

Power-Law Scaling in Protein Synthesis of a Stochastic Regulon

Emily Chapman-McQuiston¹, Chuck Yeung², and X.L. Wu¹

¹Department of Physics and Astronomy, University of Pittsburgh, PA 15260

²School of Science, The Pennsylvania State University at Erie, The Behrend College, Erie, PA 16583

(Dated: July 28, 2008)

We investigate the protein expression pattern of the *lamB* gene in *Escherichia coli* LE392. The gene product LamB is an important membrane protein for maltose transport into cells but it is also exploited by bacteriophage λ for infection. Although our bacterial population is clonal, stochastic gene expression leads to a majority population with a large receptor number and a minority population with a small receptor number. We find that LamB receptor distribution $p(n)$ of the small- n population is scale invariant, $p(n) \sim n^{-\alpha}$, where the exponent α depends on growth conditions. A heuristic model is proposed that relates the observed exponent to the protein production rate.

The occurrence of scale invariance in physics and biology often reflects an important underlying principle[1, 2, 3]. Herein, we present a novel observation of spontaneous/stochastic gene expression that gives rise to a power-law distribution $p(n) \propto n^{-\alpha}$, where n is the number of LamB receptors in individual bacteria. The experiment is made possible by a dual-colored phage labeling technique that renders the minority population quantifiable in a flow cytometer. The LamB is a maltose channel, but it is also exploited by λ phage as an infection site. We find that in addition to the main population with an average $n \simeq 500$, there is a small subpopulation ($\sim 1\%$) with a small receptor number. This small population increases the likelihood that the bacterial population as a whole survives a phage attack[4]. Surprisingly, the small- n population has a scale-invariant distribution $n^{-\alpha}$, behaving very differently from the log-normal distribution commonly seen for major proteins in a bacterium[5, 6]. We propose a model that takes into account the rate of protein synthesis and protein dilution due to cell division. This model yields a steady-state distribution $p_0(n) \sim n^{-\alpha}$ with α being a continuous function of the rate of protein synthesis.

Our experiments were carried out using *E. coli* LE392 that are sensitive to phage λ [7]. The maltose regulon of this strain is inducible; the LamB receptor number n can be varied from a few to $\sim 10^3$ depending on culture conditions[8]. The bacteria were grown in M9 minimal medium supplemented with either 0.4% glucose or 0.4% maltose. Standard protocols were used to grow the bacteria and to purify the λ phage[7]. The *lamB* is an essential gene for maltose metabolism when the substrate is present at a very low level ($\ll 10 \mu M$). However, in the current experiment, the maltose level is sufficiently high that the low receptor numbers does not impede the growth of the minority population [7, 9]. To quantify the number of LamB receptors n for individual cells, bacteria were incubated with high concentrations of fluorescently labeled λ phage (dye/phage ratio $\gtrsim 2000$), and the brightness of each bacterium was interrogated using flow cytometry (Dako Cyan ADP). Using our labeling procedure, individual phage particles were bright enough to reach the detection threshold of the flow cytometer[7], making them quantifiable by the instrument.

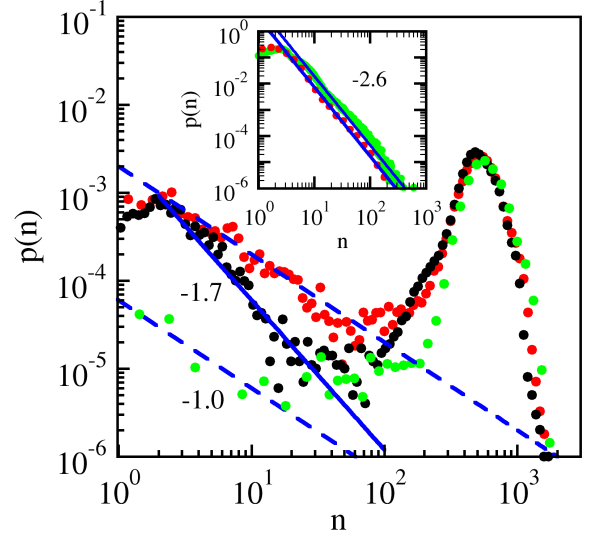


FIG. 1: Three steady-state distributions of $p_0(n)$ of bacteria grown in M9+0.4% maltose. The PDFs of the minority populations follow power laws $p_0(n) \propto n^{-\alpha}$ with $\alpha = 1$ for the dashed lines and $\alpha = 1.7$ for the solid line. The inset displays steady-state distributions of bacteria grown in M9+0.4% glucose. Since the maltose regulon is repressed, there is no majority population. The exponent for both runs was found to be $\alpha \simeq 2.6$ as shown.

We show several runs of LE392 grown in M9+maltose in Fig. 1. The steady-state LamB probability distribution function (PDF) $p_0(n)$ has two peaks, one around $n \sim 500$ and the other near $n \simeq 0$. The minority peak appears prominently in the log-log plot, but this subpopulation is only a few percent of the total population. In rare cases, the fraction is even smaller, as indicated by the lowest curve in the figure. For all runs, the minority population has a power-law distribution $p_0(n) \propto n^{-\alpha}$ that extends more than a decade in n . The measured exponent α is not constant but varies from run to run with α ranging from 1 to 2. This variation may reflect the sensitivity of α to the varying protein production rates in different bacterial cultures. On the other hand, the PDF of the majority population is quite reproducible, and has the usual log-normal form[6]. We conjecture that segre-

gation of cells into minority and majority populations is shaped by phage selection. The low n phenotype allows the bacterial population to hedge its bets due to the dual functionality of LamB receptors. These minority cells are less fit when maltose is not abundant in the environment, but their presence increases the chance that the population as a whole survives a phage attack.

We used a dual-colored phage labeling technique to investigate the role of the minority cells in the recovery of a bacterial population after a phage attack. Specifically, we are interested in the time scale of recovery, and the relationship between the minority and the majority populations. Our method consists of introducing red-labeled (AlexaFluor 633) λ phage into the bacterial population. The bacteria is exposed to the phage for 30 minutes, and the phage pressure, defined as the initial phage concentration $P(0)$, is then relieved by several washing steps[4]. The phage selectively decimates bacteria that express a large number of receptors, leaving only cells with small number of receptors. The surviving bacteria were then allowed to regrow in a fresh medium. Throughout the regrowth period, the bacteria were continuously diluted into prewarmed fresh media. This procedure ensures that the bacterial concentration is always less than 10^7 cc^{-1} , and the background phage concentration is below 10^{-5} cc^{-1} so that the probability that a phage binds to a bacterium is very small[4]. To quantify the time-dependent receptor distribution $p(n, t)$, a small sample was periodically taken from the exponentially growing culture. The receptors on individual cells were tagged with green-labeled (AlexaFluor 488) λ phage and interrogated via flow-cytometry[7]. To stringently discriminate deceased cells, the bacterial sample was also treated with propidium iodide, which specifically labels the dsDNA of infected cells making them fluoresce brightly in orange. The green channels of the flow cytometer is conditioned on the red (the killing phage) and the orange channels (the dead cells) to get rid of the background count.

Figure 2 depicts a sequence of LamB PDFs, $p(n, t)$, before and after the bacteria were exposed to λ phage. Before the phage exposure (Fig. 2a), the LamB distribution has a primary peak at $n \simeq 500$ and a secondary peak at $n \sim 0$ similar to the data in Fig. 1. The minority peak at small n in this case scales as $p_0(n) \sim n^{-1}$ as delineated by the solid lines in Fig. 2a. The majority population is quickly eliminated by the exposure to the λ phage (Fig. 2b), but cells with a small n remain unaffected. Here we observed that the n^{-1} scaling extends over more than three decades $10^0 < n < 10^3$. After the removal of phage, the majority cell population slowly grows back to its original form in ~ 8 hrs (Fig. 2(c-f)), with the distribution of the minority cells maintaining the same shape the entire time. This result is robust in that the functional form of the minority population is time invariant for a wide range of initial phage concentrations $P(0)$ so long as $P(0)$ is not large enough to erode the minority population. This observation suggests that stochastic switching between the minority and the majority states is relatively slow, indi-

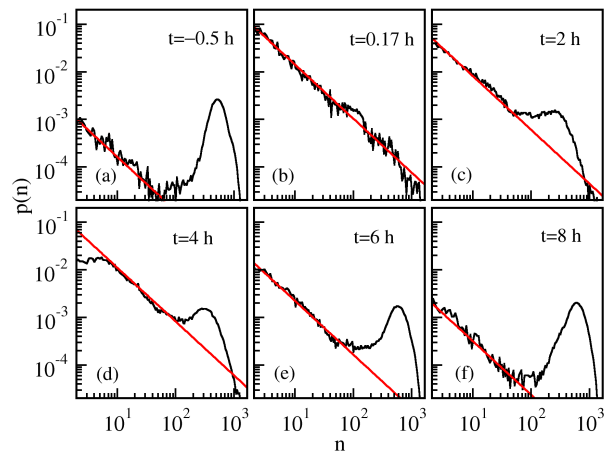


FIG. 2: LE392 bacteria at an initial concentration of $1 \times 10^7 \text{ cm}^{-3}$ and a λ phage concentration $1 \times 10^9 \text{ cm}^{-3}$ were incubated at $T = 37^\circ\text{C}$ for 30 minutes. The phage pressure was then released by washing the cells twice using pre-warmed M9+0.4% maltose. This defines $t = 0$ in the above plots. The population has two peaks before phage is introduced (a). Immediately after phage selection (b), the majority population is decimated, leaving only the low n distribution that still scales as n^{-1} over three decades in n . After the phage removal, a broad peak emerges from the n^{-1} background in ~ 2 hrs (c), and the peak sharpens and moves to large n with time (c-f). Even after 8 hrs (f), corresponding to ~ 11 generations, the majority population is still not fully recovered as indicated by the lower height of $p(n)$.

cating a weak coupling between the two populations. In other words, the minority population is able to replenish itself, maintaining the self-similar form independent of the majority population.

The maltose regulon is inactivated if LE392 is grown in M9+glucose instead of maltose. This is a part of catabolic repression resulting from a sharp drop of cAMP concentration in cells when glucose is present. However, stochastic events still lead LamB to be expressed at a very low level $\bar{n} \simeq 2$. The inset of Fig. 1 displays measurements of the steady-state LamB distribution $p_0(n)$ for this case. We observe that, over a broad range of n , $p_0(n)$ also decays as a power law with the exponent $\alpha \simeq 2.7 \pm 0.2$. In this case, the run-to-run variation of $\Delta\alpha \simeq 0.2$ is smaller than that of the maltose-grown cells. Therefore, the uninduced (or chemically repressed) bacteria and the minority population of the induced bacteria both have power-law LamB distributions $p(n) \sim n^{-\alpha}$ although with somewhat different α and $\Delta\alpha$. The larger $\Delta\alpha$ seen in the maltose-grown cells suggests that this minority population is more sensitive to external variations than uninduced cells and are hence more susceptible to environmental changes.

The population-balance approach[10] can be used to help understand the protein distributions in a bacterial population. The number of cells with n proteins at time t , $B(n, t)$, obeys a master equation:

$$\frac{\partial B(n,t)}{\partial t} = -\gamma(n)B(n,t) - \frac{\partial}{\partial n}(\nu(n)B(n,t)) + 2 \int_n^\infty \gamma(m)\Phi(n|m)B(m,t) dm, \quad (1)$$

where $\Phi(n|m)$ is the partition probability of a daughter cell inheriting n receptors from a mother cell with m receptors, and $\nu(n)$ and $\gamma(n)$ are respectively the state-dependent protein production rate and the cell division rate. The integration on the right-hand side of Eq. 1 takes into account the long-range effect of protein redistribution as a result of cell division. Based on our experiments, the division rate for the majority and minority cells are about the same, and as an approximation, we assume $\gamma(n) = \gamma$. In the steady state, a separation of variables is possible, $B(n,t) = B(t)p_0(n)$, where $B(t) = B_0 e^{\gamma t}$ is the exponentially growing bacterial population and $p_0(n)$ is the normalized steady-state PDF. In this case, Eq. 1 simplifies to:

$$2\gamma p_0(n) = -\frac{\partial}{\partial n}(\nu(n)p_0(n)) + 2\gamma \int_n^\infty \Phi(n|m)p_0(m) dm. \quad (2)$$

During cell division, daughter cells inherit proteins from mother cells. If the partition of proteins is random, we expect that the sharing of m proteins between the two daughter cells will follow a binary distribution. As a simplification, we treated $\Phi(n|m)$ as a δ -function with $\Phi(n|m) \simeq \delta(n-m/2)$. This approximation becomes progressively better with increasing m . Finally, we follow Ref.[6] and assume that the protein production rate is linear in n with $\nu(n) = cn$. This means that in the absence of cell division the proteins in an individual cell grows exponentially, $n(t) = n_0 e^{ct}$. With these assumptions, Eq. 2 becomes

$$2\gamma(p_0(n) - 2p_0(2n)) = -\frac{\partial}{\partial n}(cnp_0(n)). \quad (3)$$

To appreciate the important role of cell division in redistribution of proteins, consider a simple biologically relevant situation where a group of cells, spanning a broad range of receptor states ($0 \leq n \leq n_{max}$), produces no protein $\nu(n) \simeq 0$. This could be a result of kinetic reasons such as deficiency in some transcriptional factor or, as discussed below, a lack of a DNA loop necessary for the transcription initiation of *LamB*[11]. Although these cells produce no protein, they are genetically and physiologically identical to their cousins in the majority population and should be capable of producing a large number of proteins if this kinetic barrier is eliminated. If this protein production is rare and temporary, we would have a simple fire-and-divide scenario in which a large number of *LamB* receptors can be generated in a short burst. Cell division then dilutes the protein number, filling the

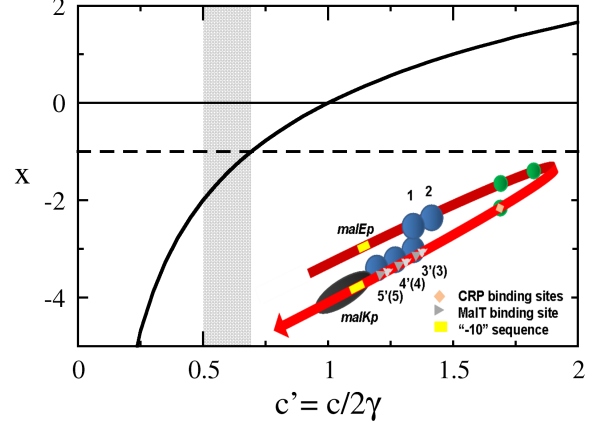


FIG. 3: The solutions to the equation $2^{x+1} = 1 + c'(x+1)$ are given by the dashed line ($x = -1$) and the solid curve. The observed exponent $-2 < \alpha < -1$ for the minority cells grown in maltose is within the shaded area, $0.5 < c' < 0.7$. The inset depicts the DNA loop for the transcription initiation of *lamB* gene. The red ribbon represents the regulatory region of *malKp*, and the big arrow is the direction of transcription when *malKp* is active. The downstream genes include *lamB* and other two proteins. The small green spheres represent CRPs that cause the DNA to bend, the large spheres are MalTs that are bound to sites 1-2 and 3'-5' as shown, and the spheroid represents the RNA polymerase. Without the formation of the loop, the three MalTs adjacent to *malKp* will bind to alternative sites 3-5 that are 3 bp upstream from the 3'-5' sites[11]. In this case, there is no contact between MalT and RNA polymerase, and *malKp* is inactive.

low n states. We can account for this mechanism by adding a term $g(n)p_0(0)$ to the right hand side of Eq. 3, where $g(n)p_0(0)$ represents a burst of proteins from the $n \simeq 0$ state. We assume that $g(n)$ is a continuous function peaked at n_{max} , which defines the upper bound of the minority population. It follows that within $0 < n < n_{max}$, the steady-state solution is $p_0(n) \sim n^{-1}$, which is consistent with some of our observations in the induced minority populations displayed in Figs. 1 and 2.

We show below that a finite protein production rate ($c \neq 0$) can significantly alter the protein PDF, but the n^{-1} distribution remains as a solution of Eq. 3. Note that if $p_0(n)$ is an homogeneous function of x degree, $p_0(n) \sim n^x$, then so is $\frac{\partial}{\partial n}(cnp_0(n))$. Substituting this power-law form into Eq. 3 gives the transcendental equation $2^{x+1} = 1 + c'(x+1)$, where $c' = c/2\gamma$ is the reduced protein production rate. This equation has two roots, one being $x = -1$ independent of c' , and the other root

being a monotonically increasing function of c' , as shown in Fig. 3. We find that the second root x is less than -1 for small protein production rate $c' < \ln 2$, indicating that $n = 0$ is an accumulation point. For $c' > \ln 2$, $x > -1$ with x becoming positive for $c' > 1$. In this case, the cumulative probability $\int_0^n p_0(n') dn'$ is infinite for large n . We interpret this to mean that the cells leave the minority state for $c' > \ln 2$. Fig. 3 also shows that x is a steep function of c' for small c' . This may be why the exponent α varies from run to run. For instance, for the maltose-grown cells, the variations seen in α ($-2 < x < -1$) only corresponds to a relatively small change in c' , i.e., $0.5 < c' < 0.7$.

It is unclear at present what mechanism gives rise to the segregation of cells into the majority and the minority populations. One possibility is due to the DNA loop required to initiate the transcription of *lamB* (see Fig. 3), which is under control of the *malKp* promoter[12]. Previous genetic studies have shown that the binding of an RNA polymerase to *malKp* requires the repositioning of three MalTs, the primary transcriptional activators, from a set of non-productive sites (3-5) to a set of productive sites (3'-5') that is staggered by 3 bp[11]. The repositioning requires the formation of a loop involving two additional MalTs at sites ~ 200 bp apart from the first three, and the binding of CRPs (cAMP receptor protein) to three sites located in the intervening region of the DNA as delineated in Fig. 3. Unlike *lac* and *ara* operons, where the DNA loop enhances repression, here it is essential for the initiation of *lamB* transcription. If the bacteria in the minority population cannot form the loop due to a deficiency either in CRP or MalT proteins, LamB will be transcribed at a very low rate[11]. Since the formation of a loop is a statistical event subject to thermal fluctuations, the ratio of the minority to the majority populations may reflect the energy difference between the open and the closed conformations. For a large population, the statistical nature of the process ensures that the ratio between the two populations will be constant as seen in our experiment[13]. Early experiments also showed that MalT expression is limited both in the

transcription and in the translation[9], indicating that maltose regulon has evolved in a way that enhances the expression noise in the population. This is in contrast with most regulatory proteins for which the system is typically constructed in a way to minimize the protein-number fluctuations. LamB, however, is exploited not only by λ but also by several other bacteriophage such as K10 and TP1. The noisy expression pattern makes biological sense because the phenotype can help the bacterial population cope with environmental fluctuations. In the uninduced state, the repressed cells still have a broad distribution of receptors, making them ready for full induction in the case that maltose or maltodextrins become available. In the fully induced case, the majority of cells can prosper in the environment but a small minority is protected from potential phage infections. This scenario is consistent with our recent observations that the heterogeneous bacterial population is more fit in an environment where λ phage is present[7].

In conclusion, we found that when *lamB* gene is fully induced in a bacterial culture, the cells segregate into two populations with very different LamB receptors numbers per bacterium. The majority population, which consists of more than 98% of bacteria, has a mean receptor number ~ 500 but the minority population has only a few receptors on average. We believe that this novel phenotype is selected for because the bacterium coevolves with viruses that exploit its receptor. We propose that the controlling mechanism is the DNA loop structure for transcriptional initiation of *lamB* gene. A noticeable feature of the minority population is that the number of proteins obeys a power-law distribution $p(n) \propto n^{-\alpha}$ with α varying from run to run in induced cell cultures. Using a simple model, we showed that the power-law scaling for $p_0(n)$ can be accounted for by a low protein production rate and a back-cascade process due to cell division. Since the model relies on very general features of protein production and cell division, we believe that our result is applicable to other proteins in the minority population.

We would like to thank C.C. Chen and D. Jasnow for useful discussions.

-
- [1] M. Pascual and F. Guinchard, Trends in Ecology and Evolution **20**, 88 (2005).
 - [2] I. Volkov, J. Banavar, S. Hubbell, and A. Maritan, Nature **424**, 1035 (2003).
 - [3] A. McKane, D. Alonso, and R. Sole, Phys. Rev. E **62**, 8466 (2000).
 - [4] E. Chapman-McQuiston, Ph.d., University of Pittsburgh (2007).
 - [5] C. Furusawa, T. Suzuki, A. Kashiwagi, T. Yomo, and K. Kaneko, arXiv:q-bio.MN/0503040 v1 (2005).
 - [6] S. Krishna, B. Banerjee, T. Ranmakrishnan, and G. Shivashankar, Proc. Nat. Acad. Sci. USA **102**, 4771 (2005).
 - [7] E. Chapman-McQuiston and X. Wu, Biophys. J. **94**, 4525 and 4537 (2008).
 - [8] M. Schwartz, *Virus Receptors (Receptors and Recognition, Series B, Vol. 7)* (Chapman and Hall, London, 1980).
 - [9] C. Chapon, The EMBO Journal **1**, 369 (1982).
 - [10] D. Ramkrishna, *Population Balances: Theory and Applications to Particulate Systems in Engineering* (Academic Press, San Diego, CA, 2000).
 - [11] E. Richet and L. Sogard-Andersen, EMBO J. **13**, 4558 (1994).
 - [12] W. Boos and H. Shuman, Microbiol. Mol. Biol. Rev. **62**, 204 (1998).
 - [13] J. Vilar and S. Leibler, JMB **331**, 981 (2003).