

Noise Minimisation in Gene Expression Switches

Diana Monteoliva⁴, Christina B. McCarthy^{2,3}, Luis Diambra^{1*}

1 Laboratorio de Biología de Sistemas, Centro Regional de Estudios Genómicos, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina,

2 Laboratorio de Metagenómica de Microorganismos, Centro Regional de Estudios Genómicos, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Florencio Varela, Argentina, **3** Departamento de Informática y Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires, Pergamino, Buenos Aires, Argentina,

4 Instituto de Física, Universidad Nacional de La Plata, La Plata, Argentina

Abstract

Gene expression is subject to stochastic variation which leads to fluctuations in the rate of protein production. Recently, a study in yeast at a genomic scale showed that, in some cases, gene expression variability alters phenotypes while, in other cases, these remain unchanged despite fluctuations in the expression of other genes. These studies suggested that noise in gene expression is a physiologically relevant trait and, to prevent harmful stochastic variation in the expression levels of some genes, it can be subject to minimisation. However, the mechanisms for noise minimisation are still unclear. In the present work, we analysed how noise expression depends on the architecture of the *cis*-regulatory system, in particular on the number of regulatory binding sites. Using analytical calculations and stochastic simulations, we found that the fluctuation level in noise expression decreased with the number of regulatory sites when regulatory transcription factors interacted with only one other bound transcription factor. In contrast, we observed that there was an optimal number of binding sites when transcription factors interacted with many bound transcription factors. This finding suggested a new mechanism for preventing large fluctuations in the expression of genes which are sensitive to the concentration of regulators.

Citation: Monteoliva D, McCarthy CB, Diambra L (2013) Noise Minimisation in Gene Expression Switches. PLoS ONE 8(12): e84020. doi:10.1371/journal.pone.0084020

Editor: Mukund Thattai, Tata Institute of Fundamental Research, India

Received: October 2, 2013; **Accepted:** November 14, 2013; **Published:** December 23, 2013

Copyright: © 2013 Monteoliva et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ldiambra@gmail.com

Introduction

Living organisms sense and respond to environmental clues in order to obtain energy resources and overcome stressful conditions. This is achieved by employing gene regulatory networks, also called gene circuits. Each circuit acts as an input-output device which is designed to be activated by a specific signal and to elicit the required response. The dose-response curve of a given genetic circuit can be described by a continuous function, also called the regulatory function, which relates to the intensity of the input signal and the magnitude of the output response. Two important aspects of the regulatory function include its sensitivity (linked to the steepness of the function) and the apparent dissociation constant (K_d , equivalent to the stimulus intensity required to obtain half the response). In gene expression, a critical feature of the output response is its inherent variability. Given the small number of molecules involved in the biochemical processing of signaling (gene transcription [1], chromatin remodeling [2], formation of transcription reinitiation complexes [2,3] and protein translation [4]), this results in variable protein concentrations across cell populations. This phenotypic variation can affect survival [3,5,6,7,8,9], differentiation [10,11,12,13,14,15,16] and also increases evolvability [17]. Thus, biochemical circuits must have evolved to maximise the overall performance of the organism. Sometimes, the evolutionary optimisation process can be constrained to obtain a robust system (i.e., insensitive to the precise values of biochemical parameters). In other cases, biochemical circuits need fine-tuned intracellular parameters and, consequently, inherent biochemical noise must be minimised.

Several studies have suggested the existence of optimisation criteria in the design of some regulatory systems [18,19,20,21,22,23]. Of course, this requires one or more functionality criteria operating on the course of evolution. Among these, cost-benefit (the trade-off between the metabolic costs of protein synthesis and the benefits of protein function [24,25]), maximisation of information transmission [21,20,26] and minimisation of biochemical noise [20,27,28], have been mentioned as functionality criteria for an optimal design.

In particular, the latter has been addressed for multistage signaling cascades in which several genes are involved [29]. Nonetheless, the existence of such a criterion operating as a design principle at single-gene level has not yet been explored. On the other hand, there have been recent advances in identifying and characterising a variety of mechanisms involved in the regulation of gene expression [30,31,32]. Nevertheless, the way in which the complex architecture of a *cis*-regulatory systems (CRS), (binding sites (BSs), transcription factors (TFs), cooperativity mechanisms, DNA-loops) orchestrates the required response is still unknown. Moreover, the evolutionary criteria operating over the latter must also be considered.

In order to achieve a clearer understanding of the principles guiding the design of complex CRS, we previously studied a stochastic model for a single TF capable of binding cooperatively to three regulatory BSs [33,34]. In these papers, we described two different cooperative binding mechanisms (CBM): the recruitment mechanism, which increases the ability for new TF recruitment, and the stabilisation mechanism, which increases the stability of

the TF-DNA bond. We also reported that, at single-gene level, the sensitivity of the output response can be due to multiple regulatory BSs for TFs that act cooperatively [33]. Moreover, we showed that cooperative interactions between TFs increase the intrinsic fluctuations associated with transcription in a mechanism dependent manner [33,34]. In such papers our study was limited to CRS with three BSs, where each TF interacts with all the TFs which are bound to DNA.

If the design principles of regulatory systems are subject to criteria that maximise sensitivity and minimise noise, our previous results suggested that there could be a trade-off between the number of BSs and the intensity of the cooperative interaction. Thus, the primary purpose of this paper was to explore the existence of this trade-off and, for this, it was necessary to generalise our previous model [33] to consider a variable number of BSs, where each TF could interact with one, two, or more TFs bound to DNA. To study this complex model we developed a small-noise approximation (SNA) framework, introduced in [35], which enables the analysis of arbitrarily complex CRS acting in a small-noise regime.

Our results showed that an increase in the number of BSs can either decrease or increase expression noise depending on the cooperativity intensity and the number of effective TF interactions. Furthermore, we found a scenario where there is an optimal trade-off between cooperativity intensity (a factor that increases noise) and the number of BSs (a factor that decreases noise). The significance of this finding is at least two-fold: from an evolutionary point of view, it represents an alternative functionality criterion mechanism based on noise minimisation, and it also contributes a design principle for synthetic biology projects.

Methods

Modeling *cis*-regulatory systems

In order to analyse the effects of tandem BS architecture on transcriptional regulation, we generalised the model used in [33]. Thus, the proposed CRS includes N regulatory sites for the same TF. Figure 1 illustrates an example of a CRS that includes three regulatory binding sites, showing the different transitions between states. The states $s = 1, 2, \dots, N + 1$ represent, respectively, states with zero, one, and N sites occupied by TFs. The states $s \geq N + 2$ correspond to the transcriptional complex formation, where all components required for transcription are assembled on the CRS. For simplicity, we consider that TFs do not bind or unbind after the formation of the transcriptional complex. Once the core transcription apparatus is formed, the synthesis of one mRNA copy begins. TFs can bind to regulatory sites with a probability which is proportional to the TF concentration c following the law of mass action for elementary reactions. TF unbinding events depend only on the kinetic constants. Thus, the elements of the transition matrix $\hat{\mathbf{T}}$ are $t_{s,s+1} = ck_{s,s+1}$ and $t_{s+1,s} = k_{s+1,s}$ for $s = 1, 2, \dots, N$, where $k_{s,r}$ are the kinetic rates.

If we consider cooperative interactions between TFs, kinetic rates are not independent because previous binding alters the actual binding or unbinding process. Known relationships between the system's kinetics and thermodynamic properties (principle of detailed balance) allow us to write the kinetic rates, $k_{s,s+1}$ and $k_{s+1,s}$ with $s = 1, 2, \dots, N$, in terms of three parameters [33]: the binding rate p , the unbinding rate q , and ϵ which represents the cooperativity intensity, i.e., ($\epsilon = e^{-\frac{\Delta G_1}{kT}}$, where ΔG_1 is the free energy of the interaction). Beyond this simplification, this relationship also allows the identification of two CBMs [33]. The first, the recruitment mechanism, occurs when the presence of already bound TFs alter DNA affinities increasing binding rates

$k_{s,s+1}$, which generates a greater ability for new TF recruitment for DNA binding (sketched in Fig. S1A). The second CBM, the stabilisation mechanism, acts when TF interaction diminishes the unbinding rate $k_{s+1,s}$, increasing the stability of the TF-DNA interaction (sketched in Fig. S1B). Thus, following [33], we can write

$$k_{s,s+1} = f_s(\epsilon)(N + 1 - s)p$$

$$k_{s+1,s} = sq, \tag{1}$$

for the first mechanism, while for the second mechanism we have

$$k_{s,s+1} = (N + 1 - s)p,$$

$$k_{s+1,s} = f_s^{-1}(\epsilon)sq, \tag{2}$$

where $f_s(\epsilon)$ establishes how the cooperative interaction of the state s affects the kinetic rates of the new binding or unbinding processes. In previous work we considered the special case of a CRS with three sites, where each TF interacted with all TFs already bound to DNA, and all interactions had the same ΔG_1 [33,34]. In this case, f_s can be written as $f_s(\epsilon) = \epsilon^{s-1}$ for all s . However, due to the spatial distribution of the BSs along the regulatory region, scenarios with a lower number of TF interactions can occur. Even though the CRS model does not include any spatial details, we can emulate several alternative CRS configurations by considering different f_s :

- (i) each TF interacts with only one bound TF, that is $f_s(\epsilon) = \epsilon$ for $s > 1$;
- (ii) each TF interacts with two bound TFs, that is $f_2(\epsilon) = \epsilon$ and $f_s(\epsilon) = \epsilon^2$ for $s > 2$;
- (iii) each TF interacts with three bound TFs, that is $f_2(\epsilon) = \epsilon$, $f_3(\epsilon) = \epsilon^2$ and $f_s(\epsilon) = \epsilon^3$ for $s > 3$.
- (iv) each TF interacts with all bound TFs, that is $f_s(\epsilon) = \epsilon^{s-1}$ for all s . Of course, $f_1(\epsilon) = 1$ for all cases, since there is no bound TF to interact in the state $s = 1$. Figure 2 illustrates three regulatory systems, with the same occupancy level, where the number of TF interactions increases from one to three.

Small-noise approximation

As other authors [36,35,33], we use the master equation approach to derive the response of a cell population to an inductive signal. The state of our system will be specified by two stochastic variables: the chemical state of the CRS s , and the number of transcripts m , where m is a positive integer and s is either $1, \dots, \mathcal{N}$. The probability to find the system in the state (s, m) , at any time t , can be written as a vector $\mathbf{P}_m(t) = (P_{1,m}(t), P_{2,m}(t), \dots, P_{\mathcal{N}}(t))$. The time evolution for this probability is governed by the following master equation:

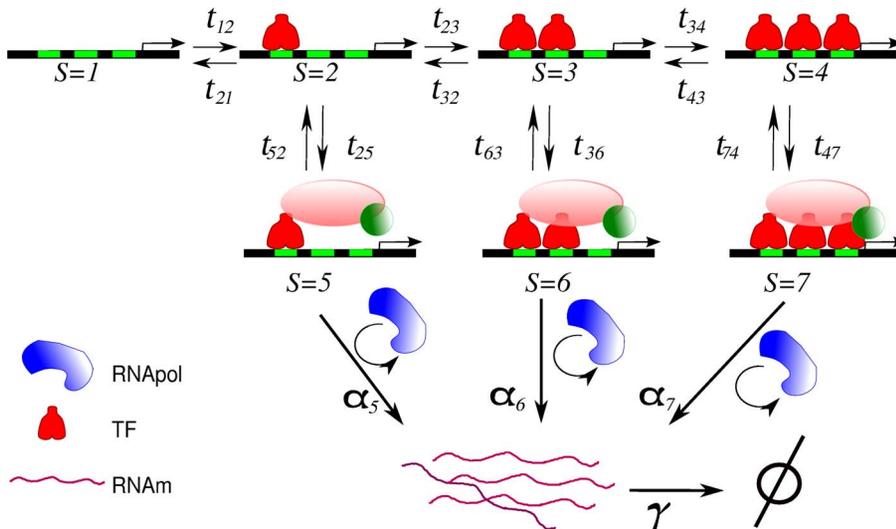


Figure 1. Sketch of the *cis*-regulatory system. For simplicity, we consider a diagram of the promoter region that includes only three regulatory BSs (green boxes) where TF proteins can bind (red molecules). The states on the top layer $s = 1, 2, \dots, N + 1$ represent states with zero, one, and N BSs occupied by TFs, respectively. After one or more TFs bind to the regulatory BSs, the regulatory system is able to initiate transcriptional complex formation, where all the components required for transcription (pink and green molecules) are assembled on the CRS. States $s \geq N + 2$ in the middle layer are those capable of recruiting RNAPol (blue protein) to synthesise mRNA. The synthesis rate (α_s) depends on the number of bound TFs. The regulatory system can make transitions from state j to state i with rate t_{ij} . The elements of the transition matrix $\hat{\mathbf{T}}$ are $t_{s,s+1} = ck_{s,s+1}$ and $t_{s+1,s} = k_{s+1,s}$ for $s = 1, 2, \dots, N$, where $k_{s,r}$ are the kinetic rates.
doi:10.1371/journal.pone.0084020.g001

$$\dot{P}_{s,m} = \alpha_s(P_{s,m-1} - P_{s,m}) + \gamma[(m+1)P_{s,m+1} - mP_{s,m}] + \sum_{r=1}^N t_{s,r}P_{r,m} \tag{3}$$

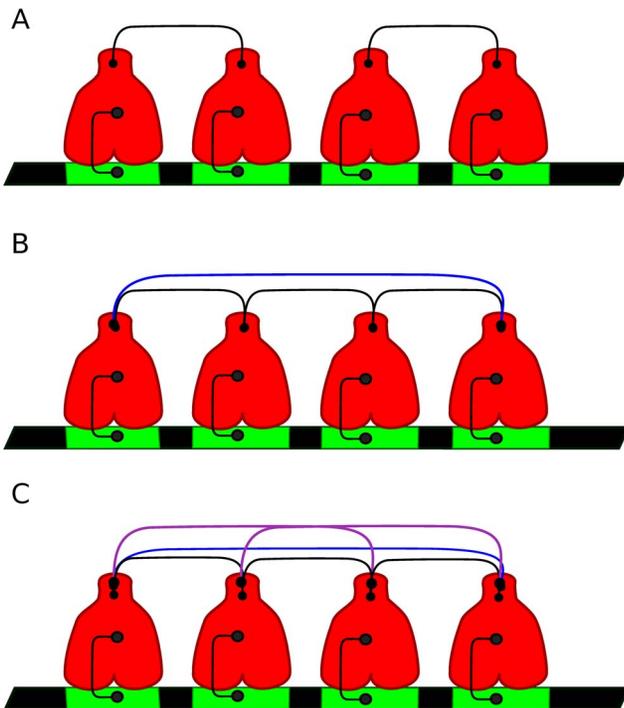


Figure 2. Different number of cooperative interactions. The schematic shows a CRS with four BSs where a bound TF can interact with: (A) only one TF, case (i), (B) two TFs (ii), (C) three other TFs, case (iii).
doi:10.1371/journal.pone.0084020.g002

where $t_{s,r}$ is the transition probability per time unit from state r to state s . The first two terms correspond to the production and degradation of mRNA, respectively. The model assumes that mRNA is synthesised at rate α_s which depends on the state s , whereas it is degraded linearly with rate γ . The last term on the right hand side of the equation (3) describes CRS dynamics.

An exact analytical description has been obtained previously for a steady state with $N = 3$ [33], but that is not possible for $N > 3$ without incorporating some approximations. In this sense, following Kepler and Elston [35], we apply two different approximations to Eq. (3): (i) when the number of transcripts is large, a diffusion approximation is used; (ii) when the CRS transition rates are large compared to the rate of production and degradation of transcripts, an SNA is possible. We now define an appropriately scaled continuous and dimensionless variable $x = \frac{m}{\langle m \rangle_*}$, where $\langle m \rangle_*$ is the mean number of transcripts in the steady state. This allows us to introduce the transformation

$$P_{s,m}(t) = \int_{(m-1/2)/\langle m \rangle_*}^{(m+1/2)/\langle m \rangle_*} \varrho_s(x,t) dx, \tag{4}$$

which defines the probability density function $\varrho_s(x,t)$. Consequently, Eq. (3) is transformed into an evolution equation for ρ_s . The use of a second order diffusion approximation and then of a first order SNA, leads to an equation for marginal density $\rho = \sum_s \varrho_s$,

$$\partial_t \rho(x,t) = \sum_{k=0}^4 M^k \partial_x^k \rho(x,t). \tag{5}$$

Neglecting terms of higher order than two on $1/m$ and further algebraic steps, allows us to find expressions for the coefficients M^k as follows: $M^0 = \gamma$, $M^1 = \gamma x + \frac{\gamma}{m} - \gamma$ and $M^2 = \frac{\gamma}{2m} x + \frac{\gamma}{2m} - \lambda \frac{R}{m^2}$, while $M^3 = M^4 = 0$ where, to simplify notation, we dropped the angle brackets and the star to denote the mean value of m in its steady state, $m = \langle m \rangle_*$. The factor R depends on the kinetic rates, $R = \sum_{ij} t_{ij}^\dagger P_{*j} \alpha_j$. (see Text S1 for a detailed derivation). λ keeps track of the SNA expansion order. Notice that noise due to CRS dynamics, through factor R , influences the term associated with diffusion in the Fokker-Planck equation, while the deterministic limit is restored when $\lambda \rightarrow 0$. Thus, the master equation can be cast into a Fokker-Planck equation for marginal density $\rho(x,t)$

$$\partial_t \rho(x,t) = -\partial_x [A(x)\rho(x,t)] + \frac{1}{2} \partial_x^2 [B(x)\rho(x,t)] \tag{6}$$

with $A(x) = \gamma(1-x)$ and $B(x) = \frac{\gamma}{m}(x+1 - \frac{2\lambda R}{m\gamma})$. An important advantage of this formulation is that the associated Fokker-Planck equation has a closed expression for steady state density in terms of a simple quadrature

$$\rho_*(x) = z \left[1 + \frac{x}{\beta} \right]^\delta \exp\left(-\frac{\delta x}{(1+\beta)}\right), \tag{7}$$

with $\beta = 1 - \frac{2R}{\gamma m}$, $\delta = 2m(1+\beta) - 1$ and z is a normalisation constant; m is indicating the average number of transcripts in the steady state. Notice that expression (7) is χ -squared and similar to the Gaussian form in the SNA region of validity. The expression (7) is general, meaning that it is valid for any CRS, whatever the dynamics and number of BSs. As solution (7) is an approximated one, an estimation of the SNA accuracy is necessary. For this, we define Δ as the ratio between $\mathcal{O}^2(\lambda)$ and R (which is proportional to $\mathcal{O}(\lambda)$). Consequently, our approximation predictions are accurate for small Δ (See Text S1 for details).

Results

In this section we validated our SNA and analysed fluctuation behaviour and sensitivity as a function of the number of binding sites and ϵ within SNA validity limits.

In order to validate the solutions obtained with the proposed approximation, we compared the transcript number analytical distributions predicted by Eq. (4), with the corresponding histograms obtained by stochastic simulation using the Gillespie method [37]. Figure 3 shows the distribution which was obtained for a CRS with three binding sites in which each TF interacts with all TFs already bound to DNA (case iv). Green color corresponds to the parameter values listed in Table 1. To highlight the effect of CRS kinetics on the approximation, we also show the distributions that were obtained with different kinetic rates, by multiplying the

elements of matrix \hat{T} by factor 10 (blue color), and by factor 100 (red color). Fig. 3A corresponds to the distribution obtained for a non-cooperative binding case (i.e. $\epsilon = 1$), while Fig. 3B and Fig. 3C correspond to recruitment and stabilisation CBMs, respectively, (with $\epsilon = 12$). As expected, our results showed excellent correspondence for fast transitions between CRS states in relation to production and degradation rates. When cooperative binding was included in the model ($\epsilon = 12$), approximation accuracy decreased. Moreover, comparison of Fig. 3B and Fig. 3C shows that the approximation for recruitment CBM is more accurate than for stabilisation CBM. This occurs because stabilisation CBM decreases unbinding rates (i.e., slows down CRS kinetics), as can be observed explicitly in Eq. (1) and Eq. (2). Furthermore, SNA accuracy increases for lower interaction intensity and for a lower number of interactions (data not shown). The comparison between SNA predictions and the corresponding simulation results in Table 2 shows that, when $\Delta < 0.5$, mean and standard deviation estimations for each distribution are reliable.

In this context, we applied SNA to study the performance of a complex CRS using the parameter values listed in Table 1. In particular, we were interested in quantifying the phenotypic noise (spread of expression levels within the cell population) of an activator switch in terms of the CRS architecture (i.e., as a function of the number of BSs and f_s), and the intensity of the cooperativity ϵ involved in the CRS. In this sense, to characterise noise expression we computed the fluctuation/signal ratio (which is known as noise) defined by $\eta = \sigma_m / \langle m \rangle$ [38], at an activator concentration c equal to the K_d of the dose-response curves. Given the SNA region of validity, hereafter both $\langle m \rangle$ and σ_m were obtained from SNA distributions with $\Delta < 0.5$.

Fig. 4 depicts two important features of the regulatory system response, as a function of N : the noise η (A) and the Hill coefficient n_H of the mean response (B). These features are presented for both CBMs: recruitment (filled circles) and stabilisation (open circles). Since the Hill coefficient is CBM-independent, curves in panel B are superimposed. Fig. 4 illustrates case (i), where each TF interacts with only one bound TF. This figure shows how η decays monotonically (Fig. 4A) while the steepness of the associated response (Fig. 4B) increases with the number of BSs N for all curves. The red and blue curves correspond to $\epsilon = 6$ and 12, respectively. Similar behaviour can be found for other ϵ values. This is illustrated by the density plot (on a plane) shown in Figure S2 for both CBMs (N, ϵ). This behaviour suggests that, when TFs interact with only one other TF, additional BSs on the CRS improve the signal/noise ratio and the sensitivity of the switch response (the latter is characterised by the Hill coefficient). However, when CRS architecture admits more than one interaction between TFs, noise, as a function of N , shows a complex behaviour. Fig. 5A depicts noise for case (ii), in which each TF interacts with two bound TFs. In this case, η behaviour depends on which CBM is acting and on the intensity of this cooperativity, ϵ . Red and blue curves correspond to $\epsilon = 6$ and 12, respectively, while the filled and open circles correspond to recruitment and stabilisation CBMs, respectively. For the cases illustrated in Fig. 5A, noise η shows a minimum around $N = 2$ when stabilisation CBM is acting, and η increases with N for $N > 2$. When recruitment CBM is acting, η can show a minimum in an ϵ dependent manner, or not, as can be observed in Fig. 5A inset. At $\epsilon = 6$ (red curve) η decays monotonically with N and there is no valley. However, for a higher intensity of cooperativity ($\epsilon = 12$), η has a minimum around $N = 6$. On the other hand, similarly to Fig. 4B, Hill coefficient associated to the steepness of the mean response, increases with the number of BSs N , as is shown in Fig. 5B. However, the steepness for case (ii) is always

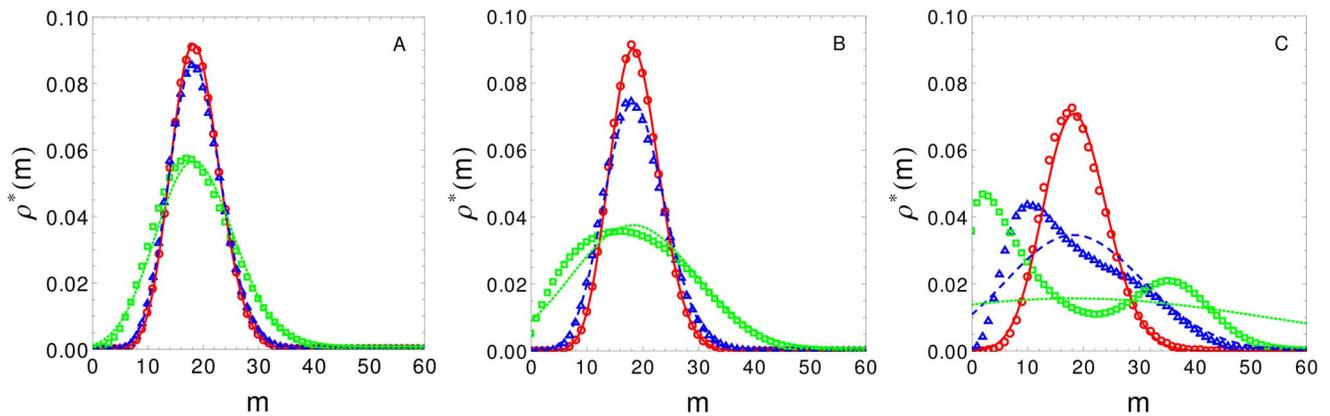


Figure 3. Transcript distributions. Comparison between exact (symbols, obtained by stochastic simulations) and analytical distributions (curves, obtained by the SNA) for three different CRS kinetic rates (red, blue and green correspond to quick, medium and slow CRS dynamics, respectively). Panels from left to right show non-cooperative, recruitment and stabilisation CBMs, respectively. In the last two cases $\epsilon = 12$. The comparison shows very good correspondence when CRS state transitions are fast in relation to production and degradation rates. This figure shows that the SNA fails to correctly predict the distribution for slow kinetics when there is TF cooperative binding, even though this is more apparent for the stabilisation CBM. doi:10.1371/journal.pone.0084020.g003

higher than for case (i). This means that Hill coefficients n_H are determined not only by the number of BSs in the regulatory system and the interaction energy between TFs, but also by the number of TF interactions which are admitted by the CRS. Noise behaviour in an ϵ dependent manner can be better illustrated by density plots on a plane (N, ϵ) , as those depicted in the bottom panels of Fig. 5. Fig. 5C and Fig. 5D show η in grey scale maps as a function of N and ϵ for both CBMs. For recruitment CBM (Fig. 5C), a valley appears when ϵ is greater than 10 and the position of the minimum changes with ϵ . For example, for $\epsilon = 10$ there is a valley around $N = 6$, whereas for $\epsilon = 12$ (blue dotted line) the valley is around $N = 6$, and for $\epsilon = 18$ the minimum position shifts to $N = 4$. For stabilisation CBM (Fig. 5D), the minimum is more apparent at a lower intensity of cooperativity. For example there is a minimum at $N = 2$ for $\epsilon = 6$ (red dotted line), although this valley also exists at smaller ϵ (data not shown).

Fig. 6 depicts noise η (A) and Hill coefficients n_H (B) as a function of N for case (iii), where each TF interacts with three bound TFs. Figs. 6B and 6D depict case (iv), where each TF interacts with all bound TFs. The red and blue curves correspond to $\epsilon = 6$ and 12, respectively, while the open and filled circles

correspond to recruitment and stabilisation CBMs, respectively. Both CBMs show a minimum in the fluctuation/signal ratio for $\epsilon = 6$. However, the number of BSs necessary for an optimal performance differ between CBMs: the minimum for recruitment CBM is around $N = 3$ or 4, depending on ϵ , while for stabilisation CBM the minimum is around $N = 2$. Thus, for many interactions, as cases (iii) and (iv), the signal/noise ratio is maximised by an architecture with two or three binding sites. Minima in η can also be found for faster CRS with weaker cooperativity (data not shown). Moreover, it can also be observed that the sensitivity of the switch response (characterised by the Hill coefficient) increases faster with N than in previous cases. In particular, for case (iv) (Fig. 6D), n_H approximation by N is accurate only for a high interaction energy and $N > 4$. Finally, Figures 4, 5 and 6 suggest that noise and steepness increase with the number of TF interactions which are admitted by the CRS.

Table 1. Parameter values.

Name	Symbol	Value
TF binding	p	0.25
TF unbinding	q	0.75
Coop. intensity	ϵ	variable
TC assembly *	$k_{s,s+N}$	$s \times 0.5$
TC disassembly *	$k_{s+N,s}$	0.5
Transcript prod.	α	1.5
Transcript degrad.	γ	0.03
Number of BS	N	variable

kinetics rates $k_{s,s+1}$ and $k_{s+1,s}$, with $s = 1, \dots, N$, were obtained in each case using the corresponding Eqs. (1) or (2), with the above values of p , q and ϵ . The time unit is min and the concentration is an arbitrary unit. * with $s = 2, \dots, N+1$.

doi:10.1371/journal.pone.0084020.t001

Table 2. SNA performance.

plot	mean value		standard deviation		Δ
	simulation	SNA	simulation	SNA	
Fig.3A green	18.75	18.75	7.01	6.94	0.134
Fig.3A blue	18.73	18.75	4.68	4.67	0.013
Fig.3A red	18.77	18.75	4.37	4.37	0.001
Fig.3B green	18.75	18.75	10.18	10.02	0.581
Fig.3B blue	18.74	18.75	5.40	5.38	0.058
Fig.3B red	18.77	18.75	4.45	4.43	0.006
Fig.3C green	18.78	18.75	25.77	14.87	12.558
Fig.3C blue	18.71	18.75	10.97	10.04	1.256
Fig.3C red	18.77	18.75	5.65	5.57	0.125

Comparison of the transcript number mean values, and its associated standard deviation, obtained by Gillespie simulations and predicted by SNA for the plots depicted in Figure 3. The last column depict the estimator for SNA accuracy Δ . doi:10.1371/journal.pone.0084020.t002

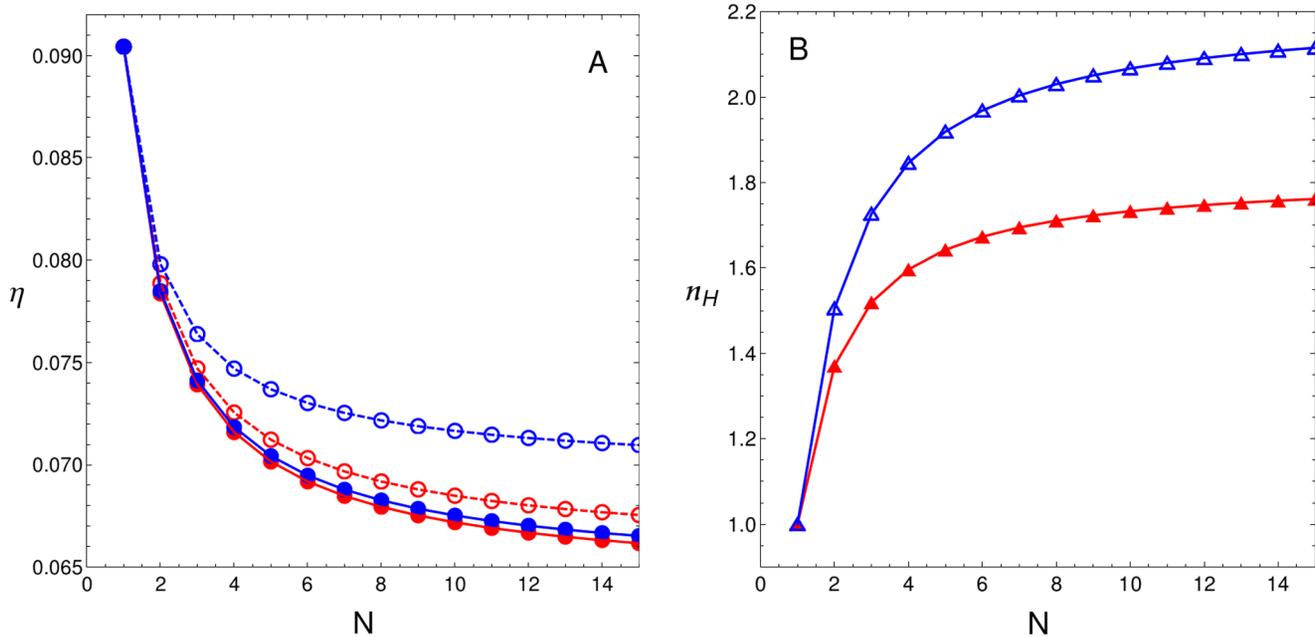


Figure 4. Case (i), TFs interact with only one other TF. When $f_s(\epsilon) = \epsilon$ for $s > 1$, noise η and steepness n_H behaviour is monotonic with respect to the number of BSs N . (A) $\eta = \sigma_m / \langle m \rangle$ as a function of N for both CBMs (filled and open circles indicate recruitment and stabilisation CBMs, respectively) with $\epsilon = 6$ (red circles) and $\epsilon = 12$ (blue circles). (B) n_H as a function of N for $\epsilon = 6$ (red triangles) and $\epsilon = 12$ (blue triangles). Curves correspond to cubic spline interpolations.
doi:10.1371/journal.pone.0084020.g004

Discussion

Regulation of gene expression is a topic of central importance in biology. Given our increasing ability to monitor gene expression levels and model its regulation, many key issues have been unraveled in the past years. An example of this is the discrete nature of the transcriptional process which impacts on the noise expression phenomenon and has promoted the extended use of stochastic modeling to study the origin of noise expression and its propagation. However, many other issues require new models and methodological approaches to be elucidated. For example, an important aspect which impacts synthetic biology, is to understand how the CRS architecture (i.e., the biological components of the regulatory system and its organisation) determines the regulatory function features and the associated fluctuations. Most of the previous studies which analysed gene product fluctuations due to transitions between CRS states, used models which considered a small number of states for the CRS. This is mainly because theoretical approaches become intractable for complex CRS [34,39], but the use of simple models can limit our understanding of gene regulatory phenomena involving complex CRS. However, modeling of complex CRS is becoming more frequent in specialised literature [39,40,41].

This paper had mainly two aims: the first was methodological and consisted in presenting an approximation of the master equation to deal with stochastic models for arbitrarily complex CRS in an analytical fashion. In this respect, we developed the SNA for a generic CRS which can include many states. Thus, we derived an explicit form for the Fokker-Plank equation from a microscopic description which can be applied to any CRS. This approximation was validated against the distribution of mRNA levels generated by a Gillespie simulation and showed very good correspondence in the unimodal regime. Other approaches have recently been developed which are capable of dealing with

complex CRS, among these, the stochastic spectral analysis [42] and the effective rate equation approach developed by Grima [43]. These approaches could deal with bimodal distributions more accurately than SNA.

The second and most important aim, was to study the impact of tandem regulatory BSs on CRS response to TF activators, using the above mentioned approximation. Namely, how noise and sensitivity depend on the number of regulatory BSs. At this point it is important to distinguish regulatory BSs from decoy BSs that competitively bind TFs. Decoy sites protect these proteins from degradation and thereby indirectly influence the expression of target promoters. The role of decoy BSs was studied recently by [44,45], who showed that protective decoys can buffer noise by reducing correlations between TFs. Hitherto, the role of the number of regulatory BSs on the regulation of expression level and its associated fluctuation, have only been addressed by a few studies. In one of these studies, experimental evidence from genetically modified mammalian cells showed that increases in the number of regulatory BSs lead to higher noise (Fig. 3B of [46]). However, the contrary effect was reported in *Drosophila*, where an increasing number of Bcd BSs decreased noise (Fig. 6 of [41]). A more recent study which analysed the expression level and noise associated with several native targets of the transcription factor Zap1 [32], suggested that the relationship between expression level and noise is a feature that characterises the CRS and is determined by CRS architecture. However, in all these studies the precise mechanisms by which noise is controlled remained unidentified.

In this context, we previously reported that noise expression behaviour depends on the acting CBM and on cooperativity intensity [33]. Our present results expand upon this analysis showing that CRS performance is a consequence of the interplay between diverse factors: CBMs, the number of regulatory BSs, and the number and intensity of TF interactions. Figures 4 and 5 show that increasing the number of regulatory BSs can lead to noise

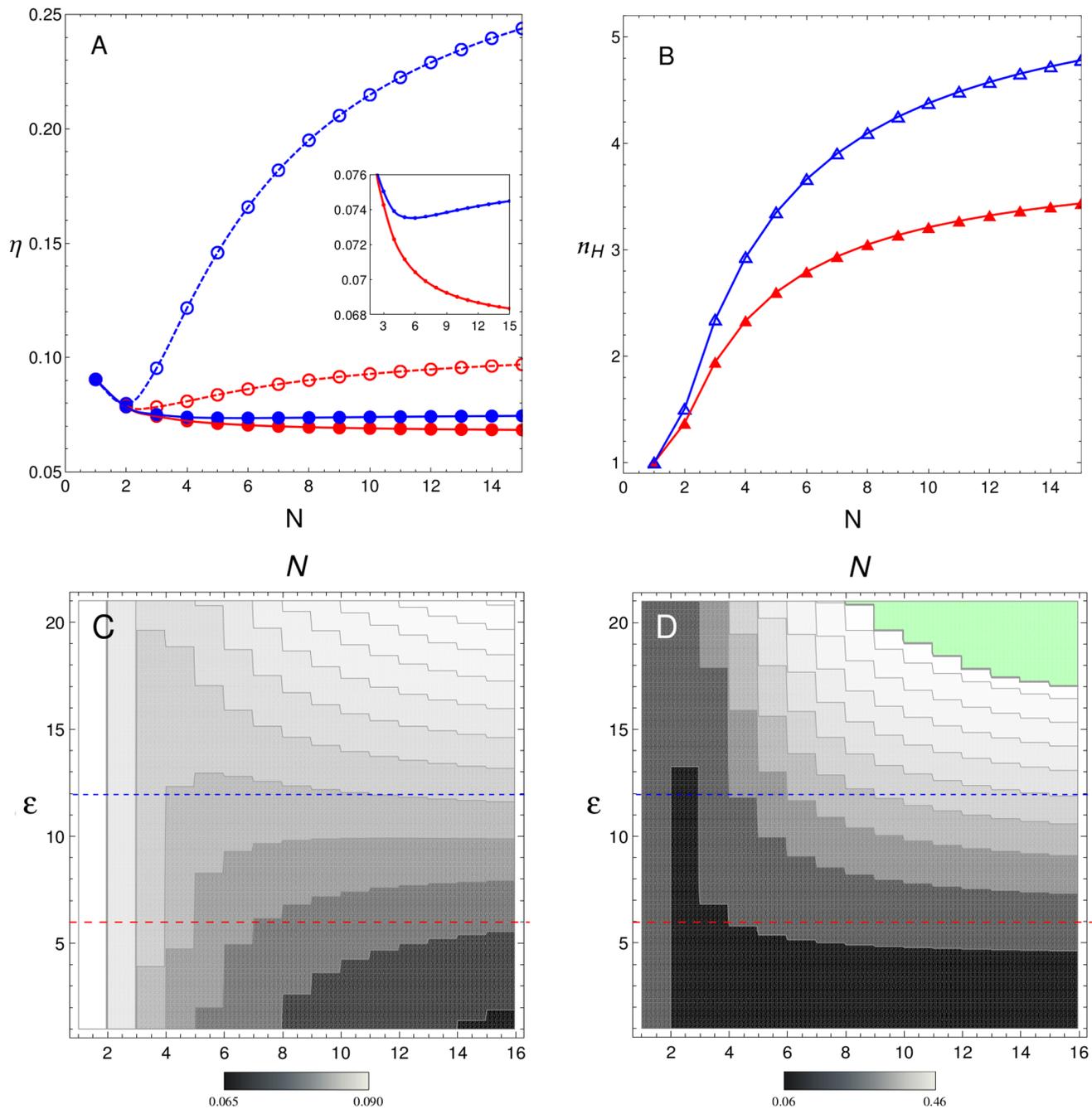


Figure 5. Case (ii), each TF can interact with up to two other TFs. When $f_2(\epsilon) = \epsilon$ and $f_s(\epsilon) = \epsilon^2$ for $s > 2$, noise η shows a more complex behaviour which depends on the acting CBM and the intensity of cooperativity. (A) $\eta = \sigma_m / \langle m \rangle$ as a function of the number of sites N for both CBMs (filled and open circles indicate recruitment and stabilisation CBMs, respectively) with $\epsilon = 6$ (red circles) and $\epsilon = 12$ (blue circles). The inset rescales the recruitment case for both ϵ -values. (B) n_H as a function of N for $\epsilon = 6$ (red triangles) and $\epsilon = 12$ (blue triangles). Curves correspond to cubic spline interpolations. Lower panels correspond to density plots of η as function of N and ϵ for recruitment CBM (C) and stabilisation CBM (D). Dotted lines indicate the values of ϵ used in panels A and B. For recruitment CBM, the density plot clearly shows the existence of a valley in η around $N = 6$ for $\epsilon > 12$ while, for stabilisation CBM, the valley in η is apparent for lower values of ϵ around $N = (2, 4)$. The green area in panel D denotes the region where $\Delta \geq 0.5$ and SNA predictions are less accurate. doi:10.1371/journal.pone.0084020.g005

reduction in a low interaction scenario. We also show that the type of acting CBM and the number of interactions play a critical role in determining how noise η depends on the number of regulatory BSs. Finally, we identified a scenario where noise presents a minimum on a plane defined by N and ϵ , where the size of the valley and its position depend on the number of interactions and

the CBM. Moreover, the position of the valley was found around intervals $[2, 6]$ for N , and $[3, 20]$ for ϵ (corresponding to ΔG_I ranging from 0.65–1.8 kcal/mol). Since the mentioned values for N and ΔG_I correspond to those found in the spectrum of real biological systems, our results suggest that evolutionary processes

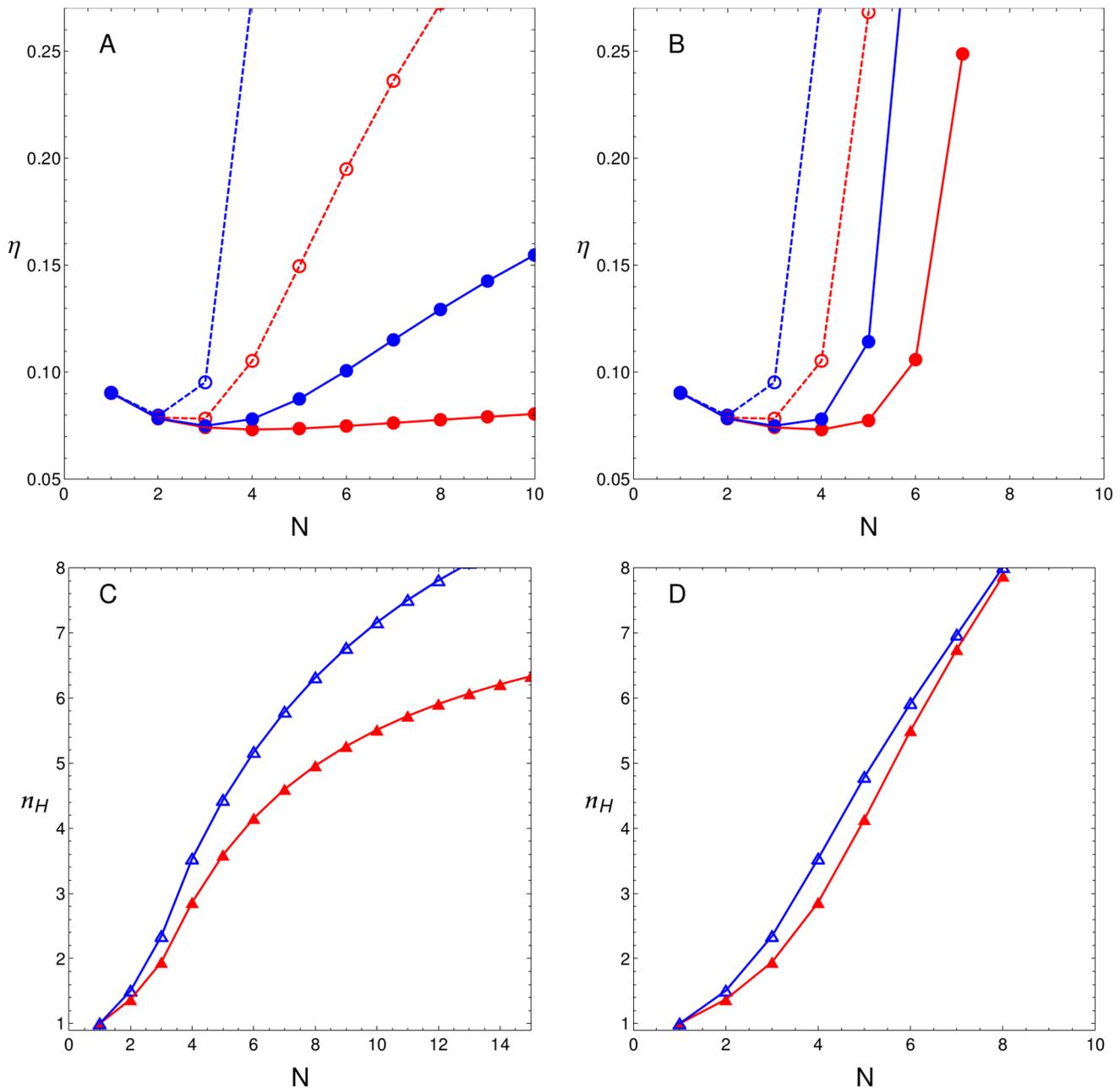


Figure 6. Cases with many interactions. Top panels: $\eta = \sigma_m / \langle m \rangle$ as a function of the number of sites N for both CBMs (filled and open circles indicate recruitment and stabilisation CBMs, respectively) with $\epsilon = 6$ (red circles) and $\epsilon = 12$ (blue circles). Bottom panels: n_H as a function of N for $\epsilon = 6$ (red triangles) and $\epsilon = 12$ (blue triangles). Panels A and C show the case in which each TF can interact with up to three other TFs, i.e., case (iii). Panels B and D show the case where each TF interacts with all available TFs (case (iv)) where $f_s(\epsilon) = \epsilon^{s-1}$ for all s . Curves correspond to cubic spline interpolations.
doi:10.1371/journal.pone.0084020.g006

are capable of adjusting these parameters to optimise noise in single-gene switches.

Supporting Information

Figure S1 Cooperative binding mechanisms. The effect of cooperative binding on binding and unbinding. (A) In the recruitment mechanism, TFs already bound to the DNA increase the ability for recruiting new TFs. (B) In the stabilisation mechanism, TF interaction diminishes the unbinding rate. ΔG_I

denotes the free energy involved in the cooperative binding; the black link represents a chemical interaction. (TIF)

Figure S2 Density plots for case (i). η as function of ϵ and N : for recruitment (panel A) and stabilisation (panel B) CBMs. Dotted lines indicate the values of ϵ used in Figs. 4A and 4B. The density plots clearly show that there is no valley in η for the explored parameters. (TIF)

Text S1 Supporting Information.

(PDF)

Acknowledgments

DM is a CICPBA researcher (Argentina), CBM is a CONICET researcher (Argentina), and LD is a CONICET researcher (Argentina).

References

- Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* 422: 633–637.
- Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* 304: 1811–1814.
- Blake WJ, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, et al. (2006) Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell* 24: 853–865.
- Ozbudak EM, Thattai M, Kurtser I, Grossman AD, Oudenaarden AV (2002) Regulation of noise in the expression of a single gene. *Nat Genet* 31: 69–73.
- Thattai M, Oudenaarden AV (2004) Stochastic Gene Expression in Fluctuating Environments. *Genetics* 167: 523–530.
- Kussell E, Leibler S (2005) Phenotypic diversity, population growth, and information in fluctuating environments. *Science* 309: 2075–2078.
- Smith MCA, Sumner ER, Avery SV (2007) Glutathione and Gts1p drive beneficial variability in the cadmium resistances of individual yeast cells. *Mol Microbiol* 66: 699–712.
- Lu T, Shen T, Bennett MR, Wolynes PG, Hasty J (2007) Phenotypic variability of growing cellular populations. *Proc Natl Acad Sci USA* 104: 18982–18987.
- Bayer TS, Hoff KG, Beisel CL, Lee JJ, Smolke CD (2009) Synthetic control of a fitness tradeoff in yeast nitrogen metabolism. *J Biol Eng* 3: 1.
- Paldi A (2003) Stochastic gene expression during cell differentiation: order from disorder? *Cell Mol Life Sci* 60: 1775–1778.
- Maamar H, Raj A, Dubnau D (2007) Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* 317: 526–529.
- Süel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB (2007) Tunability and noise dependence in differentiation dynamics. *Science* 315: 1716–1719.
- Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S (2008) Transcriptomewide noise controls lineage choice in mammalian progenitor cells. *Nature* 453: 544–547.
- Levin M (2003) Noise in gene expression as the source of non-genetic individuality in the chemotactic response of *Escherichia coli*. *FEBS Letters* 550: 135–138.
- Neiðez-Nguyen TMA, Parisot A, Vignal C, Rameau P, Stockholm D, et al. (2008) Epigenetic gene expression noise and phenotypic diversification of clonal cell populations. *Differentiation* 76: 33–40.
- Korobkova E, Emonet T, Vilar JMG, Shimizu TS, Cluzel P (2004) From molecular noise to behavioural variability in a single bacterium. *Nature* 428: 574–578.
- Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL (2007) Genetic properties influencing the evolvability of gene expression. *Science* 317: 118–121.
- François P, Siggia ED (2008) A case study of evolutionary computation of biochemical adaptation. *Phys Biol* 5: 026009.
- Gerland U, Hwa T (2009) Evolutionary selection between alternative modes of gene regulation. *Proc Natl Acad Sci USA* 106: 8841–6.
- Tostevin F, ten Wolde PR, Howard M (2007) Fundamental limits to position determination by concentration gradients. *PLoS Comput Biol* 3: e78.
- Tkačik G, Callan CG, Bialek W (2008) Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci USA* 105: 12265–70.
- Celani A, Vergassola M (2010) Bacterial strategies for chemotaxis response. *Proc Natl Acad Sci USA* 107: 1391–6.
- Mehta P, Goyal S, Long T, Bassler BL, Wingreen NS (2009) Information processing and signal integration in bacterial quorum sensing. *Mol Syst Biol* 5: 325.
- Savageau MA (1977) Design of Molecular Control Mechanisms and the Demand for Gene Expression. *Proc Natl Acad Sci USA* 74: 5647–5651.
- Chubukov V, Zuleta IA, Li H (2012) Regulatory architecture determines optimal regulation of gene expression in metabolic pathways. *Proc Natl Acad Sci USA* 109: 5127–5132.
- François P, Siggia ED (2010) Predicting embryonic patterning using mutual entropy fitness and in silico evolution. *Development* 137: 2385–2395.
- Saunders TE, Howard M (2009) Morphogen profiles can be optimized to buffer against noise. *Phys Rev E* 80: 041902.
- Sokolowski TR, Erdmann T, ten Wolde PR (2012) Mutual Repression Enhances the Steepness and Precision of Gene Expression Boundaries. *PLoS Comput Biol* 8: e1002654.
- Thattai M, van Oudenaarden A (2002) Attenuation of noise in ultrasensitive signaling cascades. *Biophys J* 82: 2943–2950.
- Segal E, Shapira M, Regev A, Pe'er D, Botstein D, et al. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 34: 166–176.
- Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, et al. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat Biotechnol* 30: 521–530.
- Carey LB, van Dijk D, Sloot PMA, Kaandorp JA, Segal E (2013) Promoter sequence determines the relationship between expression level and noise. *PLoS Biol* 11: e1001528.
- Gutierrez PS, Monteoliva D, Diambra L (2009) Role of cooperative binding on noise expression. *Phys Rev E* 80: 11914.
- Gutierrez PS, Monteoliva D, Diambra L (2012) Cooperative Binding of Transcription Factors Promotes Bimodal Gene Expression Response. *PLoS One* 7: e44812.
- Kepler TB, Elston TC (2001) Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys J* 81: 3116–3136.
- Thattai M, Oudenaarden AV (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA* 98: 8614–8619.
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81: 2340–2361.
- Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 6: 451–464.
- Sanchez A, Garcia HG, Jones D, Phillips R, Kondev J (2011) Effect of promoter architecture on the cell-to-cell variability in gene expression. *PLoS Comput Biol* 7: e1001100.
- Rieckh G, Tkačik G (2013) Noise and information transmission in promoters with multiple internal states. *arXiv:13078075*: 1–13.
- Holloway DM, Lopes FJP, Da Fontoura Costa L, Travençolo BAN, Golyandina N, et al. (2011) Gene Expression Noise in Spatial Patterning: hunchback Promoter Structure Affects Noise Amplitude and Distribution in *Drosophila* Segmentation. *PLoS Comput Biol* 7: e1001069.
- Walczak AM, Mugler A, Wiggins CH (2009) A stochastic spectral analysis of transcriptional regulatory cascades. *Proc Natl Acad Sci USA* 106: 6529–6534.
- Grima R (2010) An effective rate equation approach to reaction kinetics in small volumes: theory and application to biochemical reactions in nonequilibrium steadystate conditions. *J Chem Phys* 133: 035101.
- Burger A, Walczak AM, Wolynes PG (2010) Abduction and asylum in the lives of transcription factors. *Proc Natl Acad Sci USA* 107: 4016–4021.
- Lee TH, Maheshri N (2012) A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Mol Syst Biol* 8: 576.
- Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* 4: e309.

Author Contributions

Conceived and designed the experiments: LD. Performed the experiments: DM LD. Analyzed the data: DM LD. Contributed reagents/materials/analysis tools: DM CBM LD. Wrote the paper: DM CBM LD.