoter Fusion Library [34].

As before, the RNA and protein changes correlate ($R^2 = 0.42$, p-value <0.05, Supplementary Fig. S9). Meanwhile, the number of decreasing RNA numbers ($LFC_{RNA} < 0$) was only 7 out of 14 (50%), while the number of decreasing protein abundances ($LFC_P < 0$) was 8 out of 14 (57%). This better agreement between RNA and proteins levels using this library is expected, given that the proteins are produced from low-copy number plasmids that then a single gene integrated in the chromosome (which is expected to be noisier in RNA numbers).

Next, we compared the relationships between LFC_{RNA} and LFC_P when using each strain library (Supplementary Fig. S9). For this, we best fitted lines, and obtained $LFC_P = 0.11 \cdot LFC_{RNA} - 0.67$ for the YFP strain library and $LFC_P = 0.22 \cdot LFC_{RNA} + 0.17$ for the Promoter Fusion library. These 2 lines do not differ statistically (ANOVA F-tests p-value <0.05 for the estimated slopes and the estimated intercepts). We conclude that the two data sets are consistent, thus, as expected, both strain libraries can be used to track changes in RNA numbers from the corresponding protein levels.

Finally, we investigated how many σ^{70+38} genes exhibited reductions in their protein concentrations (rather than in their absolute abundances), when cells shifted to stationary growth. For this, we considered that the cell volume decreased in the stationary growth phase by ~17% (estimated from the cell areas, Supplementary Fig. S3B).

From the concentrations in the exponential and stationary growth (Supplementary Fig. S10), only 6 of the 14 protein concentrations decreased when changing to stationary growth. Overall, we find that most σ^{70+38} genes have their protein concentrations increased when cells shift to stationary growth, but both behaviors (increasing or

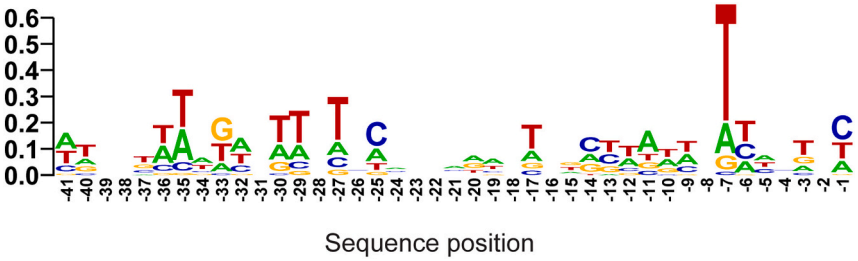

Fig. 2. Sequence logo between positions −41 and − 1 of promoter regions, with 0 being the TSS, of the 64 promoters with preference for both σ^{70} and σ^{38}, obtained as described in Section 2.4.2. The sequence logo of all other 8493 promoters in *E. coli* is shown in Supplementary Fig. S1B.

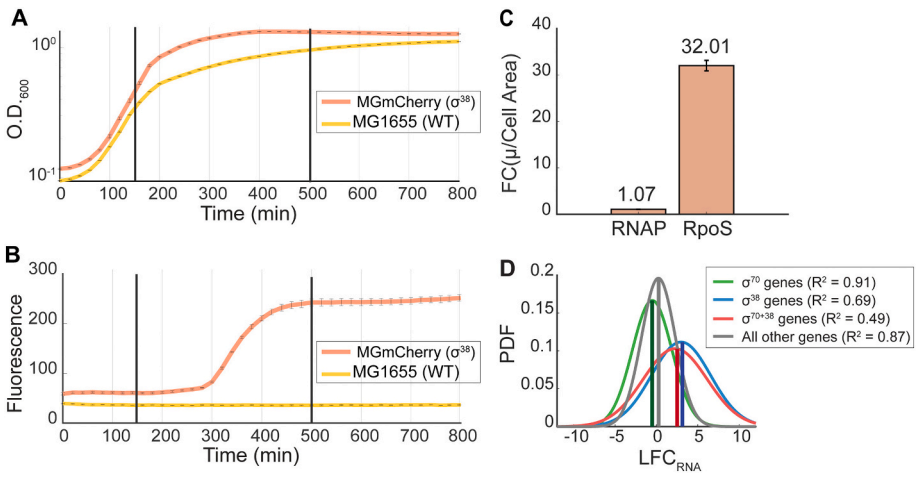

Fig. 3. Cell growth, RNAP, RpoS and genome-wide RNA levels when changing from exponential to stationary growth. (A) Optical densities 'O.D.$_{600}$' of wild-type MG1655 (control) and MGmCherry (*rpoS::mcherry*) strains (Section 2.1). (B) Mean population fluorescence of WT and MGmCherry strains. Black vertical bars show the timing of measurements in exponential (150 min) and stationary (500 min onwards) growth. At 0 min, cells were moved to fresh medium (Section 2.2). (C) Fold-change between the exponential and stationary growth phases of mean cellular fluorescence relative to cell area (FC(μ/Cell Area)) due to changes in RNAP-GFP and in RpoS-mCherry, respectively (Cell Areas are shown in Supplementary Fig. S3). Error bars are the standard error of the mean (SEM). (D) Gaussian fits to the distributions of LFC$_{RNA}$ of gene cohorts. Vertical lines mark the mean. Given the high R^2 (the coefficient of determination which a measure of the goodness of fit) of the fits, they likely capture well the shapes of the respective empirical distributions.

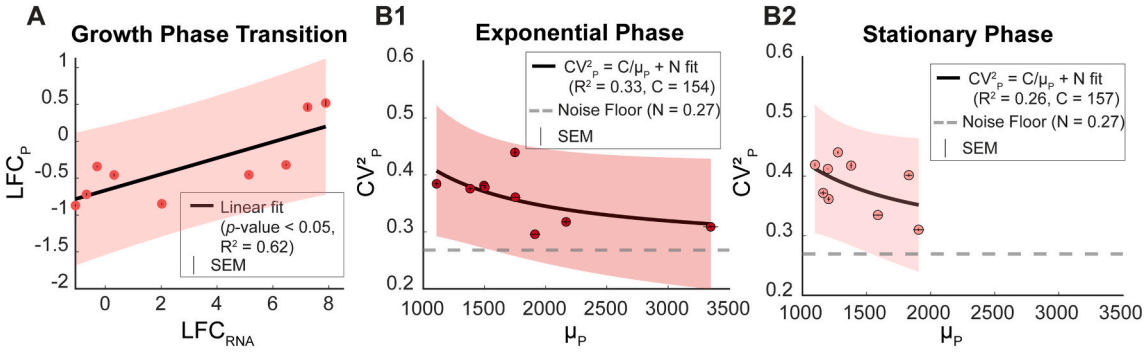

Fig. 4. Single-cell protein levels of σ^{70+38} genes in the exponential and stationary growth phases. (A) Log_2 fold-change of protein levels from exponential to stationary growth, LFC_P, plotted against the corresponding LFC_{RNA}. (B1) and (B2) CV^2_P plotted against μ_P of σ^{70+38} genes in the exponential and stationary growth phases, respectively. Error bars (small) are the standard error of mean. All figures show the best fitting curves and their 95% confidence bounds (range of values expected to contain the true mean, represented by the shadow areas). Horizontal dashed lines are the noise floors.

decreasing) can be expected.

Finally, we analyzed the variability in single-cell protein levels, prior and after the growth phase transition. In Figures 4B1 and 4B2, CV^2_P decreases quickly with μ_P for small μ_P, but slowly for high μ_P. This is well described by a function of the form: $CV_P^2 = C \times \mu_P^{-1} + N$ [33] (N is the noise floor, a lower bound on the cell-to-cell variability of protein levels in clonal populations due to extrinsic factors [43]), in agreement with [33,44,45].

3.3. Sequence-dependent model of σ^{70+38} genes

Based on the above, we proposed a sequence-dependent model for σ^{70+38} genes. We assumed that only σ^{70} is present in high numbers (in holoenzyme form) during the exponential phase and that only σ^{38} increases significantly in numbers when shifting to stationary growth, in agreement with [10,14,20,26,46–48].

The model is designed to account for competition between σ^{70} and σ^{38} to bind to RNAP core enzymes, since these exist in limited numbers [10,12,14,20,26,27]. For this, we set reactions for binding and unbinding of σ^{70} and σ^{38} to free floating RNAP core enzymes, (R1a) and (R1b), where $K^{\sigma 70}$ and $K^{\sigma 38}$ are the ratios between the respective association and dissociation rate constants:

$$\mathrm{RNAP} + \sigma^{70} \xleftrightarrow{K^{\sigma 70}} \mathrm{RNAP}.\sigma^{70} \quad \text{(R1a)}$$

$$\mathrm{RNAP} + \sigma^{38} \xleftrightarrow{K^{\sigma 38}} \mathrm{RNAP}.\sigma^{38} \quad \text{(R1b)}$$

Given (R1a) and (R1b), the RNA production dynamics depends on the ratio between σ^{70} and σ^{38} numbers if, and only if, $K^{\sigma 70}$ and $K^{\sigma 38}$ differ. The limited number of RNAPs is accounted for since, from (R1a) and (R1b) alone:

$$[RNAP_{total}] = [RNAP.\sigma^{70}] + [RNAP.\sigma^{38}]. \quad (1)$$

To introduce the dual responsiveness to σ^{70} and σ^{38}, we defined two competing, sequence-dependent reactions of transcription, (R2a) and (R2b), differing in which holoenzyme binds to the promoter. The rates $k_t^{\sigma 70}$ and $k_t^{\sigma 38}$ control the binding affinities to the holoenzymes and are expected to differ between promoters (and potentially with the transcription factors acting on the promoters, not represented for simplicity).

$$\mathrm{Pro} + \mathrm{RNAP}.\sigma^{70} \xrightarrow{k_t^{\sigma 70}(D_{\sigma 70})} \mathrm{Pro} + \mathrm{RNA} + \mathrm{RNAP} + \sigma^{70} \quad \text{(R2a)}$$

$$\mathrm{Pro} + \mathrm{RNAP}.\sigma^{38} \xrightarrow{k_t^{\sigma 38}(D^{\sigma 38})} \mathrm{Pro} + \mathrm{RNA} + \mathrm{RNAP} + \sigma^{38} \quad \text{(R2b)}$$

The reactions' rate constants are set to be sequence dependent in what concerns σ^{38} and σ^{70} dependency. In detail, $D_{\sigma 38}$ and $D_{\sigma 70}$, are the p-distances, in nucleotides, of a promoter sequence from the consensus (average) sequence of promoters with σ^{38} and with σ^{70} dependency, respectively (Section 2.4.3).

Taken together, (R1a), (R1b), (R2a) and (R2b), model the transcription kinetics of σ^{70+38} genes before, during and, after shifting from exponential to stationary growth. The rates $k_t^{\sigma 70}(D_{\sigma 70})$ and $k_t^{\sigma 38}(D_{\sigma 38})$ are dissected below.

3.4. Reduced model

The model above can be reduced since, first, the numbers of RNAP.σ^{70} and RNAP.σ^{38} in the cells are significantly larger than 1 [14,20,28,46–48]. As such, (R1a, R1b, R2a, and R2b) can be reduced to (R3a) and (R3b):

$$\text{Pro} \xrightarrow{[RNAP.\sigma^{70}]\cdot k_t^{\sigma 70}(D_{\sigma 70})} \text{Pro} + \text{RNA} + \text{RNAP} + \sigma^{70} \quad \text{(R3a)}$$

$$\text{Pro} \xrightarrow{[RNAP.\sigma^{38}]\cdot k_t^{\sigma 38}(D_{\sigma 38})} \text{Pro} + \text{RNA} + \text{RNAP} + \sigma^{38} \quad \text{(R3b)}$$

Further, one can merge (R3a) and (R3b) into a single transcription process:

$$\text{Pro} \xrightarrow{[RNAP.\sigma^{70}]\cdot k_t^{\sigma 70}(D_{\sigma 70})+[RNAP.\sigma^{38}]\cdot k_t^{\sigma 38}(D_{\sigma 38})} \text{Pro} + \text{RNA} \quad \text{(R4)}$$

It was not possible to reduce the model further (e.g., by making $D_{\sigma 70}$ a function of $D_{\sigma 38}$) since we failed to find evidence for a correlation between $D_{\sigma 70}$ and $D_{\sigma 38}$ (Supplementary Fig. S6), except for the reduced set of σ^{70+38} genes *without* known input TFs (Fig. 5C).

Finally, we included reactions for translation of RNAs into proteins Eq. (R5), and for RNA Eq. (R6) and protein Eq. (R7) decay due to degradation and dilution by cell division:

$$\text{RNA} + \text{Rib} \xrightarrow{k_{tr}} \text{RNA} + \text{Rib} + \text{P} \quad \text{(R5)}$$

$$\text{RNA} \xrightarrow{\gamma_{RNA}} \varnothing \quad \text{(R6)}$$

$$\text{P} \xrightarrow{\gamma_P} \varnothing \quad \text{(R7)}$$

3.5. Analytical solution of the reduced model

Next, we obtained an analytical solution for the expected fold change in RNA numbers of a gene whose promoter has preference for both σ^{70} as well as σ^{38}. From Eqs. (R4), (R5), (R6), and (R7) the expected RNA numbers of a σ^{70+38} gene should equal:

$$RNA = \frac{[RNAP.\sigma^{70}]\cdot k_t^{\sigma 70}(D_{\sigma 70}) + [RNAP.\sigma^{38}]\cdot k_t^{\sigma 38}(D_{\sigma 38})}{\gamma_{RNA}} \quad (2)$$

These rate constants are not expected to differ between the growth phases (as they depend on biophysical parameters, such as binding affinities, etc.). Consequently, the fold-change in RNA numbers between the growth phases should equal:

$$FC_{RNA} = \frac{RNA_{\text{sta}}}{RNA_{\text{exp}}} = \frac{[RNAP.\sigma^{70}]_{\text{sta}}\cdot k_t^{\sigma 70}(D_{\sigma 70}) + [RNAP.\sigma^{38}]_{\text{sta}}\cdot k_t^{\sigma 38}(D_{\sigma 38})}{[RNAP.\sigma^{70}]_{\text{exp}}\cdot k_t^{\sigma 70}(D_{\sigma 70}) + [RNAP.\sigma^{38}]_{\text{exp}}\cdot k_t^{\sigma 38}(D_{\sigma 38})}\cdot\Gamma \quad (3)$$

where Γ is the ratio between the RNA decay rates in the two growth phases:

$$\Gamma = \frac{\gamma_{RNA_{\text{exp}}}}{\gamma_{RNA_{\text{sta}}}} \quad (4)$$

Since $[RNAP_{total}]$ is similar in the two growth phases (Fig. 3C):

$$FC_{RNA} = \frac{\frac{[RNAP.\sigma^{70}]_{\text{sta}}}{[RNAP_{total}]}\cdot k_t^{\sigma 70}(D_{\sigma 70}) + \frac{[RNAP.\sigma^{38}]_{\text{sta}}}{[RNAP_{total}]}\cdot k_t^{\sigma 38}(D_{\sigma 38})}{\frac{[RNAP.\sigma^{70}]_{\text{exp}}}{[RNAP_{total}]}\cdot k_t^{\sigma 70}(D_{\sigma 70}) + \frac{[RNAP.\sigma^{38}]_{\text{exp}}}{[RNAP_{total}]}\cdot k_t^{\sigma 38}(D_{\sigma 38})}\Gamma \quad (5)$$

Next, considering Eq. (1), then:

$$\frac{[RNAP.\sigma^{70}]}{[RNAP_{total}]} + \frac{[RNAP.\sigma^{38}]}{[RNAP_{total}]} \approx 1 \quad (6)$$

For simplicity, let ρ_{exp} and ρ_{sta} be:

$$\rho_{\text{exp}} = \frac{[RNAP.\sigma^{38}]_{\text{exp}}}{[RNAP_{total}]} = \frac{[RNAP_{total}]\cdot[\sigma^{38}]_{\text{exp}}\cdot K^{\sigma 38}}{[RNAP_{total}]\cdot\left([\sigma^{70}]_{\text{exp}}\cdot K^{\sigma 70} + [\sigma^{38}]_{\text{exp}}\cdot K^{\sigma 38}\right)} \quad (7a)$$

$$\rho_{sta} = \frac{[RNAP.\sigma^{38}]_{\text{sta}}}{[RNAP_{total}]} = \frac{[RNAP_{total}]\cdot[\sigma^{38}]_{\text{sta}}\cdot K^{\sigma 38}}{[RNAP_{total}]\cdot\left([\sigma^{70}]_{\text{sta}}\cdot K^{\sigma 70} + [\sigma^{38}]_{\text{sta}}\cdot K^{\sigma 38}\right)} \quad (7b)$$

As such, from Eq. (5):

$$FC_{RNA} = \frac{(1-\rho_{\text{sta}})\cdot k_t^{\sigma 70}(D_{\sigma 70}) + \rho_{\text{sta}}\cdot k_t^{\sigma 38}(D_{\sigma 38})}{(1-\rho_{\text{exp}})\cdot k_t^{\sigma 70}(D_{\sigma 70}) + \rho_{\text{exp}}\cdot k_t^{\sigma 38}(D_{\sigma 38})}\cdot\Gamma \quad (8)$$

Finally, since in the exponential growth phase $[RNAP.\ \sigma^{70}] >> [RNAP.\ \sigma^{38}]$, then $\rho_{\text{exp}} \approx 0$. Thus, Eq. (8) can be simplified:

$$FC_{RNA} = \Gamma\cdot(1-\rho_{\text{sta}}) + \rho_{\text{sta}}\cdot\Gamma\cdot\frac{k_t^{\sigma 38}(D_{\sigma 38})}{k_t^{\sigma 70}(D_{\sigma 70})} \quad (9)$$

All parameters in Eq. (9) can be measured, except $k_t^{\sigma 70}(D_{\sigma 70})$ and $k_t^{\sigma 38}(D_{\sigma 38})$. Meanwhile, in the stationary phase: $[\sigma^{38}]^{\text{sta}} = 0.3\cdot[\sigma^{70}]$ [20,26,47]. Also, $K^{\sigma 70} = 5K^{\sigma 38}$ [49]. Thus:

$$\rho^{sta} = \frac{[RNAP.\sigma^{38}]_{\text{sta}}}{[RNAP_{total}]} \approx \frac{0.3\cdot[\sigma^{70}]_{\text{sta}}\cdot\frac{K^{\sigma 70}}{5}}{[\sigma^{70}]_{\text{sta}}\cdot K^{\sigma 70} + 0.3\cdot[\sigma^{70}]_{\text{sta}}\cdot\frac{K^{\sigma 70}}{5}} = 0.0566 \quad (10)$$

3.6. Promoter sequence affects the expression of σ^{70+38} genes

To model how promoter sequences (their $D_{\sigma 70}$ and/or $D_{\sigma 38}$) influence the response strength of σ^{70+38} genes to σ^{38}, we assumed basal transcription rates '$k_{t0}^{\sigma 70}$' and '$k_{t0}^{\sigma 38}$', when transcribed by RNAP.σ^{70} alone and by RNAP.σ^{38} alone, respectively.

To obtain the overall transcription rates of specific promoters, we then multiply $k_{t0}^{\sigma 70}$ and $k_{t0}^{\sigma 38}$ by a single-gene, sequence dependent function f to account for the influence of their $D_{\sigma 70}$ and $D_{\sigma 38}$:

$$k_t^{\sigma 70}(D_{\sigma 70}) = k_{t0}^{\sigma 70}\cdot f(D_{\sigma 70}) \quad \text{(11a)}$$

$$k_t^{\sigma 38}(D_{\sigma 38}) = k_{t0}^{\sigma 38}\cdot f(D_{\sigma 38}) \quad \text{(11b)}$$

To define f we considered that, in general, as $D_{\sigma i}$ increases, the promoter sequence should differ more from the 'average' sequence of

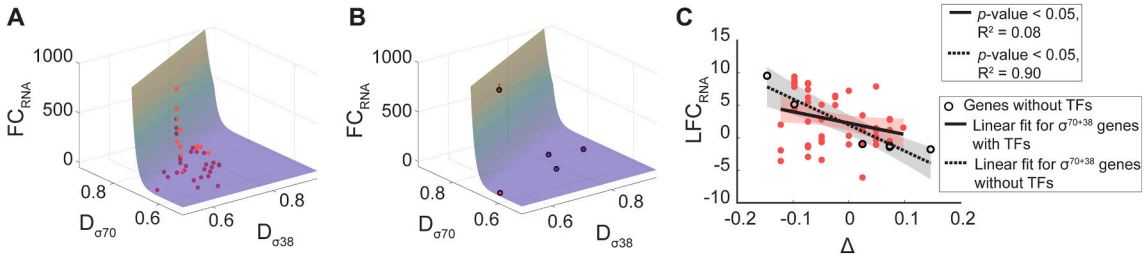

Fig. 5. Model and measurements: fold changes in RNA levels of σ^{70+38} genes plotted against their promoter *p*-distances from the consensus sequence of σ^{70} dependent promoters ($D_{\sigma 70}$) and of σ^{38} dependent promoters ($D_{\sigma 38}$). (A) Best fitting surface to FC(μ_{RNA}) assuming an exponentially decreasing function. Only σ^{70+38} genes with input TFs and FDR < 0.05 are included. Light red points are above the surface, while dark ones are below. (B) Same plot but the surfaces are those obtained from (A) and applied to σ^{70+38} genes without input TFs and FDR < 0.05. The dashed black lines depict the vertical distances between the estimated and measured FC(μ_{RNA}). (C) Scatter plot of $\Delta = D_{\sigma 38} - D_{\sigma 70}$ plotted against LFC_{RNA}. The shadows of the best fitting lines are the 95% confidence bounds.

promoters with preference for σ^i. Thus, its affinity to σ^i should decrease. If this holds true, then the promoter consensus sequence of genes with preference for a given σ factor should have strong affinity to that σ factor. Thus, we hypothesized that the consensus sequence should have the strongest affinity. If true, it follows that as $D_{\sigma 38}$ increases, the transcription rate by RNAP.σ^{38} decreases.

Having this, to model how the transcription rate of promoters recognized by σ^{70} and by σ^{38} specifically differs with $D_{\sigma i}$, we opted for an exponential function, $f(D_{\sigma i}) = e^{-m_i \cdot D_{\sigma i}}$ where $i = 38$ or 70 and m is an empirical-based constant, similar to the one in [6,29,30] for σ^{70} genes. We first best fitted this function to σ^{70+38} genes *with* input TFs (Fig. 5A), and then tested it if it could predict FC_{RNA} of σ^{70+38} genes *without* input TFs (Fig. 5B). These genes should exhibit dynamics related to their promoter sequence (and thus the model), since are not subject to known TFs interference. From Table 1, we obtained an R^2 of 0.98, suggesting that the exponential function can fit well the data (including when compared to other models (Supplementary Section S2.1)).

We further tested the model by randomly resampling (1000 times) the data from all σ^{70+38} genes into fitting and validation sets (Section 2.5). On average, the R^2 to the testing sets equaled 0.5, which is only 30% lower than the R^2 to the fitting sets. We find that the exponential model describes well the behavior of σ^{70+38} genes.

Next, from Fig. 4A, the LFC of protein numbers (LFC_P) correlates linearly with LFC_{RNA}, as expected from the model (reactions R5 to R7) (Supplementary Section S2.2). From the best fitting line, we extracted a scaling factor, α (equaling 0.1), between LFC_{RNA} and LFC_P, to be introduced in (13):

$$LFC_P = log_2\left[\Gamma\cdot(1-\rho_{sta}) + \Gamma\cdot\rho_{sta}\cdot K_0\cdot e^{-(m_{38}\cdot D_{\sigma 38} - m_{70}\cdot D_{\sigma 70})}\right]\cdot\alpha + \beta \quad (13)$$

where β, equaling −0.67, is the intercept between the y axis and the best fitting line.

Finally, we observed that for σ^{70+38} genes without input TFs (previously used to validate the surface fitted to σ^{70+38} genes with TF inputs), m_{38} and m_{70} are similar (49.04 and 52.48, respectively), unlike for genes with input TFs (Table 1). Thus, for this restricted set of genes, we defined $\Delta = D_{\sigma 38} - D_{\sigma 70}$ (Section 2.4.3) and plotted again the respective LFC_{RNA} values. From Fig. 5C, they correlate linearly.

We also searched for linear correlations between Δ and LFC_{RNA} in genes whose promoters have preference for σ^{70}, σ^{38}, σ^{24}, σ^{28}, σ^{32}, or σ^{54} (Supplementary Fig. S8A–F). No significant correlation with high R^2 was found, including in the two small cohorts of genes whose promoters have preference for two σ factors, other than σ^{70+38} (Supplementary Fig. S8G–H). Finally, we tried to fit the complete model (based on $D_{\sigma 38}$ and $D_{\sigma 70}$ separately) to these cohorts, but it was unable to fit well their behavior (Supplementary Table S6).

3.7. Expanding the model to the growth phase transition period

The model above was designed to be applied in either growth phase. Here, we expanded it to be applicable to the transition period between the growth phases, assuming σ^{38} to be the main regulatory molecule. We collected temporal data on protein numbers of three σ^{70+38} genes, specifically pstS, aidB and asr, selected to represent σ^{70+38} genes with strong, mild, and weak response strengths to the growth phase transition, respectively (Fig. 6A).

Their empirical data was plotted in Fig. 6A. Then, we applied Hill functions (given the data on σ^{38} from Fig. 3B), which fitted well the data. Moreover, we found a linear relationship between the fold changes in protein levels of the 3 genes, and the σ^{38} levels over time (Fig. 6B).

The dependency on σ^{38} levels in model is set by ρ_{sta} which is a function of the time-dependent σ^{38} levels (Eq. (7b)). Thus, given the goodness of fit of the Hill functions, we set the following time-dependent model:

$$FC_P(t) = \frac{\left(FC_P^{\max} - 1\right)\cdot t^s}{h^s + t^s} + b \quad (15a)$$

where:

$$LFC_P^{\max} = \left(log_2\left[\Gamma\cdot\left(1-\rho_{sta}^{\max}\right) + \Gamma\cdot\rho_{sta}^{\max}\cdot K_0\cdot e^{-(m_{38}\cdot D_{\sigma 38} - m_{70}\cdot D_{\sigma 70})}\right]\right)\cdot\alpha - \beta \quad (15b)$$

$$FC_P^{\max} = \frac{\left[\Gamma\cdot\left(1-\rho_{sta}^{\max}\right) + \Gamma\cdot\rho_{sta}^{\max}\cdot K_0\cdot e^{-(m_{38}\cdot D_{\sigma 38} - m_{70}\cdot D_{\sigma 70})}\right]^{\alpha}}{2^{\beta}} \quad (15c)$$

Here, ${\rho_{sta}}^{\max}$ is the final (thus, maximum) concentration of RNAP.σ^{38} relative to the total concentration of bound RNAPs. Meanwhile, ${FC_P}^{\max}$ is the expected fold change in protein numbers of a σ^{70+38} gene, which is reached after the transition to stationary growth is completed. Finally, b (the intercept) is the expression level controlled by the promoter of interest, when in the exponential growth phase. Meanwhile, s (the slope) is the response strength, and h (the half-activation coefficient) is a measure of the response timing to changes in σ^{38} levels.

3.8. Model generalization

Since we found a correlation between the promoter sequences of σ^{70+38} genes and their response to σ^{38}, it may be that genes whose promoters have different σ preferences will exhibit, in some cases, similar sequence-dependent behaviors during the stresses that they respond to. Thus, we generalized the model to be applicable to any stress and responsive gene cohort.

As a general example, we set a model for genes responsive to all seven σ factors of *E. coli*, by expanding Reactions (R1a) and (R1b) to 7 reactions as follows:

$$\text{RNAP} + \sigma^i \underset{}{\overset{K^{\sigma i}}{\longleftrightarrow}} \text{RNAP}.\sigma^i, \text{ where } i = 70, 54, 38, 32, 28, 24 \text{ or } 19 \quad \text{(R8a-R8g)}$$

Given this, we generalize Eq. (1) to account for all the σ factors in holoenzyme form:

$$[RNAP_{total}] = \sum_i \left[RNAP.\sigma^i\right], \text{where } i = 70, 54, 38, 32, 28, 24 \text{ or } 19 \quad (16)$$

Finally, we generalize Eq. R4) as follows:

$$\text{Pro} \xrightarrow{\sum_i \left[RNAP.\sigma^i\right]\times k_t^{\sigma i}(D_{\sigma i})} \text{Pro} + \text{RNA}, \text{where } i = 70, 54, 38, 32, 28, 24 \text{ or } 19 \quad \text{(R9)}$$

This general model can be tuned based on the numbers of each σ factor present in the conditions considered, and the consensus sequences to each σ factor (Supporting Table SII). In addition, following the findings in Section 3.7, it should be feasible to introduce factors to account for the timing of the changes in the transition period.

Table 1
Goodness of fit, measured by the R^2 of the surface fitting of FC_{RNA} as a function of $D_{\sigma 38}$ and $D_{\sigma 70}$. Shown are the model and best fitting parameter values, where $K_0 = \frac{k_{t0}^{\sigma 38}}{k_{t0}^{\sigma 70}}$. The surface was fitted to σ^{70+38} genes without input TFs and validated on genes σ^{70+38} genes with input TFs.

Surface equation	$FC_{RNA} = \Gamma\cdot(1-\rho_{sta}) + \Gamma\cdot\rho_{sta}\cdot K_0\cdot e^{-(m_{38}\cdot D_{\sigma 38} - m_{70}\cdot D_{\sigma 70})}$ (12)	
Coefficients	m_{38}	30.78
	m_{70}	56.90
	Γ	28.20
	K_0	3.841×10^{-9}
	ρ_{sta}	0.06 (Eq. (10))
R^2	σ^{70+38} genes without input TFs	0.98

4. Discussion

Past studies have identified differences between σ^{70+38} and σ^{38} genes (e.g., Regulon DB sets this classification from ChIP-seq and other data [32]). We additionally found a small difference in response kinetics to

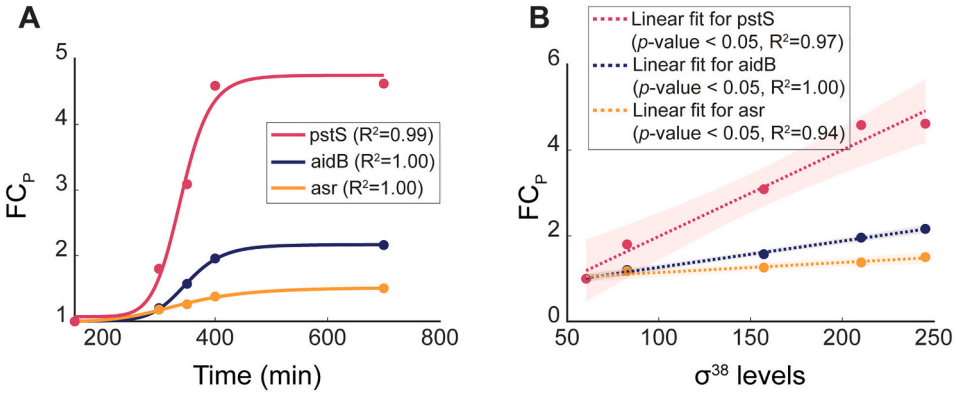

Fig. 6. Temporal changes in the fold-change of protein levels of σ^{70+38} genes as σ^{38} changes. (a) Protein levels of three σ^{70+38} genes prior to, during, and after the transition from the exponential to the stationary growth phase. The balls are the empirical mean fold changes (FC_P) in protein expression levels relative to the first-time moment. The lines are the corresponding best fitting Hill functions (parameter values in Supplementary Table S5). (b) Scatter plot of FC_P against the corresponding σ^{38} levels (data from Fig. 3B) over time. The shadows are the 95% confidence bounds.

the growth phase shift (Fig. 2D). Also, their sequence logos differ, particularly their consensus level in the ~ −35 region (Fig. 1 and Supplementary Fig. S1A). Finally, their ontologies differ with σ^{70+38} genes being more involved in respiration (Supplementary Table S4A), while σ^{38} genes are more involved in metabolic processes (Supplementary Table S4B).

From their sequence and dynamics, we proposed and validated a promoter sequence-dependent kinetic model of genes controlled by promoters responsive to both σ^{70} and σ^{38}. The model, which accurately predicts how the dynamics changes from exponential and stationary growth, is an expansion of a past model of promoters with preference for σ^{70} alone [26]. However, it has two competing reactions for transcription by RNAP when bound to σ^{70} and when bound to σ^{38}, respectively. In addition, these two reactions' rate constants are sequence-dependent, in line with the hypothesis that the consensus sequence of promoters with preference for a σ factor should provide the strongest affinity to that σ factor (in general). Further, the model is applicable during growth phase transition, based on the σ^{38} level at any given moment.

The identification of a correlation between the p-distances of the promoter sequence of σ^{70+38} genes and their response strength to the growth phase transition is, to our knowledge, a unique feature. So far, sequence dependent behaviors have not been reported for any cohort of natural promoters of *E. coli*, even under stable growth conditions. Similar relationships have only been reported for synthetic libraries of variants of single promoters (thus, sequence-restricted) [5,6,30,50].

In the future, if the model is to become a tool for engineering promoters' libraries with desired responsiveness to σ^{38} concentrations, it is imperative to test the model's predictive capacity of a promoter's responsiveness based on its sequence (i.e., from its $D_{\sigma 38}$ and $D_{\sigma 70}$). If proven successful, future synthetic promoters built in this manner could become a common component of synthetic circuits.

At the moment, it is contentious [51] whether σ factor recognition of a promoter is best modeled by a 'discrete' [27,28,52] or by a 'near-continuous' function [4]. RegulonDB reports only which σ factor (in a few cases which two σ factors) best recognize each promoter sequence. This alone does not allow implementing a near-continuous model of multiple σ factors preference for each promoter, as it would require rate constants for the recognition by each (or most) σ factor(s) of the promoter. Thus, our model is 'discrete' (reaction R4 only allows transcription by σ^{70} and σ^{38}).

Nevertheless, a near-continuous, sequence-dependent promoter recognition model might be more realistic, particularly at the genome-wide level, since σ^{70} and σ^{38} are structurally similar [53] and so are their recognition sites [4]. Also, promoters with σ^{38} dependency might not have exclusive recognition sites. Rather, their sites may just be more tolerant to deviations than the sites of σ^{70} [54,55]. Implementing a continuous model could be feasible given more empirical data, such as RNA-seq data from cells subjected to the stresses that each σ factor is responsive to. Our model could be adapted to near-continuous sequence-dependent preferences for σ factors, by adding terms in the rate constant of reaction R4, with each term representing the recognition by each σ factor, respectively, as we did for σ^{70} and σ^{38}.

One simplification of the model is the assumption that there is only one consensus sequence for σ^{70} dependent promoters and another for σ^{38} dependent promoters respectively, with the deviation from them being the only factor defining the recognition level. However, future models should consider that, first, the sequence dependence may not be monotonic. Also, other factors may interfere, such as the AT-richness between the −10 position and the TSS [53], the nucleotide length of the promoter spacers [53], the presence of a cytosine at the −13 position of the promoter [53], the sequence upstream of the promoter [56], UP elements [57], and other sequence motifs [58].

Since it remains challenging to predict if a sequence can act as a promoter and, if so, with which strength and under what regulatory mechanisms [5], the observed predictability of the dynamics from the sequence for the natural cohort of σ^{70+38} genes is of interest for three main reasons. First, these promoters and their variants (with varying p-distances from the consensus sequences of σ^{70} and σ^{38} genes) could become key components of future circuits whose predictable kinetics. Further, the circuits would likely be functional in both the exponential and stationary growth phase, with tunable responsiveness to the phase transition (given the results in Fig. 6). Finally, this approach could become a starting point for broader models of natural promoters with sequence-predictable adaptability to stresses.

However, it may prove difficult to find similar relationships between the sequences of other natural promoters and their responsiveness to other σ factors (or global regulators). E.g., we failed to find a relationship between the sequences of promoters responsive to σ^{38} alone and their response strength to σ^{38} (even for genes without TF inputs (Supplementary Fig. S8B)). This suggests that we are failing to consider some of the variables that influence the preference for this σ factor.

It may be possible to expand this model in various ways. First, it may be applicable to genes responsive to σ^{70}, σ^{54}, σ^{32}, σ^{28}, σ^{24}, or dual combinations, when the respective σ factors change numbers. Also, it should be possible to consider the influence (interference or enhancement) of TFs in the genes' responsiveness to the σ factor.

While it is expected that the model should be applicable to various bacteria, it may be further applicable to eukaryotes. Namely, it is plausible that similar mechanisms of dual global regulation exist in some eukaryotic promoters that depend on TFIID complexes of the RNA polymerase II preinitiation complex, which are highly conserved [59–61]). The transcription of these promoters only initiates after a TFIID factor binds to their TATA box, a DNA sequence specifying a start site of transcription. Such TFIID factors, in addition to a TBP (TATA binding protein), also require one TAF (TBP associated factor). Since eukaryotic cells have several different TAFs, this mechanism has similarities to σ factor regulation [62]. In the future, it should be interesting to explore our model usability for eukaryotic promoters responsive to multiple TAF

[63].

From the biological point of view, first, it should be interesting to investigate if the sequence-dependent responsiveness of σ^{70+38} genes to σ^{38} levels could explain why their promoters (from positions −41 to −1) are highly conserved (Fig. 2). Another potential reason for why they are conserved could also be that the TFs that they code for commonly serve as input TFs to essential genes [64] (2.5 times more than by chance, Fisher test p-value <0.05).

Finally, over-representation tests of the ontology [65,66] of these genes suggest that they are commonly involved in respiration (Supplementary Table S4A). In agreement, respiration is highly affected when changing from exponential to stationary growth [23,67], since aerobic respiration is reduced, while fermentation and anaerobic respiration are enhanced [18].

Moreover, of the genes associated with these biological functions, σ^{70+38} genes are among the most responsive to the growth phase transition (Supplementary Fig. S4). This suggests that they may control the most altered processes during the adaptation. Therefore, externally regulating σ^{70+38} genes may give significant control over these processes. This is particularly appealing since the control could be exerted by promoter sequence modifications (i.e., tuning p-distances) with largely predictable effects. This strategy could contribute to the engineering of synthetic circuits with tailored responses to growth phase transitions.

Data availability

RNA-seq *.fastq data and processed RNA-seq data are deposited in NCBI GEO with accession code GSE188752 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE188752). A data package was deposited in Dryad containing flow cytometry, microscopy and spectrophotometry data. The data is accessible through this link: https://doi.org/10.5061/dryad.jsxksn0b7.

Author statement

A.S.R. and V.K. conceived the study and A.S.R. directed it. I.S.C.B. planned and executed the data analysis, to which M.N.M.B., B.L.B.A., and C.S.D.P. contributed to. A.S.R., I.S.C.B. and V.K. interpreted and integrated the data. V.K., V.C., and S.D. performed measurements. A.S.R. and I.S.C.B. drafted the manuscript and supplementary. I.S.C.B. assembled the figures. All was revised by all co-authors. I.S.C.B. and V.K. contributed equally and have the right to list their name first.

Funding

This work was supported by the Jane and Aatos Erkko Foundation [10-10524-38 to A.S.R]; Finnish Cultural Foundation [00200193 and 00212591 to I.S.C.B., and 50201300 to S.D.]; Pirkanmaa Regional Fund of the Finnish Cultural Foundation to V.K.; Suomalainen Tiedeakatemia to C.S.D.P.; Tampere University Graduate Program to V.C., M.N.M.B., and B.L.B.A.; EDUFI Fellowship [TM-19-11105 to S.D]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the Tampere facility of Flow Cytometry for their service.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagrm.2022.194812.

References

[1] J. Chen, C. Chiu, S. Gopalkrishnan, A.Y. Chen, P.D.B. Olinares, R.M. Saecker, J. T. Winkelman, M.F. Maloney, B.T. Chait, W. Ross, R.L. Gourse, E.A. Campbell, S. A. Darst, Stepwise promoter melting by bacterial RNA polymerase, Mol. Cell 78 (2020) 275–288.e6, https://doi.org/10.1016/j.molcel.2020.02.017.

[2] D.F. Browning, S.J.W. Busby, Local and global regulation of transcription initiation in bacteria, Nat. Rev. Microbiol. 14 (2016) 638–650, https://doi.org/10.1038/nrmicro.2016.103.

[3] D.F. Browning, S.J.W. Busby, The regulation of bacterial transcription initiation, Nat. Rev. Microbiol. 2 (2004) 57–65, https://doi.org/10.1038/nrmicro787.

[4] B.K. Cho, D. Kim, E.M. Knight, K. Zengler, B.O. Palsson, Genome-scale reconstruction of the sigma factor network in Escherichia coli: topology and functional states, BMC Biol. 12 (2014) 1–11, https://doi.org/10.1186/1741-7007-12-4.

[5] G. Urtecho, A.D. Tripp, K.D. Insigne, H. Kim, S. Kosuri, Systematic dissection of sequence elements controlling σ70 promoters using a genomically encoded multiplexed reporter assay in Escherichia coli, Biochemistry 58 (2019) 1539–1551, https://doi.org/10.1021/acs.biochem.7b01069.

[6] T. Einav, R. Phillips, How the avidity of polymerase binding to the –35/–10 promoter sites affects gene expression, Proc. Natl. Acad. Sci. 116 (2019) 13340 LP–13345, https://doi.org/10.1073/pnas.1905615116.

[7] D.L. Jones, R.C. Brewster, R. Phillips, Promoter architecture dictates cell-to-cell variability in gene expression, Science 346 (2014) 1533–1536, https://doi.org/10.1126/science.1255301.

[8] D.M. Stoebel, K. Hokamp, M.S. Last, C.J. Dorman, Compensatory evolution of gene regulation in response to stress by Escherichia coli lacking RpoS, PLoS Genet. 5 (2009), e1000671, https://doi.org/10.1371/journal.pgen.1000671.

[9] G. Kannan, J.C. Wilks, D.M. Fitzgerald, B.D. Jones, S.S. Bondurant, J. L. Slonczewski, Rapid acid treatment of Escherichia coli: transcriptomic response and recovery, BMC Microbiol. 8 (2008) 37, https://doi.org/10.1186/1471-2180-8-37.

[10] A. Farewell, K. Kvint, T. Nyström, Negative regulation by RpoS: a case of sigma factor competition, Mol. Microbiol. 29 (1998) 1039–1051, https://doi.org/10.1046/j.1365-2958.1998.00990.x.

[11] P.E. Rouvière, A. De Las Peñas, J. Mecsas, C.Z. Lu, K.E. Rudd, C.A. Gross, rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in Escherichia coli, EMBO J. 14 (1995) 1032–1042, https://doi.org/10.1002/j.1460-2075.1995.tb07084.x.

[12] T. Dong, H.E. Schellhorn, Global effect of RpoS on gene expression in pathogenic Escherichia coli O157:H7 strain EDL933, BMC Genomics 10 (2009) 349, https://doi.org/10.1186/1471-2164-10-349.

[13] M. Rahman, M.R. Hasan, T. Oba, K. Shimizu, Effect of rpoS gene knockout on the metabolism of Escherichia coli during exponential growth phase and early stationary phase based on gene expressions, enzyme activities and intracellular metabolite concentrations, Biotechnol. Bioeng. 94 (2006) 585–595, https://doi.org/10.1002/bit.20858.

[14] H. Maeda, N. Fujita, A. Ishihama, Competition among seven Escherichia coli σ subunits: relative binding affinities to the core RNA polymerase, Nucleic Acids Res. 28 (2000) 3497–3503, https://doi.org/10.1093/nar/28.18.3497.

[15] A. Feklístov, B.D. Sharon, S.A. Darst, C.A. Gross, Bacterial sigma factors: a historical, structural, and genomic perspective, Annu. Rev. Microbiol. 68 (2014) 357–376, https://doi.org/10.1146/annurev-micro-092412-155737.

[16] J.D. Helmann, M.J. Chamberlin, Structure and function of bacterial sigma factors, Annu. Rev. Biochem. 57 (1988) 839–872, https://doi.org/10.1146/annurev.bi.57.070188.004203.

[17] R. Hengge, Proteolysis of σS (RpoS) and the general stress response in Escherichia coli, Res. Microbiol. 160 (2009) 667–676, https://doi.org/10.1016/j.resmic.2009.08.014.

[18] H. Weber, T. Polen, J. Heuveling, V.F. Wendisch, R. Hengge, Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity, J. Bacteriol. 187 (2005) 1591–1603, https://doi.org/10.1128/JB.187.5.1591-1603.2005.

[19] O. Patange, C. Schwall, M. Jones, C. Villava, D.A. Griffith, A. Phillips, J.C.W. Locke, Escherichia coli can survive stress by noisy growth modulation, Nat. Commun. 9 (2018), https://doi.org/10.1038/s41467-018-07702-z.

[20] M. Jishage, A. Iwata, S. Ueda, A. Ishihama, Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of four species of sigma subunit under various growth conditions, J. Bacteriol. 178 (1996) 5447–5451, https://doi.org/10.1128/jb.178.18.5447-5451.1996.

[21] R. Lange, R. Hengge-Aronis, The cellular concentration of the σ(S) subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability, Genes Dev. 8 (1994) 1600–1612, https://doi.org/10.1101/gad.8.13.1600.

[22] T.M. Gruber, C.A. Gross, Multiple sigma subunits and the partitioning of bacterial transcription space, Annu. Rev. Microbiol. 57 (2003) 441–466, https://doi.org/10.1146/annurev.micro.57.030502.090913.

[23] T.E. Riedel, W.M. Berelson, K.H. Nealson, S.E. Finkel, Oxygen consumption rates of bacteria under nutrient-limited conditions, Appl. Environ. Microbiol. 79 (2013) 4921–4931, https://doi.org/10.1128/AEM.00756-13.

[24] D.C. Abantika Ganguly, A comparative kinetic and thermodynamic perspective of the σ-competition model in Escherichia coli, Biophys. J. 103 (2012) 1325, https://doi.org/10.1016/j.bpj.2012.08.013.

[25] P.P. Dennis, H. Bremer, Modulation of chemical composition and other parameters of the cell at different exponential growth rates, EcoSal Plus 3 (2008), https://doi.org/10.1128/ecosal.5.2.3.

[26] V.K. Kandavalli, H. Tran, A.S. Ribeiro, Effects of σ factor competition are promoter initiation kinetics dependent, Biochim. Biophys. Acta 2016 (1859) 1281–1288, https://doi.org/10.1016/j.bbagrm.2016.07.011.

[27] I.L. Grigorova, N.J. Phleger, V.K. Mutalik, C.A. Gross, Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 5332–5337, https://doi.org/10.1073/pnas.0600828103.

[28] M. Mauri, S. Klumpp, A model for sigma factor competition in bacterial cells, PLoS Comput. Biol. 10 (2014), e1003845, https://doi.org/10.1371/journal.pcbi.1003845.

[29] J.B. Kinney, A. Murugan, C.G. Callan, E.C. Cox, Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 9158–9163, https://doi.org/10.1073/pnas.1004290107.

[30] R.C. Brewster, D.L. Jones, R. Phillips, Tuning promoter strength through RNA polymerase binding site design in Escherichia coli, PLoS Comput. Biol. 8 (2012), e1002811, https://doi.org/10.1371/journal.pcbi.1002811.

[31] K. Tanaka, Y. Takayanagi, N. Fujita, A. Ishihama, H. Takahashi, Heterogeneity of the principal sigma factor in Escherichia coli: the rpoS gene product, sigma 38, is a second principal sigma factor of RNA polymerase in stationary-phase Escherichia coli, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 3511–3515, https://doi.org/10.1073/pnas.90.8.3511.

[32] A. Santos-Zavaleta, H. Salgado, S. Gama-Castro, M. Sánchez-Pérez, L. Gómez-Romero, D. Ledezma-Tejeida, J.S. García-Sotelo, K. Alquicira-Hernández, L. J. Muñiz-Rascado, P. Peña-Loredo, C. Ishida-Gutiérrez, D.A. Velázquez-Ramírez, V. Del Moral-Chávez, C. Bonavides-Martínez, C.-F. Méndez-Cruz, J. Galagan, J. Collado-Vides, 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 (2018) D212–D220, https://doi.org/10.1093/nar/gky1077.

[33] Y. Taniguchi, P.J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, X.S. Xie, Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells, Science 329 (2010) 533–538, https://doi.org/10.1126/science.1188308.

[34] A. Zaslaver, A. Bren, M. Ronen, S. Itzkovitz, I. Kikoin, S. Shavit, W. Liebermeister, M.G. Surette, U. Alon, A comprehensive library of fluorescent transcriptional reporters for Escherichia coli, Nat. Methods 3 (2006) 623–628, https://doi.org/10.1038/nmeth895.

[35] B.P. Bratton, R.A. Mooney, J.C. Weisshaar, Spatial distribution and diffusive motion of RNA polymerase in live Escherichia coli, J. Bacteriol. 193 (2011) 5138–5146, https://doi.org/10.1128/JB.00198-11.

[36] J. Izard, C.D.C. Gomez Balderas, D. Ropers, S. Lacour, X. Song, Y. Yang, A. B. Lindner, J. Geiselmann, H. Jong, A synthetic growth switch based on controlled expression of RNA polymerase, Mol. Syst. Biol. 11 (2015) 840, https://doi.org/10.15252/msb.20156382.

[37] M.J. Chamberlin, The selectivity of transcription, Annu. Rev. Biochem. 43 (1974) 721–775, https://doi.org/10.1146/annurev.bi.43.070174.003445.

[38] G. Crooks, G. Hon, J. Chandonia, S. Brenner, NCBI GenBank FTP Site\nWebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190, https://doi.org/10.1101/gr.849004.1.

[39] T.D. Schneider, R.M. Stephens, Sequence logos: a new way to display consensus sequences, Nucleic Acids Res. 18 (1990) 6097–6100, https://doi.org/10.1093/nar/18.20.6097.

[40] M. Nei, J. Zhang, Evolutionary distance: estimation, in: Encyclopedia of Life Sciences, 2006, pp. 1–4, https://doi.org/10.1038/npg.els.0005108.

[41] T. Olsen, Hill function, MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/68935-hill-function, 2018. (Accessed 27 October 2021).

[42] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence Volume 2, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995, pp. 1137–1143, https://dl.acm.org/doi/10.5555/1643031.1643047 (accessed March 2, 2022).

[43] J. Hausser, A. Mayo, L. Keren, U. Alon, Central dogma rates and the trade-off between precision and economy in gene expression, Nat. Commun. 10 (2019) 1–15, https://doi.org/10.1038/s41467-018-07391-8.

[44] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, N. Barkai, Noise in protein expression scales with natural protein abundance, Nat. Genet. 38 (2006) 636–643, https://doi.org/10.1038/ng1807.

[45] J.R.S. Newman, S. Ghaemmaghami, J. Ihmels, D.K. Breslow, M. Noble, J.L. DeRisi, J.S. Weissman, Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise, Nature 441 (2006) 840–846, https://doi.org/10.1038/nature04785.

[46] A. Ishihama, Functional modulation of Escherichia coli RNA polymerase, Annu. Rev. Microbiol. 54 (2000) 499–518, https://doi.org/10.1146/annurev.micro.54.1.499.

[47] M. Jishage, A. Ishihama, Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of sigma 70 and sigma 38, J. Bacteriol. 177 (1995) 6832–6835, https://doi.org/10.1128/jb.177.23.6832-6835.1995.

[48] S.E. Piper, J.E. Mitchell, D.J. Lee, S.J.W. Busby, A global view of Escherichia coli Rsd protein and its interactions, Mol. BioSyst. 5 (2009) 1943–1947, https://doi.org/10.1039/B904955J.

[49] F. Colland, N. Fujita, A. Ishihama, A. Kolb, The interaction between sigmaS, the stationary phase sigma factor, and the core enzyme of Escherichia coli RNA polymerase, Genes Cells 7 (2002) 233–247, https://doi.org/10.1046/j.1365-2443.2002.00517.x.

[50] M. Liu, M. Tolstorukov, V. Zhurkin, S. Garges, S. Adhya, 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 (2004) 6911–6916, http://www.jstor.org/stable/3372027, http://www.jstor.org/stable/3372027.

[51] R. Hengge-Aronis, Recent insights into the general stress response regulatory network in Escherichia coli, J. Mol. Microbiol. Biotechnol. 4 (2002) 341–346, https://www.ncbi.nlm.nih.gov/pubmed/11931567, https://www.ncbi.nlm.nih.gov/pubmed/11931567.

[52] C.A. Gross, C. Chan, A. Dombroski, T. Gruber, M. Sharp, J. Tupy, B. Young, The functional and regulatory roles of sigma factors in transcription, Cold Spring Harb. Symp. Quant. Biol. 63 (1998) 141–155, https://doi.org/10.1101/sqb.1998.63.141.

[53] R. Hengge-Aronis, Stationary phase gene regulation: what makes an Escherichia coli promoter sigmaS-selective? Curr. Opin. Microbiol. 5 (2002) 591–595, https://doi.org/10.1016/s1369-5274(02)00372-7.

[54] A. Typas, R. Hengge, Role of the spacer between the -35 and -10 regions in sigmas promoter selectivity in Escherichia coli, Mol. Microbiol. 59 (2006) 1037–1051, https://doi.org/10.1111/j.1365-2958.2005.04998.x.

[55] A. Typas, G. Becker, R. Hengge, The molecular basis of selective promoter activation by the σs subunit of RNA polymerase, Mol. Microbiol. 63 (2007) 1296–1306, https://doi.org/10.1111/j.1365-2958.2007.05601.x.

[56] H.-J. Lim, K. Kim, M. Shin, J.-H. Jeong, P.Y. Ryu, H.E. Choy, Effect of promoter-upstream sequence on σ38-dependent stationary phase gene transcription, J. Microbiol. 53 (2015) 250–255, https://doi.org/10.1007/s12275-015-4681-8.

[57] K.S. Franco, Z. Sun, Y. Chen, C. Cagliero, Y. Zuo, Y.N. Zhou, M. Kashlev, D.J. Jin, T. D. Schneider, Escherichia coli σ38 promoters use two UP elements instead of a −35 element: resolution of a paradox and discovery that σ38 transcribes ribosomal promotersBioRxiv, 2020, https://doi.org/10.1101/2020.02.05.936344, 2020.02.05.936344.

[58] S.J. Lee, J.D. Gralla, Osmo-regulation of bacterial transcription via poised RNA polymerase, Mol. Cell 14 (2004) 153–162, https://doi.org/10.1016/s1097-2765(04)00202-3.

[59] S.R. Bhaumik, Distinct regulatory mechanisms of eukaryotic transcriptional activation by SAGA and TFIID, Biochim. Biophys. Acta 2011 (1809) 97–108, https://doi.org/10.1016/j.bbagrm.2010.08.009.

[60] M.R. Green, TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes, Trends Biochem. Sci. 25 (2000) 59–63, https://doi.org/10.1016/s0968-0004(99)01527-3.

[61] L. Tora, A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription, Genes Dev. 16 (2002) 673–675, https://doi.org/10.1101/gad.976402.

[62] J.A. Jaehning, Sigma factor relatives in eukaryotes, Science 253 (1991) 859, https://doi.org/10.1126/science.1876846.

[63] V. Haberle, A. Stark, Eukaryotic core promoters and the functional basis of transcription initiation, Nat. Rev. Mol. Cell Biol. 19 (2018) 621–637, https://doi.org/10.1038/s41580-018-0028-8.

[64] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K.A. Datsenko, M. Tomita, B.L. Wanner, H. Mori, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection, Mol. Syst. Biol. 2 (2006), https://doi.org/10.1038/msb4100050, 2006.0008-2006.0008.

[65] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (2000) 25–29, https://doi.org/10.1038/75556.

[66] Gene Ontology Consortium, The Gene Ontology resource: enriching a GOld mine, Nucleic Acids Res. 49 (2021) D325–D334, https://doi.org/10.1093/nar/gkaa1113.

[67] K.B. Andersen, K. von Meyenburg, Are growth rates of Escherichia coli in batch cultures limited by respiration? J. Bacteriol. 144 (1980) 114–123, https://doi.org/10.1128/JB.144.1.114-123.1980.