Molecular BioSystems

Cite this: DOI: 10.1039/c1mb05086a

www.rsc.org/molecularbiosystems



PAPER

# Stochastic analysis of a miRNA–protein toggle switch†

E. Giampieri,[a] D. Remondini,[a] L. de Oliveira,[a] G. Castellani[a] and P. Lió*[b]

*Received 1st March 2011, Accepted 12th May 2011*
**DOI: 10.1039/c1mb05086a**



Within systems biology there is an increasing interest in the stochastic behavior of genetic and biochemical reaction networks. An appropriate stochastic description is provided by the chemical master equation, which represents a continuous time Markov chain (CTMC). In this paper we consider the stochastic properties of a toggle switch, involving a protein compound (E2Fs and Myc) and a miRNA cluster (miR-17-92), known to control the eukaryotic cell cycle and possibly involved in oncogenesis, recently proposed in the literature within a deterministic framework. Due to the inherent stochasticity of biochemical processes and the small number of molecules involved, the stochastic approach should be more correct in describing the real system: we study the agreement between the two approaches by exploring the system parameter space. We address the problem by proposing a simplified version of the model that allows analytical treatment, and by performing numerical simulations for the full model. We observed optimal agreement between the stochastic and the deterministic description of the circuit in a large range of parameters, but some substantial differences arise in at least two cases: (1) when the deterministic system is in the proximity of a transition from a monostable to a bistable configuration, and (2) when bistability (in the deterministic system) is "masked" in the stochastic system by the distribution tails. The approach provides interesting estimates of the optimal number of molecules involved in the toggle switch. Our discussion of the points of strengths, potentiality and weakness of the chemical master equation in systems biology and the differences with respect to deterministic modeling are leveraged in order to provide useful advice for both the bioinformatician and the theoretical scientist.

## Introduction

Complex cellular responses are often modeled as switching between phenotype states, and despite the large body of deterministic studies and the increasing work aimed to elucidate the effect of intrinsic and extrinsic noise in such systems, some aspects still remain unclear. Molecular noise, which arises from the randomness of the discrete events in the cell (for example DNA mutations and repair), and experimental studies have reported the presence of stochastic mechanisms in cellular processes such as gene expression,[19,35,38] decisions of the cell fate,[33] and circadian oscillations.[34] Particularly, low copy numbers of important cellular components and molecules give rise to stochasticity in gene expression and protein synthesis, and it is a fundamental aspect to be taken into account for studying such biochemical models.[1,2] In this paper, we consider a simplified circuit that is known to govern a fundamental step during the eukaryotic cell cycle that defines cell fate, previously studied by means of a deterministic modeling approach.[3] Let us set the scene by reminding that "all models are wrong, but some are useful" (said by George Edward Pelham Box, who was the son-in-law of Ronald Fisher). Biologists make use of qualitative models through graphs; quantitative modeling in biochemistry has been mainly based on the Law of Mass Action which has been used to frame the entire kinetic modeling of biochemical reactions for individual enzymes and for enzymatic reaction network systems.[36] The state of the system at any particular instant is therefore regarded as a vector (or list) of amounts or concentrations and the changes in amounts or concentrations are assumed to occur by a continuous and deterministic process that is computed using the ordinary differential equation (ODE) approach. However, the theory based on the Law of Mass Action does not consider the effect of fluctuations. If the concentration of the molecules is not large enough, we cannot ignore fluctuations. Moreover, biological systems also show heterogeneity which occurs as a phenotypic consequence for a cell population given stochastic single-cell dynamics (when the population is not isogenic and under the same conditions). From a practical point of view, for concentrations greater than about 10 nM, we are safe using ODEs; considering a cell with a volume of $10^{-13}$ litres this corresponds to thousands of

[a] *Dip. di Fisica, Universitá di Bologna, Viale B. Pichat 6/2, 40126 Bologna, Italy*
[b] *Computer Science Dept, University of Cambridge, JJ Thomson Rd, Cambridge, UK. E-mail: pl219@cam.ac.uk*
† Published as part of a Molecular BioSystems themed issue on Computational Biology: Guest Editor Michael Blinov.

molecules that, under the poissonian hypothesis, has an uncertainty in the order of 1%. If the total number of molecules of any particular substance, say, a transcription factor, is less than 1000, then a stochastic differential equation or a Monte Carlo model would be more appropriate. Similarly to the deterministic case, only simple systems are analytically tractable in the stochastic approach, *i.e.* the full probability distribution for the state of the biological system over time can be calculated explicitly, becoming computationally infeasible for systems with distinct processes operating on different timescales. An active area of research is represented by development of approximate stochastic simulation algorithms. As commented recently by Wilkinson, the difference between an approximate and exact model is usually remarkably less than the difference between the exact model and the real biological process.[9] Given that we can see this either as an unsatisfactory state of art or as a promising advancement, we can summarise the methodological approaches as follows. Biochemical networks have been modeled using differential equations when considering continuous variables changing deterministically with time. Single stochastic trajectories have been modeled using stochastic differential equations (SDE) for continuous random variables, and using the Gillespie algorithm for discrete random variables changing with time. Another choice consists in characterizing the time evolution of the whole probability distribution. The corresponding equation for the SDE is the Fokker–Planck equation, and the corresponding equation for the Gillespie algorithm is called the Chemical Master Equation (CME).[37] Therefore, the CME could be thought as the mesoscopic version of the Law of Mass Action, *i.e.* it extends the Law of Mass Action to the mesoscopic chemistry and biochemistry, see for example ref. 47 and 48.

Here we compare the results of a stochastic *versus* deterministic analysis of a microRNA–protein toggle switch[5,6] involved in tumorigenesis with the aim of identifying the most meaningful amount of information to discriminate cancer and healthy states. We show that the stochastic counterpart of such a deterministic model has many commonalities with the deterministic one, but some differences arise, in particular regarding the number of stable states that can be explored by the system. The disagreement between the stochastic and deterministic descriptions is observed in a "ghost" effect caused by the proximity to a deterministic bifurcation,[22] and in a somehow opposite situation, in which the variance of the stable point can mask the detection of the second peak in the stationary distribution. In this paper we perform a numerical study of the complete two-dimensional model, but we consider also a simplified, biologically meaningful version of the model that allows us to calculate an exact solution, with a numerical characterization of the parameter ranges in which the two systems produce qualitatively similar results. A discussion of the possible implications of our results in a real system is described in the last section.

## I. Properties of a microRNA toggle switch

Oncogenes and tumor-suppressor genes are two pivotal factors in tumorigenesis. Recent evidences indicate that microRNAs (miRNAs) can function as tumor suppressors and oncogenes, and these miRNAs are referred to as oncomirs. miRNAs are small, non-coding RNAs that modulate the expression of target mRNAs. The biogenesis pathway of miRNAs in animals was elucidated in ref. 29. miRNAs undergo substantial processing since the nuclear transcription where two proteins play an essential role: Drosha and Dicer. Most of miRNAs are first processed into pre-miRNAs by Drosha. After export to the cytoplasm, the pre-miRNA is processed by Dicer into a small double stranded RNA (dsRNA) called the miRNA: miRNA duplex. The active strand, which is the mature miRNA, is incorporated into the RISC and binds to the target mRNA, whereas the inactive strand is ejected and degraded. In normal tissue, proper regulation of miRNAs maintains a normal rate of development, cell growth, proliferation, differentiation and apoptosis. Tumorigenesis can be observed when the target gene is an oncogene, and the loss of the miRNA, which functions as a tumor suppressor, might lead to a high expression level of the oncoprotein. When a miRNA functions as an oncogene, its constitutive amplification or overexpression could cause repression of its target gene, which has a role of the tumor suppressor gene, thus, in this situation, the cell is likely to enter tumorigenesis. miRNAs are often part of toggle switches: important examples involve gene pairs built with oncogenes and tumour suppressor genes.[7,8] Here we focus on the amplification of 13q31–q32, which is the locus of the miR-17-92. The miR-17-92 cluster forms a bistable switch with Myc and the E2F proteins.[3,14,15] The oncogene Myc regulates an estimated 10% to 15% of genes in the human genome, while the dysregulated function of Myc is one of the most common abnormalities in human malignancy.[16,18] The other component of the toggle switch is the E2F family of transcription factors, including E2F1, E2F2 and E2F3, all driving the mammalian cell cycle progression from the G1 into S phase. High levels of E2Fs, E2F1 in particular, can induce apoptosis in response to DNA damage. The toggle switch also interacts with dozens of genes (Fig. 1 depicts a portion of the whole circuitry), particularly with Rb and other key cell-cycle players. A summary of the experiments perturbing miRNA/Myc/E2F and E2F/RB behaviours has suggested the following:

- The Rb/E2F toggle switch is OFF when RB inhibits E2F, *i.e.* stopping cell proliferation; it is ON when E2F prevails and induces proliferation. Once turned ON by sufficient stimulation,

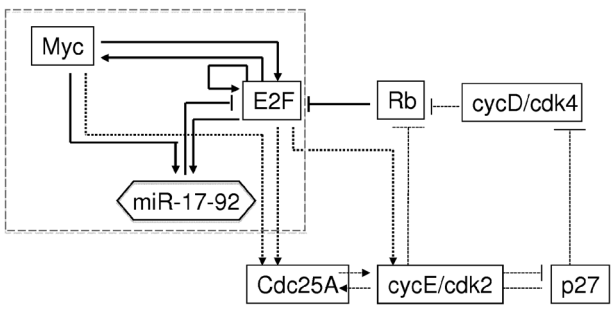


**Fig. 1** The E2F–MYC–miR-17-92 toggle switch with its biochemical environment (derived form).[3] Arrows represent activation, and bar-headed lines inhibition, respectively. The elements inside the dashed box represent the protein compound $p$ (Myc–E2F) and the miRNA cluster $m$ (miR-17-92), modelized in eqn (1) and (2).

E2F can memorize and maintain this ON state independently of continuous serum stimulation.

• The proteins E2F and Myc facilitate the expression of each other and the E2F protein induces the expression of its own gene (positive feedback loop). They also induce the transcription of microRNA-17-92 which in turn inhibits both E2F and Myc (negative feedback loop).[17]

Moreover, the increasing levels of E2F or Myc drive the sequence of cellular states, namely, quiescence, cell proliferation (cancer) or cell death (apoptosis).

Although there is increasing amount of research on cell cycle regulation, the mathematical description of even a minimal portion of the E2F, Myc and miR-17-92 toggle switch is far from trivial. Aguda and collaborators[3] have developed a deterministic model, which reduces the full biochemical network of the toggle switch to a protein (representing the E2F–Myc compound) and the microRNA-17-92 cluster (seen as a single element).

It is a 2-dimensional open system, in which $p$ represents the E2F–Myc complex and $m$ the miRNA cluster: thus no Mass Action Law holds, and the total $p$ and $m$ concentration is not conserved. The dynamics of $p$ and $m$ concentrations are described by eqn (1) and (2):

$$\dot{p} = \alpha + \frac{k_1 p^2}{\Gamma_1 + p^2 + \Gamma_2 m} - \delta p \qquad (1)$$

$$\dot{m} = \beta + k_2 p - \gamma m \qquad (2)$$

The model is the following: constitutive creation–destruction processes for $p$ and $m$ are driven by $\alpha$, $\delta$, $\beta$ and $\gamma$ parameters, while $k_1$ and $k_2$ control $p$ and $m$ state-dependent synthesis. The $\Gamma_1$ term is a kinetic (enzymatic-like) constant, while $\Gamma_2$ modulates miRNA inhibition of $p$ synthesis. The nonlinearity of $p$ in eqn (1) is a Hill coefficient ($=2$) representing a self-promotion effect driven by a sigmoidal activation curve, a very common behaviour in gene regulation systems. A Hill coefficient $>1$ can be justified by a cooperative effect of the terms involved in the compound represented by $p$: for example due to the E2F trimer, or a more complex aggregate with Myc. Although several experimental results suggest that in some cancer processes, a certain amount of interdependence and interaction among E2Fs exists, a detailed experimental investigation should be needed in order to estimate such a parameter correctly.[11–13]

All the results described in this article are very robust with respect to the choice of the specific Hill coefficient (here chosen equal to 2 for continuity with the original model in ref. 3) as long as it is larger than one (data not shown). We found a good qualitative agreement even if a different functional form was hypothesized, as long as it retained its sigmoidal-like structure.

The system can be rewritten in an adimensional form as follows:

$$\varepsilon\dot{\phi} = \alpha' + \frac{k\phi^2}{\Gamma_1' + \phi^2 + \Gamma_2'\mu} - \phi \qquad (3)$$

$$\dot{\mu} = 1 + \phi - \mu \qquad (4)$$

where the parameters are: $\alpha' = \frac{k_2}{\delta\beta}\alpha$, $k = \frac{k_1 k_2}{\delta\beta}$, $\Gamma_1' = \frac{k_2^2}{\beta^2}\Gamma_1$, $\Gamma_2' = \frac{k_2^2}{\beta\gamma}\Gamma_2$, $\varepsilon = \frac{\gamma}{\delta}$ and the change of variables is: $\phi = \frac{k_2}{\beta}p$, $\mu = \frac{\gamma}{\beta}m$ and $\tau = \gamma t$.

In the original model,[3] the rate of protein synthesis is not a function of the instantaneous concentration (as assumed in eqn (3)) but rather of its concentration at some time $\Delta$ in the past:

$$\varepsilon\dot{\phi} = \alpha' + \frac{k[\phi(\tau-\Delta)]^2}{\Gamma_1' + [\phi(\tau-\Delta)]^2 + \Gamma_2'\mu(\tau-\Delta)} - \phi(\tau). \qquad (5)$$

We will not consider such delay in our stochastic realization of the model, since it would increase system dimensionality and it does not seem necessary to obtain the features we want to characterize.

The steady state can be studied in the nondimensionalized system and, therefore, the conditions of the parameters for the existence of multiple steady states. In the resulting cubic equation:

$$\alpha' + \frac{k\phi^2}{\Gamma_1' + \phi^2 + \Gamma_2'(1+\phi)} - \phi = 0 \qquad (6)$$

the necessary (but not sufficient) conditions for the existence of 3 steady states (and thus a bistable system) are:

$$(\Gamma_2' - k) < \alpha' < \left(1 + \frac{\Gamma_1'}{\Gamma_2'}\right) \qquad (7)$$

We took advantage of the deterministic results in ref. 3 in order to consider suitable parameter ranges for our stochastic modelling (as described in the following sections).

## II. The stochastic modeling approach

The system represented by eqn (1) and (2) can be studied as a stochastic system through the Chemical Master Equation (CME) approach.[4] The resulting CME has two variables, the number of $p$ and $m$ molecules, labeled as $n$ and $m$. The temporal evolution of the probability $p_{n,m}(t)$ to have $n$ and $m$ molecules at time $t$ is described by the following equation:

$$\dot{p}_{n,m} = (\mathbb{E}_n - 1)r^n p_{nm} + (\mathbb{E}_n^{-1} - 1)g^n p_{nm} + (\mathbb{E}_m - 1)r^m p_{nm} + (\mathbb{E}_m^{-1} - 1)g^m p_{nm} \qquad (8)$$

This CME is derived under the conditions of a one-step Poisson process: $\mathbb{E}$ and $\mathbb{E}^{-1}$ are the forward and backward step operators, $g$ and $r$ are the generation and recombination terms for the $n$ and $m$ variables, as shown in superscripts.

The two generation and recombination terms associated with the $n$ and $m$ variables are respectively:

$$g^n = \alpha + \frac{k_1 n^2}{\Gamma_1 + n^2 + \Gamma_2 m}; \quad r^n = \delta n \qquad (9)$$

$$g^m = \beta + k_2 n; \quad r^m = \gamma m \qquad (10)$$

We remark that the molecule influxes into the system (represented by the $\alpha$ and $\beta$ terms) could be included in different ways in the stochastic equations, since in the deterministic equations they represent a sort of "mean field" value. As an example, molecules could be added in bursts with specific time distributions, which do not appear in the macroscopic continuous deterministic equations. We will consider the simplest approach, but the choice of different influx patterns should deserve further investigation.

### A. The one-dimensional model

We can reduce the problem from two to one dimension, by considering a different time scale for the two reactions (in particular considering $m$ as a fast variable) and thus considering the steady state solution for the $m$:

$$m = \frac{\beta + k_2 p}{\gamma} = \beta' + k' p \tag{11}$$

As a consequence we obtain a deterministic equation for the $p$ only:

$$\dot{p} = \alpha + \frac{k_1 p^2}{\Gamma' + \Gamma'' p + p^2} - \delta p \tag{12}$$

with $\Gamma' = \frac{\Gamma_2 k_2}{\gamma}$ and $\Gamma'' = \Gamma_1 + \frac{\Gamma_2 \beta}{\gamma}$. The stochastic equation for $p_n$ is thus as follows:

$$\dot{p}_n = (\mathbb{E} - 1) r_n p_n + (\mathbb{E}^{-1} - 1) g_n p_n \tag{13}$$

$$g_n = \alpha + \frac{k_1 n^2}{\Gamma' + \Gamma'' n + n^2}; \quad r_n = \delta n \tag{14}$$

A general solution can be obtained

$$p_n^s = \prod_{i=1}^{N} \frac{g(i-1)}{r(i)} p_0 = \prod_{i=1}^{N} \frac{\alpha + \frac{k_1 (i-1)^2}{\Gamma' + \Gamma''(i-1) + (i-1)^2}}{\delta i} p_0 \tag{15}$$

with an adequate normalization factor imposed on $p_0$:

$$p_0 = \frac{1}{1 + \sum_{i=1}^{N} \prod_{i=1}^{N} p_{n_1}^s} \tag{16}$$

We remark that the system is open, thus in theory $N$ is not fixed, but we can truncate the product to a sufficiently high value of $N$ obtaining a good approximation of the whole distribution. This one-dimensional system (for which an analytical solution can be obtained) will be compared to numerical simulations of the exact one-dimensional and two-dimensional systems.

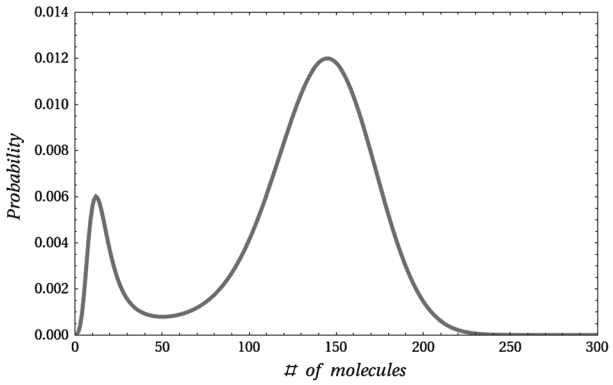


**Fig. 2** The stationary distribution for the one-dimensional space, obtained using the following parameters: $\alpha$ = 1.68 (molecule h$^{-1}$), $\beta$ = 0.202 (molecule h$^{-1}$), $\delta$ = 0.2 ($h^{-1}$), $\gamma$ = 0.2 (h$^{-1}$), $\Gamma_1$ = 10 300 (molecule$^2$), $\Gamma_2$ = 1006 (molecule), $k_1$ = 90 (molecule h$^{-1}$) and $k_2$ = 0.05 (h$^{-1}$).

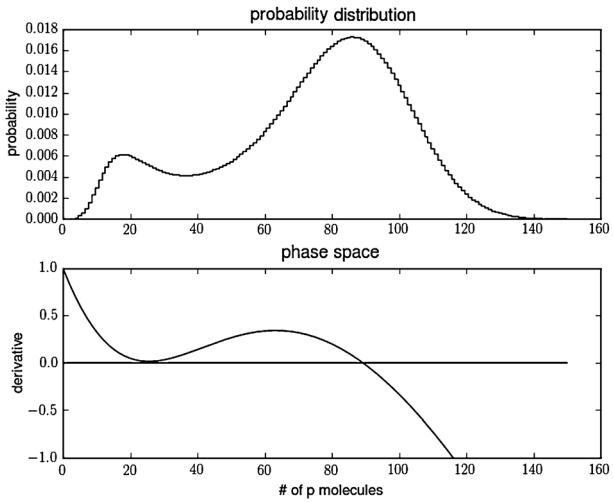


**Fig. 3** Comparison between the deterministic solution (bottom) and the stationary distribution (top) for the parameter set as in Table 1, case 3.

## III. Model analysis

### A. The stationary distribution

The one-dimensional model can show monomodal as well as bimodal stationary distributions, depending on the parameters considered. As an example, we obtain bistability with a set of parameters as shown in Fig. 2.

Thus the qualitative features of the two-dimensional deterministic model (*i.e.* the possibility of being bistable depending on the parameter range) are recovered for the one-dimensional approximation of the stochastic system. Also the two-dimensional stochastic system shows bistability for the same parameters, and they are in optimal agreement for a range of parameters in which the $\dot{m} \gg \dot{p}$ condition holds.

We also observe some remarkable differences between the deterministic and the stochastic models: there are regions in the parameter space in which the deterministic approach shows only one stable state, but in the stochastic system two maxima in the stationary distribution are observed (see Fig. 3). This difference can be explained qualitatively as follows: for the deterministic system, there are parameter values for which the system is monostable but very close to the "transition point" in which the system becomes bistable. It is known that in these situations a "ghost" remains in the region where the stable point has disappeared,[22] for which the systems dynamics has a sensible slowing down (*i.e.* when the system is close to the disappeared fixed point, it remains "trapped" for a longer time close to it, in comparison with other regions). This behaviour results in the presence of a peak in the stationary distribution of the corresponding stochastic systems, which thus remains bistable also when the deterministic system is not bistable anymore.

Another difference is observed: for some parameter values the deterministic system is bistable, but the stochastic distribution shows a clear peak for the maximum with the largest basin of attraction and the smaller peak results "masked" by the tail of the distribution around the first peak (see Fig. 4), thus resulting in a monomodal distribution with a long tail. In practice, the highest state behaves like a sort of metastable

 

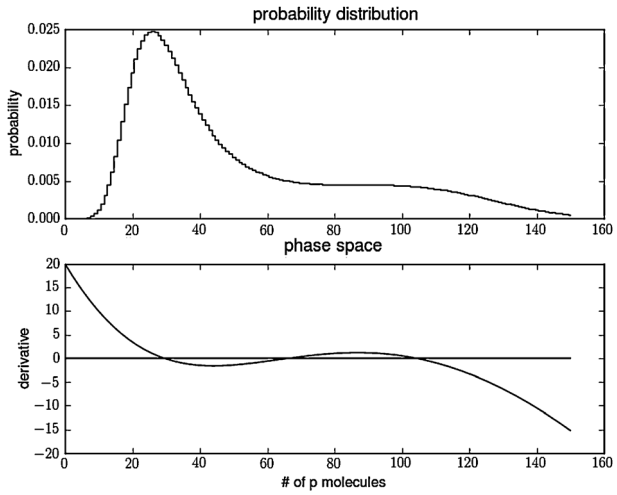


**Fig. 4** Comparison between the deterministic solution (bottom) and the stationary distribution (top) for the parameter set as in Table 1, case 4.

state, since the states of the system with a high protein level are visited only occasionally.

### B. Numerical analysis

Here we implemented numerical methods to find the stationary distribution of a CME. The most accurate is the Kernel resolution method: given the complete transition matrix of the system, it is possible to solve numerically the eigenvalue problem, obtaining the correct stationary distribution. This method, in this case, has a serious drawback: the system is of non-finite size, preventing a complete enumeration of the possible states. Even with a truncation, the system size rises in a dramatic way: the state space for a bidimensional system is of order $N^2$ if $N$ is the truncation limit, and thus the respective transition matrix is of order $N^4$. This means that even for a relatively small system (with a few hundreds of molecules) the matrix size explodes well beyond the computational limits. The only feasible resolution strategy is a massive exploration of state space by Monte Carlo methods, in which single trajectories of the system are simulated: performing this simulations long enough for several times allows us to estimate the stationary distribution.

The Monte Carlo method we chose is a modified version of the SSA (also known as the Gillespie algorithm) named logarithmic direct method,[23,25] which is a statistically correct simulation of an ergodic Markov system. It is not the fastest algorithm available, as compared to other methods like the next-reaction or the $\tau$-leap method, but it produces a correct estimation of the statistical dispersion of the final state.

For each parameter set we performed 10 simulations for about $10^6$–$10^7$ iteration steps each. The multiple simulations were averaged together for a better estimation of the stationary distribution, and they allowed also an estimation of the variance over this average distribution.

In the following we discuss four cases that describe the system behaviour for different parameter settings, shown in Table 1.

In case 1, we have a system in which the hypothesis of a time-scale separation between $m$ and $p$ is strongly satisfied. The simulation was performed up to a time limit of $10^3$: we can see how the two resulting distributions are in good agreement with the theoretical one (see Fig. 5), with the regions of higher variance of the histogram around the maxima and minima of the distribution.

**Table 1** Parameter sets for the cases considered

| Parameter | Case 1 | Case 2 | Case 3 | Case 4 |
|---|---|---|---|---|
| $\alpha$ (molecule h$^{-1}$) | 1.0 | 1.68 | 1.0 | 20.0 |
| $\delta$ (h$^{-1}$) | 1.0 | 0.20 | 0.09 | 1.19 |
| $\beta$ (molecule h$^{-1}$) | 1.0 | 0.202 | 0.0 | 1.0 |
| $\gamma$ (h$^{-1}$) | 100.0 | 0.20 | 10.0 | 1.0 |
| $k_1$ (molecule h$^{-1}$) | 30.0 | 90.0 | 12.5 | 230.0 |
| $k_2$ (h$^{-1}$) | 100.0 | 0.05 | 10.0 | 1.0 |
| $\Gamma_1$ (molecule$^2$) | 60.0 | 10 300.0 | $(72.8)^2$ | $(110.0)^2$ |
| $\Gamma_2$ (molecule) | 10.0 | 1006.0 | 10.0 | 10.0 |

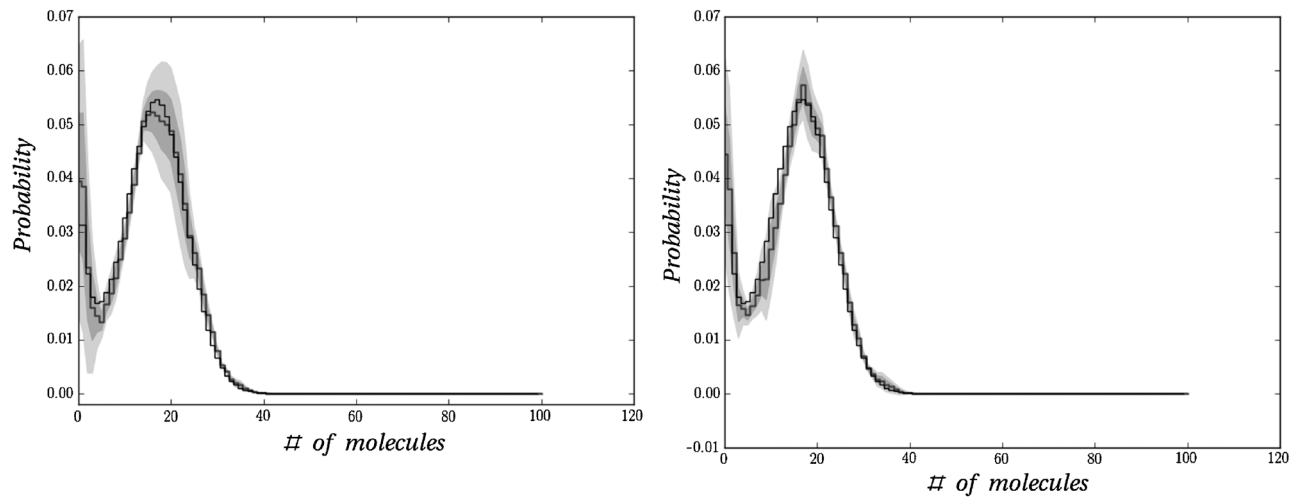


**Fig. 5** Case of good agreement between the theoretical and obtained distribution (see Table 1, case 1). Left: one-dimensional system, right: two-dimensional system. The thin black line is the theoretical distribution obtained from eqn (15). The thick dark grey line is the average of the various simulations, while the grey and light grey areas represent the range of one and two standard deviations from the average distribution.

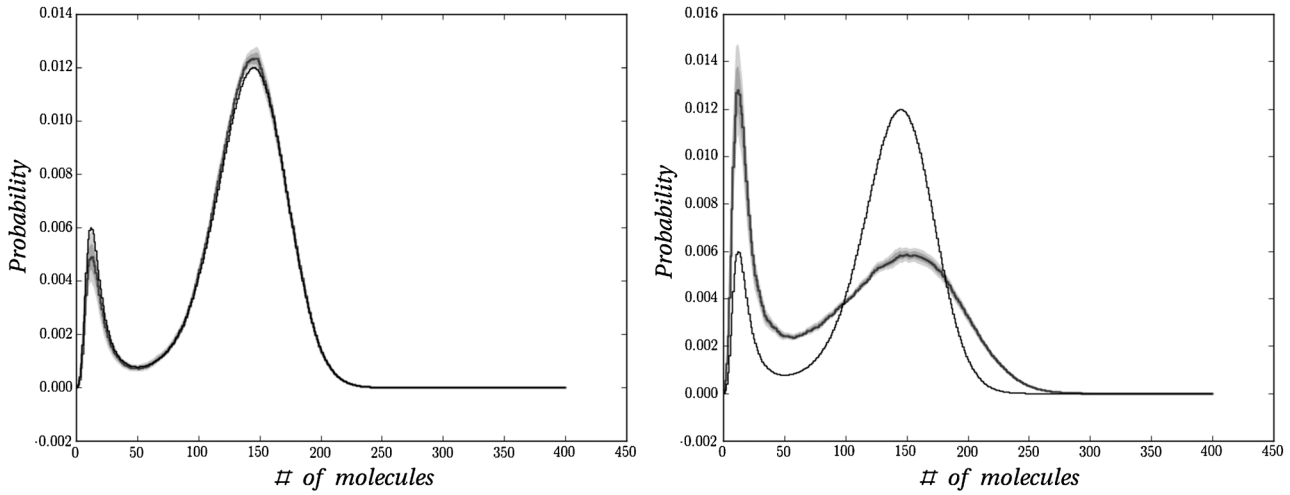


**Fig. 6** Case of poor agreement between the theoretical and obtained distribution (see Table 1, case 2). Left: one-dimensional system, right: two-dimensional system. The thin black line is the theoretical distribution obtained from eqn (15). The thick dark grey line is the average of the various simulations, while the grey and light grey areas represent the range of one and two standard deviations from the average distribution.

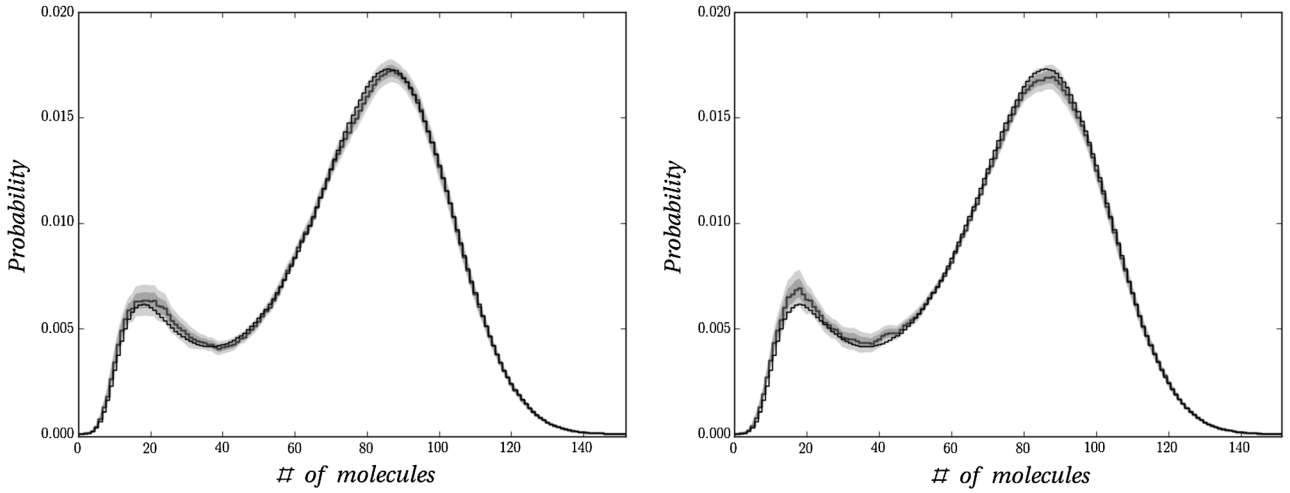


**Fig. 7** Case 3, "ghost effect": only the biggest peak comes from a deterministic stable point. Left: one-dimensional system, right: two-dimensional system. The thick dark gray line is the average of the various simulations, while the gray and light gray areas represent the range of one and two standard deviations from the average distribution.

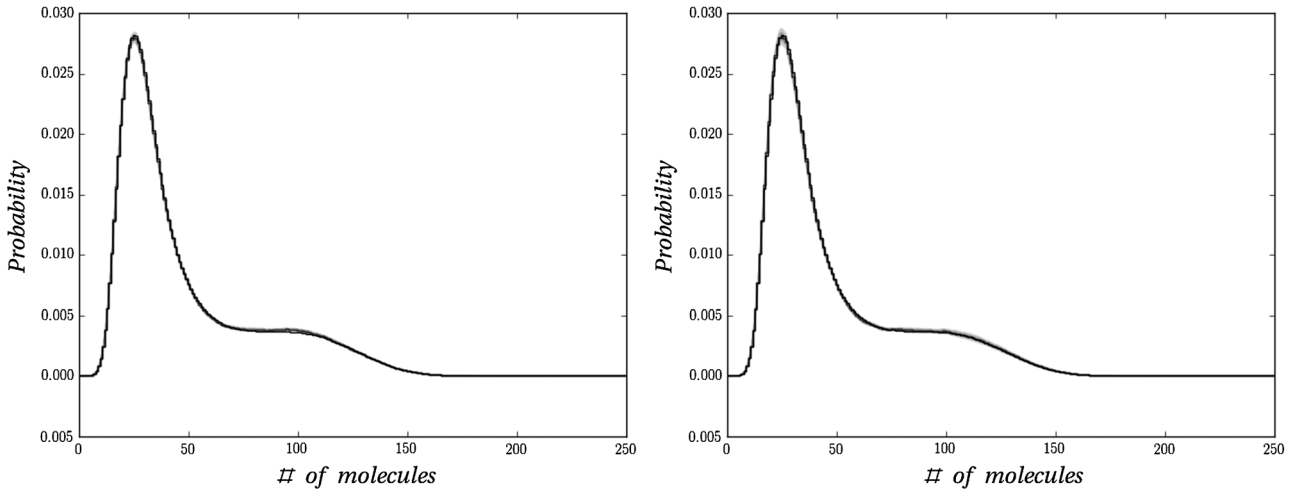


**Fig. 8** Case 4, peak masking effect (parameters as in Table 1, case 4). The deterministic system has two stable points, but only the peak related to the smallest stable point (with the largest basin of attraction) is visible. Left: one-dimensional system, right: two-dimensional system.



 

In case 2, the time-scale separation assumption does not hold, due to the very low values of $\gamma$ and $k_2$: even if this condition doesn't guarantee that the stationary state will be different from the approximate one-dimensional solution, with this set of parameters we can observe a large difference between the two distributions (Fig. 6).

In case 3, as defined before, we observe a "ghost effect" in which, even if a deterministic stable state does not exist, there is clearly a second peak in the distribution (Fig. 7). In this system the time-scale separation assumption holds, and we can see how both distributions show similar features.

In this final case (Table 1, case 4, Fig. 8) we observe another effect, in which the peak related to a deterministic stable state is masked by the tail of the stronger peak, becoming just a fat tail. Even without a strong time-scale separation for the $m$ and $p$ variables, both systems give a very similar response, evidencing that this effect is very robust. It is noteworthy that the increase of the $\gamma$ and $k_2$ values does not affect the distribution as long as their ratio is kept constant. Note that while there are several computational tools for discrete-state Markov processes such as PRISM,[42] APNNtoolbox,[45] SHARPE,[46] or Mobius,[43] there is very little for CMTC (see for instance ref. 44). Different modeling approaches for toggle switches do exist in the area of formal methods (see for example ref. 26 and 27).

## IV. Discussion and conclusion

We have studied a stochastic version of a biochemical circuit (the toggle switch) that is supposed to be involved in cell cycle control, with implications for the onset of severe diseases such as cancer, consisting of a gene cluster (Myc–E2F) and a miRNA cluster (miR-17-92). This cluster has been reported in a very large number of cancer types: particularly in different types of lymphomas, glioma, non-small cell lung cancer, bladder cancer, squamous-cell carcinoma of head and neck, peripheral nerve sheath tumor, malignant fibrous histiocytoma, alveolar rhabdomyosarcoma, liposarcoma and colon carcinomas. This huge variety of cancer stresses the centrality of this toggle switch and suggests that advancements in its modelling could lead to insights into differences between these cancers. This aim is still far but we are delighted to report that our modeling approach shows important results inching to that direction. First of all, many features are recovered, as observed for the deterministic version of the same system, also by means of a further approximation that reduces the system to a unique variable: in this case the system can be treated analytically, and compared to the one- and two-dimensional numerical simulations.

The stochastic approach, that is the exact approach when the number of molecules involved is low, shows a different behaviour than the deterministic one in two situations we have observed. It is noteworthy that the number of molecules involved shows some agreement with the estimates in ref. 31 and ref. 30 for other miRNA-systems (see also ref. 32). The cell volume is assumed to be about $10^{-13}$ litres, then 1 nM = 100 molecules.

First, bistability in the stochastic system (namely, the possibility of having two stable states, one associated with a resting and the other with a proliferative cell state) is observed also in situations in which the corresponding deterministic system is monostable, and this can be explained by the presence of a "ghost" state in the deterministic system that is strong enough to produce a second peak in the stationary distribution of the stochastic model.

Secondly, there are situations in which the peak for the stochastic distribution related to the highest level of expression (with parameter values for which the deterministic system is bistable) is masked by the tail of the distribution of the lowest-expression maximum (that is related to the largest basin of attraction in the deterministic model), making the "proliferative state" appear almost as a scarcely visited metastable state. This is an interesting behaviour that should be further investigated in real experimental data of protein concentration and gene expression related to the biochemical circuit considered. The "metastable" and the "fully" bimodal distributions could be associated with healthy and tumoral cell states, respectively, because the highest "proliferative" state has different properties in the two cases. From a biological point of view, such state, being associated with a dysregulated, disease-related conditions, could actually represent a compendium of several dysregulated states.

We argue that the deterministic approach to this biochemical circuit is not capable of characterizing it completely, and the stochastic approach appears more informative: further features unique to the stochastic model could be obtained by considering different time patterns for the molecular influxes to the system, and this point in our opinion should deserve more investigation in a future work. MicroRNAs (miRNAs) express differently in normal and cancerous tissues and thus are regarded as potent cancer biomarkers for early diagnosis. We believe that the potential use of oncomirs in cancer diagnosis, therapies and prognosis will benefit from accurate cancer mathematical models.

Given that miR-17-92 seems to act as both oncogene and tumor suppressor through decreasing the expression levels of anti-proliferative genes and proliferative genes, this behavior is suggestive of a cell-type dependent toggle switch. Therefore, fitting experimental data could provide insights into differences among cancer types and on which cell type is behaving differently. The fitting of experimental data with respect to models with different values for the Hill coefficients could also be interesting towards understanding better the chemistry/physics of the real microRNA system. Moreover, the comparison between the shape of the expression distributions of the genes/proteins involved in the circuit (and not only the average expression) considering normal and tumoral cells for different cell types should provide experimental evidence for the different behaviour described from a theoretical point of view in our work, namely the possibility that normal and tumoral cells are in different proliferative "stationary states".

## Acknowledgements

D.R. acknowledges Bologna University 'Progetto Strategico d'Ateneo 2006' funding.

## References

1 M. Samoilov, S. Plyasunov and A. P. Arkin, Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations, *PNAS*, 2005, **102**(7), 2310–2315.

2 G. C. Castellani, A. Bazzani and L. N. Cooper, Toward a microscopic model of bidirectional synaptic plasticity, *PNAS*, 2009, **106**(33), 14091–14096.
3 B. D. Aguda, Y. Kim, M. G. Piper-Hunter, A. Friedman and C. B. Marsh, MicroRNA regulation of a cancer network: consequences of the feedback loops involving miR-17-92, E2F, and Myc, *PNAS*, 2008, **105**(50), 19678–19683.
4 N. G. Van Kampen, *Stochastic processes in physics and chemistry*, Elsevier, North Holland, 2007.
5 T. S. Gardner, C. R. Cantor and J. J. Collins, Construction of a genetic toggle switch in *Escherichia coli*, *Nature*, 2000, **403**, 339–342.
6 H. Kobayashi, M. Kaern, M. Araki, K. Chung, T. S. Gardner, C. R. Cantor and J. J. Collins, Programmable cells: Interfacing natural and engineered gene networks, *PNAS*, 2004, **101**(22), 8414–8419.
7 L. P. Lim, N. C. Lau, P. Garrett-Engele, A. Grimson and J. M. Schelter, *et al*., Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs, *Nature*, 2005, **433**, 769–773.
8 J. Lu, G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak and A. A. Ferrando, *et al*., MicroRNA expression profiles classify human cancers, *Nature*, 2005, **435**, 834–838.
9 D. J. Wilkinson, Stochastic modelling for quantitative description of heterogeneous biological systems, *Nat. Rev. Genet*., 2009, **10**, 122–133.
10 G. Yao, T. J. Lee, S. Mori, J. R. Nevins and L. You, A bistable Rb–E2F switch underlies the restriction point, *Nat. Cell Biol*., 2008, **10**, 476–482.
11 K. Fujiwara, I. Yuwanita, D. P. Hollern and E. R. Andrechek, Prediction and genetic demonstration of a role for activator E2Fs in Myc-induced tumors, *Cancer Res*., 2011, **71**, 1924–1932.
12 P. Trikha, N. Sharma, R. Opavsky, A. Reyes, C. Pena, M. C. Ostrowski, M. F. Roussel and G. Leone, E2f1-3 are critical for myeloid development, *J. Biol. Chem*., 2011, **286**, 4783–4795.
13 E. Lazzerini Denchi and K. Helin, E2F1 is crucial for E2F-dependent apoptosis, *EMBO Rep*., 2005, **6**(7), 661–668.
14 K. A. O'Donnell, E. A. Wentzel, K. I. Zeller, C. V. Dang and J. T. Mendell, c-Myc-regulated microRNAs modulate E2F1 expression, *Nature*, 2005, **435**(7043), 839–843.
15 M. J. Bueno, I. P. de Castro and M. Malumbres, Control of cell proliferation pathways by microRNAs, *Cell Cycle*, 2008, **7**, 3143–3148.
16 D. Remondini, B. OConnell, N. Intrator, J. M. Sedivy, N. Neretti, G. C. Castellani and L. N. Cooper, Targeting c-Myc-activated genes with a correlation method: detection of global changes in large gene expression network dynamics, *PNAS*, 2005, **102**(19), 6902–6906.
17 Y. Sylvestre, V. De Guire, E. Querido, U. K. Mukhopadhyay, V. Bourdeau, F. Major, G. Ferbeyre and P. Chartrand, An E2F/miR-20a autoregulatory feedback loop, *J. Biol. Chem*., 2007, **282**, 2135–2143.
18 H. A. Coller, J. J. Forman and A. Legesse-Miller, Myced Messages: Myc induces transcription of E2F1 while inhibiting its translation *via* a microRNA polycistron, *PLoS Genet*., 2007, **3**, e146.
19 M. B. Elowitz, M. J. Levine, E. D. Siggia and P. S. Swain, Stochastic Gene Expression in a Single Cell, *Science*, 2002, **297**, 1183–1186, DOI: 10.1126/science.1070919.
20 E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman and A. van Oudenaarden, Regulation of Noise in the Expression of a Single Gene, *Nat. Genet*., 2002, **31**, 69–73, DOI: 10.1038/ng869.
21 W. J. Blake, M. Kaern, C. R. Cantor and J. J. Collins, Noise in Eukaryotic Gene Expression, *Nature*, 2003, **422**, 633–637, DOI: 10.1038/nature01546.
22 S. H. Strogatz, *Nonlinear dynamics and chaos*, Perseus Books, 1994.
23 D. T. Gillespie, Exact Stochastic Simulation of Coupled Chemical Reactions, *J. Phys. Chem*., 1997, **81**(25), 2340–2361.
24 X. Wang and X. Wang, Systematic identification of microRNA functions by combining target prediction and expression profiling, *Nucleic Acids Res*., 2006, **34**, 1646–1652.
25 H. Li, L. Petzold, *Logarithmic Direct Method for Discrete Stochastic Simulation of Chemically Reacting Systems*, *Technical Report*, 2006.
26 G. Bella, P. Lio, Formal Analysis of the Genetic Toggle CMSB 2009, 7th Conference on Computational Methods in Systems Biology, Sept. 2009, Bologna, Italy, P.Degano and R.Gorrieri (editors), pages 96–110, LNBI 5688, Springer.
27 G. Bella, P. Lio, Analysing the microRNA-17-92/Myc/E2F/RB Compound Toggle Switch by Theorem Proving NETTAB 2009, 9th Workshop on Network Tools and Applications in Biology, June 2009, Catania, Italy. P.Romano, *et al*. (editors), pages 59–62.
28 W. C. Cho, Oncomirs: the discovery and progress of microRNAs in cancers, *Mol. Cancer*, 2007, **6**, 60.
29 D. P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell*, 2004, **116**, 281–297.
30 Lee P. Lim, Nelson C. Lau, Earl G. Weinstein, Aliaa Abdelhakim, Soraya Yekta, Matthew W. Rhoades, Christopher B. Burge and David P. Bartel, The microRNAs of Caenorhabditis elegans, *Genes Dev*., 2003, **17**(8), 991–1008.
31 Ho-Man Chan, Lai-Sheung Chan, Ricky Ngok-Shun Wong and Hung-Wing Li, Direct Quantification of Single-Molecules of MicroRNA by Total Internal Reflection Fluorescence Microscopy, *Anal. Chem*., 2010, **82**, 6911–6918.
32 Aaron Arvey, Erik Larsson, Chris Sander, Christina S Leslie and Debora S Marks, Target mRNA abundance dilutes microRNA and siRNA activity, *Mol. Syst. Biol*., 2010, **6**, 363.
33 A. Arkin, J. Ross and H. H. McAdams, Stochastic kinetic analysis of developmental pathway bifurcation in phage-infected *E. coli* cells, *Genetics*, 1998, **149**, 1633–1648.
34 N. Barkai and S. Leibler, Biological rhythms: Circadian clocks limited by noise, *Nature*, 2000, **403**, 267–268.
35 H. H. McAdams and A. Arkin, Stochastic mechanisms in gene expression, *PNAS*, 1997, **94**, 814–819.
36 E. W. Lund, Guldberg and waage and the law of mass action, *J. Chem. Educ*., 1965, **42**, 548–550.
37 J. Liang and H. Qian, Computational cellular dynamics based on the chemical master equation: A challenge for understanding complexity, *Comput. Sci. Technol*., 2010, **25**, 154–168.
38 T. B. Kepler and T. C. Elston, Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations, *Biophys. J*., 2001, **81**, 3116–3136.
39 C. V. Rao and A. P. Arkin, Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm, *J. Chem. Phys*., 2003, **118**, 4999–5010.
40 J. E. Ferrell and W. Xiong, Bistability in cell signaling: How to make continuous processes discontinuous, and reversible processes irreversible, *Chaos*, 2001, **11**, 227–236.
41 J. Gunawardena, Multisite protein phosphorylation makes a good threshold but can be a poor switch, *Proc. Natl. Acad. Sci. U. S. A*., 2005, **102**, 14617–14622.
42 M. Z. Kwiatkowska, G. Norman, D. Parker, *Prism 2.0: A tool for probabilistic model checking*, In QEST, 2004, pp. 322–323.
43 D. Daly, D. D. Deavours, J. M. Doyle, P. G. Webster, W. H. Sanders, *Mobius: An extensible tool for performance and dependability modeling*, In Computer Performance Evaluation/TOOLS, 2000, pp. 332–336.
44 F. Didier, T.A. Henzinger, M. Mateescu, and V. Wolf. Approximation of event probabilities in noisy cellular processes. In Proc. of CMSB09, volume 5688 of LNCS, pages 173188. Springer, 2009.
45 P. Buchholz, J.-P. Katoen, P. Kemper and C. Tepper, Model-checking large structured Markov chains, *J. Log. Algebr. Program*., 2003, **56**(1–2), 69–97.
46 C. Hirel, R. A. Sahner, X. Zang, K. S. Trivedi, *Reliability and performability modeling using sharpe 2000*, *In Computer Performance Evaluation/TOOLS*, 2000, pp. 345–349.
47 Verena Wolf, Rushil Goel, Maria Mateescu and Thomas A. Henzinger, Solving the chemical master equation using sliding windows, *BMC Syst. Biol*., 2010, **4**, 42.
48 Hong Qian and Lisa M. Bishop, The Chemical Master Equation Approach to Nonequilibrium Steady-State of Open Biochemical Systems: Linear Single-Molecule Enzyme Kinetics and Nonlinear Biochemical Reaction Networks, *Int. J. Mol. Sci*., 2010, **11**, 3472–3500.


 