# Binary and graded evolution in time in a simple model of gene induction

Marcin Tabaka[1] and Robert Hołyst[1,2,*]

[1]*Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka 44/52, 01-224 Warsaw, Poland*

[2]*Department of Mathematics and Natural Sciences, College of Science, Cardinal Stefan Wyszyński University, Dewajtis 5, 01-815 Warsaw, Poland*

(Received 19 February 2010; published 8 November 2010)

We solve analytically the model of gene expression induction which consists of three steps: gene activation, gene products synthesis, and product degradation. The solution is given as a time-dependent probability distribution for gene products. Following the distribution in time from the inactive state of the gene to the stationary state we observe binary or graded response depending solely on the ratio $r$ of the gene activation rate to the rate of the gene product degradation. If $r \ll 1$ the response is binary and the continuous transition from binary to graded response occurs between $r=0.1$ and $r=1$. Therefore, if binary response is observed during relaxation to steady state, then the activation rate constant must be smaller than the degradation rate constant.

DOI: 10.1103/PhysRevE.82.052902 PACS number(s): 87.10.Mn, 87.18.Tt

Single-cell measurements of gene expression (inherently random) have resulted in establishing experimentally sources, regulation, and function of noise in gene product syntheses in both prokaryotic [1–3] and eukaryotic [4–7] organisms. Moreover, our understanding of nature of stochasticity is strongly supported by theoretical studies [8–15]. Intriguing effect of the stochasticity in the gene expression on the phenotypic variability arises upon an induction of gene expression. A response of cells can be graded or binary [16–20]. For graded response (homogeneous and rheostatic) the distribution of gene products alters smoothly to higher production levels during induction. The binary response (heterogeneous, stochastic, and all or none) is manifested by the bimodal distribution of products with two characteristic peaks: one for base level of gene expression and another one for stationary level of gene expression.

Phenotypic diversification, as an effect of noisy gene expression, plays a key role in the survival of cell population in a fluctuating environment [21,22]. Cell fraction of maximal gene product level is larger for binary response than for graded response. Therefore, for high stress threshold the binary response is advantageous, because instantly after stress signal appearance there is a subpopulation adapted to new conditions. Graded response, on the contrary, is favored for low stress thresholds.

We present analytically tractable model for the induction of gene expression. We construct a master equation describing the evolution of gene products for which we find time-dependent probability distributions. We show that the gene products, after induction, reach their stationary state in a graded or binary evolution. The type of the response depends on the relationship between degradation rates of gene products and rates of gene activation. The binary response does not require any additional mechanisms and can appear even for the most simplified model of the gene expression induction.

Recently, Iyer-Biswas *et al.* [15] derived a time-dependent generating function for probability distribution of gene products, but did not consider the problem of binary and graded evolution in time. Our Brief Report is complementary to theirs. Additionally, we find directly the probability distribution of gene product numbers in time. Peccoud and Ycart [23] considered the problem of the induction of gene expression analytically. They considered the possibility of the consecutive transitions from active to inactive gene state, but in the analytical solution they neglected the degradation of gene products and did not analyze the possible response patterns. Shahrezaei and Swain [14] derived time-dependent distributions of the proteins and messenger RNA (mRNA) number for the two-stage model (synthesis of mRNA and proteins) in which the gene is always active. In such a case the response is always graded in the mRNA and protein distributions. The emergence of bimodality in steady-state gene product distributions for previously analyzed models [14,15,23] is due to consecutive silencing of genes after activation. Here, the distribution of gene products is always unimodal in the steady state, but the response of the system after gene induction exhibits two modes of response (graded and binary) upon approach to the steady state.

The model consists of the following steps:

$$I \xrightarrow{k_a} A \circlearrowright \xrightarrow{k_s} P \xrightarrow{k_d} \emptyset, \tag{1}$$

where $I$ and $A$ are the inactive and active forms of gene, respectively. $P$ denotes gene products. A circle arrow $\circlearrowright$ is used to emphasize that gene product synthesis takes place continuously after transition to the active form of gene. The active form is not blocked in the process of product synthesis. Gene activation, gene product synthesis, and degradation are modeled as first-order processes with probabilities $k_a$, $k_s$, and $k_d$ per unit time, respectively. The activation of gene is irreversible in this model, i.e., the transition $A \rightarrow I$ is forbidden. Such a unidirectional transition is a natural choice for the induction of gene expression.

At any instant of time, the system, described by Eq. (1), is in the state specified by the ordered pair $(o,p)$. $o$ and $p$ are

*Corresponding author; holyst@ichf.edu.pl

 

copy numbers of the active gene and gene products, respectively. The scheme of reactions in Eq. (1) imposes the following transitions between states and their probabilities per unit time:

$$(o,p) \xrightarrow{k_a(N-o)} (o+1,p),$$

$$(o,p) \xrightarrow{k_s o} (o,p+1),$$

$$(o,p) \xrightarrow{k_d p} (o,p-1). \tag{2}$$

Since the probabilities of transitions are constant per unit time, the waiting time for transitions is exponentially distributed, and the system satisfies the Markov property [24]. The probability that the system is in the state $(o,p)$ at time $t$, denoted by $f_{o,p}(t)$, satisfies the master equation

$$\frac{df_{o,p}(t)}{dt} = -f_{o,p}(t)[k_a(N-o)+k_s o+k_d p]+f_{o-1,p}(t)\{k_a[N-(o-1)]\}+f_{o,p-1}(t)(k_s o)+f_{o,p+1}(t)[k_d(p+1)], \tag{3}$$

where $N$ is the copy number of the genes and $f_{o-1,p}(t)=0$ for $o=0$. Division of both sites of Eq. (3) by $k_d$ simplify significantly the notation. Therefore, the following analysis is carried out in terms of the parameters

$$K=\frac{k_s}{k_d}, \quad r=\frac{k_a}{k_d}, \quad \tau=k_d t. \tag{4}$$

The variability of the gene products within cell or cell population is through moments of probability distributions. We follow also this approach. The probability distribution of gene products is equal to the Poisson distribution in the stationary state. The Poisson distribution has its variance equal to the mean; therefore, a proper measure of variability is given by the dispersion index. The dispersion index is equal to the time-dependent variance divided by the time-dependent mean:

$$D(\tau)=\frac{\langle p^2\rangle(\tau)-\langle p\rangle^2(\tau)}{\langle p\rangle(\tau)}. \tag{5}$$

We calculate the dispersion index by the following method. Equation (3) is a linear difference-differential equation. Therefore, the moments can be calculated strictly by multiplying Eq. (3) by $o$, $p$, $o^2$, $p^2$, and $op$ and summing over all values of $o$ and $p$:

$$\sum_o \sum_p f_{o,p}(\tau) o^x p^y = \langle o^x p^y\rangle(\tau), \tag{6}$$

because the hierarchy of equations is closed. The mean level of gene products is given by

$$\langle p\rangle(\tau)=\begin{cases} \dfrac{KN[(e^{-r\tau}-1)+r(1-e^{-\tau})]}{(r-1)} & \text{if } r\neq 1 \\ KN[1-e^{-\tau}(1+\tau)] & \text{if } r=1, \end{cases} \tag{7}$$

and the dispersion index for gene products can be presented as follows:

$$D(\tau)=1+Kh(r,\tau), \tag{8}$$

where

$$h(r,\tau)=\begin{cases} \dfrac{2e^{-\tau}\left\{e^{-\tau}\left[1+\tau\left(1+\frac{1}{2}\tau\right)\right]-1\right\}}{e^{-\tau}(1+\tau)-1} & \text{if } r=1 \\ \dfrac{e^{-2\tau}(3-2\tau+e^{-2\tau}-4e^{-\tau})}{(2e^{-\tau}-e^{-2\tau}-1)} & \text{if } r=2 \\ \dfrac{e^{-r\tau}[2(r-1)^2+(r-2)(e^{-r\tau}-2re^{-\tau})]-re^{-2\tau}}{(r-1)(r-2)[r(e^{-\tau}-1)+(1-e^{-r\tau})]} & \text{otherwise.} \end{cases} \tag{9}$$

$D(\tau)$ is a linear function of $K$, and $D(\infty)=1$ since $\lim_{\tau\to\infty} h(r,\tau)=0$.

We find solution of Eq. (3) using the following procedure. We assume that $N=1$, i.e., there is one copy of the gene in the system, and at $\tau=0$ the gene is inactive. The probability density function of times required for activation of the gene is exponential and given by $g(\tau)=re^{-r\tau}$. Having the solution of the master equation for the probability that there are $p$ gene products at time $\tau$ [denoted by $f'_p(\tau)$] for the system in which the gene is always active, the solution of Eq. (3) is equivalent to the time convolution

$$f_{1,p}(\tau)=(g*f'_p)(\tau)=\int_0^\tau g(\tau')f'_p(\tau-\tau')d\tau', \tag{10}$$

where $f'_p(\tau)$ satisfies the following master equation (for the system of continuous expression of gene):

$$\frac{df'_p(\tau)}{d\tau}=f'_{p-1}(\tau)K+f'_{p+1}(\tau)(p+1)-f'_p(\tau)(K+p). \tag{11}$$

The assumption that at time $\tau=0$ there are no gene products in the system leads to the following distribution:

$$f'_p(\tau) = \frac{1}{p!} e^{K(e^{-\tau}-1)}[K(1-e^{-\tau})]^p. \quad (12)$$

Equation (12) is the Poisson distribution, for which the mean is equal to $K(1-e^{-\tau})$. By making the substitution $x=e^{-(\tau-\tau')}$, Eq. (10) becomes

$$f_{1,p}(\tau) = \frac{K^p}{p!} e^{-K} r e^{-r\tau} I_p, \quad (13)$$

where $I_p = \int_{e^{-\tau}}^{1} x^{-r-1} e^{Kx} (1-x)^p dx$. Using the binomial theorem for $(1-x)^p$, the integral $I_p$ can be written in terms of the generalized incomplete gamma function [25]

$$I_p = \sum_{j=0}^{p} \binom{p}{j} (-1)^j (-K)^{r-j} [\Gamma(j-r, -Ke^{-\tau}, -K)]. \quad (14)$$

Total probability, denoted by $f_p$, that there are $p$ gene products present and the gene is in active or inactive state is equal to

$$f_p(\tau) = f_{0,p} + f_{1,p}(\tau). \quad (15)$$

The first term in Eq. (15) is nonzero only for $p=0$. From Eq. (3) and under the assumption that the gene is in the inactive state at $\tau=0$, we get $f_{0,p}(\tau)=\delta_p e^{-r\tau}$, where $\delta_p$ is 1 for $p=0$ and zero otherwise.

When the gene is always active, the probability distribution of gene products [Eq. (12)] is given by the Poisson distribution. Therefore, the dispersion index does not change with time and the mode of the response is graded. On the other hand the gene being inactive at the start of the induction can produce both modes of the response. Using previously derived probability distribution [Eq. (15)] we characterize the modes of gene expression induction. The graded and binary responses in time depend solely on $r$, i.e., on the ratio of the gene activation rate constant to the gene product degradation or deactivation rate constant. In fact, in Eq. (13) the characteristic time does not depend on the constant $K$. Therefore, the shape of the probability distribution functions is the same for various values of $K$. We will show the range of $r$ where each response is dominant.

In order to pinpoint the graded versus binary response we plot the probability distributions for gene product numbers for various values of the ratio $r=k_a/k_d$ [see Fig. 1(a)] and few values of time during the gene induction. In Fig. 1(a) we show the changes in the dispersion index as a function of time $\tau$ (vertical axis) and ratio $r$ (horizontal axis). We show also contours corresponding to 0.1 and 0.9 of the steady-state level of the gene products. These two contours superimposed on the dispersion index show where actually the changes in gene product numbers are significant. For cases (c) and (e) the gene activation rates are higher than the gene product deactivation rates [$r>1$; Figs. 1(c) and 1(e)], which results in the graded response (a single peak in the distribution which moves from zero toward the peak at the steady state with almost constant dispersion index). For cases (d) and (f) [$r<1$; Figs. 1(d) and 1(f)] the response is undoubtedly bi-

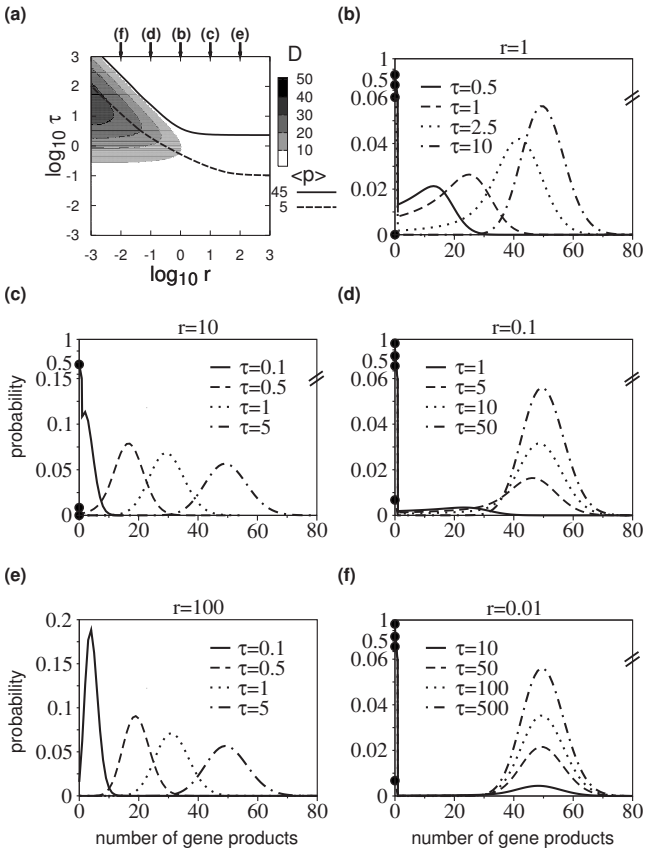


FIG. 1. The distribution of gene product numbers $f_p(\tau)$ during the induction of gene expression within cell population. The probability distributions pinpoint to the (c),(e) graded; (d),(f) binary; and (b) intermediate responses. The steady-state level of gene products is fixed at $K=k_s/k_d=50$ and $N=1$. Initially the cells do not have any gene products and the gene is inactive. The activation of gene starts at time $\tau=0$ for each cell. (a) Changes in the dispersion index in time $\tau$ as a function of the ratio $r=k_a/k_d$. The continuous lines depict the mean level of gene products $\langle p\rangle$ equal to 0.1 and 0.9 of the steady-state level. The letters above picture correspond to the values of $r$ at which the probability distributions of gene product numbers are calculated and presented in (b)–(f). (b)–(f) Probability distributions of gene product numbers at different times after start of the activation of gene expression. By means of filled circles we mark the probabilities that no gene products are present.

nary. There are two peaks in the probability distribution: one at $p=0$ and another one at the $p$ value corresponding to the steady state. The bimodal distribution in the binary response leads to a very large dispersion index. In the binary response the peak at $p=0$ decreases in the course of time, and the second peak grows until the steady state is reached. In the steady state all distributions are unimodal. The behavior of the system, during induction for case (b) [i.e., $r=1$; Fig. 1(b)], reveals the intermediate response, resembling the graded response. The characteristic feature of the response in this case is the long-tailed distributions with quite high probabilities of intermediate gene product levels within cell population. The similar long-tailed distributions are present also for $r<1$ at the beginning of the induction [e.g., continuous line in Fig. 1(d)]. The most striking feature revealed by our analysis is that the transition between the graded and

binary responses is quite sharp and takes place between $r=1$ and $r=0.1$. It puts a very strong constraint on the more complex models of the gene induction.

For example, Pirone and Elston [19] considered a biologically realistic model of the gene expression regulation for the cytokine interleukin-2 with the *lacZ* as a reporter gene. They showed in computer simulations that the response in that system changes from the binary to graded upon increase in the transcription factor binding constant. The system investigated by Pirone and Elston [19] revealed the binary response (see Figs. 7(A)–7(C) in [19]). We calculated the factor $r$ for their system and got $r=\{0.2,0.4,0.5\}$ for three studied inducer concentrations, i.e., the range of $r$ recognized in our analysis as giving the binary response. They noticed the graded pattern (see Fig. 9 in [19]) for the conditions which correspond, as we have checked, to $r=4$.

Kierzek *et al.* [20] studied by computer simulations a response mode of the multiparameter model of a two-component system. This system is comprised of a membrane-bound histidine kinase and a partner response regulator that positively autoregulates its own synthesis and is an activator for a reporter gene. It appeared that such a system exhibits binary response for the reporter. However, modification abolishing autoregulation results in graded response. Positive feedback loop extends the regulator response time [26], i.e., the delay from an activation of histidine kinase until the regulator concentration approaches the activation threshold for the reporter promoter. It explains the observed change in the response mode since the effective activation rate and thus $r$ decrease for a positively autoregulated two-component system.

The characteristic time scales for gene activation differ significantly between prokaryotic and eukaryotic organisms. In eukaryotes, promoter activation requires remodeling of the chromatin structure which can take hours or longer [27]. Moreover, the induction is highly sophisticated and requires integration of many signals and correlated activations of many transcriptional modules [28]. The mean half-life of proteins and mRNAs determined in *S. cerevisiae* is equal to 43 min [29] and 23 min [30], respectively. Since $r \ll 1$, a binary pattern of gene induction is widely observed in eukaryotic organisms [16,18,31]. In prokaryotes, the promoters are easily accessible for transcription factors, and thus the activation of genes usually is on time scales of seconds or minutes [32–34]. Therefore, the presence of the graded induction patterns among prokaryotic cells should be common since such a case corresponds to $r>1$.

This project was operated within the Foundation for Polish Science Team Programme co-financed by the EU European Regional Development Fund Grant No. TEAM/2008-2/2 and also by the Human Frontiers Science Program Organization. R.H. acknowledges the "Mistrz" grant from FNP.

[1] A. Becskei and L. Serrano, Nature (London) **405**, 590 (2000).
[2] E. Ozbudak, M. Thattai, I. Kurtser, A. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002).
[3] M. Elowitz, A. Levine, E. Siggia, and P. Swain, Science **297**, 1183 (2002).
[4] W. Blake, M. Kaern, C. Cantor, and J. Collins, Nature (London) **422**, 633 (2003).
[5] J. Raser and E. O'Shea, Science **304**, 1811 (2004).
[6] D. Volfson, J. Marciniak, W. Blake, N. Ostroff, L. Tsimring, and J. Hasty, Nature (London) **439**, 861 (2006).
[7] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, Nat. Genet. **38**, 636 (2006).
[8] H. McAdams and A. Arkin, Proc. Natl. Acad. Sci. U.S.A. **94**, 814 (1997).
[9] T. Kepler and T. Elston, Biophys. J. **81**, 3116 (2001).
[10] M. Thattai and A. van Oudenaarden, Proc. Natl. Acad. Sci. U.S.A. **98**, 8614 (2001).
[11] P. Swain, M. Elowitz, and E. Siggia, Proc. Natl. Acad. Sci. U.S.A. **99**, 12795 (2002).
[12] J. Paulsson, Nature (London) **427**, 415 (2004).
[13] N. Friedman, L. Cai, and X. S. Xie, Phys. Rev. Lett. **97**, 168302 (2006).
[14] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. U.S.A. **105**, 17256 (2008).
[15] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, Phys. Rev. E **79**, 031911 (2009).
[16] S. Fiering, J. P. Northrop, G. P. Nolan, P. S. Mattila, G. R. Crabtree, and L. A. Herzenberg, Genes Dev. **4**, 1823 (1990).
[17] M. S. Ko, J. Theor. Biol. **153**, 181 (1991).
[18] M. Louis and A. Becskei, Sci. STKE **2002**, pe33 (2002).
[19] J. Pirone and T. Elston, J. Theor. Biol. **226**, 111 (2004).
[20] A. Kierzek, L. Zhou, and B. Wanner, Mol. Biosyst. **6**, 531 (2010).
[21] M. Acar, J. T. Mettetal, and A. van Oudenaarden, Nat. Genet. **40**, 471 (2008).
[22] D. Fraser and M. Kaern, Mol. Microbiol. **71**, 1333 (2009).
[23] J. Peccoud and B. Ycart, Theor Popul. Biol. **48**, 222 (1995).
[24] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier Science & Technology, Amsterdam, 2007).
[25] http://functions.wolfram.com/GammaBetaErf/Gamma3/
[26] U. Alon, Nat. Rev. Genet. **8**, 450 (2007).
[27] W. Muller, D. Walker, G. Hager, and J. McNally, J. Cell Biol. **154**, 33 (2001).
[28] B. Lemon and R. Tjian, Genes Dev. **14**, 2551 (2000).
[29] A. Belle, A. Tanay, L. Bitincka, R. Shamir, and E. K. O'Shea, Proc. Natl. Acad. Sci. U.S.A. **103**, 13004 (2006).
[30] Y. Wang, C. Liu, J. Storey, R. Tibshirani, D. Herschlag, and P. Brown, Proc. Natl. Acad. Sci. U.S.A. **99**, 5860 (2002).
[31] S. Fiering, E. Whitelaw, and D. Martin, BioEssays **22**, 381 (2000).
[32] J. Elf, G.-W. Li, and X. S. Xie, Science **316**, 1191 (2007).
[33] J. A. Megerle, G. Fritz, U. Gerland, K. Jung, and J. O. Rädler, Biophys. J. **95**, 2103 (2008).
[34] M. Tabaka, O. Cybulski, and R. Hołyst, J. Mol. Biol. **377**, 1002 (2008).