Functional characteristics of gene expression motifs with single and dual strategies of regulation

Mainak Pal¹, Sayantari Ghosh² and Indrani Bose¹

¹ Department of Physics, Bose Institute 93/1, Acharya Prafulla Chandra Road, Kolkata-700009, India
² Amity Institute of Applied Sciences, Amity University, Kolkata, India

E-mail: sghosh@kol.amity.edu and indrani@jcbose.ac.in

Keywords: microRNAs, four-gene motif, mutual information, linear noise approximation, bistability and hysteresis

Abstract

Transcriptional regulation by transcription factors and post-transcriptional regulation by microRNAs constitute two major modes of regulation of gene expression. While gene expression motifs incorporating solely transcriptional regulation are well investigated, the dynamics of motifs with dual strategies of regulation, i.e., both transcriptional and post-transcriptional regulation, have not been studied as extensively. In this paper, we probe the dynamics of a four-gene motif with dual strategies of regulation of gene expression. Some of the functional characteristics are compared with those of a two-gene motif, the genetic toggle, employing only transcriptional regulation. Both the motifs define positive feedback loops with the potential for bistability and hysteresis. The four-gene motif, contrary to the genetic toggle, is found to exhibit bistability even in the absence of cooperativity in the regulation of gene expression. The four-gene motif further exhibits a novel dynamical feature in which two regions of monostability with linear threshold response are separated by a region of bistability with digital response. Using the linear noise approximation, we further show that the coefficient of variation (a measure of noise), associated with the protein levels in the steady state, has a lower magnitude in the case of the four-gene motif as compared to the case of the genetic toggle. We next compare transcriptional with post-transcriptional regulation from an information theoretic perspective. We focus on two gene expression motifs, Motif 1 with transcriptional regulation and Motif 2 with post-transcriptional regulation. We show that amongst the two motifs, Motif 2 has a greater capacity for information transmission for an extended range of parameter values.

1. Introduction

A living cell contains several thousands of genes the expression of a large fraction of which is regulated [1]. Two major strategies for the regulation of gene expression are transcriptional and post-transcriptional regulation, each with a distinctive mode of operation. In the case of transcriptional regulation, regulatory molecules bind the DNA to either activate or repress the initiation of transcription and thereby control the amount of proteins synthesised during translation. Post-transcriptional regulation is brought about by small non-coding RNAs (small RNAs/microRNAs) which bind the messenger RNAs (mRNAs) of the target gene resulting in the degradation of the mRNAs and/or the inhibition of translation [2, 3]. Small RNAs regulate bacterial gene expression whereas microRNAs are functional in eukaryotic cells.

The study of large-scale gene regulatory networks has so far been mostly confined to regulatory networks [4, 5] which take into account only transcriptional regulation. The connectivity structure of these networks reveals the existence of a number of gene expression motifs or substructures, e.g., positive and negative feedback loops, single input modules and feed forward loops. In recent years, similar motifs have been identified in genome-scale regulatory networks with dual strategies of regulation, i.e., the regulatory interactions include both transcriptional and post-transcriptional regulation [6–11]. Amongst the various types of motifs, the positive feedback loop has the potential for the generation of multistability, i.e., the coexistence of more than one stable steady state for the same
parameter values. Bistability, with two stable steady states, provides the basis for cell-fate decisions between two alternative fates as in the case of cell differentiation in which a progenitor cell faces an either-or decision in choosing between two distinct cell lineages. A biological switch utilises bistability for the generation of the so-called ‘OFF’ and ‘ON’ states between which the switch operates. A well-known example of a biological switch is the genetic toggle, a two-gene motif, in which the two genes mutually inhibit each other’s expression through transcriptional regulation (figure 1(a)) [12]. The mutual inhibition (double negative feedback) defines a positive feedback loop and in terms of the steady state concentrations $[P_1]$ and $[P_2]$ of the two proteins, the stable steady states correspond to $[P_1] \gg [P_2]$ and $[P_2] \gg [P_1]$ respectively. The two-gene motif appears extensively in several natural genetic networks including the ones governing cell-fate decisions [13].

A number of recent studies [11, 14–16] have identified a pair of interconnected mutual-inhibition feedback circuits employing dual strategies of regulation. In a single feedback circuit, a microRNA and a transcription factor (TF) mutually repress each other’s synthesis. The microRNAs involved in the two feedback loops are miR-34 and miR-200 whereas the corresponding TFs are SNAIL and ZEB respectively. The two interlinked feedback loops constitute the core of a gene regulatory network governing the epithelial to mesenchymal transition (EMT). The transition brings about the conversion of epithelial to mesenchymal cells characterised by the loss of cell-cell adhesion and enhanced cell mobility. EMT plays crucial roles in the development of embryos and the repair of tissues. Aberrantly regulated transitions result in cancer metastasis with primary tumor cells losing cell-cell adhesion and becoming migratory. Beyond two-gene motifs, there is limited knowledge about more complex positive feedback loops with dual strategies of regulation. Figure 1(b) provides a natural example of such a motif involving four genes [17, 18]. The genes die-1 and cog-1 synthesise the transcriptional regulators Die-1 and Cog-1 respectively. Die-1 (Cog-1) activates the expression of the lys-6 (mir-273) gene synthesising the microRNA lys-6 (mir-273). The microRNA lys-6 (mir-273) targets the mRNA of the cog-1 (die-1) gene. The meanings of the symbols $m_i$, $s_i$, $p_i$ ($i=1,2$) are explained in the text. Both the feedback loops in (a) and (b) define a positive feedback loop. The arrow signs indicate activation and the hammerhead symbols represent repression.

The major objective of our paper is to investigate the functional characteristics of simple gene...
expression motifs with single and dual strategies for the regulation of gene expression. In the first part of our study (section 2), we undertake a comparative study of the gene expression motifs shown in figures 1(a) and (b). We consider both deterministic and stochastic dynamics of the motifs in order to identify the distinctive operational features in the steady state. In the second part of our study (section 3), we consider two one-gene motifs designated as Motif 1 and Motif 2, in Motif 1, TFs repress the expression of the target gene, whereas Motif 2 describes the post-transcriptional regulation of the target gene expression by microRNAs. A regulatory module may be represented as an input–output device with the regulatory molecules defining the input and the gene expression levels constituting the output [19–21]. Due to the stochastic nature of gene expression [22–24], the input and output are not single levels but are best described as distributions around average values. The variability (noise) corrupts the fidelity of information transmission from the input to the output as the one-to-one correspondence between the input and the output is lost. The information-theoretic quantity mutual information (MI) provides knowledge of the amount of information that the value of one random variable yields about that of another. The transmission of information through an input–output device can equivalently be represented as information flow through a channel. The channel capacity is defined to be the MI maximised over all possible input distributions and is a measure of the maximum amount of information that can be transmitted through the channel. In section 3 of the paper, we compute the channel capacity for the Motifs 1 and 2 to analyze transcriptional and post-transcriptional gene regulation from the viewpoint of information theory. Section 4 of the paper contains concluding remarks.

2. Dynamics of two-gene and four-gene motifs

We first consider a four-gene motif with dual strategies of gene regulation, a specific example of which is shown in figure 1(b). In the motif, there are altogether four genes, two of which synthesise the mRNAs $m_1$ and $m_2$ and the other two synthesise the microRNAs $s_1$ and $s_2$. The mRNAs $m_1$ and $m_2$ are translated to yield the proteins $p_1$ and $p_2$. The microRNA $s_2$ binds the mRNA $m_i$ ($i = 1, 2$) and targets it for degradation. The protein $p_1$ ($p_2$) activates the synthesis of the microRNA $s_2$ ($s_1$). The deterministic dynamics of the motif are described by the following set of differential equations:

$$\frac{d[m_i]}{dt} = \alpha_{m_i} - \mu_{m_i}[m_i][s_i] + \mu_{s_i}[s_i] - \gamma_{m_i}[m_i],$$  

(1)

$$\frac{d[s_i]}{dt} = \alpha_i + \frac{\kappa[p_i]}{R_i + [p_i]} - \mu_{s_i}[m_i][s_i] + k_1[s_i] - \gamma_{si}[s_i],$$  

(2)

$$\frac{d[p_i]}{dt} = \alpha_{p_i} - \gamma_{pi}[p_i],$$  

(3)

$$\frac{d[m_2]}{dt} = \alpha_{m_2} - \mu_{m_2}[m_2][s_2] + \mu_{s_2}[s_2] - \gamma_{m_2}[m_2] - \gamma_{si}[m_2]s_i,$$  

(4)

$$\frac{d[s_2]}{dt} = \alpha_{s_2} + \frac{\tau_2[p_2]}{R_2 + [p_2]} - \mu_{s_2}[m_2][s_2] + \mu_{s_1}[s_1] - \gamma_{si}[s_2],$$  

(5)

$$\frac{d[s_1]}{dt} = \mu_{m_2}[m_2][s_2] - \mu_{s_2}[s_2] - k_2[s_2] - \gamma_{si}[s_1],$$  

(6)

$$\frac{d[p_2]}{dt} = \alpha_{p_2} - \gamma_{pi}[p_2].$$  

(7)

The symbols $m_i$, $s_i$, $c_i$, $p_i$ ($i = 1, 2$) denote the molecular types as well as numbers and the quantities within third brackets represent the concentrations. The modelling of the post-transcriptional regulation is based on the mathematical models proposed in [25, 26]. The models derive their strength from the fact that the theoretical results closely match the experimental observations in single cell measurements.

The mRNAs are transcribed at the rates $\alpha_{m_1}$ and $\alpha_{m_2}$ and constitutively degraded at the rates $\gamma_{m_1}[m_1]$ and $\gamma_{m_2}[m_2]$. Proteins $p_i$ are translated from the free mRNA $m_i$ at the rate $\alpha_{p_i}[m_i]$ and degrade at the rate $\gamma_{p_i}[p_i]$ ($i = 1, 2$). The microRNA $s_i$ binds the target mRNA $m_i$ to form the mRNA-microRNA complex $c_i$ ($i = 1, 2$) which results in an accelerated degradation of the mRNA so that the translation of the mRNA into proteins is not possible. The bound complex $c_i$ ($i = 1, 2$) forms with the rate constant $\mu_{s_i}$ and dissociates into free mRNA and free microRNA with the rate constant $\gamma_{s_i}$. The mRNA $m_i$ in the bound complex $c_i$ degrades with the additional rate $k_i[s_i]$ ($i = 1, 2$) while the microRNAs are recycled to the pool of free microRNAs after the mRNA degradation is completed. The bound complex $c_i$ has a natural degradation rate with rate constant $\gamma_i$ ($i = 1, 2$). The free microRNA $s_1$ is synthesised at the rate $\alpha_{s_1}$ and degraded at the rate $\gamma_{s_1}[s_1]$ ($i = 1, 2$). Protein $p_2$ ($p_1$) activates the synthesis of $s_1$ ($s_2$) with $t_1$ ($t_2$) being the maximum rate of synthesis and $R_1$ ($R_2$) being the corresponding Michaelis–Menten-type constant. The dynamics of the four-gene motif exhibit bistability in the steady state in an extended parameter regime. For the sake of simplicity, we consider the symmetric situation $\gamma_{m_1} = \gamma_{m_2} = \gamma_{m}, \alpha_{s_1} = \alpha_{s_2} = \alpha_s, \gamma_{s_1} = \gamma_{s_2} = \gamma_s, \alpha_{p_1} = \alpha_{p_2} = \alpha_p, \gamma_{p_1} = \gamma_{p_2} = \gamma_p, \mu = \mu_1 = \mu_2, k_1 = k_2 = k$ and $\gamma_1 = \gamma_2 = \gamma$. Figures 2(a)–(d) show the steady state concentrations $[p_{1sf}], [s_{1df}], [p_{2sf}]$ and...
respectively as a function of the mRNA synthesis rate constant $a_{m1}$. The parameter values used for the steady state solutions are displayed in table 1. The solid (dashed) lines represent stable (unstable) steady states. The bistability in each case is accompanied by hysteresis, i.e., the transitions from the lower to the upper stable steady state and from the upper to lower stable steady state occur at different values of the bifurcation parameter. The four-gene motif (figure 1), similar to the genetic toggle (figure 1(a), defines a positive feedback loop in which the genes transcribing the mRNAs $m_1$ and $m_2$ mutually inhibit each other’s expression. Consequently, the concentrations of the proteins $p_1$ and $p_2$ cannot be simultaneously high in a stable steady state. If the concentration of $p_1$ is high then that of $p_2$ is low and vice versa. A high concentration of $m_i$ (corresponding protein concentration is also high) is incompatible with a high concentration of the microRNA $s_i$ as $s_i$ targets $m_i$ for degradation ($i = 1, 2$) blocking protein synthesis in the process. The two stable steady states thus correspond to $[p_{1st}], [s_{1st}], [p_{2st}], [s_{2st}]$ as a function of the mRNA synthesis rate constant $a_{m1}$. The solid (dashed) lines represent stable (unstable) steady states. The essential requirement for the generation of bistability is the presence of positive feedback combined with an ultrasensitive response [27]. The most common origin of ultrasensitivity may be ascribed to cooperativity in the

| Parameter | Figure 2 | Figure 3(a) | Figure 3(b) | Figure 4 | Figure 5 |
|-----------|----------|-------------|-------------|----------|----------|
| $a_{m1}$  | —        | —           | —           | 15.00    | —        |
| $a_{m2}$  | 15.00    | 30.00       | 15.00       | 0.950    | —        |
| $\gamma_0$ | 0.200  | 0.200       | 0.200       | 0.800    | —        |
| $\gamma_1$ | 1.000   | 1.000       | 1.000       | 1.000    | 1.000    |
| $\gamma_2$ | 0.080  | 0.080       | 0.080       | 0.080    | 0.080    |
| $\gamma_3$ | 0.100   | 0.100       | 0.100       | 0.100    | 1.400    |
| $b_0$     | 20.00    | 20.00       | 20.00       | 20.00    | 20.00    |
| $R$       | 40.00    | 40.00       | 40.00       | 40.00    | 20.00    |
| $\mu_1$   | 0.200    | 0.200       | 0.200       | 0.200    | 0.200    |
| $\mu_2$   | 0.001    | 0.001       | 0.001       | 0.001    | 0.010    |
| $k$       | 0.100    | 0.100       | 0.100       | 0.100    | 0.010    |
| $\tau$    | 0.200    | 0.200       | 0.200       | 0.200    | 0.800    |
regulation of gene expression. A less explored source of ultrasensitivity lies in molecular sequestration [28, 29] in which the activity of a biomolecule A is compromised through sequestration due to the formation of an inactive complex with another biomolecule B. In the case of the four-gene motif, the regulatory links constitute a positive feedback loop. The model exhibits bistability even in the absence of cooperativity in gene expression. The ultrasensitive response occurs through molecular sequestration with A representing the mRNA of the target gene and B representing the microRNA. In the case of small RNA/microRNA-regulated gene expression, the ultrasensitivity arising from molecular sequestration is in the form of a linear threshold behaviour [3, 25]. We consider the post-transcriptional regulation of the expression of a target gene by a microRNA. The target gene is not expressed if the mRNA synthesis rate \(a_m\) is lower than a threshold value set by the microRNA synthesis rate \(a_s\). In this case, all the mRNA molecules form bound complexes with the microRNAs. For \(a_m > a_s\), the number of mRNA molecules is larger than that of the microRNAs and some free mRNAs are available for translation into proteins. The expressed protein level is linearly proportional to \(a_m - a_s\). The linear threshold behaviour has been demonstrated experimentally in the cases of both small RNA and microRNA-regulated gene expression [3, 25]. Figure 3(a) shows an example of linear threshold behaviour in the region of monostability of the four-gene motif with the parameter values given in table 1. The linear threshold behaviour survives even in the presence of dual strategies of regulation of gene expression. The threshold location is now defined by a relationship more complex than \(a_m = a_s\). Figure 3(b) exhibits how the steady state concentration of the protein \(p_1\) changes in the enlarged parameter space \(a_m - a_s\) from one region of monostability to another such region via a region of bistability. The linear threshold behaviour indicates reversible response, whereas the bistable response is digital (low-or-high) and hysteretic in character. Figure 3(b) captures the evolution of the steady state response of the four-gene motif from the linear threshold monostable to the digital bistable to finally the linear-threshold monostable response once again. This type of steady state response has not been reported earlier in the case of microRNA-mediated regulation of gene expression. The parameter values for the figure are given in table 1. An examination of figure 3(b) shows that as the rate constant \(a_m\) decreases from a high value, the value of the upper bifurcation point (the value of \(a_m\) at which an abrupt transition from the lower to the upper branch of stable steady states) also decreases. This is an expected result as the genes synthesising the mRNAs \(m_1\) and \(m_2\) at the rates \(a_m\) and \(a_m\) respectively, mutually inhibit each other’s expression. Large values of \(a_m\) favour high (low) concentration of \(m_2\) (\(m_1\)). Thus, as the value of \(a_m\) decreases, the transition to the stable steady state with high \(p_1\) concentration becomes more favourable, i.e., the upper bifurcation point shifts to lower values of \(a_m\). Steady state behaviour similar to that shown in figure 3(b) is obtained if
instead of the parameter space $\alpha_m - \alpha_{m'}$, one considers the parameter space $\alpha_m - \alpha_{m''}$.

In figure 4, we display the results of a parameter sensitivity analysis with respect to some of the parameters, $\alpha_m$, $\gamma_{m}$, $\alpha_p$, $\gamma_{p}$, in the absence (figure 4(a)) and in the presence (figure 4(b)) of cooperativity in the transcriptional regulation of gene expression. The cooperativity is taken into account by replacing the protein concentrations $[p_1]$ and $[p_2]$ in the second terms of equations (2) and (6) by $[p_1]^n$ and $[p_2]^n$ respectively, where $n$ is the Hill Coefficient. Similarly, the constants $R_1$ and $R_2$ are replaced by $R_1^n$ and $R_2^n$. The set of parameter values shown in table 1 serves as the reference set. The value of each of the four parameters was varied separately in steps up to values that are 100 fold greater and lower than the reference value listed in table 1. Figure 4 shows the log-fold change in parameter value within the range $-2$ to $+2$ for a specific parameter (marked on the x-axis) and identifies the region of bistability, shaded black, in the parameter range. As figure 4(a) clearly demonstrates, bistability occurs even in the absence of cooperativity. With cooperativity taken into account ($n = 2$), the region of bistability becomes more extensive (figure 4(b)). In this case, both cooperativity and molecular sequestration contribute to the generation of the ultrasensitive response. In the presence of cooperativity, the steady state response becomes sharper in the region of monostability.

One important advantage of microRNA-mediated regulation of gene expression is that microRNAs (as well as small RNAs) are known to suppress fluctuations in the target gene’s protein levels [3, 26, 30, 31]. Cell fate decisions are, in general, controlled by bistable genetic circuits [24]. The two stable steady states are of key regulatory proteins. The stable steady states correspond to the two stable phenotypes that a cell has to choose between. Fluctuations in the protein levels can bring about undesirable transitions between the alternative states resulting in a loss in the stability of the phenotypes. A robust biological switch operating between two stable steady states eliminates to a large extent the random transitions between the states. A major outcome of stochastic gene expression is that protein production occurs in random bursts rather than in a continuous manner [22, 23, 32]. A single mRNA has a finite lifetime during which the mRNA is repeatedly translated to yield a burst of proteins. The average size $b$ of a protein burst is given by $b = \frac{\alpha_p}{\gamma_{m}}$ where $\alpha_p$ is the translation rate and $\gamma_{m}$ is the average lifetime of the mRNA. In the case of post-transcriptional regulation, the microRNAs promote mRNA degradation ($\gamma_{m}$ is increased) and/or suppress translation ($\alpha_p$ decreases). The outcome is a reduction in the protein burst size, i.e., diminished protein fluctuations. In a two-step model of gene expression, the noise at the protein level is quantified in terms of the coefficient of variation ($CV_p$) which is the ratio of the standard deviation and mean $\langle p \rangle$ [33]. The expression for $CV_p$ in the steady state is $CV_p^2 = \frac{1 + b}{\langle p \rangle}$. A reduction in the average burst size lowers the amount of relative fluctuations in the protein level. We compute the $CV_p$ for both the two-gene and four-gene motifs, shown in figure 1, to ascertain which motif is less noisy. We use the linear noise approximation (LNA) to the master equation [34–36] governing the stochastic dynamics of the motifs. In the following, we outline the major steps in the computational procedure. We consider N distinct chemical species which participate in R chemical reactions. The concentrations of the chemical species are given by the variables $x_i$, $i = 1, \ldots, N$. The state of the dynamical system is represented by the
vector $x = (x_1, x_2, \ldots, x_N)^T$ where $T$ denotes the transpose. The state changes as a function of time due to the occurrence of the chemical reactions. Let $S$ be the stoichiometric matrix with elements $S_{ij}$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, R$. The interpretation of $S_{ij}$ is that the number of molecules of the chemical species $i$ changes from $X_i$ to $X_i + S_{ij}$ when the $j$th reaction occurs. One also defines a reaction propensity vector $f(x) = (f_1(x), \ldots, f_R(x))^T$. The deterministic dynamics of the system are generated by the rate equations
\[
\frac{dx_i}{dt} = \sum_{j=1}^{R} S_{ij} f_j(x) \quad (i = 1, 2, \ldots, N).
\]
The steady state vector $x_s$ is determined from the condition $\dot{x} = 0$, i.e., $f(x_s) = 0$.

In the steady state and under the LNA, the covariance of the fluctuations about the deterministic steady state is given by the fluctuation-dissipation (FD) relation
\[
JC + (JC)^T + D = 0,
\]
where $J$ is the Jacobian matrix, $C = \langle \delta x \delta x^T \rangle$ is the covariance matrix, the diagonal elements of which are the variances, and $D$ is the diffusion matrix. The matrix $D$ is given by $D = S \text{ diag}(f(x)) S^T$, where diag$(f(x))$ is represented by a diagonal matrix with the elements $f_j(x)$, $j = 1, 2, \ldots, R$. Once the matrices $J$ and $D$ are specified, the elements of the covariance matrix, $C_{ij} = \langle \delta x_i \delta x_j \rangle$ ($i, j = 1, 2, \ldots, N$), with $C_{ij}(i = 1, 2, \ldots, N)$ denoting the variance, can be calculated using the FD relation (10). Under the LNA, the mean value of $x_i$ is given by the deterministic steady state value so that the CV associated with $x_i$ can be computed. In the case of the two-gene (four-gene) motif $N = 4$ ($N = 8$). In the appendix, the differential rate equations governing the deterministic dynamics of the two-gene motif, the genetic toggle, are displayed in equations (A.1)–(A.4) The set of reaction schemes for both the two-gene and four-gene motifs and the matrices $S$ and $f(x)$ are also shown. Figure 5 exhibits the plots of the CV versus the protein mean value in the steady state in the cases of both the genetic toggle and the four-gene motif. The first motif employs only transcriptional regulation, whereas the second motif incorporates the dual strategies of both transcriptional and post-transcriptional regulation of gene expression. The motif with dual strategies of regulation is clearly less noisy than the motif with only transcriptional regulation. The parameter values for the four-gene motif are given in table 1. The parameters for the genetic toggle are $j_{M_1} = 0.9$, $\delta_{M_1} = 0.4$, $\delta_{M_1} = 0.4$, $j_{P_1} = 1.0$, $\delta_{P_1} = 1.0$, $\delta_{P_2} = 0.6$, $\beta_1 = 70.0$, $\beta_2 = 70.0$, $G_1 = 0.01$, $G_2 = 0.01$ ($j_{M_1}$ is the parameter which is varied). In stochastic gene expression, the frequency $a$ of the transcriptional protein bursts is given by $a = \frac{\sigma_m}{\gamma_P}$, where $\sigma_m$ is the transcription rate and $\gamma_P$ the protein degradation rate constant [37]. As discussed in [30], transcriptional regulation modifies $\sigma_m$ thereby changing the burst frequency $a$. The steady state mean protein level $<p>_0 = ba$. Since $CV_p^2 = \frac{b^2}{a^2}$, for the same degree of repression brought about transcriptionally or post-transcriptionally, the micro-RNAs are more effective in reducing the $CV_p$ as they diminish the burst size $b$.

3. Regulation strategies: information theory perspective

In this section, we compute the channel capacities of the input–output devices, representing Motifs 1 and 2 respectively. The MI between the regulatory molecule distribution $p(x)$ and the output protein distributions $p(y)$ is given by
\[
I(x; y) = \int dxdy \ p(x, y) \log_2 \frac{p(y|x)}{p(y)},
\]
where the variable $x$ represents the concentration of the regulatory molecules and the variable $y$ denotes the concentration of the proteins synthesised by the target gene, $p(y|x)$ is the conditional probability distribution and the joint distribution $p(x, y) = p(y|x)p(x)$. The MI is measured in bits and has a clear physical interpretation. For the value of MI = 1 bits, $2^I$ levels of $y$ can be distinguished, in the presence of noise in the channel, given the value of the input $x$. Equivalently since the MI is a symmetric function of $x$ and $y$, $2^I$ input levels can be distinguished given the value of the output $y$. In the small noise approximation [19, 21, 38], the channel capacity for information transmission is given by the MI maximised over all possible distributions, $p(y)$, of the output signal. The optimal distribution has the form
\[
P_{\text{opt}} = \frac{1}{Z} \sigma_y,
\]
where $\sigma_y$ (the standard deviation) is a measure of noise in the output and $Z$ normalises the distribution, i.e.
where \( Z \) is the maximum value of \([p] \) corresponding to the unrepressed condition which has values in the range \([0, 1]\). The quantity \( b_0 = \alpha_p/\gamma_m \) (in the notation of section 2 and with the suffixes omitted) is the average number of proteins synthesised from a mRNA during its lifetime, in the absence of any regulation. The quantity \( R_s = 1 - \frac{1}{[p]} = 1 - \bar{F} \) denotes the repression strength of the microRNA-mediated regulation with \( 0 \leq R_s \leq 1 \). The values \( R_s = 0 \) and 1 indicate no repression and full repression respectively of protein synthesis. The quantity \( F \) is given by

\[
F = 1 - \frac{s_s}{s_{tot}},
\]

where \([s]\) is the concentration of free microRNAs and \([s_{tot}]\) is the total concentration of microRNAs, a fraction of which is free and the rest form bound complexes with the mRNAs. From (13), \( Z \) is given by the integral

\[
Z = \int_{0}^{\infty} \frac{dp}{\sqrt{p_0 \bar{F} + b_0 [p] (1 - R_s) \frac{(1 - \bar{F})^2}{(1 - R_s \bar{F})^2}}}
\]

The value of the protein concentration \([p_0]\) is \( \gamma_m = 0.0058, \alpha_p = 0.1155 \) and \( \gamma_p = 0.0002 \) (the parameter values are the same as reported in [33]). From (14) and (17), the channel capacity \( I_{RF} = 2.91 \) for \( F \simeq 0 \) and \( I_{RF} = 0.74 \) for \( F \simeq 1 \). For \( F \simeq 0 \), the term \( R_s \bar{F} \) in equation (17) may be neglected, i.e., the factor \((1 - R_s \bar{F})\) is replaced by 1. Similarly, for \( F \simeq 1 \), 1 - \( R_s \bar{F} \) is replaced by 1 - \( R_s \). In the case of Motif 1, \( \sigma_F \) is given by [19, 33]

\[
\sigma_F = \sqrt{p_0 (1 + b_0)}.
\]

4. Summary and discussion

Regulatory networks governing gene expression dynamics are characterised by the appearance of different types of motifs. The motifs emerged through evolutionary processes because of their functional superiority over more random substructures. In this
paper, we study a four-gene motif with dual strategies of regulation to identify the distinctive functional features which enhance its performance as a biological switch and/or its role in cellular decision making in comparison with the genetic toggle, a two-gene motif employing transcriptional regulation. The four-gene motif is a component of the gene regulatory network controlling the cell fate decision between two alternative neuronal fates in C. elegans. The motif is found to be bistable over a wide parameter regime (figure 4). Bistability is obtained on inclusion of the effect of molecular sequestration of $\sigma^E$ by its anti-$\sigma$ factor RseA [39]. We now briefly discuss some other examples of bistability without cooperativity. Bacterial two component systems regulate global responses to various types of stress. A two-component system consists of a sensor kinase (SK) and a response regulator (RR). The phosphorylated SK transfers the phosphate group to the RR which mediates the response to the stress signal. In general, the SK has a bifunctional character, it can act as a kinase in the phosphorylated state and as a phosphatase (dephosphorylating the RR) in the unphosphorylated state. The two-component system can exhibit bistability if the unphosphorylated SK and RR form a dead-end complex [40] and/or the phosphatase activity of the SK is much reduced (abolished) [41]. A synthetic circuit constructed in E. coli with autoregulating T7 RNA polymerase involves a non-cooperative positive feedback which by itself cannot generate bistability. The expression of T7 RNA slows down cell growth thus reducing protein dilution and creating an implicit positive feedback loop. The two loops, in combination, can give rise to bistability [42]. Several toxin–antitoxin operons are regulated according to the toxin/antitoxin ratio through a mechanism termed ‘conditional cooperativity’. The conditional regulation can give rise to bistability for a wide range of parameter values [43]. Additionally, stochastic effects can generate bistability/bimodal gene expression without the requirement of cooperativity in the regulation of gene expression [44, 45]. As discussed in section 2, cooperativity in the regulation of gene expression has the effect of extending the region of bistability, a feature pointed out in several earlier studies [46–48]. The ultrasensitivity in the case of post-transcriptional regulation is in the form of a linear threshold.

Figure 6. Coefficient of variation (CV) versus $p$ for Motif 1 (dashed line) and Motif 2 (solid line) for (a) $F \leq 0$ and (b) $F \geq 1$.

Figure 7. Plot of $I_{opt}$ versus $F$. The dotted line corresponds to $F = F_0 = 0.8961$ above which $I_{opt}^{TF} = 1.74$ is greater than $I_{opt}$.
behaviour [3, 25]. The four-gene motif, incorporating both transcriptional and post-transcriptional regulation also exhibits linear threshold behaviour (figure 3(a)). The theoretical prediction could be tested in the actual experimental setting of [17] or by constructing a synthetic genetic circuit, as in the case of the genetic toggle [12]. The linear threshold behaviour, by setting a threshold for activated response, prevents spurious switchings to the active state. In the latter state, the response is analog, i.e., graded. The threshold is further tunable through the modifications of the mRNA and microRNA synthesis rate constants.

The four-gene motif, because of its enlarged parameter space, exhibits a novel steady state response (figure 3(b)) which has not been reported earlier. The region of bistability separates two regions of monostability. In this region, response is digital because of the discontinuous changes in response at the bifurcation points. The response further exhibits hysteresis whereas the response in the monostable regions is reversible. The four-gene motif thus combines the advantages of the linear threshold behaviour with bistability in order to mediate cellular decision making or act as a biological switch. The linear threshold behaviour provides the basis for an ultrasensitive switch with reversible response. The genetic toggle exhibits bistability in an extended parameter region with two monostable regions separating the region bistability. In the monostable regions, the response is graded and featureless. The motifs studied in [11, 14–16] exhibit tristability and thus can function as a three-way switch in contrast to the two-way switch behaviour of the motifs studied in our paper. The two-gene and four-gene motifs studied in the paper exhibit monostability or bistability depending on the parameter regime. Motifs which include negative feedback loops and/or introduce time delays in gene expression can exhibit oscillations in protein concentrations [49, 50]. Small RNAs have been shown to establish delays and temporal thresholds in gene expression [51] as well as generate oscillations in protein concentrations [52]. Motifs governed by transcriptional regulation are also known to exhibit oscillations [49, 50]. It will be of interest to compare the functional characteristics of motifs incorporating transcriptional and post-transcriptional modes of gene regulation respectively, in generating oscillatory dynamics.

Post-transcriptional regulation mediated by microRNAs has the desirable feature of reducing the variability or noise in the target gene protein levels [3]. The reduction is mainly due to the diminished burst sizes in protein production. Reduction in protein noise, as compared to the case of unregulated gene expression, has been demonstrated experimentally in the cases of simple motifs with dual strategies of regulation of gene expression [30, 31]. A more recent study [26] combines mathematical modeling with single cell experiments to establish that microRNA-regulated gene expression results in lower intrinsic noise as compared to the case of unregulated gene expression. We compute the CV of the protein levels of a target gene in the steady state of the four-gene motif using the FD relation (10). The CV is considerably lower than that in the case of the genetic toggle (figure 5). The four-gene motif, employing dual strategies of regulation, has greater noise-reduction capability than the genetic toggle incorporating only transcriptional regulation.

In section 3 of the paper, we have compared transcriptional regulation (Motif 1) with post-transcriptional regulation (Motif 2) from an information theoretic perspective. The channel capacity \( I_{\text{opt}} \) has been calculated in the small noise approximation and considering only the intrinsic component of the total noise at the output. Figures 6(a) and (b) show the plots of CV versus \( F \) for Motifs 1 and 2 with \( F \approx 0 \) and \( F \approx 1 \) respectively. In the first (second) case, the CV of Motif 2 is lower (higher) than that of Motif 1. The parameter \( F \) is a measure of the fraction of microRNA molecules bound to the mRNAs. The CV of Motif 2 is found to be lower than that of Motif 1 over an extended range of \( F \) values. This is reflected in the plot of the channel capacity \( I_{\text{opt}} \) of Motif 2, versus \( F \) in figure 7. For \( F < F_{c} = 0.8961 \), \( I_{\text{opt}} \) is greater than \( I_{\text{opt}} \) of Motif 1. The case \( F = 0 \) corresponds to the experimental situation reported in [26]. In this limit, \( I_{\text{opt}} = 2.91 \) and \( I_{\text{opt}}^{\text{TF}} = 1.74 \). Instead of being limited to ON/OFF switching, transcriptional and post-transcriptional regulation could in principle designate \( 2^{I_{\text{opt}}} \sim 2–4 \) and \( 2^{I_{\text{opt}}^{\text{TF}}} \sim 6–8 \) distinct levels, respectively, of gene expression. Because of a higher channel capacity, post-transcriptional regulation of gene expression is more efficient in information transmission than transcriptional regulation over an extended region of \( F \) values. As \( F \) approaches the value \( F_{c} \), the difference between \( I_{\text{opt}} \) and \( I_{\text{opt}}^{\text{TF}} \) decreases but even a fractional bit difference in the channel capacity can give rise to non-trivial consequences in the ability of the motifs to appropriately respond to the input signal [53]. The computation of the channel capacity in the small noise approximation requires only a knowledge of the variance \( \sigma_{y}^{2} \) of the output protein distribution.

We have not considered the effect of extrinsic noise in the computation of \( \sigma_{y} \). As shown in [26], microRNA-regulated gene expression has decreased total noise at low expression but increased total noise at high expression as compared to the case of unregulated gene expression. The experimental results in [26] confirm the dominant contribution of intrinsic noise to the total noise over a considerable range of expression levels. In this range of expression levels, post-transcriptional regulation appears to be more efficient than transcriptional regulation in the transmission of information.
Acknowledgments

MP acknowledges the support by UGC, India, vide sanction Lett. No. F.2-8/2002(SA-I) dated 23.11.2011. IB acknowledges the support by CSIR, India, vide sanction Lett. No. 21(0956)/13-EMR-II dated 28.04.2014.

Appendix

Genetic toggle:
The differential rate equations governing the dynamics of the genetic toggle are:

\[
\frac{d[M_i]}{dt} = j_{M_i} - \delta M_i [M_i] + \frac{\beta_i}{K_i^2 + [P_i]^2} \quad (A.1)
\]

\[
\frac{d[P_i]}{dt} = j_{P_i} [M_i] - \delta P_i [P_i] \quad (A.2)
\]

\[
\frac{d[M_j]}{dt} = j_{M_j} - \delta M_j [M_j] + \frac{\beta_j}{K_j^2 + [P_j]^2} \quad (A.3)
\]

\[
\frac{d[P_j]}{dt} = j_{P_j} [M_j] - \delta P_j [P_j]. \quad (A.4)
\]

The symbols \(M_i (i = 1, 2), P_i (i = 1, 2)\) represent the concentrations of the mRNAs and proteins respectively. The first terms in the equations (A.1)–(A.4) denote unregulated synthesis rates and the second terms the degradation rates of the biomolecules. The third terms in (A.1) and (A.3) represent the mutual repression of transcriptional initiation.

The composite reactions considered are:

\[
M_1 \xrightarrow{j_{M_1}} [M_1] \xrightarrow{\delta M_1} [M_1] + 1 \quad (A.5)
\]

\[
M_1 \xrightarrow{\varepsilon_{M_1}(M_1)} \xrightarrow{\delta M_1} [M_1] \xrightarrow{\delta M_1} [M_1] + 1 \quad (A.6)
\]

\[
P_1 \xrightarrow{j_{P_1}} [P_i] \xrightarrow{\delta P_1} [P_i] + 1 \quad (A.7)
\]

\[
P_1 \xrightarrow{\varepsilon_{P_1}(P_1)} \xrightarrow{\delta P_1} [P_i] \xrightarrow{\delta P_1} [P_i] + 1 \quad (A.8)
\]

and similar reactions for \(M_2, P_2\). The elements of the reaction propensity vector \(f\) appear over the arrows. The stoichiometric matrix

\[
S = \begin{bmatrix}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
\end{bmatrix} \quad (A.9)
\]

The reaction propensity vector

\[
f = (j_{M_1} + \frac{\beta_i}{K_i^2 + [P_i]^2}) [M_i] \xrightarrow{\delta M_i} [M_i] \xrightarrow{j_{P_i}} [P_i] \xrightarrow{\delta P_i} [P_i] \\
\]

\[
\frac{d[M_1]}{dt} + \frac{\beta_1}{K_1^2 + [P_1]^2} [M_1] \xrightarrow{j_{M_1}} [M_1] \xrightarrow{\delta M_1} [M_1] \xrightarrow{j_{P_1}} [P_1] \xrightarrow{\delta P_1} [P_1] \quad (A.10)
\]

The entries over the arrows in the reaction set (A.11)–(A.32) constitute the successive elements the column vector \(f\) (reaction propensity vector). The stoichiometric matrix \(S\) is a 8 × 22 matrix the elements of which are specified by the reaction scheme (A.11)–(A.32).

References

[1] Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walters P 2002 Molecular Biology of the Cell (New York: Garland Science)
[2] Ambros V 2004 Nature 431 350
[3] Levine E and Hwa T 2008 Curr. Opin. Microbiol. 11 574
[4] Alon U 2006 An Introduction to Systems Biology: Design Principles of Biological Circuits (Boca Raton, FL: CRC)
[5] Alon U 2007 Nat. Rev. Genet. 8 450
[6] Martinez N I and Walhout A J M 2009 Bioessays 31 435
[7] Shalgi R, Lieber D, Oren M and Pilpel Y 2007 PloS Comput. Biol. 3 e131
[8] Tsang J, Zhu J and van Oudenaarden A 2007 Mol. Cell 26 753
[9] Riba A, Bosia C, El Baroudi M, Olino L and Caselle M 2014 PloS Comput. Biol. 10 e1003490
[10] Nitzan M, Shimoni Y, Rosoloi O, Margalit H and Biham O 2015 Phys. Rev. E 91 052706
[11] Jolly M K, Boarreto M, Huang B, Jia D, Lu M, Onuchic J N, Levine H and Ben-Jacob E 2015 Frontiers Oncol. 9 155
[12] Gardner T S, Cantor C R and Collins J J 2000 Nature 403 339
[13] Heinäniemi M et al 2013 Nat. Methods 10 577
[14] Lu M, Jolly M K, Levine H, Onuchic J N and Ben-Jacob E 2013 Proc. Natl Acad. Sci. 110 18141
[15] Tian X-I, Zhang H and Xing J 2013 Biophys. J. 105 1079
[16] Bracken C P, Gregory P A, Kolesnikoff N, Bert A G, Wang I, Shannon M F and Goodall G J 2008 Cancer Res. 68 7846
[17] Johnston R J, Chang S, Etchberger J F, Oritz C O and Hobert O 2006 Proc. Natl Acad. Sci. 103 12449
[18] Bowerman B 2006 Curr. Biol. 16 R1039
[19] Tkáčik G and Walczak A M 2011 J. Phys. Condens. Matt. 23 153102
[20] Rhee A, Cheong R and Levchenko A 2012 Nat. Rev. Genet. 13 577
[21] Kærn M, Elston T C, Blake W J and Collins J J 2005 Proc. Natl Acad. Sci. 102 15666
[22] Friedman N, Cai L and Xie X S 2006 Proc. Natl Acad. Sci. 103 620
[23] Raj A and van Oudenaarden A 2008 Cell 135 216
[24] Balázsi G, van Oudenaarden A and Collins J J 2011 Cell 144 910
[25] Mukherji S, Ebert M S, Zheng G Y X, Tsang J S, Sharp P A and van Oudenaarden A 2011 Nat. Genet. 43 854
[26] Schmiedel J M, Klemm S L, Zheng Y, Sahay A, Blüthgen N, Marks D S and van Oudenaarden A 2015 Science 348 128
[27] Chen D and Arkin A P 2012 Mol. Syst. Biol. 8 620
[28] Buchler N E and Louis M 2008 J. Mol. Biol. 384 1106
[29] Buchler N E and Cross F R 2009 Mol. Syst. Biol. 5 272
[30] Osella M, Riba A, Testorri A, Corá D and Caselle M 2014 Frontiers Genet. 5 345
[31] Siciliano V, Garulli I, Fracassi C, Criscuolo S, Venture S and di Bernardo D 2013 Nat. Commun. 4 2364
[32] Yu J, Xiao J, Ren X, Lao K and Xie X S 2006 Science 311 1600
[33] Thattai M and van Oudenaarden A 2001 Proc. Natl Acad. Sci. 98 8616
[34] van Kampen N G 1992 Stochastic Processes in Physics and Chemistry (Amsterdam: Elsevier)
[35] Elf J and Ehrenberg M 2003 Genome Res. 13 2475
[36] Pal M, Pal A K, Ghosh S and Bose I 2013 Phys. Biol. 10 036010
[37] Friedman N, Cai L and Xie X S 2006 Phys. Rev. Lett. 97 168302
[38] Bialek W 2012 Biophysics: Searching for Principles (Princeton, NJ: Princeton University Press)
[39] Tiwari A, Balázsi G, Gennaro M L and Igoshin O A 2010 Phys. Biol. 7 036005
[40] Igoshin O A, Alves R, Michael A and Savageau M A 2008 Mol. Microbiol. 68 1196
[41] Ram S and Goaliain M 2013 PLoS Genet. 9 e1003706
[42] Tari C, Marguet P and You L 2009 Nat. Chem. Biol. 5 842
[43] Cataudella I, Sneppen K, Gerdes K and Mitarai N 2013 PLoS Comput. Biol. 9 e1003174
[44] Lipshtat A, Loinger A, Balaban N Q and Biham O 2006 Phys. Rev. Lett. 96 181101
[45] Ochab-Marcinek A and Tabaka M 2010 Proc. Natl Acad. Sci. 107 22096
[46] Tiwari A and Igoshin O A 2012 Phys. Biol. 9 055003
[47] Huang S, Guo Y Y P, May G and Enver T 2007 Dev. Biol. 305 635
[48] Chickarmane V, Enver T and Peterson C 2009 PLoS Comput. Biol. 5 e1000268
[49] Tiana G, Krishna S, Pigolotti S, Jensen M H and Sneppen K 2007 Phys. Biol. 4 R1
[50] Tiana G and Jensen M H 2013 Phil. Trans. R. Soc. A 371 20120469
[51] Legewie S, Dienst D, Wilde A, Herzel H and Axmann I M 2008 Phys. Rev. E 78 011910
[52] Liu D, Chang X, Liu Z, Chen L and Wang R 2011 PLoS One 6 e17029
[53] Tabbas O P and Jayaprakash C 2014 Phys. Biol. 11 046004