# Graded and Binary Responses in Stochastic Gene Expression

Rajesh Karmakar and Indrani Bose*
Department of Physics
Bose Institute
93/1, Acharya Prafulla Chandra Road, Kolkata-700 009, India.
* Author to be contacted; e-mail: indrani@bosemain.boseinst.ac.in

**Abstract**
Recently, several theoretical and experimental studies have been undertaken to probe the effect of stochasticity on gene expression (GE). In experiments, the GE response to an inducing signal in a cell, measured by the amount of $mRNAs/proteins$ synthesized, is found to be either graded or binary. The latter type of response gives rise to a bimodal distribution in protein levels in an ensemble of cells. One possible origin of binary response is cellular bistability achieved through positive feedback or autoregulation. In this paper, we study a simple, stochastic model of GE and show that the origin of binary response lies exclusively in stochasticity. The transitions between the active and inactive states of the gene are random in nature. Graded and binary responses occur in the model depending on the relative stability of the activated and deactivated gene states with respect to that of $mRNAs/proteins$. The theoretical results on binary response provide a good description of the "all-or-none" phenomenon observed in an eukaryotic system.

**Keywords**: gene expression, graded and binary responses, stochastic binary response, "all-or-none" phenomenon, probability density, activation

## 1. Introduction

Gene expression (GE) is the central activity in a living cell. The two major steps in GE, transcription and translation, involve several biochemical reactions. The time evolution of this system of reactions or events is not a continuous process as molecular population levels in a reacting system change only by discrete amounts. Furthermore, the time evolution is not deterministic as the biochemical events underlying GE are probabilistic in nature, i.e., the timing of the biochemical events cannot be predicted with certainty. For example, the binding/unbinding of $RNA$ polymerase ($RNAP$) at the promoter region of $DNA$ and that of regulatory molecules at the operator regions are probabilistic processes. The discrete, probabilistic nature of the

biochemical events may be ignored in the limit of large numbers of participating biochemical molecules. In this case, the biochemical reactions/events occur at much higher frequencies and fluctuations around the mean levels of biomolecules participating in GE are small. Thus, the time evolution of the system of reactions may be treated to be continuous and deterministic as in the traditional differential rate equation approach. In a living cell, the number of biomolecules involved in GE is often small so that a stochastic rather than deterministic description provides the more correct picture. In recent years, there is a growing realization that stochasticity plays an important role in determining the outcome of biochemical processes in the cell [1, 2]. Stochastic effects in GE explain the pronounced cell-cell variation observed in isogenic populations. A cell may have the option of proceeding along one of two possible developmental pathways. The pathway selection is probabilistic and the cell fate depends on the particular choice of pathway. Thus, even a clonal population of cells can give rise to two distinct subpopulations in the course of time. The randomization of pathway choice leads to diversity and increases the likelihood of survival of organisms in widely different environments. A well-known example of the two way-choice is that of lysis-lysogeny in *E. Coli* [3].

The effect of stochasticity (randomness/noise) is prominent at the level of an individual cell and can be masked due to ensemble averaging in a population of cells. Single cell experiments provide evidence that GE in a cell occurs in abrupt stochastic bursts [4, 5, 6]. In more recent experiments, a quantitative measure of noise associated with GE has been obtained in both prokaryotic as well as eukaryotic cells [7, 8, 9]. A large number of theoretical studies address the origin and consequences of stochasticity in GE [3, 10, 11, 12, 13, 14, 15, 16, 17, 18]. Thus, the notion of stochasticity in GE is well established both theoretically and experimentally.

Regulation of GE in a cell is achieved in a manifold of ways which increase in complexity from prokaryotic to eukaryotic cells. In the prokaryotic systems, regulation is achieved by the binding of regulatory molecules (repressors or activators) to the operator regions of $DNA$. In eukaryotes, the activator molecules are known as transcription factors (TFs). Intra- and extra- cellular inducing signals activate the TFs which then bind to appropriate enhancer sequences on the $DNA$. The GE response to an inducing signal in an individual cell may be graded or binary. Response is quantified by the amount of $mRNAs/proteins$ synthesized. In graded response, the output varies continuously as the amount of input stimulus is varied till the

steady state is reached. In binary response, alternatively termed the "all-or-none" phenomenon, the output has a binary character, i.e., GE occurs at either a low or a high level and expression at intermediate levels is rare. This gives rise to a bimodal distribution in protein levels in an ensemble of cells. Several experiments on both prokaryotic and eukaryotic cells establish the binary character of GE [6, 9, 19, 20, 21, 22, 23]. Binary response may be ascribed to bistability which implies existence of two stable steady states with low and high protein levels of GE. One way of achieving bistability is through positive feedback or autocatalysis in which the protein product of GE promotes further GE either directly or via intermediates. The *lac* operon in *E. Coli* is an example of a model system in which autocatalytic induction gives rise to the "all-or-none" phenomenon in GE [19, 23, 24, 25, 26]. Beckskei et al. [23] have demonstrated that positive feedback can generate binary response in a synthetic eukaryotic gene circuit. In eukaryotic transcription, enhancers activate the usually weak eukaryotic promoters. There is now strong experimental evidence that in some systems enhancers do not affect transcription rate but rather increase the probability of a gene synthesizing proteins at a high level [27, 28, 29, 30]. In a population of cells, enhancers increase the number of cells expressing at a high level but not the level of expression per cell.

Binary response in GE can have a purely stochastic origin. Kepler and Elston [10] provide examples of stochastic binary response (SBR), i.e., binary response induced by noise. A simple model of SBR shows a binary distribution of $mRNA$ levels in an ensemble of cells [31]. A recent model of eukaryotic GE suggests that fluctuations in the binding of TFs to $DNA$ can explain graded and binary responses observed in inducible GE [32]. Fast chemical kinetics is responsible for a graded response whereas slow kinetics leads to a binary output. The "all-or-none" phenomenon observed in some eukaryotic systems does not involve positive feedback processes explicitly [4, 5, 6]. On the other hand, protein synthesis in these systems occurs in stochastic bursts. Since the effect of stochasticity is prominent in these systems, it is reasonable to conjecture that the "all-or-none" phenomenon (binary response) observed in these systems is a manifestation of stochasticity. In this paper, we consider a simple, stochastic model of GE studied earlier [13, 17, 33]. We show that graded and binary responses occur naturally in the model depending on the relative stability of activated and deactivated gene states with respect to that of $mRNAs/proteins$. Binary response, obtained in the model, arises solely due to stochasticity and not due to positive feedback processes. We

further show that our model gives a good description of the "all-or-none" phenomenon observed in an eukaryotic system [6].

## 2. Stochastic model of GE

In the minimal model of GE, a gene can be in two possible states: inactive ($G$) and active ($G^*$). Random transitions occur between the states $G$ and $G^*$ according to the first order kinetics

$$G \underset{k_d}{\overset{k_a}{\rightleftharpoons}} G^\star \overset{j_p}{\longrightarrow} p \overset{k_p}{\longrightarrow} \Phi \tag{1}$$

where $k_a$ and $k_d$ are the activation and deactivation rate constants. In the active state $G^*$, transcription is initiated followed by translation and protein synthesis. The separate processes are combined into a single step of protein ($p$) synthesis with the rate constant $j_p$. The protein degrades with the rate constant $k_p$ and the degradation product is represented as $\Phi$. If cell division is taken into account, the protein decay rate has two components, one the degradation rate and the other the dilution rate of proteins due to cell growth and division. In this case, $k_p$ denotes the rate constant for protein decay.

In inducible GE systems, the activation of a gene is brought about by an activator $S$, say, $TFs$. The reaction scheme in the presence of $S$ is given by

$$G + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} G\_S \underset{k_d}{\overset{k_a}{\rightleftharpoons}} G^\star \overset{j_p}{\longrightarrow} p \overset{k_p}{\longrightarrow} \Phi \tag{2}$$

where $G\_S$ represents the bound complex of $G$ and $S$. The reaction scheme in equation (2) can be generalized by including direct transitions between $G$ and $G^*$. The rate constants for transitions from $G$ to $G^*$ and $G^*$ to $G$ are $k_{on}$ and $k_{off}$ respectively. For eukaryotic systems, the rate constant $k_{on}$ has a very low value as activating TFs, $S$, are required in most cases for transistion to the active state $G^*$. The reaction scheme is given by

$$G + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} G\_S \underset{k_d}{\overset{k_a}{\rightleftharpoons}} G^*, \qquad G \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} G^*$$

$$G^* \xrightarrow{j_p} p \xrightarrow{k_p} \Phi \tag{3}$$

For eukaryotic systems, the initiation of transcription by *RNA polymerase* II generally requires a prior assembly of TFs on the enhancer regions of the target gene. This state of the gene is represented by $G\_S$ in the reaction scheme 3. The activating TFs, $S$, facilitate the formation of the transcription initiation complex which is bound to the promoter region of $DNA$ and consists of general TFs, other factors and *RNA polymerase* II. The gene is now in the active state $G^*$ and $RNAP$ starts transcription after disengaging itself from the initiation complex through the key step of phosphorylation. The general TFs are then released allowing for the initiation of a new round of transcription with another $RNAP$ molecule. In the simple reaction scheme 3, this is respresented by a return after transcription initiation to the intermediate complex $G\_S$ and subsequent return to the active state $G^*$.

If $n_G$ be the total concentration of genes then

$$n_G = [G] + [G\_S] + [G^*] \tag{4}$$

where $[G]$, $[G\_S]$, and $[G^*]$ denote the concentrations of genes in the states $G$, $G\_S$, and $G^*$ respectively. Using the method of King and Altman [34], the fractions of genes in the inactive, intermediate and active states are given by

$$\begin{aligned}
\frac{[G]}{n_G} &= \frac{k_a\, k_{off} + k_2\, k_{off} + k_d\, k_2}{A} \\
\frac{[G_S]}{n_G} &= \frac{k_d\, k_1\, S + k_{on}\, k_d + k_{off}\, k_1\, S}{A} \\
\frac{[G^*]}{n_G} &= \frac{k_a\, k_1\, S + k_a\, k_{on} + k_2\, k_{on}}{A}
\end{aligned} \tag{5}$$

respectively, where

$$A = k_a k_1 S + k_a k_{on} + k_2 k_{on} + k_d k_1 S + k_{on} k_d + k_{off} k_1 S + k_a k_{off} + k_2 k_{off} + k_d k_2 \tag{6}$$

From equation (5), one can further write

$$[G^*] = \frac{\frac{n_G\, k_a\, (\frac{S}{k_s} + \frac{1}{k}) + k_{on}}{(1 + \frac{1}{k} + \frac{S}{k_s})}}{\frac{k_a\, (\frac{S}{k_s} + \frac{1}{k}) + k_{on}}{(1 + \frac{1}{k} + \frac{S}{k_s})} + \{k_d + \frac{k_a/k' + k_{off}\, (1 + \frac{S}{k_s})}{(1 + \frac{1}{k} + \frac{S}{k_s})}\}}$$

$$= \frac{n_G\, k_a^{'}}{k_a^{'}+k_d^{'}} \tag{7}$$

where

$$k_a^{'} = \frac{k_a\,(\frac{S}{k_s}+\frac{1}{k})+k_{on}}{(1+\frac{1}{k}+\frac{S}{k_s})}; \qquad k_d^{'} = k_d + \frac{k_a/k^{'}+k_{off}\,(1+\frac{S}{k_s})}{(1+\frac{1}{k}+\frac{S}{k_s})} \tag{8}$$

Also,

$$k_s = \frac{k_2}{k_1}, \qquad k = \frac{k_2}{k_{on}} \quad and \quad k^{'} = \frac{k_2}{k_{off}} \tag{9}$$

In the reaction scheme 1, the steady state concentration of genes in the active state is given by

$$[G^*] = \frac{n_G\, k_a}{k_a + k_d} \tag{10}$$

Expressions (7) and (10) are identical in form with $k_a$ and $k_d$ replaced by $k_a^{'}$, $k_d^{'}$. The equivalence relations in equation (8) enable one to map the reaction scheme 3 onto the simpler scheme 1 while calculating various quantities. Use of the simpler reaction scheme leads to greater mathematical tractability. The half-lives of the active and inactive states of the gene in the reaction scheme 3 are given by $T_a^{'} = log2/k_a^{'}$ and $T_d^{'} = log2/k_d^{'}$ respectively. Since $k_a^{'}$ and $k_d^{'}$ are given by equation (8), the half-lives are dependent on $k_a$, $k_d$ as well as $S$, the concentration of TFs.

We now consider a simple stochastic model corresponding to reaction scheme 1. The results we derive hold true for the more complicated reaction scheme 3 but with $k_a$, $k_d$ replaced by $k_a^{'}$, $k_d^{'}$ (equation (8)). At this point, one can ask about the validity of the equivalence relations (equations (7) and (8)) in the stochastic case. Use of the relations is justified only if the fluctuations in the concentration $S$ of the activator molecules are ignored. Exact validity can be established by deriving expressions for variance from the Master Equations (treating $S$ to be constant) corresponding to the reaction schemes 1 and 3. This has been done for the simpler case $k_{on} = 0$, $k_{off} = 0$ (in the general case, these rate constants are much smaller than the activation and deactivation rate constants $k_a$ and $k_d$). The expressions for variance in the reaction schemes 1 and 3, are found to be identical with $k_a$, $k_d$ in scheme 1 replaced by $k_a^{'}$, $k_d^{'}$ in scheme 3. In the model, the only stochasticity arises from random transitions of a gene between the inactive and active states as in the minimal model of Cook et al. [13]. Protein synthesis from the active

gene and protein degradation occur in a deterministic manner. In each state of the gene, the concentration of proteins evolves deterministically according to the equation

$$\frac{dx}{dt} = \frac{j_p}{X_{max}} z \; - \; k_p \, x \; = \; f(x, z) \tag{11}$$

where $z = 1(0)$ when the gene is in the active (inactive) state and $x = \frac{X}{X_{max}}$, $X$ and $X_{max}$ being the protein concentration at time t and the maximum protein concentration respectively. The variable $x$ thus denotes protein concentration normalized by the maximum possible concentration. The latter quantity is equal to the protein concentration in the steady state if the gene is always in the active state, i.e, deactivation processes are disallowed. We note that $X_{max} = \frac{j_p}{k_p}$. Let $p_j(x,t)$ $(j = 0,\, 1)$ be the probability density function when $z = j$. The total probability density function is

$$p(x,t) = p_0(x,t) + p_1(x,t) \tag{12}$$

The rate of change of probability density is given by

$$\frac{\partial p_j(x,t)}{\partial t} = -\frac{\partial}{\partial x}[f(x,j)\, p_j(x,t)] + \sum_{k \neq j} [W_{kj}\, p_k(x,t) - W_{jk}\, p_j(x,t)] \tag{13}$$

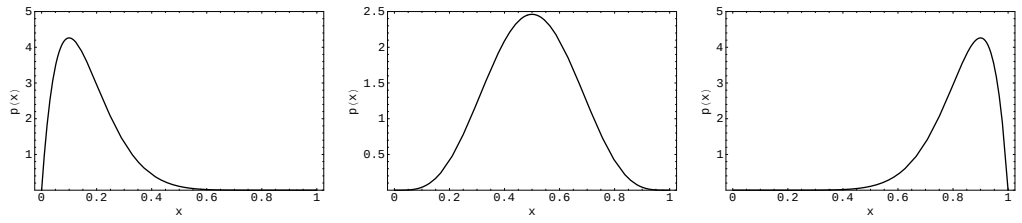
where $W_{kj}$ is the transition rate from the state $k$ to the state $j$ and $W_{jk}$ is the same for the reverse transition. The first term in equation (13) is the so called "transport" term representing the net flow of the probability density. The second term represents the gain/loss in the probability density due to random transitions between the state $j$ and other accessible states. In the present case, equation (13) gives rise to the following two equations:

$$\frac{\partial p_0(x,t)}{\partial t} = -\frac{\partial}{\partial x}(-k_p \, x \, p_0(x,t)) + k_d \, p_1(x,t) - k_a \, p_0(x,t) \tag{14}$$

$$\frac{\partial p_1(x,t)}{\partial t} = -\frac{\partial}{\partial x}\{(\frac{j_p}{X_{max}} - k_p \, x) \, p_1(x,t)\} + k_a \, p_0(x,t) - k_d \, p_1(x,t) \tag{15}$$

The steady state distribution ($\frac{\partial p_0(x,t)}{\partial t} = 0, \; \frac{\partial p_1(x,t)}{\partial t} = 0$) is given by

$$p(x) = N \, x^{(\frac{k_a}{k_p} - 1)} (1 - x)^{(\frac{k_d}{k_p} - 1)} \tag{16}$$




(a) (b) (c)

Figure 1: Plot of $p(x)$ versus $x$ (Case I, $r_1,\ r_2 > 1$): 1A ($r_2 >> r_1$), 1B ($r_2 = r_1$) and 1C ($r_1 >> r_2$) respectively.

where $N$, the normalization constant, is given by the inverse of a beta function [14]

$$N = \frac{1}{B(\frac{k_a}{k_p},\ \frac{k_d}{k_p})} \tag{17}$$

# 3. Graded and binary responses

The graded and binary responses to varying concentrations of $S$ can be understood by considering two limiting cases of the steady state distribution of protein levels $p(x)$ (equation (16)): $\frac{k_a}{k_p},\ \frac{k_d}{k_p} > 1$ (Case I) and $\frac{k_a}{k_p},\ \frac{k_d}{k_p} < 1$ (Case II). Figure 1 shows plots of $p(x)$ versus $x$ corresponding to Case I. The mean value of $x$ is given by

$$< x > = \int_0^1 x\ p(x)\ dx \tag{18}$$

$$= \frac{j_p}{k_p}\frac{k_a}{k_a + k_d} \tag{19}$$

Define $r_1 = \frac{k_a}{k_p}$ and $r_2 = \frac{k_d}{k_p}$. Figures 1A, 1B and 1C correspond to $r_2 >> r_1 > 1$, $r_1 = r_2 > 1$ and $r_1 >> r_2 > 1$ respectively. In the presence of an inducing stimulus, $k_a$ and $k_d$ are replaced by $k_a^{'}$ and $k_d^{'}$ (equation (8)). As $r_1 = \frac{k_a^{'}}{k_p}$ increases, the mean protein level increases from a lower to a higher value. The increase in $r_1$ can be brought about by increasing the concentration of $S$. Thus the mean protein level is a continuous function of

$[S]$, i.e., a graded response is obtained. Saturation level is reached when $\frac{S}{k_s}$ in equation (8) is $>> 1$ so that $k_a^{'} = k_a$ and $k_d^{'} = k_d + k_{off}$ .

Figure 2 shows plots of $p(x)$ versus $x$ corresponding to Case II, i.e., $r_1 < 1$, $r_2 < 1$. In this case $p(x)$ is peaked at a low (zero) value of $x$ ($r_1 << r_2$, figure 2A), a high value of $x$ ($r_1 >> r_2$, figure 2D) or at both low and high values of $x$ (figures 2B and 2C). Thus, in a cell GE predominantly occurs at low and/or high levels and protein production at intermediate levels is negligible. Again, in the presence of an inducing stimulus, $S$, $r_1 = \frac{k_a^{'}}{k_p}$ and $r_2 = \frac{k_d^{'}}{k_p}$ can be changed by changing the concentration $[S]$. The response in this case is not graded as the mean protein level is not a continuous function of $[S]$ but has only low and high values. Figures 2B and 2C correspond to SBR and bifurcation from a unimodal probability distribution function (figure 2A) to a bimodal one is brought about by varying $r_1$ and $r_2$. SBR gives rise to the "all-or-none" phenomenon in GE. In experiments on a population of cells, a fraction of cells is found to be in the state with low (zero) level of GE and another fraction is in the state with high level of GE. The fraction of cells in which protein synthesis occurs at intermediate levels is small. In the cases when $r_1 > 1$, $r_2 < 1$ and $r_1 < 1$, $r_2 > 1$, unimodal responses are obtained. In the first case, GE occurs at a high level and in the second case, GE occurs at a low level. The response is not graded in the presence of an inducing stimulus.

The graded and binary responses to an inducing stimulus are a natural outcome of stochastic gene activation and deactivation processes. If the gene is always in the inactive state ($z = 0$ in equation (11)), the mean protein level corresponds to $x = 0$. If the gene is in the active state ($z = 1$) and no deactivation processes are allowed, the mean protein level is given by $\frac{j_p}{k_p}$ and $x = 1$ corresponding to maximum protein synthesis. When stochastic GE is considered, i.e., random activation/deactivation processes are taken into account, two possibilities arise. If the activation and deactivation rates are faster than the protein degradation rate, an average protein level is obtained due to the accumulation of proteins over random transitions between the values $x = 0$ and $x = 1$. In the opposite case, i.e., when the activation and deactivation rates are slower than the protein degradation rate, the mean protein level is either $x = 0$ or $x = 1$ depending on whether the gene is in the inactive or the active state. The half-life of each such state is larger than that of synthesized proteins so that in each case sufficient time is available

for the mean protein level to attain its particular steady state value. Due to the relatively larger protein degradation rate, there is no accumulation of proteins over the random transitions so that observed protein levels are predominantly at $x = 0$ and $x = 1$.

Ko [11] has considered a stochastic model for gene induction and has shown using computer simulation that two types of response are possible. GE in the model is switched on and off due to the binding and unbinding of the TF-complex at the gene. Stochasticity is introduced into the model because of the probabilistic nature of the binding and unbinding events. An unstable transcription complex causes a "homogeneous" level of gene induction while a stable transcription complex gives rise to a "heterogeneous" level. The homogeneous case is analogous to graded response and the binary response is an example of heterogeneous response. In the detailed stochastic model studied by Pirone and Elston [32], fluctuations in TF binding are shown to explain graded and binary responses to an inducing stimulus. A binary pattern of GE is obtained when the enhancer-state fluctuations (caused by the binding and unbinding of TFs) are slow whereas faster enhancer-state fluctuation give rise to a graded response. The conclusions are arrived at by using a combination of approximate analytical methods and numerical techniques like Monte Carlo (MC) simulation based on the Gillespie Algorithm. The role of operator fluctuations in transcriptional regulation has been studied by Kepler and Elston [10] using the Master Equation Approach. In the limit of large protein abundance, an equation similar to equation (13) is obtained. Again, the interpretation is the same. In each state of the operator, the protein concentration evolves deterministically but there are random transitions between the two states of the operator corresponding to the occupation and unoccupation of the operator region by an activator. The analysis is not, however, extended further.

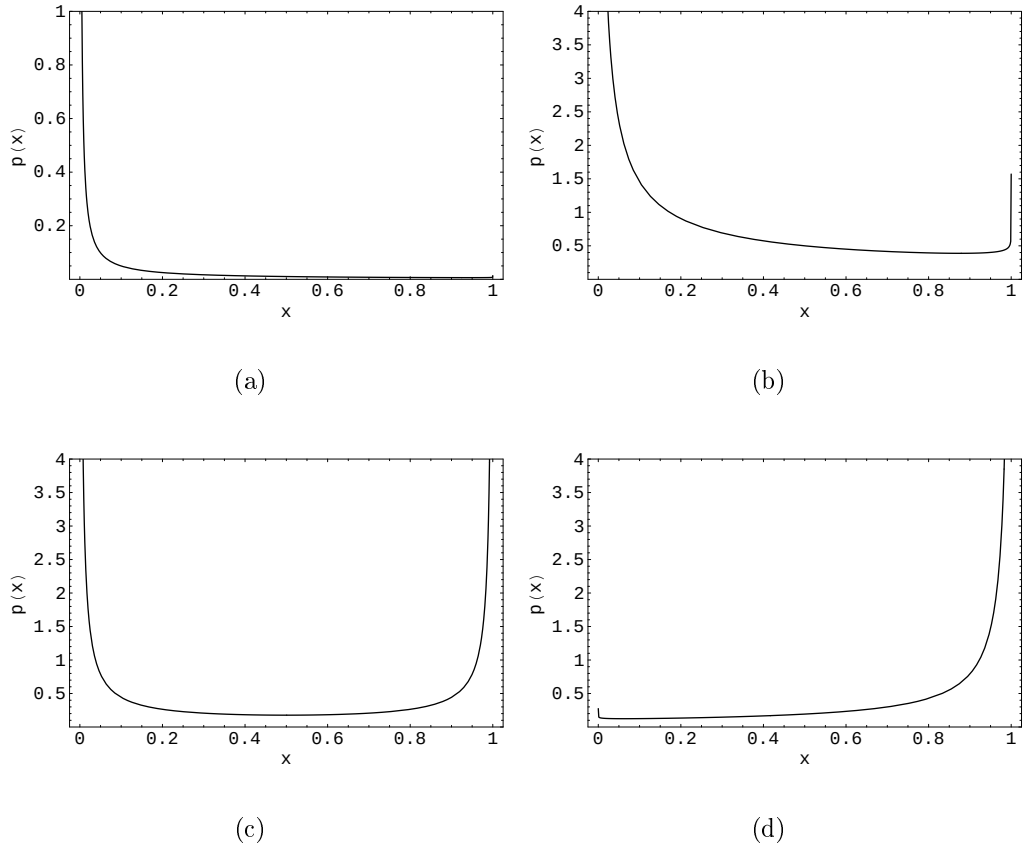



Figure 2: Plot of $p(x)$ versus $x$ (Case II, $r_1$, $r_2 < 1$): 2A ($r_2 >> r_1$), 2B ($r_2 > r_1$), 2C ($r_2 = r_1$), and 2D ($r_1 >> r_2$) respectively.

Discovery of biologically active molecules, say, drugs involves testing the effect of such molecules on appropriate targets. Membrane receptors are the largest class of drug targets. Drugs interacting with receptors are broadly of two types: antagonists and agonists. Antagonists block receptor activity whereas agonists binding to the receptors enhance cellular activity. The binding triggers a series of biochemical events which lead to a change in cellular activity. The change can be linked to the expression of a reporter gene so that detection and quantification of the response to agonist-induced receptor activation are possible. Figure 3 shows a cartoon of how the reporter gene conveys information regarding receptor activation [6]. A cascade of intracellular processes are initiated by the binding of an agonist to the receptor.

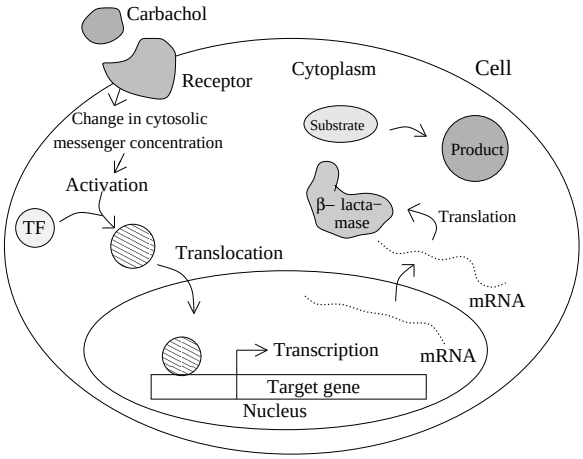



Figure 3: Schematic diagram showing the action of carbachol.

This is accompanied by a change in the concentrations of messengers in the cell. As a result, TFs are activated which then translocate to the nucleus. The TFs bind to the target gene and initiate expression of both this gene as well the reporter gene. The $mRNA$, generated from the reporter gene, is translated into the reporter enzyme. The reporter enzyme catalyzes the conversion of substrates into detectable products.

The scenario depicted in figure 3 provides the basis for high-throughput screening of pharmaceutical candidate drugs in living mammalian cells [6]. The reporter gene used is $\beta - lactamase$ the protein product of which hydrolyzes a substrate giving rise to a large shift in fluorescence emission wavelength. Cells in which the reporter gene is not expressed or expressed at a very low level appear fluorescent green whereas reporter-positive cells with a high level of GE appear fluorescent blue. The activation of cellular processes is brought about by the binding of the agonist carbachol to the muscarinic receptor. In the experiment, the percentages of blue, blue-green (intermediate level of GE) and green cells are measured by flow cytometry with varying carbachol dose and also as a function of time after stimulation by carbachol. The major finding is that as the carbachol dose increases from a low to high value, the fraction of green cells (low level of GE) decreases and that of blue cells (high level of GE) increases. The percentage of blue-green cells remains fairly low throughout. This is a manifestation of the "all-or-none" phenomenon, i.e., binary response in GE.

We now show that the simple stochastic model studied by us provides a good quantitative fit to the experimental data of Zlokarnik et al. Since the probability density function $p(x)$ of protein levels is known (equation (16)), one can calculate experimentally measurable quantities like the dose-response function. Figure 4 shows the experimental data points corresponding to fractions of blue + blue-green (depicted by solid triangles) and blue (depicted by solid squares) cells versus $log\,(C)$, where $C$ is the carbachol concentration. The fraction of blue-green cells is given by the difference in data points belonging to the two curves. The remaining cell fractions describe green cells. The concentration of activated TFs ($S$ in our model) may be taken to be proportional to the concentration $C$ of carbachol and in our theoretical dose-response curves (solid lines in figure 4), $[S]$ replaces $C$. From equation (16), the steady state probability of finding a cell with $x$ (mean protein level divided by maximum protein concentration), greater than a threshold value $x_{thr}$ is

$$p(x > x_{thr}) = 1 - \frac{\int_0^{x_{thr}} x^{(\frac{k_a}{k_p}-1)}\,(1-x)^{(\frac{k_d}{k_p}-1)}\,dx}{\int_0^1 x^{(\frac{k_a}{k_p}-1)}\,(1-x)^{(\frac{k_d}{k_p}-1)}\,dx} \tag{20}$$

$$= 1 - \frac{k_p\, x_{thr}^{\frac{k_a}{k_p}}\, {}_2F_1[1-\frac{k_d}{k_p}, \frac{k_a}{k_p}, 1+\frac{k_a}{k_p}, x_{thr}]}{k_a\, B(\frac{k_a}{k_p},\, \frac{k_d}{k_p})} \tag{21}$$

where ${}_2F_1(a, b, c; z)$ is the hypergeometric function [35]. In our model, we assume that a cell is in a state with high level of GE if the mean protein level in the steady state is greater than a fraction of 0.9 of the maximum protein concentration i.e., $x > 0.9$. By setting $x_{thr} = 0.9$ in equation (20) and replacing $k_a$, $k_d$ by the effective rate constants $k_a^{'}$, $k_d^{'}$ (equation (8)), one can calculate $p(x > x_{thr})$ for various values of $S$. The probability $p(x > x_{thr})$ can also be interpreted as the fraction of cells in a cell population with $x > x_{thr}$. The theoretical dose-response curve obtained in this manner gives a good fit to the experimental data points (solid squares in figure 4) for the parameter values (in arbitrary units) $k_2 = 1.6 \times 10^{-4}$, $k_{on} = 1.2 \times 10^{-6}$, $k_{off} = 1.32 \times 10^{-4}$, $k_s = 1.6 \times 10^{-6}$, $k_p = 1$, $k_a = 0.17$ and $k_d = 0.0465$. The data points in this case correspond to the fraction of cells in a high level of GE (blue cells). A cell is assumed to be in a state with low level of GE if $x$ is $< x_{thr} = 0.1$ (green cells). A cell is in a state with intermediate level of GE when $0.1 < x < 0.9$ (blue-green cells). The cell fraction in the last case

is given by

$$p(0.1 < x < 0.9) = \frac{\int_{0.1}^{0.9} x^{(\frac{k_a}{k_p}-1)} (1-x)^{(\frac{k_d}{k_p}-1)} dx}{\int_{0}^{1} x^{(\frac{k_a}{k_p}-1)} (1-x)^{(\frac{k_d}{k_p}-1)} dx} \tag{22}$$

with $k_a$, $k_d$ replaced by $k_a^{'}$ and $k_d^{'}$. The fraction of blue + blue-green cells is computed from an expression similar to 22 but with the integration limits (0.1, 0.9) in the numerator replaced by (0.1, 1.0). The calculated dose-response curve gives a good fit to the experimental data points (solid triangles in figure 4). The two curves in figure 4 have been obtained for the same set of parameter values using Mathematica. The good quantitative agreement between our theoretical results and experimental data indicates that the stochastic model of GE considered by us captures the essential features of stochastically induced binary response in GE.

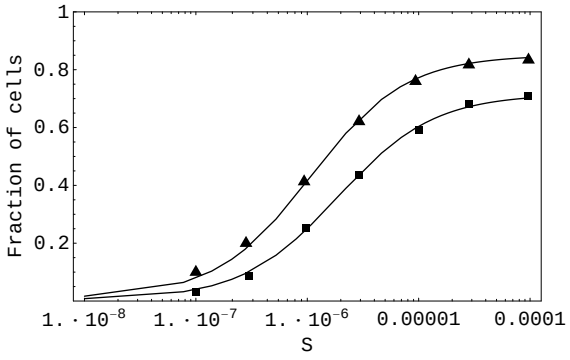



Figure 4: Fractions of blue + blue-green (upper curve) and blue (lower curve) cells versus $S$ in a semi-logarithm plot. The experimental points are depicted by solid triangles signs and solid squares. The parameter values used for the fitting are mentioned in the text.

The reaction scheme 3 doesnot fulfill detailed balance, i.e, equilibrium conditions. The steady state concentrations in equation (5) are derived from the more general non-equilibrium conditions of stationarity. Ref. [36] provides examples of reaction schemes related to GE which violate detailed balance conditions. One of these involves the phenomenon of stochastic focussing (SF) in which signal fluctuations sharpen the response in a regulated

process. SF appears to be an out-of-equilibrium effect which is absent if the reaction scheme is constrained by detailed balance. Modifications of reaction scheme 3 which preserve detailed balance are possible and it will be of interest to determine whether SBR can occur in both equilibrium and out-of-equilibrium scenarios.

# 4. Conclusion and outlook

In this paper, we have considered a simple stochastic model of GE and demonstrated that stochasticity provides the basis for graded and binary responses to inducing signals. The sole ingredients of the minimal model of stochastic GE studied in this paper are gene activation, deactivation, protein synthesis and degradation, each of which involves a number of biochemical events. Stochasticity in this model is associated only with the gene activation and deactivation processes whereas protein synthesis and degradation are assumed to occur in a deterministic manner. A deterministic description of protein synthesis is justified when the number of proteins produced is large. This is the situation in the experiment by Zlokarnik et al. [6] in which proteins per cell are a few thousands in number. For smaller protein numbers, the inclusion of stochasticity during protein synthesis and degradation is expected to blur the GE responses but the major conclusions of the paper still remain valid. The processes of transcription and translation in the model are not treated separately but lumped together in a single protein synthesis step. In an eukaryotic cell, combining transcription and translation into a single step may be considered to be a drastic approximation. One can study the effect of stochastic gene activation and deactivation on the transcription process itself and focus on $mRNA$ synthesis rather than proteins in reaction schemes 1-3. This type of approach highlights the quantal nature of transcription with bursts of $mRNAs$ being produced in a probabilistic manner in agreement with experimental observations [9, 30]. In fact, the value $k_p = 1$ is more appropriate if $k_p$ is interpreted as the $mRNA$, rather than protein decay constant. The mathematical analysis and conclusions are the same as before since protein production is linked to $mRNA$ synthesis. Despite the limitations of the model, it contains the important features necessary for an explanation of the stochastically induced "all-or-none" phenomenon observed in some eukaryotic systems. The model results give a good description of the experimental data of Zlokarnik et al. [6] and are expected to be of relevance in explaining the binary response in GE observed in other eukaryotic

systems [4, 5]. The probabilistic nature of gene activation and deactivation processes is crucial to explain how graded and binary responses in GE occur in the model. The stochastic origin of binary response is distinctive from the binary response brought about by positive feedback processes. Experiments designed to probe the stochastic origins of graded and binary responsPlot of $p(x)$ versus $x$ (Case I, $r_1$, $r_2 > 1$): 1A ($r_2 >> r_1$), 1B ($r_2 = r_1$) and 1C ($r_1 >> r_2$) respectively.es, are needed for a clearer understanding of the role of stochasticity in such responses. The stochastic model of GE,corresponding to reaction scheme 2, has earlier been studied to explore the stochastic origins of haploinsufficiency [13, 17, 37]. The model studied in the paper is a modification of the earlier model. The simplicity of the models allows for mathematical analysis and helps in identifying the origins of phenomena associated with stochastic GE. The knowledge and insight gained from the study of simple models like the present one provide necessary inputs to develop more detailed and realistic models of GE.

**ACKNOWLEDGEMENT**

The Authors thank Gregor Zlokarnik and David Hume for helpful comments. R. K. is supported by the Council of Scientific and Industrial Research, India under Sanction No. 9/15 (239) / 2002 - EMR - 1.

# References

[1] McAdams H H and Arkin A 1999 Trends Genet. 15 65-69

[2] Rao C V, Wolf D M and Arkin A P 2002 *Nature* **420** 231-37

[3] Arkin A, Ross J and McAdams H H 1998 Genetics 149 1633-48

[4] Karttunen J and Shastri N 1991 Proc. Natl. Acad. Sci. U. S. A. 88 3972-76

[5] Negulescu P A, Shastri N and Cahalan M D 1994 Proc. Natl. Acad. Sci. U. S. A. 91 2873-77

[6] Zlokarnik G, Negulescu P A, Knapp T E, Mere L, Burres N, Feng L, Whitney M, Roemer K and Tsien R Y 1998 Science 279 84-88

[7] Ozbudak E M, Thattai M, Kurtser I, Grossman A D and van Oudenaarden A 2002 Nature Genet. 31 69-73

[8] Elowitz M B, Levine A J, Siggia E D and Swain P S 2002 Science 297 1183-86

[9] Blake W J, Kærn M, Cantor C R, and Collins J J 2003 Nature 422 633-37

[10] Kepler T B and Elston T C 2001 Biophys J. 81 3116-36

[11] Ko M S H 1991 J. Theor. Biol. 153 181-94

[12] McAdams H H and Arkin A 1997 Proc. Natl. Acad. Sci. U. S. A. 94 814-19

[13] Cook D L, Gerber A N and Tapscott S J 1998 Proc. Natl. Acad. Sci. U. S. A. 95 15641-46

[14] Hasty J, Pradines J, Dolnik M, and Collins J J 2000 Proc. Natl. Acad. Sci. U. S. A. 97 2075-80

[15] Thattai M and van Oudenaarden A 2001 Proc. Natl. Acad. Sci. U. S. A. 98 8614-19

[16] Swain P S, Elowitz M B and Siggia E D 2002 Proc. Natl. Acad. Sci. U. S. A. 99 12795-800

[17] Bose I and Karmakar R 2004 in Biology of genetic dominance Edited by Veitia R A (USA : Landes Bioscience), Chapter 6

[18] Sasai M and Wolynes P G 2003 Proc. Natl. Acad. Sci. U.S.A 100 2374-79

[19] Novick A and Weiner M 1957 Proc. Natl. Acad. Sci. U. S. A. 43 553-66

[20] Maloney P C and Rotman B 1973 J. Mol. Biol. 73 77-91

[21] Siegele D A and Hu J C 1997 Proc. Natl. Acad. Sci. U. S. A. 94 8168-72

[22] Rossi F M V, Kringstein A M, Spicher A, Guichert O M and Blau H M 2000 Mol. Cell. 6 723-28

[23] Becskei A, Seraphin B, and Serrano L 2001 EMBO J. 20 2528-35

[24] Carrier T A and Keasling J D 1999 J. Theor. Biol. 201 25-36

[25] Vilar J M G, Guet C C and Leibler S 2003 J. Cell. Biol. 161 471-76

[26] Ozbudak E M, Thattai M, Lim H N, Shraiman B I, van Oudenaarden A 2004 Nature 427 737-40

[27] Walters M C, Fiering S, Eidemiller J, Magis W, Groudine M, and Martin D I K 1995 Proc. Natl. Acad. Sci. U. S. A. 92 7125-29

[28] Weintraub H 1988 Proc. Natl. Acad. Sci. U. S. A. 85 5819-23

[29] Fiering S, Whitelaw E, and Martin D I K 2000 Bioessays 22 381-87

[30] Hume D A 2000 Blood 96 2323-28

[31] Roy S, Bose I and Manna S S 2001 Int. J. Mod. Phys. C. 12 413-20

[32] Pirone J R and Elston T C 2004 J. Theor. Biol. 226 111-21

[33] Peccoud J and Ycart B 1995 Theor. Pop. Biol. 48 222-34

[34] Cornish-Bowden A 1976 Principles of Enzyme Kinetics, Butterworth, London 28-50

[35] Gradshteyn I S and Ryzhik I M 1980 Table of Integrals, Series and Products. Academic Press.

[36] Berg O G, Paulsson J and Ehrenberg M 2000 Biophysical Journal 79 2944-53

[37] Veitia R A 2002 BioEssays 24 175-84