Stochastic resonance in a self-repressing gene with transcriptional memory

Jingkui Wang, Quentin Thommen, & Marc Lefranc

Université Lille 1, Laboratoire de Physique des Lasers, Atomes, Molécules, CNRS, UMR 8523, UFR de Physique, Bât. P5, F-59655 Villeneuve d'Ascq, France
quentin.thommen@univ-lille1.fr

Abstract. Biochemical reaction networks are subjected to large fluctuations due to small molecule numbers, yet underlie reliable biological functions. Most theoretical approaches describe them as purely deterministic or stochastic dynamical systems, depending on which point of view is favored. Here, we investigate the dynamics of a self-repressing gene using an intermediate approach based on a moment expansion of the master equation, taking into account the binary character of gene activity. We thereby obtain deterministic equations which describe how nonlinearity feeds back fluctuations into the mean-field equations, providing insight into the interplay of determinism and stochasticity. This allows us to identify a region of parameter space where fluctuations induce relatively regular oscillations.

The simplest gene regulatory network is formed by a single gene which is regulated by its own protein. In the case of negative feedback (the gene is repressed by the protein), it serves as a paradigmatic genetic oscillator, which has for example been proposed as a model for the somite clock [1]. Accordingly, its dynamics has been actively investigated throughout mathematical biology [2,3,4,5,6,7,8,9]. The self-repressing gene reaction network involves four chemical species : the free gene G, RNA M, protein P and the DNA-protein complex GP. These molecular actors interact via the following biochemical reactions :

$$\begin{array}{cccc} \mathrm{G+P} & \underset{\theta}{\overset{\alpha}{\rightleftharpoons}} & \mathrm{GP} & ; \quad \mathrm{G} \xrightarrow{\lambda} \mathrm{G+M} \\ \mathrm{GP} & \xrightarrow{\mu} & \mathrm{GP+M} & ; \quad \mathrm{M} \xrightarrow{\beta} \mathrm{M+P} \\ \mathrm{M} & \xrightarrow{\delta_M} & \varnothing & ; \quad \mathrm{P} \xrightarrow{\delta_P} \varnothing \end{array} \tag{1}$$

Parameter α (resp., θ) is the DNA-protein binding (resp., unbinding) rate, δ_m (resp., δ_p) is the linear mRNA (resp., protein) degradation rate, λ (μ) is the transcription rate of unbound (resp., bound) gene and β is the translation rate.

Besides its biological significance, the self-regulating gene is interesting from a mathematical viewpoint because it combines two qualitatively different types of variables. On the one hand, proteins and RNAs may be present in high copy numbers, and can therefore be described by macroscopic variables in the large volume limit. On the other hand, the gene, which is a DNA fragment, is a single molecule. We assume that it can only be in two distinct states, bound or unbound, and therefore must be described mathematically by a binary variable. In most theoretical treatments, the difficulty of including this binary variable is circumvented by assuming that fluctuations of the gene state are fast compared to transcription, translation, and degradation. Under this aproximation, the binary gene state gene is replaced by a continuous variable which is the average occupancy time.

To explore the rich dynamics induced by fluctuations associated to gene state switching, a stochastic treatment is needed. A key point is that fluctuations interact with nonlinearities and modify the mean-field behavior, because averaging does not commute with evaluating a nonlinear function (the definition of variance is the simplest example of this fact). In particular, fluctuations may destabilize a system towards an oscillatory behavior. To capture the role of fluctuations in a simple setting, we study a model where nonlinearity only occurs in the transcriptional regulation by a monomer, RNA and protein degradation being linear. The most general probabilistic description of the chemical reaction network (1) is provided via the chemical master equation. If $P_{g,m,p}(t)$ denotes the probability to find a bound (resp., unbound)

gene, represented by $g = 0$ (resp., $g = 1$), accompanied by m ARN and p protein copies at time t, its time evolution is governed by the following master equation in the limit where θ/α is large and μ is zero :

$$\begin{aligned}\frac{d}{dt}P_{g,m,p} = (-1)^g\left[\alpha p P_{g,m,p} - \theta P_{g,m,p}\right] + \delta_{g,1}\lambda\left[P_{1,m-1,p} - P_{g,m,p}\right] + \beta m\left[P_{g,m,p-1} - P_{g,m,p}\right] \\ +\delta_m\left[(m+1)\,P_{g,m+1,p} - m\;P_{g,m,p}\right] + \delta_p\left[(p+1)\,P_{g,m,p+1} - pP_{g,m,p}\right].\end{aligned} \tag{2}$$

The asymptotic probability distribution satisfying (2) has been widely investigated [10,11] but gives only a static picture of the dynamics, averaged over time. To better understand the influence of stochastic fluctuations on the temporal dynamics of the self-repressing gene, we reformulate the master equation as an infinite hierarchy of coupled differential equations whose variables are the joint cumulants of random variables g, m and p. To be specific, the first-order and second-order joint cumulants of random variables x and y are the averages and the covariances

$$\langle x\rangle = \sum_{g,m,p} x P_{g,m,p}; \quad \Delta_{x,y} = \langle xy\rangle - \langle x\rangle\langle y\rangle,$$

while the third-order joint cumulants are defined by

$$K_{x,y,z} = \langle xyz\rangle - \langle x\rangle\langle y\rangle\langle z\rangle - \langle x\rangle\Delta_{y,z} - \langle y\rangle\Delta_{x,z} - \langle z\rangle\Delta_{x,y}.$$

By rescaling parameters and joint cumulants according to

$$\begin{aligned}&r_t = \tfrac{1}{\delta_m}; \quad r_g = 1; \quad r_m = \tfrac{\delta_p\theta}{\beta\alpha} \quad r_p = \tfrac{\theta}{\alpha}; \quad \Theta = r_t\,\theta; \quad \Lambda = \lambda\tfrac{r_t}{r_m} \quad \delta = \tfrac{\delta_p}{\delta_m}, \\ &T = r_t\,t; \quad X = r_x\,\langle x\rangle; \quad \Delta_{X,Y} = r_x\,r_y\Delta_{x,y}; \quad K_{X,Y,Z} = r_x\,r_y\,r_z K_{x,y,z} \qquad x,y,z \in \{g,m,p\};\end{aligned} \tag{3}$$

the following normalized time evolution equations for averages and covariances are obtained :

$$\begin{aligned}&\tfrac{d}{dT}P = \delta\,(M-P); && \tfrac{d}{dT}\Delta_{P,P} = 2\delta\,(\Delta_{P,M} - \Delta_{P,P}); \\ &\tfrac{d}{dT}M = \Lambda G - M; && \tfrac{d}{dT}\Delta_{M,M} = \Lambda\left[2\Delta_{G,M} + \mu G\right] - 2\Delta_{M,M}; \\ &\tfrac{d}{dT}G = \Theta\,(1 - G - GP - \Delta_{G,P}); && \tfrac{d}{dT}\Delta_{P,M} = \Lambda\Delta_{G,P} - (\delta+1)\,\Delta_{P,M} + \delta\Delta_{M,M}; \\ &\tfrac{d}{dT}\Delta_{G,P} = \delta\left[\Delta_{G,M} - \Delta_{G,P}\right] - \Theta\left[K_{G,P,P} + G\Delta_{P,P} + (P+1)\,\Delta_{G,P}\right] \\ &\tfrac{d}{dT}\Delta_{G,M} = \Lambda G\,(1-G) - \Delta_{G,M} - \Theta\left[K_{G,M,P} + G\Delta_{P,M} + (P+1)\,\Delta_{G,M}\right].\end{aligned} \tag{4}$$

Since G is a binary variable, $\Delta_{G,G} = G(1-G)$ and is slaved by G. Because of the nonlinearity, Eqs. (4) do not form a closed system as the time derivatives of joint cumulants involve joint cumulants of higher order. In particular, Eqs. (4) do not constrain the two third-order joint cumulants $K_{G,M,P}$ and $K_{G,P,P}$ which are unspecified. As is well known, the simplest way to truncate the hierarchy of moment equations is to set all second-order cumulants to zero [12]. This neglects all fluctuations and leads to deterministic rate equations for the time averages. Here, these equations predict that only stable stationary states can be observed in the system studied.

Incorporating fluctuations in the dynamics requires truncating the hierarchy at a higher order. As a first approximation, it would seem natural to set third-order joint cumulants to zero, which amounts to represent the stochastic variables by Gaussian probability distribution functions whose means and variances vary with time and interact with each other. In our case, however, this is not necessarily a correct approach because of the coexistence of macroscopic variables (g and p) interacting with a binary (microscopic) variable (g). In the following, we consider two limiting cases depending on the value of the gene response time.

In the first case, we assume that the unbinding and binding rates are very large compared to other dynamical rates and keep their ratio constant ($\Theta \to \infty$, θ/α constant). The gene remains bound or unbound for very short amounts of time, during which mRNA and proteins copy numbers can be considered as constant. RNA and protein levels keep a memory of many previous state switching cycles, and reach a stationary state with an expected gaussian distribution. In this case the third-order cumulants $K_{G,M,P}$

and $K_{G,P,P}$ vanish so that Equations (4) become closed. In the limit where the overall transcription rate is large ($\Lambda \gg 1$), the stationary state is given by

$$G^* \simeq \Delta^*_{GY} \simeq \frac{1}{\sqrt{\Lambda}}\left(1+\frac{\delta}{2\Theta\,(1+\delta)}\right); \qquad \Delta^*_{GX} \simeq \frac{-1}{\sqrt{\Lambda}}\left(1+2\frac{\delta}{\Theta\,(1+\delta)}\right); \tag{5a}$$

$$P^* \simeq M^* \simeq \Delta^*_{PP} \simeq \Delta^*_{MP} \simeq \Delta^*_{MM} \simeq \sqrt{\Lambda}\left(1+\frac{\delta}{2\Theta\,(1+\delta)}\right). \tag{5b}$$

where we include the correction to first order in Θ^{-1}. An interesting finding is that this correction only depends on the combination $\Theta(1+\delta)/\delta$. A linear stability analysis then indicates that the stationary state is always stable, in agreement with the rate equation approximation.

Conversely, let us assume that the gene reacts infinitely slowly ($\Theta \to 0$). The dynamics is then driven by the gene jumping between two states according to a Poisson process. When the gene is active ($G_{ON} = 1$), protein and RNA levels quickly converge to high level states $M_{ON} = P_{ON} = \Lambda$. When the gene is inactive, protein and RNA levels are low ($M_{OFF} = P_{OFF} = 0$). At the end of an ON/OFF cycle, the system is always in the same state, with no memory of previous cycles. Protein temporal profiles feature a sequence of spikes, distributed in time according to a Poisson process.

In this limit case, averages, covariances and third-order joint cumulants can easily be derived thanks to the high correlation between variables. The gene is active during a time $t_{ON} = 1/(\theta\Lambda)$ and inactive during a time $t_{OFF} = 1/\theta$ so that its average activity is $G^* = t_{ON}/(t_{ON}+t_{OFF}) = 1/(1+\Lambda)$ and $P^* = M^* = G^*\Lambda$. If the transcription rate is large ($\Lambda \gg 1$), the different variables scale according to :

$$G^* \simeq \Lambda^{-1}; \quad P^* = M^* = \Delta^*_{G,P} = \Delta^*_{G,M} \simeq 1; \quad \Delta^*_{P,P} = \Delta^*_{M,M} = \Delta^*_{P,M} = K^*_{G,P,P} = K^*_{G,M,P} \simeq \Lambda. \tag{6}$$

If we further assume that $\Theta\Lambda \ll \delta, 1$, the following reduced system uncouples from the other equations, regardless of whether the third-order joint cumulants are vanishing or not :

$$\begin{aligned} \tfrac{d}{dt}P &= \delta\,(M-P); & \tfrac{d}{dt}\Delta_{G,P} &= \delta\,(\Delta_{G,M}-\Delta_{G,P});\\ \tfrac{d}{dt}M &= \Lambda G - M; & \tfrac{d}{dt}\Delta_{G,M} &= \Lambda G\,(1-G) - \Delta_{G,M}; \end{aligned} \tag{7}$$

$$\tfrac{d}{dt}G = \Theta\,(1-G-GP-\Delta_{G,P}).$$

Besides the averages, this system involves the covariances of the gene state variable with protein and mRNA levels. The stationary state of Eqs. (7) is given by $G^* = \Lambda^{-1}$, with all other variables equal to 1, thus satisfying scaling (6) exactly. The dynamical behaviour of Eqs. (7) can be studied by carrying out a stability analysis around the fixed point. It reveals that unlike with the mean-field equations, the system exibits a Hopf bifurcation leading to oscillatory behaviour, revealing that it can be destabilized by the stochastic fluctuations. Under the approximation $\Lambda \gg 1$, the oscillation criterion is simply

$$\mathcal{H}(\Theta) = 4\Theta^2 + \Theta\left[2\,(1+\delta) - \frac{\delta}{(1+\delta)}\Lambda\right] + \delta \;< 0. \tag{8}$$

If the gene is infinitely slow ($\Theta \to 0$), the criterion is never satisfied because $\mathcal{H}(0) = \delta > 0$. For intermediate values of Θ, the quantity $\mathcal{H}$ can be become negative provided Λ is high enough and δ is not too large compared to 1. However this can only occur when $\Lambda\Theta \geq \delta + 1$, which a priori conflicts with the assumptions under which the reduced model (7) has been derived. We therefore carry out stochastic simulations of the chemical model (1) to assess the relevance of the two truncations of the moment equation hierarchy considered here : Eqs. (7) or Eqs. (4) with third-order cumulants set to zero.

We first study how the average value of the gene activity as a function of Θ is reproduced by the two truncations. Fig. (1a) shows how this average depends on Θ and δ due to the presence of fluctuations, an effect which is not captured by the mean-field equations. An important result is that the reduced parameter $\Theta\left(1+\delta^{-1}\right)$ which appeared in (5) is indeed the main parameter controlling the fixed point location throughout the range of Θ explored, as curves obtained for various values of θ and δ superimpose relatively well. The limit values predicted by (6) and (5) for the two extreme regimes $\Theta \to 0$ and $\Theta \to \infty$,

respectively, are indeed recovered. Fig. 1b displays the average gene activity predicted from the fixed point of Eqs. (4) with the third-order cumulants set to zero. The agreement with stochastic simulations is fairly good : the limit values in Fig. 1a and Fig. 1b are identical, the same global shape with a maximum around $\Theta = 1/(1+\delta^{-1})$ is observed and the evolution of the maxima with δ is correctly reproduced. The main discrepancy is that the transition from the fast to the slow gene regime is more abrupt in Fig. 1b than in Fig. 1a. In contrast to this, the fixed point of Eqs. (7) does not depend on Θ nor δ, thus incorrectly predicts a constant gene activity (Fig. 1c). At this stage, the model (4) with vanishing third-order cumulants correctly captures the effect of fluctuations but not the simpler model (7).

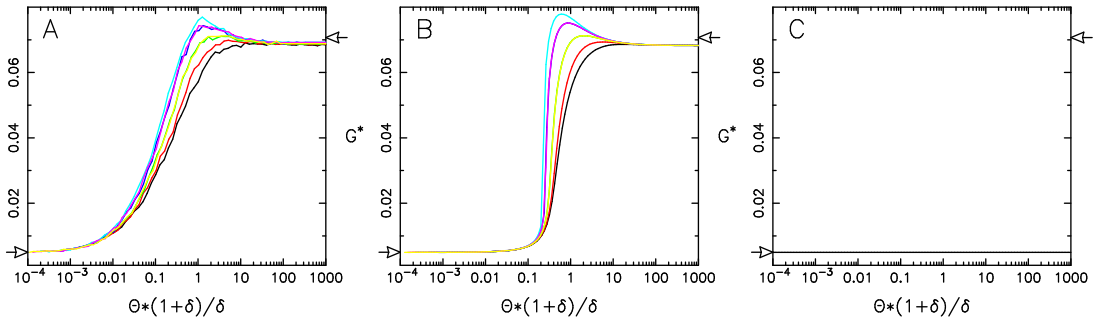

Figure 1. Average gene activities as a function of the reduced parameter $\boldsymbol{\Theta}\left(\mathbf{1}+\delta^{-\mathbf{1}}\right)$ (a) Numerical estimation using stochastic simulations with parameter values : $\theta/\alpha = 100$, $\Lambda = 200$, $\lambda = \beta$. Each color corresponds to a given value of δ, varying from 10^2 (black) to 10 (cyan). (b) Fixed point value of gene activity in model (4) with vanishing third-order joint cumulants. (c) Fixed point value of gene activity in model (7). Arrows indicate the gene activity limiting values given by (5) for large Θ and by (6) for small Θ.

Let us now consider the dynamics of protein level fluctuations. Without feedback, gene switching is a purely Poissonian process. Protein levels follow the gene state with a characteristic time scale determined by protein and mRNA lifetimes. With feedback, the probability of switching evolves rapidly in time as protein levels increase (gene is active) or decrease (gene is inactive). This feedback may reduce the stochasticity arising from gene switching, with protein peaks occur more regularly.

The regularity of a stochastic oscillatory behavior is often quantified using a temporal autocorrelation function, which is sensitive to reproducibility both in time and in amplitude. However, temporal regularity is certainly more relevant than amplitude regularity for biological protein signals. The highly nonlinear response of many signaling cascades can protect them against fluctuations in amplitude, for example by saturating output above an input threshold. A standard technique for assessing temporal regularity is to divide the state space into two regions I and II and to study the distribution of the times where the system leaves I to enter II. It is often useful to require a minimal excursion in region II to avoid spurious transitions induced by noise. Here, we detect events where the protein level crosses successively the mean protein level P^* and the $P'^* = P^* + 0.25\sqrt{\Delta^*_{pp}}$ level before falling back below the mean protein level.

Given the list of times where the system transits from low to high protein levels, we compute the probability of detecting n transitions within a time interval of fixed duration. The probability distribution is then characterized by its variance to mean ratio (Fano factor). This method is inspired by how the temporal distribution of photons from a light source is generally characterized, with the event of interest being a photon detection. A Fano factor close to unity indicates that transition times follow a Poison distribution. A Fano factor greater (less) than unity indicates a super-Poissonian (resp., sub-Poissonian) distribution corresponding to a bunching (resp., anti-bunching) of transition events. Transition anti-bunching can be viewed as a stochastic counterpart of deterministic oscillations.

Fig. 2 shows stochastic simulations of the chemical reaction network (1) for a slow, an intermediate and a fast gene, as well as the probability distribution of the number n of transitions within a given time window. As expected, protein spikes in the slow gene case (Fig. 2a) are slaved to the switching process, leading to a Poisson probability distribution for n (Fig. 2d) and accordingly a unity Fano factor. In the

intermediate gene response time case (Fig. 2b), protein spikes are mostly antibunched (see red circles). The probability distribution of spike number is gaussian-like (Fig. 2e), the Fano factor being around 0.35. This anti-bunching degrades in the case of a fast gene (Fig. 2c) with the Fano factor rising to 0.5. Thus, we oberve a resonance effect involving the time in which the gene responds to protein variations and the time during which previous gene states are remembered, which is controlled by the protein and RNA decay rates.

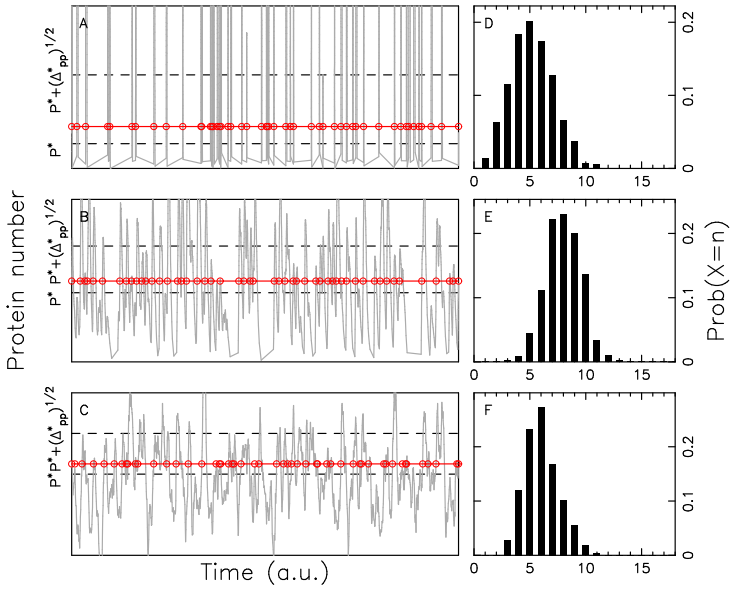

Figure 2. Dynamical behaviour. Time evolution of protein copy number for $\Lambda = 200$, $\delta = 1$ and $\theta = 5.10^{-3}$ (a), $\theta = 0.5$ (b), $\theta = 500$ (c). The dashed lines indicate mean protein level and mean protein level plus variance. Red lines correspond to the high trigger level and spiking events are indicated by red circles. Probabilities of observing n spikes during a given time window for each of the three regimes (d)-(f).

We have studied systematically how the Fano Factor depends on the gene unbinding rate Θ and the relative protein decay rate δ in stochastic simulations of reaction network (1). As Fig. (3) shows, regularity of protein spikes is enforced when (1) the decay rates δ_p and δ_m are comparable ($\delta \sim 1$) and (2) the reduced parameter $\Theta * \left(1 + \delta^{-1}\right)$ is close to unity. Quite remarkably, the parameter space region where protein spikes are more regularly spaced is extremely well approached with the region where the reduced model (7) displays deterministic oscillations. This suggests that this model captures well the dynamical interaction of mean-field variables and fluctuations, although it did not reproduce satisfactorily the average gene activities in Fig. 2. This probably indicates that the dynamically important joint cumulants are those involving the gene state variable. This is not surprising given that gene state remains binary in all limits and is thus the most stochastic variable.

Conclusion

In conclusion, we have studied the stochastic time evolution of the self-repressing gene and characterized the regularity of protein spikes using a Fano-like indicator. This allowed us to evidence a dynamical resonance phenomenon where a more regular time evolution of protein concentration is observed for certain values of the relative protein degradation rate and of the gene response time. Average quantities, on one hand, and the location of the resonance in parameter space, on the other hand, are reproduced separately by two reduced deterministic models obtained from a truncation of the moment equations

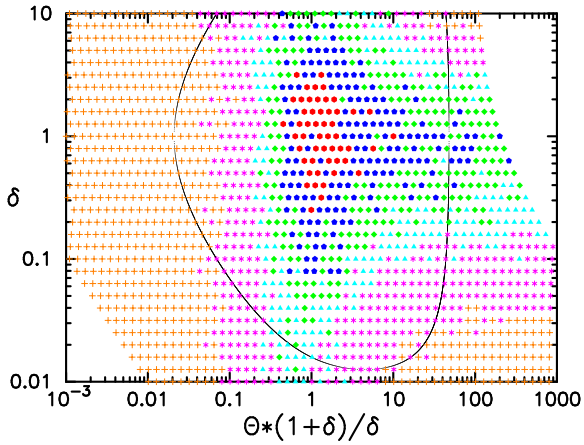

Figure 3. Fano factor. Dependency of the Fano factor F quantifying spiking regularity on δ and $\Theta * \left(1+\delta^{-1}\right)$. Stochastic simulations of network (1) have been carried out with $\Lambda = 200$; $\beta = \lambda$; $\theta/\alpha = 100$. Different values of the Fano factor F are indicated by red hexagons ($F < 0.35$), blue pentagons ($0.35 < F < 0.4$), green diamonds ($0.4 < F < 0.45$), cyan triangles ($0.45 < F < 0.5$), magenta stars ($0.5 < F < 0.7$), and orange crosses ($0.7 < F$). The black line encloses the region where the reduced model (7) oscillates.

hierarchy. It remains to combine these two models to reach a global description of how averages and fluctuations interact through nonlinearity in the self-repressing gene circuit.

Références

1. H. Hirata, S. Yoshiura, T. Ohtsuka, Y. Bessho, T. Harada *et al.*, Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop, *Science*, **298**, 840 (2002)
2. B. C. Goodwin, Oscillatory behavior of enzymatic control processes, *Advances in Enzyme Regulation*, **3**, 425 (1965)
3. J. S. Griffith, Mathematics of cellular control processes I. negative feedback to one gene., *Journal of Theoretical Biology*, **20**, 202 (1968)
4. R. D. Bliss, P. R. Painter & A. G. Marr, Role of feedback inhibition in stabilizing the classical operon, *Journal of Theoretical Biology*, **97**, 177 (1982)
5. A. Goldbeter, A model for circadian oscillations in the Drosophila period protein (per), *Proceedings of the Royal Society of London B*, **261**, 319 (1995)
6. J.-C. Leloup, D. Gonze & A. Goldbeter, Limit cycle models for circadian rhythms based on transcriptional regulation in Drosophila and Neurospora, *Journal of Biological Rhythms*, **14**, 433 (1999)
7. M. H. Jensen, K. Sneppen & G. Tiana, Sustained oscillations and time delays in gene expression of protein hes1, *FEBS Letters*, **541**, 176 (2003)
8. N. A. M. Monk, Oscilatory expression of hes1, p53 and nk-κb driven by transcriptional time delays, *Current Biology*, **13**, 1409 (2003)
9. J. Lewis, Autoinhibition with transcriptional delay : a simple mechanism for the zebrafish somitogenesis oscillator, *Current Biology*, **13**, 1398 (2003)
10. J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak *et al.*, Self-regulating gene : An exact solution, *Physical Review E*, **72**, 051907 (2005)
11. D. Lepzelter, H. Feng & J. Wang, Oscillation, cooperativity, and intermediates in the self-repressing gene, *Chemical Physical Letters*, **490**, 216 (2010)
12. N. G. van Kampen, *Stochastic processes in physics and chemistry* Elsevier (2007)