CONTROL OF STOCHASTIC GENE EXPRESSION BY A NONLINEAR BIOLOGICAL OSCILLATOR

Federica Colombo
Division of Genetics and Cell Biology
San Raffaele Scientific Institute
Italy
colombo.federica@hsr.it

Nacho Molina
Institut de Génetique et de Biologie Moléculaire et Cellulaire
France
nacho.molina@igbmc.fr

Samuel Zambrano
Division of Genetics and Cell Biology
San Raffaele University and Scientific Institute
Italy
zambrano.samuel@hsr.it

Abstract

Understanding the dynamical processes by which cells translate external stimuli into an adequate response is a fundamental problem in biology that can be addressed through the methods of physics and nonlinear dynamics. Transcription activators play a key role in this system: they are proteins that become active under certain external stimuli and trigger the expression of genes encoding for the proteins required by the cells to provide an adequate response. Their activity is commonly controlled by different negative feedbacks that regulate the duration and strength of their activations. This can give rise to pulses in the activity of the transcription activator and these systems can be referred to as "genetic oscillators": a paradigmatic example is NF-κB, that can be modelled as a nonlinear stochastic oscillator. We recently showed that a simplified model of this genetic circuit can be analyzed combining ideas from dynamical systems theory and stochastic processes. Here we use those tools to show that it can also reproduce the dynamical patterns of gene expression that this genetic oscillator can produce. Furthermore, we use our approach to characterize the dynamics of our simple model in cancer cells, where the tight regulation of this circuit is lost by the effect of mutations, and show how the dynamical patterns of gene expression are disrupted. These insights can have implications in cancer biology.

Key words

Control of molecular systems, Control of oscillations, Stochastic systems.

1 Introduction

The central dogma of molecular biology establishes that each gene is transcribed into an (m)RNA, which is then translated into a protein: these are the fundamental steps of what we call here *gene expression*. However, not all the genes can be expressed simultaneously at arbitrarily high level by a cell, partly because this process requires energy and this is necessarily a limited resource, partly because cells need different proteins (which are the molecules that eventually perform the different tasks required by cells) at different times, depending on their function (e.g. a neuron versus a keratinocyte) or in response to different external signals.

For this reason, it is known that cells control gene expression through finely regulated genetic circuits, that are called this way in analogy with the electronic circuits. The regulation in these circuits relies on the fact that some genes can be activated and repressed by certain proteins, which are called transcription *activators* and *repressors* respectively. Some of those genes can themselves encode for proteins that are either activators or repressors, or can regulate the activity of such proteins, giving rise to different positive and negative feedbacks and regulations. Indeed, it has been shown that transcription activators and repressors form a network of mutual activation-repression with motifs of interconnection that enable the cell to perform different functions [Alon 2007].

In recent years, thanks to live cell imaging techniques which allow us to look at the evolution in time of the amount of selected proteins in single cells, it is becoming clear that these circuits can display a rich dynamics, including oscillators, even under relatively simple stimuli[Levine, Lin and Elowitz 2013]. Since activation and inactivation of genes is governed by a small number of molecules (we have typically two copies of each gene and few hundreds of RNAs produced per gene) that operate amidst strong thermal noise, gene expression is essentially a stochastic process [Elowitz, Levine, Siggia and Swain 2002]. A full understanding of how genetic circuits operate in this context and are able to provide the right gene expression for a given stimulus is a fundamental task.

We have been interested in the transcription factor NF-κB. This is a transcriptional activator regulated by a negative feedback: NF-κB controls also the expression of an inhibitor that makes it inactive by relocating it into the cytoplasm of the cell [Hoffmann, Levchencko,

Scott and Baltimore 2002] (recall that the DNA, and hence the genes, is in the cell nucleus). Live cell imaging shows that this mechanism produces oscillations in the nuclear concentration of NF-κB [Zambrano, Bianchi and Agresti 2014]. We [Zambrano, Bianchi, Agresti and Molina 2015] recently proposed a simple model of this genetic circuit and showed the importance of stochasticity of gene activation in generating pulses of activity, that we can refer to loosely as oscillations (even if we do not find a limit cycle in this model). In parallel, using a combination of biochemical and microscopy approaches, we experimentally observed that different genes activated by NF-κB have different expression dynamics: some transcripts will follow the oscillations of the activator NF-κB, some will just increase to reach an asymptotic state, and some will display a behavior between these two extremes [Zambrano, De Toma, Piffer, Bianchi and Agresti 2016]

A question that remains open is whether our simplified stochastic model can produce patterns of gene expression similar to those observed experimentally. On the other hand, we are currently interested in how the NF-κB pathway is deranged in different cancer models, such as in multiple myeloma. In cancer, typically mutations in different genes lead to new proteins that are either inactive, or not functional, or in some cases hyper-active. In this particular type of cancer, the related mutations in the circuit regulating NF-κB have the effect of hyperactivating NF-κB upon external stimulation [Annunziata, Davis, Demchenko et al. 2007]. Such hyperactivation is correlated with the overexpression of all the genes that are under the control of this activator, in particular those related with cell survival and proliferation. Here we show that our model is able to reproduce qualitatively such behaviour, providing insights on its origin from a dynamical systems point of view and leading to predictions on how gene expression might be disrupted in this system.

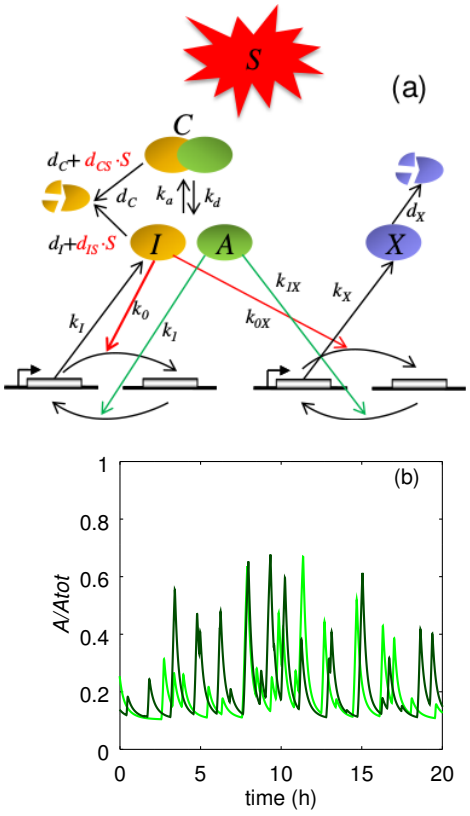

Figure 1. (a) A scheme of the model of the genetic circuit considered here. The activator A and the inhibitor I do also control the expression of the gene G_X. (b) Two examples of the evolution in time of the activator A using simplified hybrid simulations. The initial parameters are set as follows: $k_a = 5.1 \cdot 10^{-6} s^{-1}$; $k_d = 1.4 \cdot 10^{-3} s^{-1}$; $d_I = 2.1 \cdot 10^{-5} s^{-1}$; $d_C = 4.2 \cdot 10^{-5} s^{-1}$; $d_{IS} = 2.6 \cdot 10^{-9} s^{-1}$; $d_{CS} = 1.1 \cdot 10^{-8} s^{-1}$; $k_I = 1.1 \cdot 10^{1} s^{-1}$; $d_X = 2.1 \cdot 10^{-5} s^{-1}$; $k_X = 1.1 \cdot 10^{1} s^{-1}$; $k_1 = 5.8 \cdot 10^{-7} s^{-1}$; $k_{1X} = 5.8 \cdot 10^{-7} s^{-1}$; $k_0 = 1.4 \cdot 10^{-7} s^{-1}$; $k_{0X} = 1.4 \cdot 10^{-7} s^{-1}$. $A_{tot} = 10^4$ and $S = 10^5$.

2 Hybrid model of a biological oscillator

The model that we consider in this paper is shown in Fig. 1 (a), and was already described in detail in [Zambrano, Bianchi, Agresti and Molina 2015]. For the sake of completeness, we describe briefly here the processes involved and, for each of them, we provide in parenthesis the name of the biochemical rate that regulates how fast they take place. Letters here represent different biochemical species, but also the amount (in number of copies) of each of them.

The core of the model is formed by a transcription activator A (or simply an *activator*) analogous to NF-κB, hence able to activate the transcription of genes, and by an inhibitor I. Such inhibitor can bind A (k_a) and form a complex C that cannot activate (neither repress) the activation of the genes. Spontaneous dissociation of the complex (k_d) is unlikely but also possible. The total amount of the activator A_{tot} (free plus bound to the inhibitor) is considered to be constant in analogy with NF-κB [Zambrano, Bianchi and Agresti 2014]. Mathematically, $C + A = A_{tot}$

The protein I is synthetized (k_I) when the gene is in its active state G, but not when it is inactive $\bar{G}$. Notice that, for the sake of simplicity, we skip the transcription process. As for NF-κB, the negative feedback in this system arises since the activator can activate G (k_1). Importantly, the inhibitor can also lead to the direct inactivation of G (k_0). The inhibitor spontaneously degrade, both when free (d_I) and when forming the complex with A (d_C).

We simplify the modelling of the complex process by which an external signal leads to the activation of the transcription activator by simply assuming that the degradation of the inhibitor in the complex C can be enhanced by the presence of the signal S (d_{CS}), a process that also leads to an additional degradation of the free inhibitor (d_{IS}). Hence, when an external signal S appears, the transcription activation is set free and able to start doing its job. We assume that the concentration

of S is constant in time.

Due to the fact that there is a discrete number of biochemical species and that biochemical reactions take place with a probability that depends on the abundance of the reactants, the classical way to gain insights on the dynamics of this type of system is to use a stochastic formalism [Gillespie 1977]. However, as we showed in [Zambrano, Bianchi, Agresti and Molina 2015], an alternative approach is to use what we call a *hybrid model*, by which we model using ordinary differential equations the biochemical species with high numbers of copies, whereas we use stochastic processes for those with low copy numbers. Applied to this system, this means that since the species A and I come in high copy numbers their evolution in time can be considered continuous and the dynamics can be described by these differential equations:

$$\frac{dA}{dt} = -k_a \cdot A \cdot I + (k_d + d_C + d_{CS} S) \cdot (A_{\text{tot}} - A) \quad (1)$$

and

$$\frac{dI}{dt} = -k_a \cdot A \cdot I + k_d \cdot (A_{\text{tot}} - A) - (d_I + d_{IS} S) I + k_I \cdot G(t), \quad (2)$$

two equations that are driven by the stochastic process $G(t)$. This process reproduces the evolution in time of the gene, of which we have just one copy and cannot be faithfully reproduced by a continuous variable, so we allow it to switch between 1 (G) and 0 ($\bar{G}$) according to the rule:

$$\bar{G} \underset{k_0 I(t)}{\overset{k_1 A(t)}{\rightleftarrows}} G. \quad (3)$$

We showed [Zambrano, Bianchi, Agresti and Molina 2015] that this kind of hybrid simulation gives results that are very close to those of the full stochastic simulation (i.e., when all the biochemical reactions are considered stochastic processes and not just the gene switching) for a wide parameter range around those used by default (given in the caption of Fig. 1). On the other hand, it also allows us to interpret the evolution in time of the system using tools of the dynamical systems theory, as we will also see below. In Fig. 1 (b) we can see the reported stochastic spiky pulses of A. Although strictly speaking the system is not an oscillator, since it lacks a limit cycle, we will refer to this dynamics in what follows either as pulses or as oscillations.

The first question that we want to address now is how this system will regulate the dynamics of a gene X under the control of A and I.

3 Gene expression dynamics

The situation that we want to consider now is the one represented in Fig.1(a), in which a prototypical gene

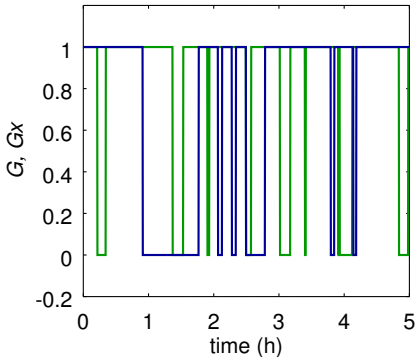

Figure 2. The gene state of G (green line) and G_X (blue line). Even if the rates of activation and inactivation are the same ($k_1 = k_{1X}; k_0 = k_{0X}$), the on-off switching of the two genes will differ due to the intrinsic stochasticity of gene expression.

G_X is under the control of the activator A and the inhibitor I. This is the typical situation for hundreds of genes, for example, in the NF-κB system [Zambrano, De Toma, Piffer, Bianchi and Agresti 2016]. In our case, this means that the gene is activated by A and inactivated by I. As the inhibitor, X will be produced at certain rate k_X and spontaneously degrade at a rate d_X. Since in principle X comes also in high copy numbers, its dynamics will be well approximated by a continuous variable, so we can model its dynamics using our hybrid simulation approach as

$$\frac{dX}{dt} = -d_X X + k_X \cdot G_X(t). \quad (4)$$

where $G_X(t)$ will switch stochastically between 1 (G_X) and 0 ($\bar{G_X}$) in a process governed by the dynamics of A and I according to the rules

$$\bar{G_X} \underset{k_{0X} I(t)}{\overset{k_{1X} A(t)}{\rightleftarrows}} G_X. \quad (5)$$

Notice that the gene G and G_X will switch differently even if the rates of activation and inactivation of these genes are exactly the same (Fig. 2): this is part of the inherent stochasticity in gene expression in single cells [Elowitz, Levine, Siggia and Swain 2002]. On the other hand, starting from low values, it will reach at most a value of $max(X) = k_X/d_X$.

We find that it is possible to obtain different gene expression dynamics, acting on the deterministic part of the equation (4) by changing the degradation parameter d_X, which is the degradation of X. The X dynamics shows different dynamical patterns of expression (Fig. 3): in Fig. 3 (a) we see an example of a gene displaying large oscillations, whereas in Fig. 3 (b) oscillations cover a lower range of value which is much smaller in Fig. 3 (c) (notice the change in the y-axis scale). This

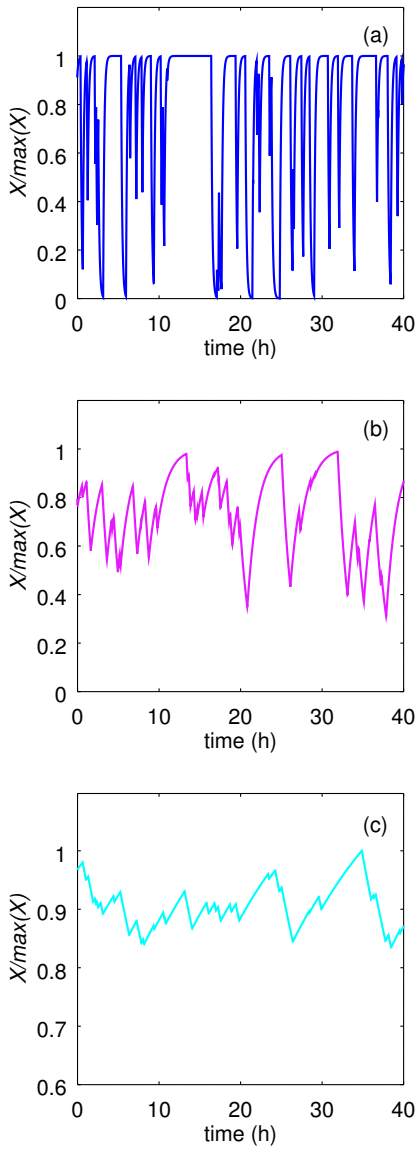

Figure 3. Simulated X dynamics using our hybrid model. The model gives different dynamical patterns. Simulations were performed by keeping all the parameters equal to those regulating the expression of the gene G, but changing d_X: (a) $d_X = 2.1 \cdot 10^{-3} s^{-1}$; (b) $d_X = 2.1 \cdot 10^{-4} s^{-1}$; (c) $d_X = 2.1 \cdot 10^{-5} s^{-1}$. Notice that X is normalized with the maximum amount of X produced during each simulation ($max(X)$=$\frac{k_X}{d_X}$).

is qualitatively similar to the patterns experimentally observed in [Zambrano, De Toma, Piffer, Bianchi and Agresti 2016], where a different mathematical model also suggested that degradation was the key parameter to produce the different gene expression patterns. This result allows us to gain confidence on the ability of the model to reproduce the behaviours observed in "healthy" conditions. We will now address how the dynamics of the activator changes in pathological conditions, when certain network parameters are changed, and how it affects the expression of genes governed by the same parameters as those in Figs. 3 (a)-(c).

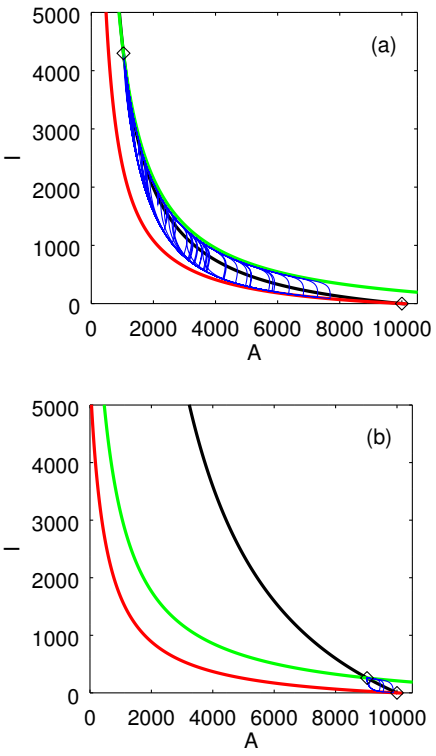

Figure 4. Nullclines for $dA/dt = 0$ (black line) and for $dI/dt = 0$ when $G(t) = 1$ (green line) and when when $G(t) = 0$ (red line), as well as typical trajectories (blue thin line). The nullclines intersect giving rise to two fixed points in each situation (black diamonds). The nullclines and trajectories are plotted for (a) normal and (b) pathological parameters. It is possible to see that the activator makes wide oscillations between the fixed points for the normal condition and smaller oscillations around high A values in the pathological condition.

4 Oscillations and gene expression in pathological conditions

It is possible now to describe how the dynamics of the network are modified in a pathological situation. As an example, for the transcription factor NF-κB, it is known that certain mutations lead to an excessive activation of the system upon an external signal S [Annunziata, Davis, Demchenko et al. 2007]. This is due to the presence of mutations that typically increase the activity of the proteins that trigger the activation of NF-κB when certain signals are detected by the cell receptors. The only straightforward way to reproduce this with our model is just to consider that the rates d_{CS} and d_{IS}, those responsible for the degradation of the inhibitor due to an external signal, would be strongly increased.

As proposed [Zambrano, Bianchi, Agresti and Molina 2015], to understand the effect of such increase, we can look at the nullclines of the dynamical system given by eqs. (1) and (4) i.e. the combinations of A and I for which $\frac{dA}{dt} = 0$ and $\frac{dI}{dt} = 0$. Notice that we have three of such nullclines, since there are two possible

nullclines for (4): the one for which $G(t)$=1 and the one for which $G(t) = 0$. For the parameter of the "healthy" conditions, those considered in the previous sections, we can see that this set of three nullclines gives two fixed points, see Fig. 4 (a), and the system essentially switches between the two fixed point as $G(t)$ switches between 0 and 1. This gives the pulsed behavior of A that we can observe in Fig. 1.

We can now simulate the pathological condition by multiplying d_{CS} and d_{IS} by a factor 10 and observe how this is reflected in the nullclines in Figure 4 (b). In particular, it is evident that the nullclines move in such a way that the two new fixed points between which the trajectories switch as the gene switches between active and inactive state become closer and around a value of A close to A_{tot}, the maximum value that the free activator A can reach. A comparison of the dynamics for both type of situations is provided in Fig. 5 (a): the pulsed physiological dynamics of the activator are wider (Fig. 5 (a), green line) compared to the pathological conditions (Fig. 5 (a), red line) where oscillations are mainly lost. The long periods between the activation and inactivations in the latter case are due to the fact that high levels of A and low levels of I of these new fixed points make the gene inactivation very unlikely as compared to gene activation.

We can now explore the consequences of changes in the dynamics of our genetic circuit in the dynamical patterns of gene expression that it produces. To do so, we use the same parameters k_{1X}, k_{0X}, k_X and d_X that gave rise to the 3 patterns of gene expression in the "healthy" conditions (obtained by changing the degradation rate d_X) shown in Fig. 3, but now we consider their evolution when A and I evolve as in the pathological condition. The results are shown in Fig. 5 (b). Compared to Fig. 3, we can see that the oscillating dynamics are almost lost and all the gene expression values tend to the maximum amount of X synthetized in each simulation. As already highlighted in the pathological nullclines, the higher probability of having gene activation rather than gene inactivation has an impact also on X synthesis, that escapes from the negative feedback control of the genetic circuit and is kept at high values.

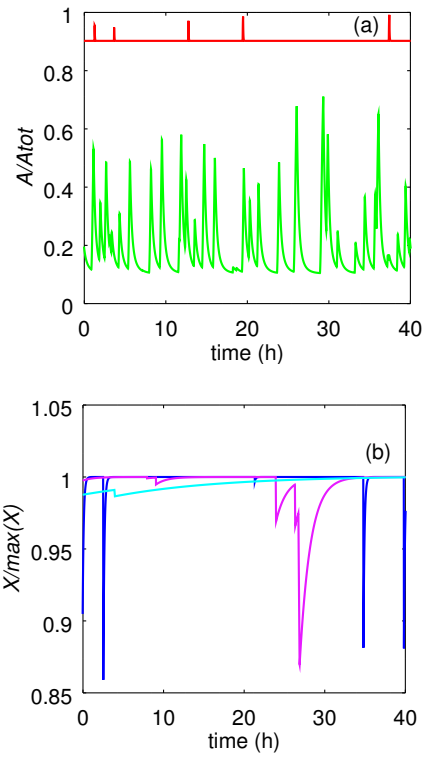

Figure 5. (a) Pulsed physiological dynamics of the activator A (green line). The oscillating behavior is due to the stochastic switching of $G(t)$ from 0 and 1. In pathological conditions (red line), simulated with the increase of the rates d_{CS} and d_{IS} by an order of magnitude, the oscillations of A are mainly lost and its value is always around the maximum value, as predicted by the nullclines analysis. (b) Simulated X dynamics by our hybrid model in a pathological cell. Simulations were performed by imposing a 10 fold increase of d_{CS} and d_{IS} and keeping the same parameteres for the gene expression as in Fig 3 (blue line= $d_X = 2.1 \cdot 10^{-3} s^{-1}$; purple line= $d_X = 2.1 \cdot 10^{-4} s^{-1}$; cyan line= $d_X = 2.1 \cdot 10^{-5} s^{-1}$). The different patterns of gene expression observed in healthy conditions are lost, giving just small fluctuations below the maximum value (notice the y-axis scale).

5 Conclusions

Starting from a description of our hybrid model of a simple biological oscillator with a negative feedback [Zambrano, Bianchi, Agresti and Molina 2015], we have shown that the expression of a gene controlled by our activator-inhibitor system can follow different patterns, which resemble the ones experimentally for NF-κB [Zambrano, De Toma, Piffer, Bianchi and Agresti 2016], a paradigmatic example of this type of systems. In some pathological conditions, mutations can overactivate the system upon external signals: by changing the parameters related to the signal S, we have shown that in our model the oscillations of the transcriptional activator of interest are mainly lost and the system is essentially always active. This is also reflected in the disruption of the gene expression patterns that this simple genetic circuit can generate: the patterns of gene expression obtained in the normal conditions are essentially lost. The biological implications might be huge, since we found for NF-κB that the patterns of gene expression, from oscillating to slowly accumulating, were related with specific functions and hence an incorrect dynamics of production of the proteins involved can lead to malfunction. This can be the case of genes related to cell survival and proliferation, that are known to be expressed in anomalously high amounts in multiple myeloma, where NF-κB is often deregulated [Annunziata, Davis, Demchenko et al. 2007]. The next step is to try to experimentally verify if this is the case for cancer cells and to envision ways to correct this be-

haviour, something that might have a positive therapeutic impact when dealing with this type of tumour.

Acknowledgements

F.C. and S.Z. were supported by AIRC IG Grant 18687. S.Z. thanks the support of a fellowship by San Raffaele University. N.M. was supported by CNRS.

6 Bibliography

References

Alon, U. (2007). *An Introduction to Systems Biology*. CRC. Boca Raton.

Hoffmann, A., Levchencko, A., Scott, M., Baltimore, D. (2002) The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science*, **298**, pp. 1241-1245.

Levine, J.H., Lin, Y., Elowitz, M.B. (2013) Functional roles of pulsing in genetic circuits. *Science*, **342**, pp. 1193-1200.

Elowitz, M.B., Levine, J.H., Siggia, E.D., Swain, P.S. (2002) Stochastic gene expression in a single cell. *Science*, **297**, pp. 1183-1186.

Zambrano, S., Bianchi, M.E., Agresti, A. (2014) High-Throughput Analysis of NF-κB Dynamics in Single Cells Reveals Basal Nuclear Localization of NF-κB and Spontaneous ACtivations of OScillations. *PLoS ONE*, **9:3**, e90104.

Zambrano, S., De Toma, I., Piffer, A., Bianchi, M.E., Agresti, A. (2016) NF-κB oscillations translate into functionally related patterns of gene expression. *eLife*, **5:e09100**, pp. 1-38.

Annunziata, C.M., Davis, R.E.,Demchenko, Y., Bellamy, W., Gabrea, A., Zhan, F., Lenz, G., Hanamura, I., Wright, G., Xiao, W., Dave, S., Hurt, E.M., Tan, B., Zhao, H., Stephens, O., Santra, M., Williams, D.R., Dang, L., Barlogie, B., Shaughnessy, J.D., Kuehl, W.M., Staudt, L.M. (2007) Frequent engagement of the classical and alternative NF-kappaB pathways by diverse genetic abnormalities in multiple myeloma. *Cancer Cell*, **2**, pp. 115-130.

Zambrano, S., Bianchi, M.E., Agresti, A., Molina, N. (2015) Interplay between stochasticity and negative feedback leads to pulsed dynamics and distinct gene activity patterns. *Phys. Rev. E*, **92**, pp. 022711.

Gillespie, D. T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, pp. 2340-2361.

Zambrano, S., Bianchi, M.E., Agresti, A. (2014) A simple model of NF-κB dynamics reproduces experimental observations. *J. Theor. Biol*, **347**, pp. 44-53.