# Gene Expression Changes with Minor Effects on the Population Average Have Major Effects on the Occurrence of Cells with Extreme Protein Concentrations

Mikkel Skjoldan Svenningsen,[a] Szabolcs Semsey,[b] Namiko Mitarai[a]

[a]The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
[b]Department of Biology, University of Copenhagen, Copenhagen, Denmark

**ABSTRACT** The cell-to-cell heterogeneity in a bacterial population provides a rich response to environmental changes and robust survival of an isogenic population. Especially, the rare, extreme phenotypes can be important for survival under transient lethal conditions. We analyze the probability of having an extremely high or low protein level in a stochastic model of gene expression. The fraction of rare state cells defined as the cells in the tails of distributions is found to be highly sensitive to small changes of the mean protein level. The result highlights the importance of relatively weak changes to the mean for the occurrence of rare phenotypes.

**KEYWORDS** distribution, noise, rare events

A large population of bacteria enables us to detect rarely occurring phenotypes experimentally. Examples of rare phenotypes include the well-known bacterial persistence that occurs in as rarely as 1 in $10^6$ cells and allows survival at otherwise lethal doses of antibiotics (1) or sporulation in a subpopulation of *Bacillus subtilis* under strong starvation stress (2). More broadly, a small subpopulation of cells containing much less or much more of a given protein than the population average can have a significant consequence in the overall population's growth and survival, by allowing them to avoid phage attack (3, 4), to survive in the presence of antimicrobials (5–7), or to respond faster to nutrient shifts (8–10). Such heterogeneity of phenotypes can occur due to inevitable noise in gene expression (11), or it can evolve as a regulated bet-hedging strategy: i.e., investing a subpopulation into various phenotypes that are beneficial for different environments in order to survive a sudden environmental change (12, 13).

It has been extensively studied how cells' behaviors depend on the noise in gene expression (14, 15). Typically the noise is characterized by using the ratio between the standard deviation and the mean (11, 16) or the Fano factor (15, 16), which characterize the "typical" deviation of the protein concentrations from the mean concentrations. There have also been attempts to characterize the full distribution based on mathematical models (17–19) and experiments (16, 19).

For a typical protein distribution (16, 19), a seemingly small change in the protein distribution shape can have a paramount impact on the probability of rare concentrations. Because this is associated with the tails of the protein distributions, we refer to this as the "tail effect." However, despite this observation and the biological importance of rare phenotypes, the dependence of the extreme fractions on the various parameters in gene expression has received little attention.

Here, we revisit the protein distribution in an isogenic population in a simple mathematical model, with a focus on the extreme events that occur with small probability. We analyze the fraction of "rare-state cells," which have a rare number

Address correspondence to Namiko Mitarai, mitarai@nbi.dk.

Rare, extreme phenotypes can be important for survival under transient lethal conditions. We show that gene expression changes with minor effects on the population average have major effects on the occurrence of cells with extreme protein concentration.
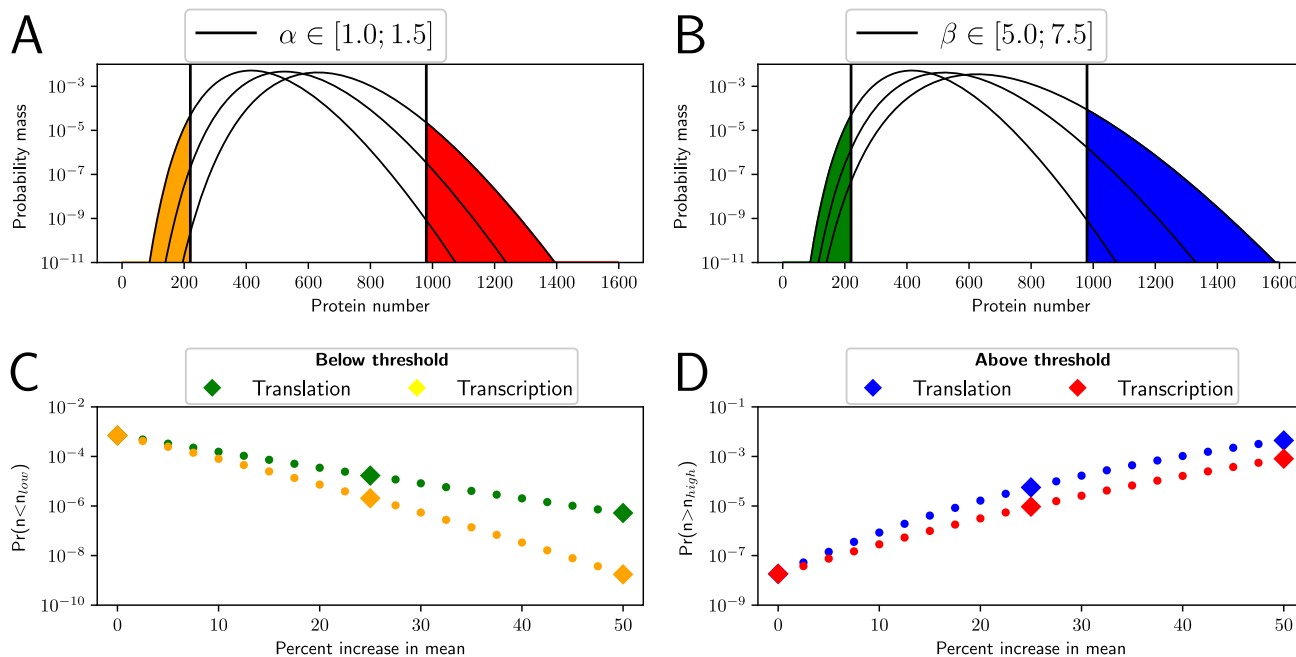
**FIG 1** The occurrence of rare states is strongly dependent on the mean protein number. We assume the *E. coli* volume is ~1 μm³; hence, one protein per cell corresponds to the concentration of 1 nM. Unless otherwise noted, parameters are fixed to $\alpha = 1/\text{min}$, $\gamma_m = 0.35/\text{min}$, $\gamma_p = 1/30/\text{min}$, and $\beta = 5/\text{min}$. (A) Protein distribution with 3 different values of the transcription rate ($\alpha = 1/\text{min}$, 1.25/min, and 1.5/min). A high threshold and low threshold are indicated by vertical lines and are the same for both panels A and B. (B) Protein distribution with 3 different values of the translation rate ($\beta = 5/\text{min}$, 6.25/min, and 7.5/min). The three different means are the same in panels A and B. (C) The probability to be below a threshold, upon changing the transcription rate, $\alpha$ (yellow), or the translation rate, $\beta$ (green), to have a different protein mean. The effect of changes in transcription has a stronger impact than changes in translation. (D) The probability to be above the threshold, upon changing the transcription rate, $\alpha$ (red), or the translation rate, $\beta$ (blue), to have a different protein mean. The three distributions depicted in panels A and B are represented by diamonds in panels C and D.

of a given protein compared to the rest of the population. In a bacterial culture, the cell count can easily exceed $10^9$ cells/ml, meaning that the rare events that occur with a probability as low as 1 in $10^9$ can still be observed. Although they are rare and often transient, these events can play a crucial role if the rare state of interest provides advantages in survival. We demonstrate the strong sensitivity of the fraction of the rare-state cells to the various parameters that result in changes in the mean.

## MODELING

**The rare-state cells in the steady state.** We consider a simple model of stochastic gene expression (19, 20):

$$\text{DNA} \xrightarrow{\alpha} \text{mRNA} \xrightarrow{\beta} \text{protein} \qquad (1)$$
$$\phantom{\text{DNA}} \downarrow\gamma_m \quad\; \downarrow\gamma_p$$
$$\phantom{\text{DNA}} \varnothing \qquad\;\; \varnothing$$

Here, mRNAs are transcribed at rate $\alpha$ and degraded at rate $\gamma_m$, each mRNA is translated to produce proteins at rate $\beta$, and the proteins are degraded or diluted at rate $\gamma_p$. We use parameters typical for the bacterium *Escherichia coli* (20), where mRNA lifetime is much shorter than the protein dilution time: hence multiple proteins are produced as a burst after one mRNA is transcribed. Because this "burstiness" strongly affects the protein number fluctuation (16, 17), the transcription and the translation contribute differently to the distribution. For that reason, changes in the protein distribution tails are analyzed for both changes in the transcription rate, $\alpha$, and the translation rate, $\beta$. To quantify the dependence of the distribution tail on the model parameters, we numerically integrate the corresponding master equation (see Methods) to obtain the distributions.

The example protein number distributions in the steady state are shown in Fig. 1. In Fig. 1A, we show the three distributions, where the protein mean number is altered by
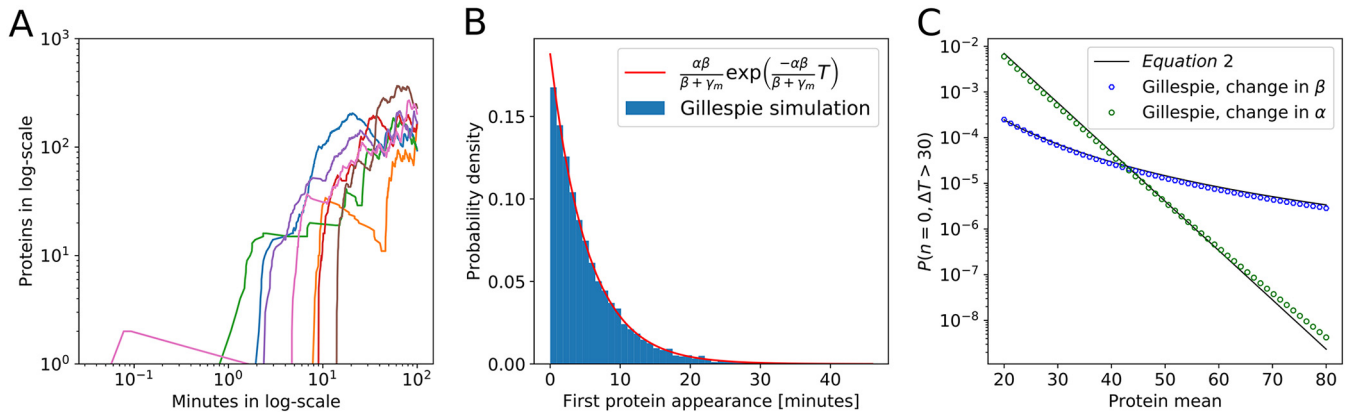
**FIG 2** Probability to have zero proteins for the duration of a cell generation time. Unless otherwise noted, parameters are set to be $\alpha = 0.2$/min, $\beta = 10$/min, $\gamma_m = 0.35$/min, and $\gamma_p = (1/30)$/min. (A) Example trajectories of protein numbers per cell over time when the cell had initially zero mRNA and proteins. (B) Distribution of time intervals before the first protein production. The distribution resembles the exponential distribution expected for the waiting for the first mRNA to appear. (C) Probability of a cell having zero proteins for more than 30 min when the mean number of proteins harbored by cells in the population is altered by varying the transcription rate, $\alpha$ (green), or the translation rate, $\beta$ (blue). Gillespie simulations (symbols) agree well with the analytical expression in equation 2, with $\tau_c = 30$ min.

changing the transcription rate, $\alpha$, without changing other parameters. For comparison, in Fig. 1B, we show the distribution of the protein number with the same means, but by changing the translation rate, $\beta$.

The different outcomes with regard to the probability of tail events are due to their different degrees of burstiness. Increasing the average by more frequent transcription (higher $\alpha$) averages out the noise from bursty production, while the increase of $\beta$ results in more protein production per mRNA, increasing the burstiness and hence increasing the frequency of the rare tail events.

To quantify the rare states, we draw two thresholds, $n_{low}$ and $n_{high}$. We see that the fractions of the population beyond the thresholds change drastically upon changing the mean moderately. This sensitivity is depicted quantitatively in Fig. 1C by plotting the probability to have protein below $n_{low}$ (or above $n_{high}$ [Fig. 1D]) as a function of the mean number of proteins.

For all the cases, we observe change of several orders of magnitude in the extreme population fraction upon changing the mean by only up to 50%. The difference in the $\alpha$ and $\beta$ dependence is consistent with the fact that an increase of $\beta$ results in a stronger increase of noise than an increase of $\alpha$: The rare states below a threshold with increasing $\beta$ do not decrease as steep as the case with increasing $\alpha$, even though the population mean is going away from the threshold $n_{low}$, because the distribution is widening more with $\beta$. Similarly, the widening distribution with increasing $\beta$ increases the probability to be above $n_{high}$, steeper than the case with increasing $\alpha$.

The sensitivity of the tails to the parameter change can easily also be confirmed in regulated systems. We summarize the effect in autorepressed systems in the supplemental material.

**Residence time in a rare state.** The duration a cell stays in the extreme state is often also important. A good example is the survival of a phage attack by bacteria not expressing the phage receptors (3); if a cell can stay in a zero-phage-receptor state for one cell generation time under exposure to phage, that cell will be able to give rise to two uninfected new cells. This motivates us to examine the probability to have zero proteins for a time scale of a cell generation.

Figure 2A depicts several trajectories of protein number in a single cell over time that has zero protein and zero mRNA at time zero obtained by the Gillespie method (21). With the present parameter set where the mean protein production per mRNA $\dfrac{\beta}{\gamma_m}$ is significantly larger than 1, most cells will produce a protein soon after the first mRNA is produced. This means that the first protein production is dominated by a Poisson

process with the rate $\frac{\alpha\beta}{\gamma_m+\beta}$, which corresponds to the rate of mRNA production, $\alpha$, with a correction for the probability to produce a protein before degrading the mRNA, $\frac{\beta}{\gamma_m+\beta}$. Thus, the probability to stay in the zero-protein state decays exponentially with $t$ as $e^{\frac{-\alpha\beta}{\gamma_m+\beta}t}$ (Fig. 2B).

Figure 2C shows the probability to have zero proteins for a duration of $\Delta T$ longer than $\tau_c = 30$ min, $P(n = 0, \Delta T > \tau_c)$, in the steady state obtained by Gillespie simulation. Only cells with zero proteins at time zero are considered, and each of these cells that resides in the zero-protein state for 30 min or more is included. We see that when the protein mean is reduced from 80 to 20, such a population fraction increases several orders of magnitude. We again see the exponential dependence on $\alpha$, while there is weaker dependence on $\beta$. Qualitatively this is natural, because we already know that the steady-state probability to be below a threshold value depends exponentially on $\alpha$ as in Fig. 1C, and the duration of time that a bacterium has zero proteins depends exponentially on $\alpha$. Using the steady-state probability to have zero proteins in reference 18, the approximate form of the distribution (see the supplemental material for derivation) is

$$P(n = 0, \Delta T > \tau_c) = \left(\frac{1}{1 + \beta\gamma_m}\right)^{\frac{\alpha}{\gamma_p}}\exp\left(\frac{-\alpha\beta}{\gamma_m + \beta}t\right) \tag{2}$$

For $\beta >> \gamma_m$, expression 2 has a weak dependence on $\beta$ and a strong dependence on $\alpha$, confirming a much stronger effect from changes in the transcription rate compared to changes in the translation rate. Equation 2 is compared to the Gillespie simulation in Fig. 2C.

## CONCLUSIONS

We have demonstrated that the fraction of cells in a population that has an extremely high or low protein level is remarkably sensitive to changes in the protein production parameters that do not affect the mean protein level of the population much. When the mean protein level is changed by a transcriptional or translational regulation by 50%, the fraction of the cells in the rare state can easily change by several orders of magnitude. This insight underscores a need to consider the potentially large effects on the tails of a distribution of cells, when a weak regulatory effect on the population average is observed, if the cells in the tail of the distribution have an interesting phenotype.

For example, in transcriptome-wide studies that analyze changes in gene expression in response to an external stimulus, the threshold for detection of the expression change is often chosen to be relatively high (e.g., 2-fold) to avoid the experimental noise. However, the present analysis shows that even a 10% change in the mean protein level can be enough to cause an order of magnitude change in the tail population fraction. This gap between the response in the mean and the tail may make it challenging to reveal the regulatory links responsible for the rare phenotypes.

An interesting example of the tail effect is the observation that bacterial quorum-sensing signals repress the mean λ phage receptor level by only 40%, but the fraction of cells in the quorum-sensing-induced bacterial population that survived incubation with $\lambda_{vir}$ increased by more than 3-fold (22). Although not as pronounced as the change of several orders of magnitude reported here, it is likely that the stronger effect on the survival than the mean is partly due to the distribution tail effect.

Of course, we should be cautious about the relationship between the rare protein concentrations in a cell and a rare phenotype.

For example, the bacterial persistence against antibiotics can be a global effect on the cellular physiology, and the probability of a single protein taking an extreme value may be less relevant. At the same time, the extremely low fraction of persisters indicates that they belong to a tail of some kind of distribution, and in general, the fraction of the extreme in a distribution tends to be sensitive to the parameter change.

The growth phase has an effect of several orders of magnitude on the persister frequency (23), which can be partly due to the tail effect.

Another important point to consider when applying the tail effect concept to phenotypes is how the threshold between a main population and a rare phenotype is determined. In this study, we chose a hard threshold between the bulk and a rare protein concentration, where plus/minus a single protein at the threshold makes the difference. The tail effect is a general feature of most probability distributions with a hard threshold, so it is not surprising to find it in protein distributions. Even though the hard threshold is highly idealized, we point out here that approximately hard thresholds could possibly be observed in biological systems. The evasion of a phage attack is a good candidate, where the lack of a receptor protein should protect the cell from a lethal infection. Other candidates are regulations by heterocomplex formation with strong binding, as seen, for example, in type II toxin-antitoxin systems (24) and regulation of translation by small RNAs (25), where one component exceeding another in numbers changes the phenotype, resulting in ultrasensitivity.

Obviously, softer thresholds can obstruct the tail effect. In the supplement, we show this for thresholds defined by a Hill function associated with complex formation, an exponential function associated with a constant probability per protein for phenotype change, or by having an even softer log-scale threshold where a difference of an order of magnitude in the protein level is required for the phenotype change. Naturally, the softer the threshold, the weaker the effect, and the sensitivity of the tail to the mean disappears for a log-scale threshold. The notion of a tail effect can thus aid in understanding the biological function of various types of thresholds, where hard thresholds lead to ultrasensitive dependence on the mean protein distribution, whereas rare phenotypes given by soft thresholds are less sensitive to small changes.

## METHODS

The master equation for the probability $P(m,n;t)$ to have $m$ mRNAs and $n$ proteins at time $t$ is given by

$$\frac{\partial}{\partial t}P(m,n;t) = \alpha[P(m-1,n;t) - P(m,n;t)] + \gamma_m[(m+1)P(m+1,n;t) - mP(m,n;t)]$$
$$+ \beta[P(m,n-1;t) - P(m,n;t)] + \gamma_p[(n+1)P(m,n+1;t) - nP(m,n;t)]$$

The distributions were obtained by numerically integrating the master equation over time. Maximum $m$ and $n$ in computation were chosen to be large enough so that the probability to take the maximum values stayed zero. The probability distribution for protein number is given by $P(n;t) = \Sigma P(m,n;t)$.

All solutions are found with the initial condition that the cell contains both zero mRNAs and zero proteins of the gene in question: i.e., all solutions are found for the initial condition that $P(m,n;t) = \delta_{m,0}\delta_{n,0}$, where $\delta_{i,j}$ is the Kronecker delta function. The steady-state value is found at 500 min or more.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSphere.00575-18.

**TEXT S1**, PDF file, 0.2 MB.
**FIG S1**, EPS file, 0.2 MB.
**FIG S2**, EPS file, 0.1 MB.
**FIG S3**, EPS file, 0.1 MB.
**FIG S4**, EPS file, 0.1 MB.
**FIG S5**, PDF file, 0.04 MB.
**FIG S6**, EPS file, 0.1 MB.

## ACKNOWLEDGMENTS

## REFERENCES

1. Harms A, Maisonneuve E, Gerdes K. 2016. Mechanisms of bacterial persistence during stress and antibiotic exposure. Science 354:aaf4268. https://doi.org/10.1126/science.aaf4268.
2. Maamar H, Raj A, Dubnau D. 2007. Noise in gene expression determines cell fate in *Bacillus subtilis*. Science 317:526–530. https://doi.org/10.1126/science.1140818.
3. Chapman-McQuiston E, Wu XL. 2008. Stochastic receptor expression allows sensitive bacteria to evade phage attack. Part I. Experiments. Biophys J 94:4525–4536. https://doi.org/10.1529/biophysj.107.120212.
4. Chapman-McQuiston E, Wu XL. 2008. Stochastic receptor expression allows sensitive bacteria to evade phage attack. Part II. Theoretical analyses. Biophys J 94:4537–4548. https://doi.org/10.1529/biophysj.107.121723.
5. Wakamoto Y, Dhar N, Chait R, Schneider K, Signorino-Gelo F, Leibler S, McKinney JD. 2013. Dynamic persistence of antibiotic-stressed mycobacteria. Science 339:91–96. https://doi.org/10.1126/science.1229858.
6. Pu Y, Zhao Z, Li Y, Zou J, Ma Q, Zhao Y, Ke Y, Zhu Y, Chen H, Baker MAB, Ge H, Sun Y, Xie XS, Bai F. 2016. Enhanced efflux activity facilitates drug tolerance in dormant bacterial cells. Mol Cell 62:284–294. https://doi.org/10.1016/j.molcel.2016.03.035.
7. Gerdes K, Semsey S. 2016. Pumping persisters. Nature 534:41–42. https://doi.org/10.1038/nature18442.
8. Ozbudak EM, Thattai M, Lim HN, Shraiman BI, Van Oudenaarden A. 2004. Multistability in the lactose utilization network of Escherichia coli. Nature 427:737–740. https://doi.org/10.1038/nature02298.
9. Semsey S, Jauffred L, Csiszovszki Z, Erdossy J, Stéger V, Hansen S, Krishna S. 2013. The effect of LacI autoregulation on the performance of the lactose utilization system in Escherichia coli. Nucleic Acids Res 41:6381–6390. https://doi.org/10.1093/nar/gkt351.
10. Fritz G, Megerle JA, Westermayer SA, Brick D, Heermann R, Jung K, Rädler JO, Gerland U. 2014. Single cell kinetics of phenotypic switching in the arabinose utilization system in E. coli. PLoS One 9:e89532. https://doi.org/10.1371/journal.pone.0089532.
11. Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic gene expression in a single cell. Science 297:1183–1186. https://doi.org/10.1126/science.1070919.
12. Veening J-W, Smits WK, Kuipers OP. 2008. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol 62:193–210. https://doi.org/10.1146/annurev.micro.62.081307.163002.
13. De Jong IG, Haccou P, Kuipers OP. 2011. Bet hedging or not? A guide to proper classification of microbial survival strategies. Bioessays 33:215–223. https://doi.org/10.1002/bies.201000127.
14. Raser JM, O'Shea EK. 2005. Noise in gene expression: origins, consequences, and control. Science 309:2010–2014. https://doi.org/10.1126/science.1105891.
15. Munsky B, Neuert G, Van Oudenaarden A. 2012. Using gene expression noise to understand gene regulation. Science 336:183–187. https://doi.org/10.1126/science.1216379.
16. Taniguchi Y, Choi PJ, Li G, Chen H, Babu M, Hearn J, Emili A, Xie XS. 2010. Sensitivity in single cells. Science 329:533–538. https://doi.org/10.1126/science.1188308.
17. Friedman N, Cai L, Xie XS. 2006. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Phys Rev Lett 97:1–4. https://doi.org/10.1103/PhysRevLett.97.168302.
18. Shahrezaei V, Swain PS. 2008. Analytical distributions for stochastic gene expression. Proc Natl Acad Sci U S A 105:17256–17261. https://doi.org/10.1073/pnas.0803850105.
19. Krishna S, Banerjee B, Ramakrishnan TV, Shivashankar GV. 2005. Stochastic simulations of the origins and implications of long-tailed distributions in gene expression. Proc Natl Acad Sci U S A 102:4771–4776. https://doi.org/10.1073/pnas.0406415102.
20. Sneppen K. 2014. Models of life. Dynamics and regulation in biological science. Cambridge University Press, Cambridge, United Kingdom.
21. Gillespie DT. 1977. Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81:2340–2361. https://doi.org/10.1021/j100540a008.
22. Høyland-Kroghsbo NM, Mærkedahl RB, Svenningsen SL. 2013. A quorum-sensing-induced bacteriophage defense mechanism. mBio 4:e00362-12. https://doi.org/10.1128/mBio.00362-12.
23. Harms A, Fino C, Sørensen MA, Semsey S. 2017. Prophages and growth dynamics confound experimental results with antibiotic-tolerant persister cells. mBio 8:e01964-17. https://doi.org/10.1128/mBio.01964-17.
24. Rotem E, Loinger A, Ronin I, Levin-Reisman I, Gabay C, Shoresh N, Biham O, Balaban NQ. 2010. Regulation of phenotypic variability by a threshold- based mechanism underlies bacterial persistence. Proc Natl Acad Sci U S A 107:12541–12546. https://doi.org/10.1073/pnas.1004333107.
25. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL. 2004. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae. Cell 118:69–82. https://doi.org/10.1016/j.cell.2004.06.009.