Ohm's Law for emergent gene expression under fitness pressure

Marta Ciechonska[#1], Marc Sturrock[#2], Alice Grob[1], Gerald Larrouy-Maumus[1], Vahid Shahrezaei*[3] and Mark Isalan*[1]

[1]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK
[2]Department of Physiology, Royal College of Surgeons in Ireland, Dublin, Ireland
[3]Department of Mathematics, Imperial College London, London, United Kingdom

*To whom correspondence should be addressed:
Mark Isalan (m.isalan@imperial.ac.uk) and Vahid Shahrezaei (v.shahrezaei@imperial.ac.uk),
Imperial College London, London SW7 2AZ, UK
[#]Contributed equally

Abstract (192/200 words max)

Bacteria likely encounter more environmental conditions and stresses than specific hard-wired gene regulation networks could ever evolve to control. One potential solution is that fitness pressures may lead to phenotypic natural selection upon variations in gene expression, as long as the genes are essential for cell survival and growth[1-11]. Such mechanisms of emergent gene expression or "*gene expression according to need*" remain poorly understood. Here we show that an antibiotic resistance gene displays dose-responsive gene upregulation that is proportional to the applied stress (fitness pressure) supplied by antibiotic exposure. Drawing an analogy with Ohm's law in electricity (V=IR), we propose that fitness pressure acts similarly to voltage (V), output gene expression to current (I), and resistance (R) to the constraints of cellular machinery, thus describing a linear relationship between gene expression and environmental pressure. Mathematical modelling shows that positive feedbacks between growth and gene expression, and stochasticity at the level of both mRNA and protein expression, are necessary and sufficient for this phenomenon. Overall, we find that the efficiency of the survival gene (antibiotic resistance enzymatic activity) and the amount of fitness pressure (antibiotic) tune the magnitude of the apparent emergent regulation.

Main text: (1675/1500 words max)

Fitness-induced stochastic switching[12], occurs when cells apparently switch spontaneously to a fitter state of gene expression, thus eliminating the high cost of evolving and maintaining a specialised sensory apparatus. Examples include switching within a bistable genetic circuit[1], stress response regulator pulsing[13] or upregulation of an essential gene required for metabolism, upon changing environmental conditions[2]. The specific mechanisms and conditions driving such changes in gene expression are poorly understood, but phenotypic selection of a heterogenously expressed gene could result in an apparent regulation at the population level[2,14]. Such emergent gene expression (EGE) might be important for regulating expression of a fitness-inducing gene according to need. Therefore, we applied a combination of synthetic biology, quantitative experiments and mathematical modelling to search for conditions and mechanisms underpinning this phenomenon.

To isolate EGE, we applied a fitness pressure, using antibiotic challenge with the protein synthesis inhibitor chloramphenicol (Cm), and analysed phenotypic adaptation output based on changing expression levels of chloramphenicol acetyl transferase (CAT). We integrated wild type (WT) and mutant versions of *cat-gfp* fusion constructs, driven by either constitutive or inducible promoters, into the *E. coli* genome (Fig. 1a). We expected that as Cm concentration increased, cells expressing higher levels of GFP-CAT would be more fit; consequently, *cat* expression and GFP fluorescence in the entire population would rise (Fig. 1b). Additionally, we reasoned that strains expressing less-active mutant versions of CAT would require higher expression levels than WT, in order to acetylate and neutralize equal amounts of antibiotic. We first challenged populations where cells constitutively express the WT *gfp-cat* cassette or mutant (T172A, ~46% of k_{cat} WT activity[15]) and collected growth

and GFP fluorescence data. When adding more Cm, there was a delay in the dilution of the resistance protein as the cells grew, resulting in higher GFP-CAT per cell during the entire growth phase (Fig. 1c,d). As expected, the effect was far more pronounced in the mutant *cat* strain (Fig. 1d). The resulting Cm-dependent increase in GFP-CAT per cell can be visualised by normalising to the case without antibiotics (Fig. 1e). Antibiotic challenge thus resulted in a delay in returning to steady state dynamics for increasing amounts of Cm, during which time increased gene expression was evident, thus overcoming growth attenuation. For constitutive promoters, continuous expression of CAT results in the acetylation (deactivation) of the antibiotic and leads to reaching stationary phase rapidly. EGE relies on phenotypic selection that requires cell population growth and a sustained fitness pressure, so this accounts for the observed transient effect here (validated in our modelling, discussed below). Importantly, using a weaker mutant such as CAT_{T172A} increases the time during which the phenomenon occurs.

To enhance EGE further, we employed an arabinose-inducible system where *cat* expression was weakly induced only at the time of Cm challenge. As expected, this resulted in slight upregulation of WT *gfp-cat* expression (Fig 1f), but resulted in strong, non-transient, Cm-dependent increases in mutant GFP-CAT_{T172A} production, well above the amount induced by arabinose alone (Fig. 1g). Strikingly, in all cases the increase in fitness gene (CAT) expression was linearly correlated to rising fitness pressure (Cm concentration) (Fig. 1h). As shown later (cf. Fig. 3 and Extended Data, Fig. 5), this linear relationship also holds in the constitutive case that exhibits transient EGE. We find it helpful to think about this observation using an analogy taken from electrical conductivity. In Ohm's law, current (I) is proportional to voltage (V, electrical pressure), with resistance (R) being the opposition to electron flow.

Here, by analogy, the fitness pressure (Cm concentration, "voltage") drives a proportional increase in the CAT-GFP expression ("current") with the metabolic and resource cost of gene expression accounting for the “resistance”, while the slope of the graph (1/R, Fig. 1h) gives the “conductance” of the system, or the propensity of the cell to increase EGE per unit fitness pressure.

We applied a series of controls to test the EGE hypothesis. These included addition of a constitutively-expressed CAT to the GFP-CAT_{T172A}-expressing strain, under the same conditions of Cm treatment, which relieved the fitness pressure and switched off EGE (Extended Data Fig. 1a). Furthermore, GFP expression did not increase upon Cm treatment in a strain encoding a functionally inactive GFP-CAT_{H193A} mutant controlled by pBad, suggesting that EGE selects only those cells which contain increased amounts of a functional fitness-conferring protein, and that Cm does not directly activate the promoter (Extended Data Fig. 1b). We also found the effect is reversible and reproducible over several short rounds of antibiotic challenge and washout, reflecting the inherent flexibility of the mechanism (Extended Data Fig. 2). In addition, RT-qPCR analysis showed that *gfp-cat* transcripts were specifically upregulated by increasing Cm; housekeeping genes, sigma factors, and other antibiotic response genes[16] remained constant (Extended Data Fig. 3). This suggests that phenotypic selection based on Cm-induced EGE is specific to the *gfp-cat* gene.

To understand the mechanisms underlying EGE, we established the minimal requirements necessary to recapitulate EGE *in silico*. We constructed agent-based models that include growth and division of cells that contain stochastic gene expression of a fitness-inducing

gene[17]. We find that a simple model of constitutive stochastic gene expression, with constant transcription and translation, is not able to produce EGE behaviour (in contrast to an earlier study that did not incorporate protein dilution strictly)[9]. As gene expression at the transcription or translation level may be regulated by cellular growth rate[18], we performed a comparison of 10 models of gene expression, exploring a range of assumptions for gene expression and growth regulation. We used Approximate Bayesian Computation (ABC)[19] to perform model selection and identify which models were the best at exhibiting EGE. The simplest model to explain this behaviour included constitutive gene expression, dilution and a global positive feedback coupling translation rate to cell growth (Extended Data Table 1). Positive feedback extends the life-time of protein fluctuations beyond the dilution time set by the cell cycle[20]. We also found that the models explicitly required both transcription and translation processes to produce EGE, as the model ignoring mRNAs does not produce high enough levels of noise for the phenotypic selection to act upon. Intuitively, population-level noise in gene expression ensures existence of cells with a range of fitness, including higher levels for faster division; positive feedback produces a memory effect in the daughter cells, thus making phenotypic selection and EGE possible.

Since global links from both transcription and translation to cell growth are established[4,21,22], we included these in the full model of the inducible expression of CAT protein (Fig. 2a and Supplementary data Fig. S5). A corresponding model was developed for the case of constitutive expression of CAT protein (Extended Data Fig. 4 and Supplementary data Fig S1). In both models, we assumed CAT acetylates the antibiotic Cm, and that Cm impedes cell growth and translation[5]. We used ABC to constrain the parameters of our models to gene expression and growth kinetics data from our microplate experiments, which resulted in

best fits (Fig. 2b and Supplementary Information Video 1). Importantly, without further fitting, the inducible model was able to predict a unimodal shift to the right in the distribution of CAT protein in single cells (Fig. 2c), which we have validated with flow cytometry (Fig. 2d). The model similarly predicts experimentally-observed increasing *gfp-cat* mRNA levels (Fig. 2e) (Extended Data Fig. 3b). Also, the model allowed us to capture the time-course of external antibiotic depletion as observed by mass spectrometry (Fig. 2f).

The constitutive model enabled similar accurate predictions (Extended Data Fig. 5, and Supplementary Information Video 2) and showed that the observed transient upregulation of CAT genes with increasing doses of antibiotics is the result of a combination of delayed dilution and phenotypic selection. EGE is transient in the constitutive case because there is not a sustained phenotypic selection of growing cells; antibiotic is gradually acetylated and growth slows down as cells enter stationary phase. EGE in the inducible pBAD system is also eventually reversible but the model predicts that this occurs on the time-scale at which the pBAD promoter deactivates, therefore we observe a more sustained effect (Supplementary data Fig. S11 and S12). This matches the reversibility seen experimentally when cycling between the presence and absence of Cm (Extended Data Fig. 2). Furthermore, by setting the pBAD promoter activity to a constant, we recover the dynamics of the constitutively expressed model (Supplementary Fig. S8b), showing the two models are compatible. Ultimately, through Bayesian model comparison we show that inclusion of global positive feedbacks in the inducible system is required to reproduce the observed EGE (Extended Data Table 1; Supplementary Information Section 4).

Transient, stochasticity-based, fitness-induced gene expression effects in antibiotic resistance have been suggested in several experimental situations[14,23-25], but tend to be difficult to pin-point and analyse. In this study, we have used lower-activity fitness gene mutants to enhance EGE and observe and quantify dose-responses for the first time. In parallel, using mathematical modelling we revealed that to observe maximal EGE magnitude, one needs to use intermediate fitness gene strength (CAT_{T172A}) and intermediate pressure (Cm) (Fig. 3ab). The interplay between the strength of the fitness-inducing gene and the fitness pressure thus results in islands of maximal emergent gene expression, within fitness parameter landscapes (Fig. 3a).

Our mathematical and experimental results show that EGE is a global phenomenon of gene expression that can be observed with some degree of tuning of the relevant parameters and it is likely to be of relevance for any fitness-conferring genes. It should be noted that even small changes in the expression of highly-active fitness genes (such as enzymes) may be crucially important for cells' biology, even though such EGE is difficult to detect with current experimental methods[26]. Mathematical modelling reveals these relationships and indeed confirms that an intermediate-activity mutant maximises EGE and gives a model output that fits remarkably well with the experimental Ohm's law-like linear framework (Fig. 3b).

Overall, our results indicate that the linear relationship between fitness pressure and gene expression relies on growth feedbacks, stochasticity and fitness gene activity. We propose that evolution could tune the parameters of a cellular system in the operating regime of EGE, providing need-based expression.

Materials and methods

Cell growth

All *E. coli* strains were cultured in lysogeny broth (LB) and, when indicated, 0.005% (w/v) L-(+)-arabinose (Sigma) (stock concentration 5% (w/v) in water) and chloramphenicol (Cm) (Sigma) (stock concentration 50mg/ml in ethanol), which were stored at -20°C and added to the medium at the beginning of each experiment. Chloramphenicol was diluted so that 2µl of an intermediate concentration was added to 148µl of cells in LB per well in 96-well plates. All experiments were inoculated from 5ml overnight cultures grown in LB without antibiotics at a starting Absorbance$_{600}$ (A_{600}) of 0.01 as measured on the Tecan F200 PRO microplate reader. The microplate reader format was used for all experiments; cells were grown at 37°C with orbital shaking at 335.8 RPM with an amplitude of 1.5mm. For the Rounds experiment, 2µl of each culture, grown for 12hrs, was transferred to 148µl of fresh LB media (± arabinose and Cm as indicated) to generate an exact replica of the parent plate with the diluted cultures, which were then growth again under the same conditions for 12hrs.

Cassette and strain construction

The *catI* gene was cloned from pTKIP-cat which was a gift from Edward Cox & Thomas Kuhlman (Addgene plasmid # 41065 ; http://n2t.net/addgene:41065 ; RRID:Addgene_41065). Gibson assembly (NEB) was used for all cloning steps, and all constructs were transformed into Top10 *E. coli* (Invitrogen). The Biobrick part B0034 RBS-GFP-Gly4Ser-CAT construct was inserted downstream of the AraC-pBad cassette in the pCola vector; pCola and GFP were obtained from Schaerli et al. 2014[27]. Lox sites were inserted to flank a kanamycin resistance gene which was then cloned downstream of the GFP-CAT cassette to aid in selection of genomic integrations. For constitutive expression, the AraC gene and pBad promoter were replaced by the Biobrick promoter J23100

upstream of the GFP-CAT-lox-kan-lox cassette in the pCola vector. Site directed mutagenesis was performed to generate the CAT-T_{172}A and H_{193}A mutants within the inducible and constitutive expression cassettes. All expression cassette sequences used in this study are reported in Figure 17 in Supplementary Information. A set of plasmids for the main constructs, along with maps and sequences, were deposited in Addgene (IDs are listed in **Supplementary Table X**).

The expression cassettes were integrated into the *intC* locus of the *E. coli* strain MK01 (genotype: F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, Δ(araH-araF)570(::FRT), ΔaraEp-532::FRT, ϕP_{cp8}-araE535, rph-1, Δ(rhaD-rhaB)568, hsdR514, ΔlacI)[28]. The strain was modified to decrease biofilm formation by knocking out the *flu* and *fim* genes. Briefly, MK01 cells were transformed with the pRed/ET expression plasmid (Gene Bridges kit K006). Transformants were grown up and recombinase expression was induced as described previously[29]. A kanamycin resistance cassette flanked by lox sites was amplified using primers containing sequences homologous to the 5' and 3' regions of the *flu* gene and electroporated into the recombinase-expressing MK01 cells. Recombinants were selected on LB-agar containing 15µg/ml kanamycin (Sigma). Successful integration was confirmed via amplification and sequencing of the *flu* locus. The kanamycin resistance cassette was removed by transforming the cells with *Cre* recombinase (Gene Bridges, 706-Cre) according to the manufacturer's instructions. This sequence was repeated in order to remove the *fim* locus, and subsequently, to introduce the various CAT-GFP expression cassettes into the *intC* locus. Genomic integration primers used in this study are reported in Table 2 in Supplementary Information.

Fluorescence and Absorbance measurements

All experiments where GFP fluorescence and Absorbance$_{600}$ (A_{600}) were measured were performed in 96-well PS, flat bottom, µclear, black plates (Greiner Bio-One), with n=3 technical replicates per treatment, and n=3 biological replicates unless stated otherwise. GFP fluorescence was measured at ex485nm/em535nm and a gain value of 25. Measurements were taken at 15min intervals.

Flow cytometry

MK01 cells carrying the *gfp-cat-*T_{172}*A* genomic integration cassette were induced with 0.005% arabinose and treated with 0 and 5µg/ml Cm, and grown using the microplate reader with GFP and A_{600} measurements performed as above. At 12hr, 4% paraformaldehyde (Sigma) in PBS was added to each well to a final concentration of 2%, and pipetted up and down to mix. Cells were stored at 4°C in the dark for one to three days. Flow cytometry was performed on a BD Fortessa Analyzer (BD Biosciences) and sample data was analysed using FlowJo (v10) Software.

RT-qPCR

MK01 cells carrying the *gfp-cat-*T_{172}*A* genomic integration cassette were induced with 0.005% arabinose and treated with 0 and 5µg/ml Cm, and grown using the microplate reader with GFP and A_{600} measurements performed as above with n=4 technical replicates. Cells were harvested at 1hr time intervals for 11hrs. Briefly, 600µl of culture of both 0 and 5µg/ml Cm treatments was removed from the plate, added to 1200µl of RNAprotect (Qiagen) in 2mL Eppendorf tubes, and processed according to the manufacturer's protocol. All samples were stored at -80°C until the end of the time course. RNA was extracted using

the RNeasy mini kit (Qiagen). The extracted RNA was treated with TURBO DNA-*free* Kit (Invitrogen) and cDNA was generated using the SuperScript IV First-Strand Synthesis System (Invitrogen). The LightCycler 480 SYBR Green I Master kit (Roche) was used as the qPCR master mix, and the experiments were performed on the Roche 480 LightCycler Instrument II. Housekeeping genes used in this study include *idnT*, *hcaT*, and *cysG*[30], and were used to quantify *gfp* mRNA expression at each time point. The delta-delta Ct method was used to determine differences in gene expression, with significance determined using the unpaired Student's t-test. The mean Ct value of the housekeeping genes was also used to normalise Ct values of the control genes *rpoD*, *rpoH*, *rpoE*, *rpoN*, *acrB*, *pntB*, *oppA*, and *cyoC*. Primer sequences used for pPCR amplification are reported in Table 3 in Supplementary Information.

Mass Spectrometry

100µl of culture medium was mixed with 100µL of a solution containing a mixture of acetonitrile, methanol and water (40:40:20, v/v/v). After centrifugation at 17,000 x g, at 4°C, for 10 mins, 100µl of the supernatant was mixed 100µl of a solution of acetonitrile containing 0.2% acetic acid. After vortexing and centrifugation at 17,000 x g, at 4°C, for 10 mins, 100µl of the supernatant was loaded into LC-MS vials prior to analysis.

Aqueous normal phase liquid chromatography was performed using an Agilent 1290 Infinity II LC system equipped with binary pump, temperature-controlled auto-sampler (set at 4°C) and temperature-controlled column compartment (set at 25°C), containing a Cogent Diamond Hydride Type C silica column (150 mm × 2.1 mm; dead volume 315 µl). A flow-rate of 0.4 ml/min was used. Elution of polar metabolites was carried out using solvent A (0.2% acetic acid in deionized water (Resistivity ~ 18 MΩ cm), and solvent B (acetonitrile and 0.2%

acetic acid). Mass spectrometry was carried out using an Agilent Accurate Mass 6545 QTOF apparatus. Nozzle Voltage and fragmentor voltages were set at 2,000 V and 100 V, respectively. The nebulizer pressure was set at 50 psig and the nitrogen drying gas flow rate was set at 5 l/min. The drying gas temperature was maintained at 300°C. Data were collected in the centroid mode in the 4 GHz (extended dynamic range) mode[31] and the values were normalized to the starting concentration measurement at 0hrs.

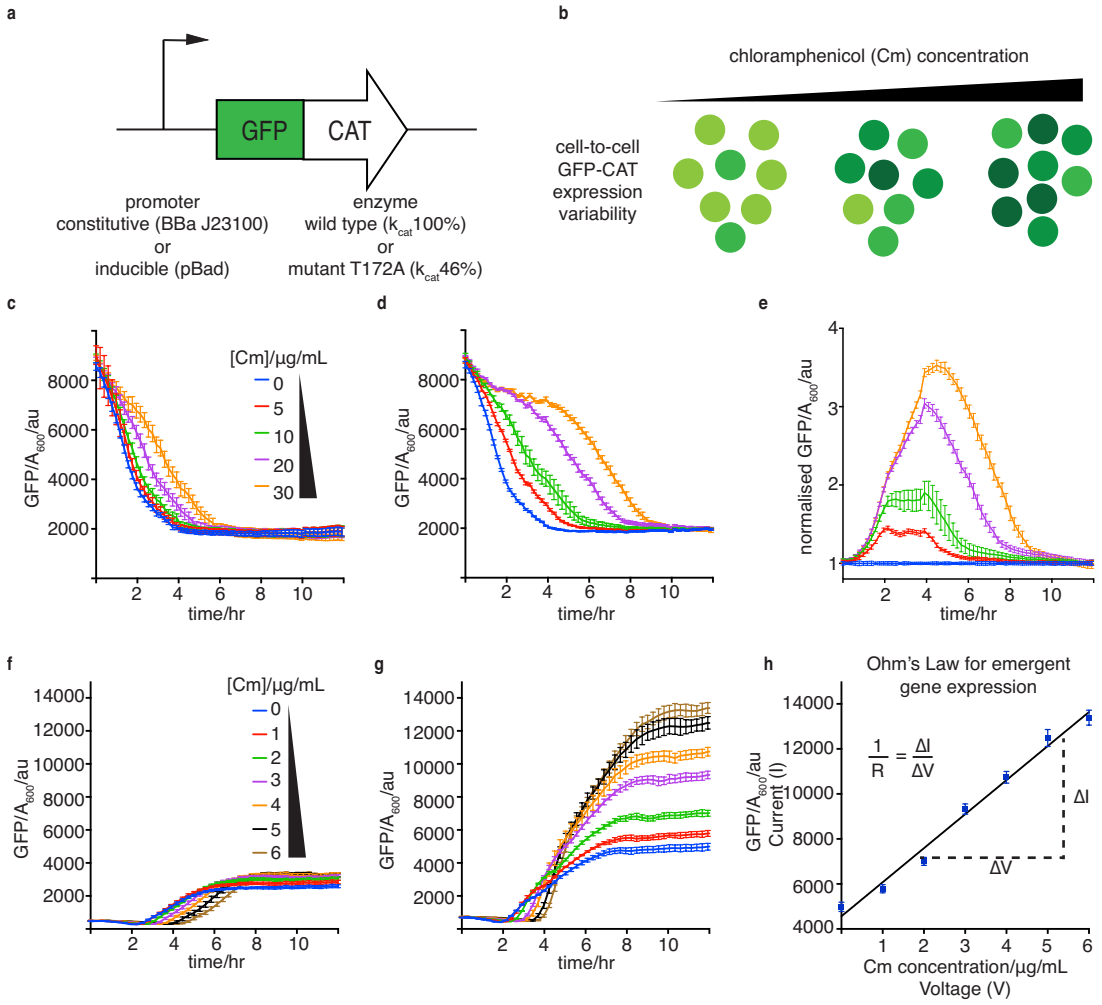


Fig. 1: Emergent gene expression (EGE) increases linearly with rising antibiotic concentration in analogy to Ohm's law.

a, Schematic of GFP-chloramphenicol acetyl transferase (CAT) expression cassettes used in this study with constitutive (BBaJ23100) or inducible (pBad) promoters driving expression of wild type (wt) or mutant $T_{172}A$ CAT (relative acetylation efficiencies of k_{cat}100% and k_{cat} 46% respectively). **b**, Schematic of emergent gene expression for antibiotic resistance. GFP-CAT concentration per cell is indicated by a range of light to dark green; three separate populations are shown where higher GFP-CAT expression is spontaneously selected in the presence of increasing chloramphenicol (Cm) concentration, to increase cell fitness. **c**,**d** Emergent GFP-CAT expression with a constitutive promoter. Graphs show mean (± SDm) GFP expression (GFP/A_{600} per well) in populations constitutively expressing genome-integrated wt GFP-CAT (c) and mutant $T_{172}A$ GFP-CAT (d). These are treated with 0, 5, 10, 20, or 30 µg/mL Cm, to induce EGE, and monitored for 12 hrs; n=3 biological replicates. **e** Constitutively expressed mutant $T_{172}A$ GFP-CAT expression shown in (d) normalised to the control 0 µg/mL Cm condition. **f**,**g** Emergent GFP-CAT expression with a weakly-induced pBAD promoter (0.005% arabinose). Graphs show GFP/A_{600} per well in populations expressing genome-integrated wt GFP-CAT (f) and mutant $T_{172}A$ GFP-CAT (g). These are treated with 0,1,2,3,4,5, or 6 µg/mL Cm to induce EGE, and monitored for 12 hrs; n=3 biological replicates. **h**, Ohm's law analogy for emergent gene expression, where GFP-CAT expression is analogous to current (I), Cm concentration to voltage (V), and the slope of the linear relationship between these, representing cellular propensity to increase EGE, is analogous to conductance (1/R). The data points are the maximal emergent gene expression values in (f).

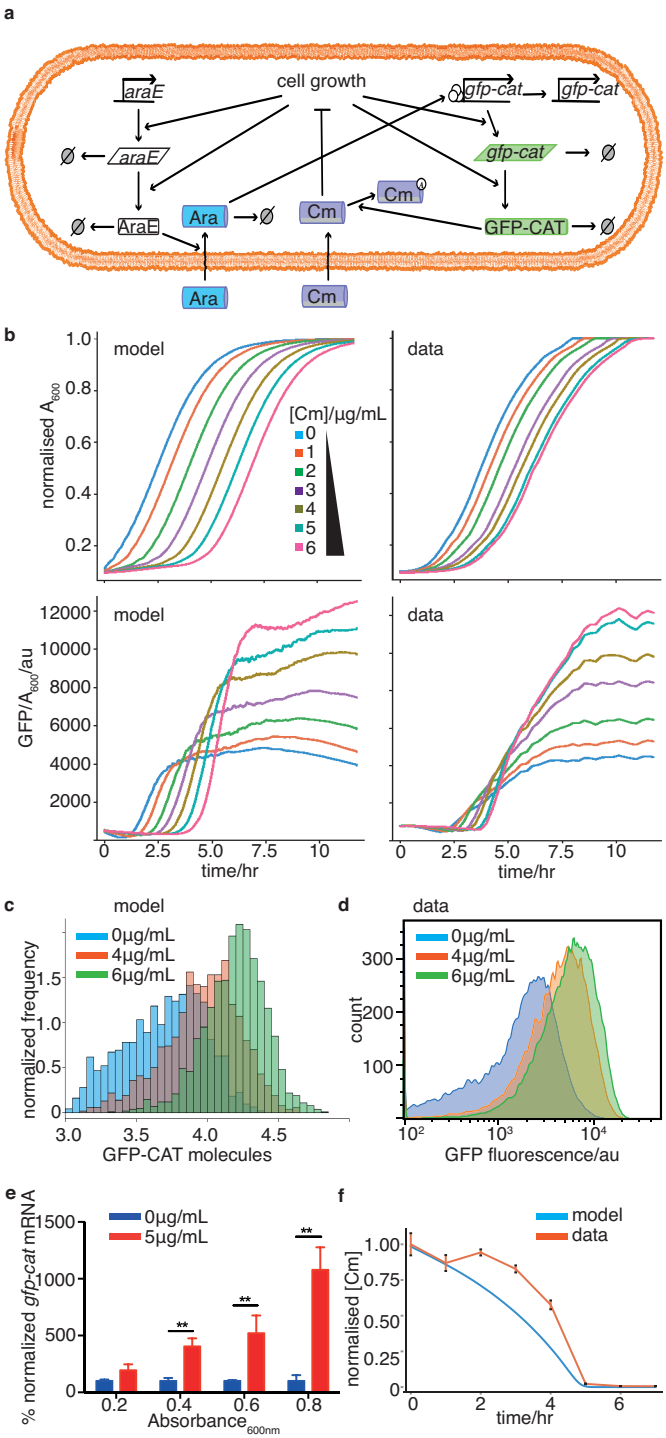


Fig. 2: Experimental verification of the computational emergence model.
a, Schematic for mathematical model of the inducible promoter system for a single cell. The reactions capture GFP-CAT expression and degradation, AraE expression and degradation, arabinose import, degradation and activation of the pBAD promoter, as well as chloramphenicol import and acetylation by GFP-CAT. Intracellular dynamics are coupled to a second model scale which captures cell division, partitioning and logistic growth of the cell population (more details in Supplementary Material Section 3). **b,** Comparing model outputs with experimental data for growth (normalised A_{600} data) and GFP-CAT molecules per cell (GFP/A_{600}), for increasing Cm concentrations, over 12 hrs. Best fit parameters are used from ABC parameter inference and initial conditions (initial number of cells and GFP expression

levels) are taken from the experimental data displayed in right hand panels. mRNA levels are assumed to be zero initially. **c**,**d**, Comparing model outputs with experimental data for population distributions of emergent gene expression. GFP-CAT molecule distributions are shown for increasing Cm concentrations. X-axis (c) is displayed on a log10 scale and y-axis is scaled such that the total area of the histograms sum to 1. The flow cytometry analysis (d) of the genome-integrated GFP-CAT-T_{172}A mutant matches the model prediction (c). Cells in (d) were weakly induced with 0.005% arabinose, treated with 0, 4 or 6 µg/mL Cm for 12hrs, and fixed in 2% paraformaldehyde; n=3 biological replicates. **e**, RT-qPCR assay of Gfp-cat-T172A mRNA expression in populations induced with 0.005% arabinose, treated with 0 (blue) or 5 (red) µg/mL Cm, and harvested at A_{600} of 0.2, 0.4, 0.6 and 0.8 (Mean ± SEm; n=3 biological replicates). Asterisks represent p values: ** = $p<0.01$. **f**, Cm concentration in cultures of the genome-integrated GFP-CAT-T_{172}A mutant strain, measured by LC-MS at one hour intervals (orange). This is compared with the predicted Cm concentration from simulation of the refined inducible promoter model (Supplementary Information, Fig. S17). Both time series are normalised to their maximal values.

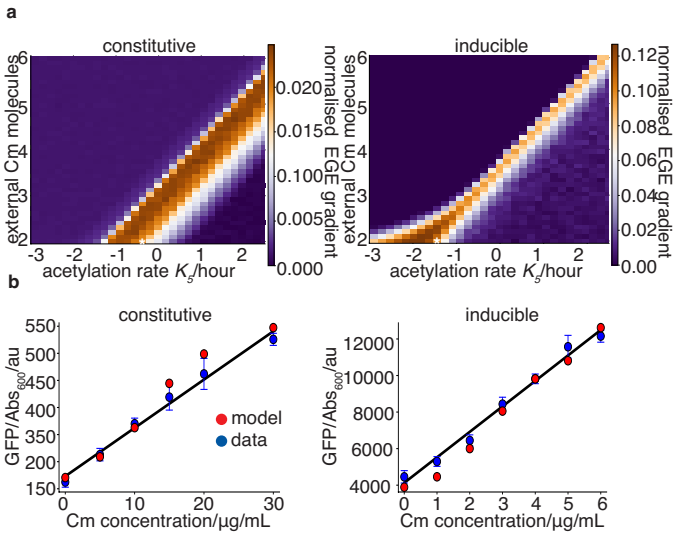
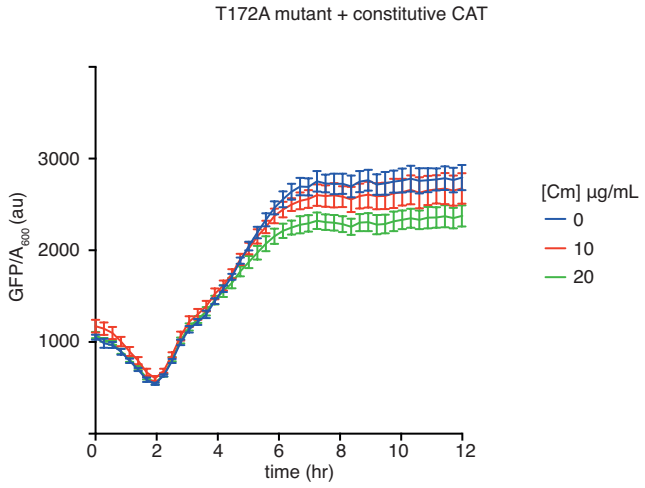
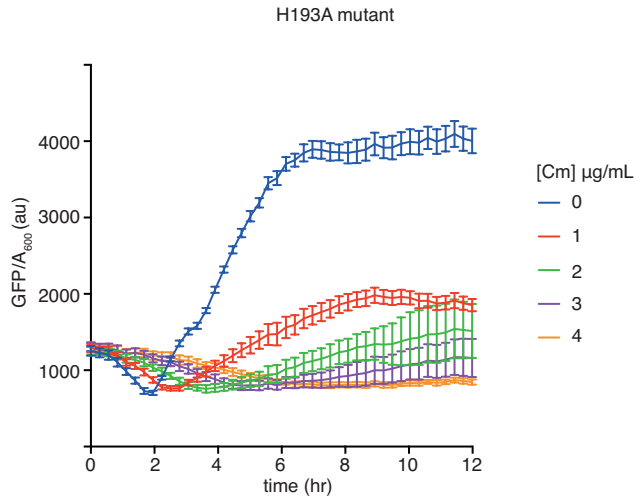


Fig. 3: Bands of maximal emergent gene expression in parameter space.
a, Heatmaps showing effect of varying acetylation rate and chloramphenicol (Cm) dosage on a normalised emergent gene expression (EGE) gradient, for inducible and constitutive promoter models. Six different simulations are computed per model, corresponding to cells being treated with Cm (left panel: 0C, 5C, 10C, 15C, 20C or 30C; right panel: 0C, C, 2C, 3C, 4C, 5C or 6C; where C is the number of external Cm molecules). Linear regression is applied to the GFP-CAT levels yielded from these simulations for each time point, to obtain the maximum gradient. Asterisks highlight the parameter combinations found to provide the best fits to the microplate reader data. **b,** Linear regression fits of microplate reader data showing linearity of GFP expression relationship ('Ohm's law') with Cm treatment dosage, for both inducible and constitutive cases. Left panel shows mean (± SDm) GFP expression (GFP/A_{600} per well) in populations treated with 0, 5, 10, 15, 20, or 30 µg/mL Cm at t = 3 hrs (blue points) with linear regression fit overlaid (black solid line). These experimental data are compared with the predicted mean GFP-CAT molecules per cell for constitutive promoter model simulations at same time point (red points). The right panel is similar, for populations treated with 0, 1, 2, 3, 4, 5 or 6 µg/mL Cm at t = 12 hrs (blue points), compared with model predictions (red points).

a

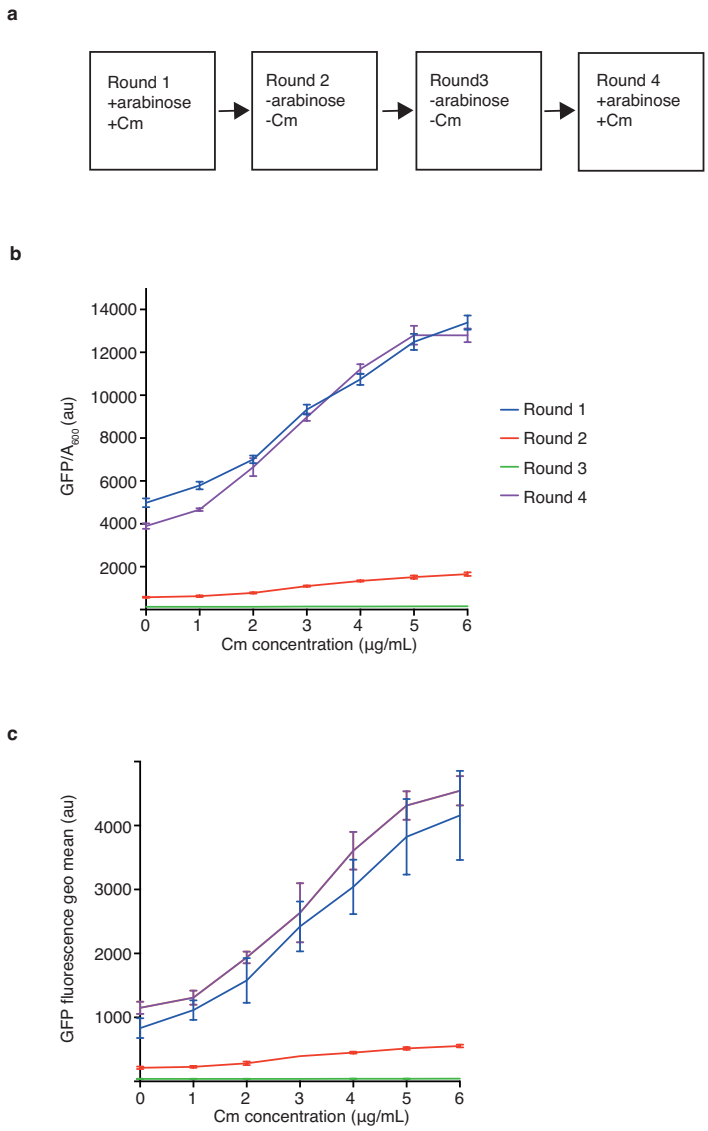


b

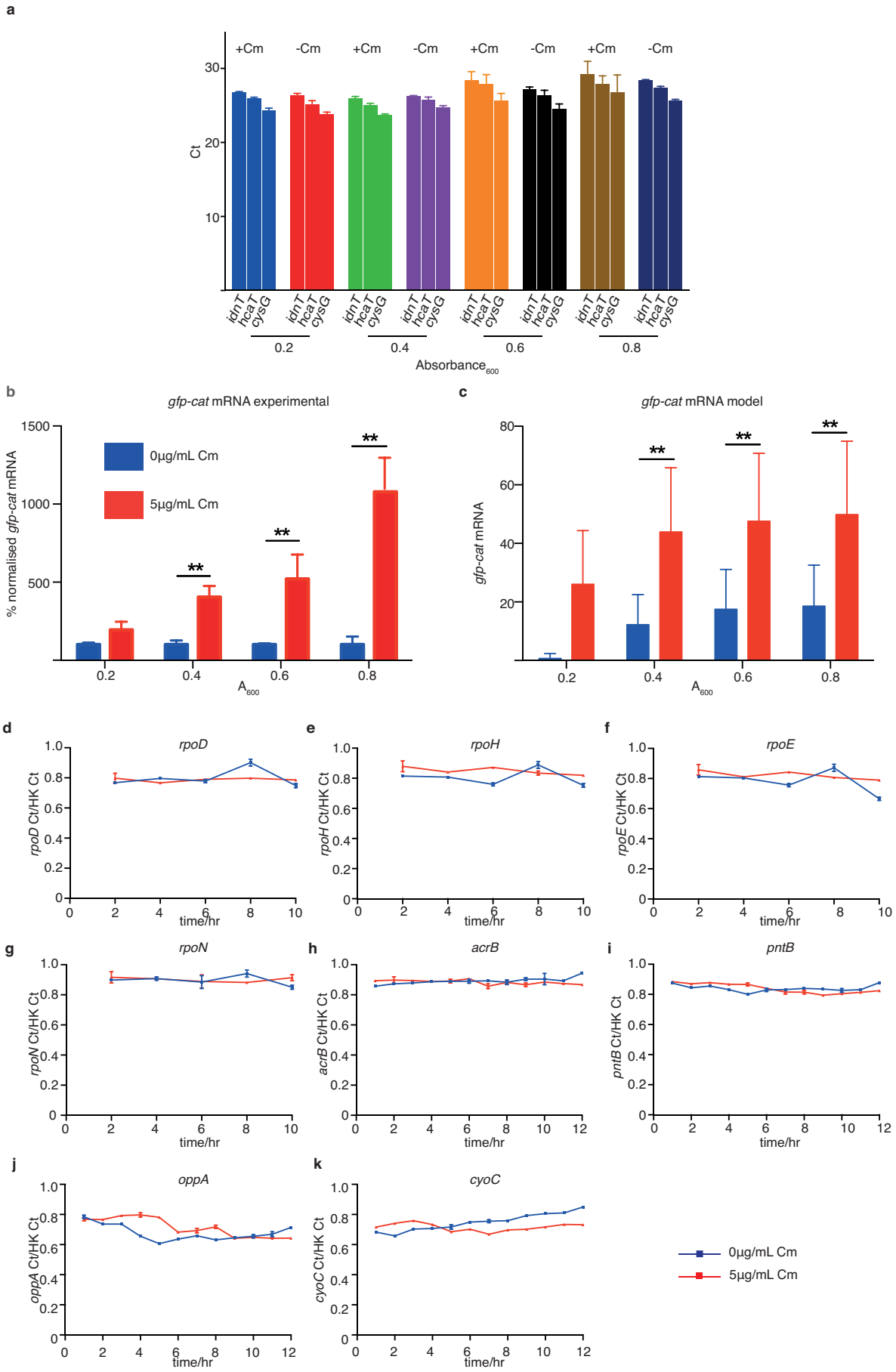


Extended Data Fig. 1: Chloramphenicol treatment does not result in increased gene expression of GFP-CAT cassette under conditions of relieved fitness pressure or inactivity of CAT enzyme.

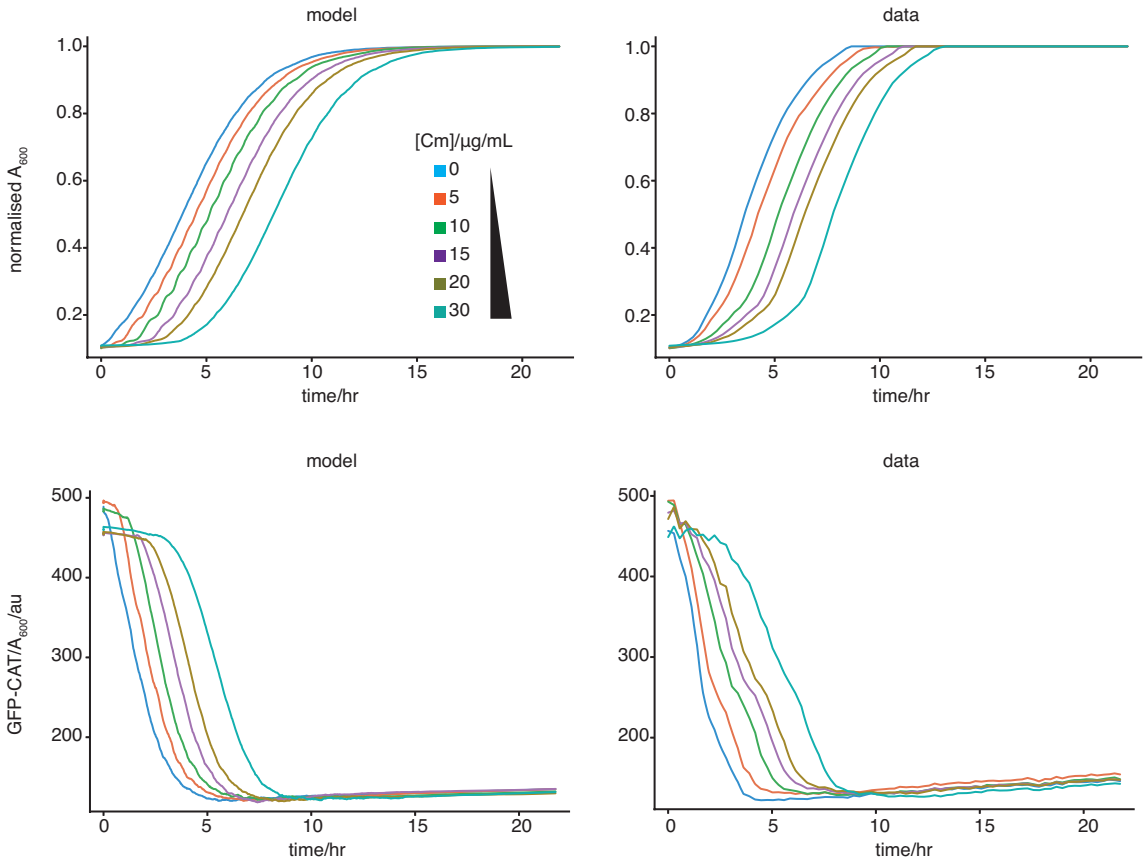
a, The *E. coli* strain MK01 intC::pBad-B0034-GFP-CAT-T_{172}A was transfected with an extra plasmid constitutively expressing the CAT gene to neutralise the fitness selection pressure on the integrated antibiotic resistance gene. Cells were induced with arabinose and treated as before (Fig. 1f), with 0, 10 and 20 µg/mL Cm, but no longer displayed emergent gene expression. Mean (± SDm) GFP expression (GFP/A_{600} per well) was monitored for 12hrs, representative of n=3 biological replicates. **b,** An inactive GFP-CAT no longer upregulates gene expression upon Cm challenge. The *E. coli* strain MK01 intC::pBad-B0034-GFP-CAT-H_{193}A (inactive CAT mutant) was induced with 0.005% arabinose and treated with 0, 1, 2, 3, and 4 µg/mL Cm. Mean (± SDm) GFP expression (GFP/A_{600} per well) was monitored for 12 hrs; n=3 biological replicates.



Extended Data Fig. 2: Emergent gene expression (EGE) is reversible and repeatable upon rounds of growth after stressor addition, removal, and re-addition.
a, Schematic of experimental design. Round 1, 150 µl of *E. coli* MK01 intC::pBad-B0034-GFP-CAT-T_{172}A culture was induced with 0.005% arabinose and treated with 0,1,2,3,4,5, or 6 µg/mL Cm, to apply different fitness pressures on the GFP-CAT. GFP expression and growth were monitored for 12hrs. Starting cultures for Round 2 were obtained from the 12hr time point wells of the Round 1 experiment, where 2µl of culture was used to inoculate 148µl LB media, now in the absence of arabinose or Cm (to test EGE reversibility). GFP expression and growth were monitored for 12hrs. Starter culture for Round 3 was obtained as above from the 12hr time point of Round 2 and grown in the same manner, in the absence of arabinose or Cm. Round 4 was inoculated from the 12hr time point of Round 3 and cells were cultured while restoring the presence of 0.005% arabinose and 0,1,2,3,4,5, or 6 µg/mL Cm. **b**, Experimental results of reversibility experiment showing mean (± SDm) GFP expression (GFP/A_{600} per well), monitored for 12hrs of Rounds 1-4; n=3 biological replicates. **c,** Comparable flow cytometry data for the plate reader experiments in (b). Every condition in each round of growth was fixed with 2% paraformaldehyde at 12hrs, analysed by flow cytometry and graphed as the geometric mean of each culture in Rounds 1-4; n=3 biological replicates.

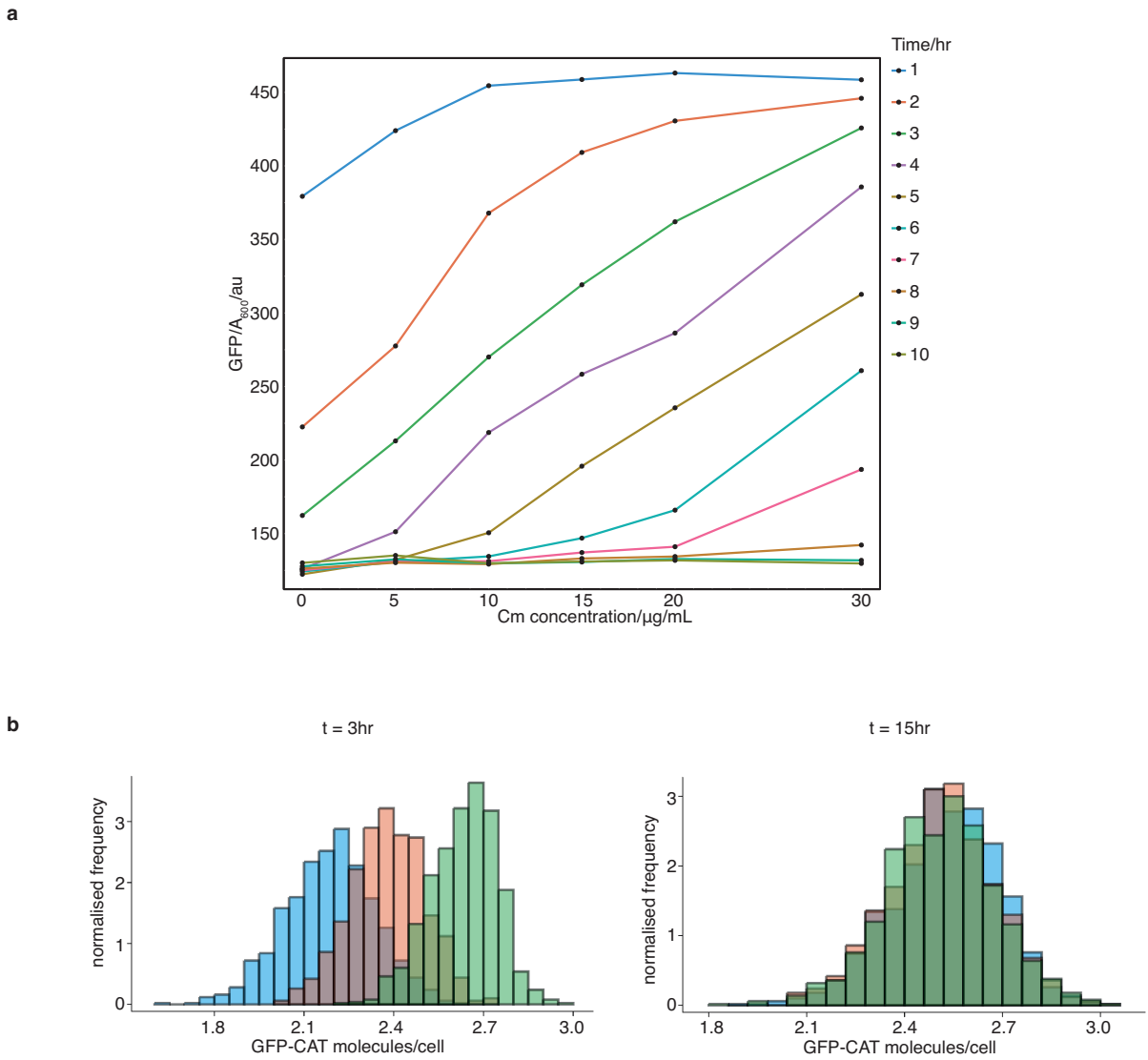

a
+Cm
-Cm
Ct
idnT
hcaT
cysG
0.2
0.4
0.6
0.8
Absorbance600
b
gfp-cat mRNA experimental
% normalised gfp-cat mRNA
0µg/mL Cm
5µg/mL Cm
A600
c
gfp-cat mRNA model
gfp-cat mRNA
d
rpoD
rpoD Ct/HK Ct
time/hr
e
rpoH
rpoH Ct/HK Ct
f
rpoE
rpoE Ct/HK Ct
g
rpoN
rpoN Ct/HK Ct
h
acrB
acrB Ct/HK Ct
i
pntB
pntB Ct/HK Ct
j
oppA
oppA Ct/HK Ct
k
cyoC
cyoC Ct/HK Ct
0µg/mL Cm
5µg/mL Cm

Extended Data Fig. 3: Emergent gene expression-associated upregulation is specific to *gfp-cat* mRNA and is predicted by the model without further fitting.
RT-qPCR assay showing (**a**) housekeeping gene *idnT*, *hcaT*, and *cysG* transcript raw Ct values used to analyse *gfp-cat-T172A* expression in cells induced with 0.005% arabinose, treated with 0 or 5 µg/mL Cm and harvested at A_{600} of 0.2, 0.4, 0.6, and 0.8, mean (± SEm, n=3 biological replicates) and (**b**), *gfp-cat-T172A* expression normalised to the housekeeping genes in (**a**), Cm concentrations are indicated as 0 (blue) or 5 (red) µg/mL in each sample (± SEm, n=3 biological replicates). Asterisks represent p values: ** = p<0.01; delta-delta Ct value analysis was used for statistical analysis and values within comparison groups were normalised to the control 0 µg/mL Cm treatment. (**c**) *gfp-cat* mRNA values from simulations of the inducible promoter model treated with 0 (blue) or 5 (red) µg/mL Cm for cell populations corresponding to A_{600} of 0.2, 0.4, 0.6 and 0.8. Asterisks represent p values: ** = p<0.01; Two-sample Kolmorgorov-Smirnov test was used. mRNA Expression of sigma factors (**d**) *rpoD*, (**e**) *rpoH*, (**f**) *rpoE*, (**g**) *rpoN* as well as genes with reported potential transcriptional level fluctuations in the presence of antibiotics[16] including (**h**) *acrB*, (**i**) *pntB*, (**j**) *oppA*, and (**k**) *cyoC*. Ratios of transcript Ct values to housekeeping transcript Ct values are reported (± SDm, n=1 biological replicate).



Extended Data Fig. 4: Comparison of constitutive expression model of emergent gene expression with experimental data (promoter BBaJ23100 with mutant $T_{172}A$ GFP-CAT).
Time series showing total number of cells from simulation of a constitutive promoter model normalised by carrying capacity (upper left panel), corresponding experimental mean A_{600} data (upper right panel), mean number of GFP-CAT molecules per cell from simulation of

the constitutive promoter model (lower left panel) and corresponding experimental mean GFP expression data (lower right panel), for cells treated with 0, 5, 10, 15, 20 or 30 µg/mL Cm, for a time period of 21 hrs. Best fit parameters are used from ABC parameter inference and initial conditions (initial number of cells and GFP expression levels) are taken from data displayed in right hand side panels. mRNA levels are assumed to be zero initially.



Extended Data Fig. 5: Transient upregulation of GFP-CAT observed under experimental and model conditions (constitutive emergent gene expression model). **a**, The *E. coli* strain MK01 intC::J23100-B0034-GFP-CAT-T_{172}A was treated with 0, 5, 10, 20, or 30 µg/mL Cm and mean (± SDm) GFP expression (GFP/A_{600} per well) was monitored for 10hrs, representative of n=3 biological replicates, with each line corresponding to a different time point. Linear regression analysis of these lines reveals linearity is maximised at t = 3hrs. **b**, Left panel shows GFP-CAT distributions from simulations of the constitutive promoter model at t = 3 hr, displayed in histograms for 2, 4 or 6 µg/mL Cm (coloured in blue, red and green respectively). Right panel shows GFP-CAT distributions from simulations of constitutive promoter model at t = 15 hr displayed in histograms for 2, 4 or 6 µg/mL Cm (coloured in blue, red and green respectively). X-axes are displayed on log10 scale and y-axes are scaled such that the total area of the histograms sum to 1.

Extended Data Table 1: Summary of ABC parameter study of 10 different potential models of EGE.

Model	1	2	3	4	5	6	7	8	9	10
mRNA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
protein	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓
biased partitioning	✗	✗	✗	✗	✓	✗	✗	✗	✓	✗
global transcription feedback	✗	✗	✗	✓	✗	✓	✗	✓	✗	✗
global translation feedback	✗	✗	✗	✗	✗	✗	✓	✓	✗	✗
constitutive	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗
regulated	✗	✗	✗	✓	✗	✗	✗	✗	✗	✓
Ohm's Law	✗	✗	✗	✗	✗	✗	✓	✓	✗	✓
max ratio	1.0	1.0	1.0	1.0	1.0	1.0	1.75	1.8	1.0	2.1
no. of parameters	2	3	3	4	3	4	4	5	4	5

The upper part of the table lists the different model assumptions made. A tick indicates that the model (for a given column) contains an assumption (given by the row) while a cross indicates that it does not. The lower part (highlighted in gray) summarises the model properties. The first row indicates whether each model is capable of producing a linear relationship between the fitness protein and the fitness pressure. The second row shows the maximum mean ratio of the fitness protein to the reference protein for each model. The final row shows the number of parameters for each model.

References

1 Kashiwagi, A., Urabe, I., Kaneko, K. & Yomo, T. Adaptive response of a gene network to environmental changes by fitness-induced attractor selection. *PLoS One* **1**, e49, doi:10.1371/journal.pone.0000049 (2006).

2 Tsuru, S. *et al.* Adaptation by stochastic switching of a monostable genetic circuit in Escherichia coli. *Mol Syst Biol* **7**, 493, doi:10.1038/msb.2011.24 (2011).

3 Klumpp, S. & Hwa, T. Bacterial growth: global effects on gene expression, growth feedback and proteome partition. *Curr Opin Biotechnol* **28**, 96-102, doi:10.1016/j.copbio.2014.01.001 (2014).

4 Klumpp, S., Zhang, Z. & Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**, 1366-1375, doi:10.1016/j.cell.2009.12.001 (2009).

5 Deris, J. B. *et al.* The innate growth bistability and fitness landscapes of antibiotic-resistant bacteria. *Science* **342**, 1237435, doi:10.1126/science.1237435 (2013).

6 Ho, W. C. & Zhang, J. Evolutionary adaptations to new environments generally reverse plastic phenotypic changes. *Nat Commun* **9**, 350, doi:10.1038/s41467-017-02724-5 (2018).

7 van Boxtel, C., van Heerden, J. H., Nordholt, N., Schmidt, P. & Bruggeman, F. J. Taking chances and making mistakes: non-genetic phenotypic heterogeneity and its consequences for surviving in dynamic environments. *J R Soc Interface* **14**, doi:10.1098/rsif.2017.0141 (2017).

8 Lin, W. H. & Kussell, E. Complex Interplay of Physiology and Selection in the Emergence of Antibiotic Resistance. *Curr Biol* **26**, 1486-1493, doi:10.1016/j.cub.2016.04.015 (2016).

9 Mora, T. & Walczak, A. M. Effect of phenotypic selection on stochastic gene expression. *J Phys Chem B* **117**, 13194-13205, doi:10.1021/jp403231f (2013).

10 Thomas, P. Making sense of snapshot data: ergodic principle for clonal cell populations. *J R Soc Interface* **14**, doi:10.1098/rsif.2017.0467 (2017).

11 Martins, B. M. & Locke, J. C. Microbial individuality: how single-cell heterogeneity enables population level strategies. *Curr Opin Microbiol* **24**, 104-112, doi:10.1016/j.mib.2015.01.003 (2015).

12 Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075-2078, doi:10.1126/science.1114383 (2005).

13 Patange, O. *et al.* Escherichia coli can survive stress by noisy growth modulation. *Nature Communications* **9**, 5333, doi:10.1038/s41467-018-07702-z (2018).

14 Nozoe, T., Kussell, E. & Wakamoto, Y. Inferring fitness landscapes and selection on phenotypic states from single-cell genealogical data. *PLoS Genet* **13**, e1006653, doi:10.1371/journal.pgen.1006653 (2017).

15 Lewendon, A. & Shaw, W. V. Transition state stabilization by chloramphenicol acetyltransferase. Role of a water molecule bound to threonine 174. *J Biol Chem* **268**, 20997-21001 (1993).

16 Suzuki, S., Horinouchi, T. & Furusawa, C. Prediction of antibiotic resistance by gene expression profiles. *Nat Commun* **5**, 5792, doi:10.1038/ncomms6792 (2014).

17 Bertaux, F., Marguerat, S. & Shahrezaei, V. Division rate, cell size and proteome allocation: impact on gene expression noise and implications for the dynamics of genetic circuits. *R Soc Open Sci* **5**, 172234, doi:10.1098/rsos.172234 (2018).

18 Shahrezaei, V. & Marguerat, S. Connecting growth with gene expression: of noise and numbers. *Curr Opin Microbiol* **25**, 127-135, doi:10.1016/j.mib.2015.05.012 (2015).
19 Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M. P. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface* **6**, 187-202, doi:10.1098/rsif.2008.0172 (2009).
20 Austin, D. W. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439**, 608-611, doi:10.1038/nature04194 (2006).
21 Erickson, D. W. *et al.* A global resource allocation strategy governs growth transition kinetics of Escherichia coli. *Nature* **551**, 119-123, doi:10.1038/nature24299 (2017).
22 Klumpp, S. & Hwa, T. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc Natl Acad Sci U S A* **105**, 20245-20250, doi:10.1073/pnas.0804953105 (2008).
23 El Meouche, I., Siu, Y. & Dunlop, M. J. Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells. *Sci Rep* **6**, 19538, doi:10.1038/srep19538 (2016).
24 Garcia-Bernardo, J. & Dunlop, M. J. Tunable stochastic pulsing in the Escherichia coli multiple antibiotic resistance network from interlinked positive and negative feedback loops. *PLoS Comput Biol* **9**, e1003229, doi:10.1371/journal.pcbi.1003229 (2013).
25 Sanchez-Romero, M. A. & Casadesus, J. Contribution of phenotypic heterogeneity to adaptive antibiotic resistance. *Proc Natl Acad Sci U S A* **111**, 355-360, doi:10.1073/pnas.1316084111 (2014).
26 Baumstark, R. *et al.* The propagation of perturbations in rewired bacterial gene networks. *Nat Commun* **6**, 10105, doi:10.1038/ncomms10105 (2015).
27 Schaerli, Y. *et al.* A unified design space of synthetic stripe-forming networks. *Nat Commun* **5**, 4905, doi:10.1038/ncomms5905 (2014).
28 Kogenaru, M. & Tans, S. J. An improved Escherichia coli strain to host gene regulatory networks involving both the AraC and LacI inducible transcription factors. *J Biol Eng* **8**, 2, doi:10.1186/1754-1611-8-2 (2014).
29 Isalan, M. *et al.* Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452**, 840-845, doi:10.1038/nature06847 (2008).
30 Zhou, K. *et al.* Novel reference genes for quantifying transcriptional responses of Escherichia coli to protein overexpression by quantitative PCR. *BMC Mol Biol* **12**, 18, doi:10.1186/1471-2199-12-18 (2011).
31 Rebollo-Ramirez, S. & Larrouy-Maumus, G. NaCl triggers the CRP-dependent increase of cAMP in Mycobacterium tuberculosis. *Tuberculosis (Edinb)* **116**, 8-16, doi:10.1016/j.tube.2019.03.009 (2019).

Acknowledgements

This work was funded by Wellcome Trust UK New Investigator Award No. WT102944 and was supported by a Royal Society International Exchanges Cost Share Award (IE170297). We thank Miles Priestman for technical assistance and Anthony Bowman for sharing the script for ABC-SMC model selection in Julia language. We also thank Martin Buck, Philipp Thomas and Peter Swain for comments on the manuscript.