Multiple binding sites for transcriptional repressors can produce oscillations and enhance noise suppression

Iván M. Lengyel\(^1,2\) and Luis G. Morelli\(^1,2,3\)

\(^1\) IBioBA, Max Planck Society Partner Institute, Godoy Cruz 2390, C1425PQD, Buenos Aires, Argentina
\(^2\) Departamento de Física, FCExN UBA, Ciudad Universitaria, 1428 Buenos Aires, Argentina
\(^3\) morelli@ibioba-mpsp-conicet.gov.ar

(Dated: March 18, 2016)

Cells may control fluctuations in protein levels by means of negative autoregulation, where transcription factors bind DNA sites to repress their own production. Theoretical studies have assumed a single binding site for the repressor, while in most species it is found that multiple binding sites are arranged in clusters. We study a stochastic description of negative autoregulation with multiple binding sites for the repressor. We find that increasing the number of binding sites induces noisy oscillations of gene products. By tuning the threshold for repression, we show that multiple binding sites can also suppress fluctuations. Our results highlight possible roles for the presence of multiple binding sites of negative autoregulators.

PACS: 87.18.Tt 87.16.Yc 87.16.dj 05.40.-a

The state of living cells is determined by the molecules they produce and their numbers [1]. Cells can control the production of molecules by means of gene regulatory networks [2]. In these networks, transcription factors are key proteins that bind DNA to activate or repress synthesis [1, 2]. Negative autoregulation is a common component of these networks in which a transcription factor binds a specific site in the DNA and prevents its own synthesis [2]. In \(E.\ coli\), about 40\% of transcription factors are negatively autoregulated [3, 4, 2].

It is thought that negative autoregulation can speed up response times [2, 4] and tame fluctuations [5, 6, 7]. Since the number of some molecules in the cell can be small and the process of molecular synthesis is subject to fluctuations in the cellular environment, the resulting molecule numbers are in general noisy [8, 7, 9, 10]. Such circuits can suppress noise by regulating the number of molecules that a cell produces, permitting or repressing synthesis depending on the relative amount of the molecule [5, 6]. However, negative autoregulation can also boost fluctuations if transcription factor binding is too strong [11]. A balance of timescales is key for negative autoregulation to be able to act as a noise suppressor [12, 13]. Furthermore, it has been argued that the cost of suppressing noise by negative autoregulation may be high due to fluctuations in intermediate signalling events [14]. Such intermediate events may also introduce effective delays and cause oscillations in molecule numbers [15, 16, 17].

Despite its ubiquity, the effects of noise on negative autoregulation are still poorly understood. Theoretical studies often assume a single binding site for a transcriptional repressor or an effective Hill function. However, multiple binding sites for the same transcription factor are known to form clusters within regulatory domains [18, 19, 20, 21]. This feature has been widely observed in \(E.\ coli\) [20], eukaryotes [22, 23], invertebrates [24] and vertebrates [25]. In fact more than half of the human genes contain clusters of binding sites for the same transcription factor [25]. This suggests that multiple binding sites for transcription factors may play an important role in gene regulation [25, 26].

In a deterministic system, multiple binding sites can enhance non-linear effects [27]. In a stochastic description, their occupation can have multiple states and stochastic switching will occur between these states. This could set an additional timescale which affects fluctuations. Here we study how fluctuations are affected by multiple binding sites, in the framework of a stochastic binding theory in which a single transcription factor represses its own production, Fig. 1.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Negative autoregulation with multiple binding sites. Protein (blue circles) is synthesised from the information in the gene (blue stripe) at a rate \(r(m)\). Protein binds to the \(N\) binding sites (orange) at a rate \(k_b\) and falls off at a rate \(k_u\). Bound proteins repress production when more than \(M\) proteins are bound (blunted arrow). Free and bound protein are degraded at the same rate \(d\).}
\end{figure}
with the number $m$ of DNA bound proteins, $r = r(m)$, Fig. 1. We focus on noncooperative binding, that is the affinity of proteins for binding sites $k_b$ is not affected by how many molecules are bound already. Bound proteins fall off from the binding site with a rate $k_u$ per molecule. Both bound and free protein have a finite half-life and decay at a rate $d$ per molecule.

The statistics of the stochastic process can be described by the probability distribution $P(n, m, t)$ to have $n$ free proteins and $m$ bound proteins at time $t$ [28]. This probability distribution obeys the master equation

$$
\frac{dP(n, m, t)}{dt} = r(m)(P(n - 1, m, t) - P(n, m, t)) \\
+ k_b((N - (m - 1))(n + 1)P(n + 1, m - 1, t)) \\
- k_b((N - m)nP(n, m, t)) \\
+ k_u((m + 1)P(n - 1, m + 1, t) - mP(n, m, t)) \\
+ d((n + 1)P(n + 1, m, t) - nP(n, m, t)) \\
+ d((m + 1)P(n, m + 1, t) - mP(n, m, t)).
$$

The first line accounts for molecule production, the second and third for the binding process, the fourth for the unbinding process, and the two last lines for degradation of free and bound molecules. The regulatory function $r(m)$ can be adjusted to describe different regulatory mechanisms. In this work we focus on threshold regulation, in which synthesis occurs at a rate $r_0$ while there are $M$ or less bound proteins, and it is fully repressed when there are more than $M$ bound proteins:

$$r(m) = \begin{cases} 
    r_0 & \text{if } 0 \leq m \leq M \\
    0 & \text{if } M + 1 \leq m \leq N. 
\end{cases}
$$

In general, changing parameters in the system will affect the steady state, in particular the mean value of the total number of proteins $\langle n_T \rangle = \langle n + m \rangle$. We work under the assumption that autoregulation controls the level of proteins to keep it at some functional value in steady state [2]. To meaningfully assess the effects of adding multiple binding sites we fix $\langle n_T \rangle$. To achieve this, we adjust the synthesis rate $r_0$ to compensate the changes introduced by adding multiple binding sites. In the following we choose the value $\langle n_T \rangle = 20$ for illustration [11, 29, 13].

**Methods.** We generate trajectories satisfying equation (1) by means of a standard Gillespie algorithm [30, 31]. We obtain the free $n(t)$, bound $m(t)$ and total $n_T(t) = n(t) + m(t)$ number of proteins. The total number of proteins $n_T$ is useful to compare theoretical results with experiments which do not distinguish free and bound molecules. We use this approach to analyse the dynamics of various quantities and approximate steady state statistics. We also employ a complementary approach to solve the master equation (1), similar to [32]. The master equation can be written in the form $dP(t)/dt = AP(t)$ where $A$ is a matrix and $P(t)$ is a vector of probabilities with elements $P(n, m, t)$ in a truncated space $m = 0, 1, \ldots, N, n = 0, 1, \ldots, n_{\text{max}}$. The components of the matrix $A$ are the different rates of the master equation (1). Each line of the matrix $A$ connects state $(n, m)$ to other states $(\tilde{n}, \tilde{m})$. We obtain the steady-state probability distribution $P(n, m)$ in the truncated space by solving the system $AP = 0$ with the constraint $\sum_{m, n} P = 1$. From these probabilities $P(n, m)$ we can compute the statistics of $n$, $m$ and $n_T$ in steady state.

**Oscillations.** We first study the case in which a single bound protein fully represses production, $M = 0$, and vary the number of binding sites $N$. Deviations of the total number of proteins $n_T$ from its mean value $\langle n_T \rangle$ increase with $N$, Fig. 2(a-c). Adding binding sites to the system leads to bursty kinetics, with increasing burst size and consequently longer decay times. This is reflected on an increasing Fano factor, defined as the variance to mean ratio $F = \sigma^2_T/\langle n_T \rangle$, Fig. 2(d).

These bursts occur with some temporal regularity. The time interval between two consecutive maxima naturally define burst duration $T$. Burst duration becomes longer while the relative dispersion of the distribution around the mean decreases with increasing $N$ (Supplementary Information), as reflected in the coefficient of variation $CV = \sigma_T/T$, Fig. 2(e).
Fano Factor defined as the ratio between bandwidth $\Delta f$ and main frequency $f_0$, $Q = \Delta f/f_0$. Higher values of $Q$ indicate a narrower frequency spectrum and a more defined characteristic time [15]. We fit a Lorentzian distribution to the frequency spectrum to obtain the quality factor from numerical simulations (Supplementary Information). We find that $Q$ grows with $N$ above $Q = 1$, indicating that noisy oscillations occur in the system.

Thus, in contrast to common wisdom about deterministic systems [33, 34], we find here that noisy oscillations are possible even in the absence of explicit delays or intermediate steps in the regulatory loop. Multiple binding sites introduce an effective timescale in the system, the time taken to exponentially decay from the maximum to the minimum value of total number of proteins (Supplementary Information). In addition, the multiple binding sites act as a buffer for the bound proteins: as long as the occupation number $m$ is larger than one, there can be fluctuations in $m$ and the system is still repressed.

Noise suppression. We now turn to the effects of changing the regulatory threshold $M$ for a fixed value of total number of binding sites $N$. As $M$ changes, we adjust the synthesis rate $r_0$ accordingly to keep the mean value $\langle n_T \rangle$ fixed. Increasing the regulatory threshold $M$, fluctuations decrease to a minimum and then rise again, Fig. 3(a)-(d).

This minimum occurs when synthesis fluctuations are driven by the binding and unbinding of proteins (Supplementary Information). For lower values of $M$, fluctuations around the mean value of bound proteins $\langle m \rangle$ are not sufficient to give rise to synthesis events because a small number of molecules is enough to overcome the low threshold. For higher values of $M$, fluctuations that bring $m$ below threshold are more frequent and synthesis rate is slower, resulting in a lax control of the mean value of proteins $\langle n_T \rangle$.

Noise suppression with multiple binding sites outperforms the single binding site control system, see its smaller variance and Fano factor, Fig. 3(c),(e). This stronger noise suppression occurs for a range of $M$ values, Fig. 3(f). Even in cases in which single binding site regulation $N = 1$ increases fluctuations [11, 13, 35], the presence of multiple binding sites may bring noise below the level for unregulated synthesis. Moreover, the turnover of molecules does not change with the threshold value, that is the number of molecules degraded per unit time is constant for different values of $M$, Fig. 3(g). This indicates that this noise suppression mechanism comes at no extra cost for the system.

Oscillations and noise suppression are robust. So far we focused on a set of parameters in which the binding and unbinding rates are equal, $k_b = k_u$. We now relax this constraint and analyse the effects of adding multiple binding sites $N$ and changing the regulatory threshold $M$, Fig. 4. In the three different cases spanning three orders of magnitude, adding multiple binding sites generates oscillations for low $M$, Fig. 4(a)-(c). The smaller the ratio $k_u/k_b$, the larger the Quality factor and the region in $\{N, M\}$ space where oscillations are enhanced. Increasing the binding affinity generates longer decay times of the oscillations together with a larger memory effect.

The noise suppression mechanism discussed above is present for a wide range of $k_u/k_b$ ratios, Fig. 4(d)-(f). We introduce the relative Fano factor $F_R$ defined as the Fano factor normalised to its single binding site value ($N = 1$ and $M = 0$). Values of $F_R$ smaller than one reveal a region in $\{N, M\}$ space in which noise suppression outperforms the single binding site case,
Figure 4: Oscillations and noise suppression are robust to different binding / unbinding ratios. (a-c) Quality factor maps for different numbers of binding sites \( N \) and regulatory threshold \( M \). (d-f) Relative Fano factor \( F_R \). Orange lines correspond to \( F_R = 1 \). Parameters: (a) \( k_u = 10, k_b = 1 \), (b) \( k_u = k_b = 335 \) and (c) \( k_u = 1, k_b = 10 \). In all the panels \( d = 1 \).

For any number of binding sites \( N \) there is a regulatory threshold \( M \) that reduces fluctuations. As the ratio \( k_u/k_b \) increases the region of noise suppression becomes larger and the relative suppression effect stronger. This is in line with the finding that weak binding is critical for the fidelity of autoregulation [13].

Discussion. We studied the stochastic dynamics of a negatively autoregulated gene with multiple binding sites for the repressor, Fig. 1. We showed that increasing the number of binding sites induces noisy oscillations of gene products, Fig. 2. This is at odds with the result for deterministic systems, where explicit time delays or multiple intermediate steps are required together with strong nonlinearity in order to generate biochemical oscillations [33, 34]. The system we study here does not include such intermediate steps or delays, but the stochastic dynamics of multiple binding sites occupation introduces an additional effective timescale that triggers the oscillations. Thus, our results establish a paradigm for biochemical oscillators. The vertebrate segmentation clock [36] may be a good model system to test our results. This genetic oscillator acting during embryonic development is thought to be driven by negative autoregulation of \( her/Hes \) genes, which have several regulatory binding sites for the repressors [18, 37, 38]. Although it is thought that autoregulation of such \( her \) genes involves effective delays [39, 37], it is possible that the presence of multiple binding sites enhances oscillations. This hypothesis could be tested for example by interfering with binding sites [40].

For a fixed number of binding sites we showed that tuning the threshold for autoregulation, noise suppression can be enhanced, Fig. 3. Noise suppression appears to be strongest when synthesis events are directly driven by fluctuations in binding and unbinding of transcription factors. This noise suppression occurs without increasing the product turnover, suggesting a mechanism that could decrease the high cost of regulatory control systems [14]. Noise can be tamed even when single binding site regulation aggravates fluctuations with respect to unregulated synthesis [11, 13, 35]. Thus, noise suppression can be rescued from this situation without changing the affinity of the molecules for the binding sites.

Cellular control systems use negative autoregulation in different contexts either to generate oscillations [36, 41, 42], or to control the level of some target molecules by suppressing number fluctuations [5, 6, 2]. Multiple binding sites, occurring in different natural regulatory systems [25, 20, 18, 37], may enhance these functions. While there may be other mechanisms that perform these feats, as for example cooperativity [43] and cooperative binding [44, 45, 19], we argue that adding binding sites to the regulatory region of a gene might be a simpler and more straightforward evolutionary path [2]. Short local duplications of the genome can easily increase the copy number of binding sites for transcription factors [26]. Microsatellites, which are tandem arrays of multiple copies of a short sequence [26, 46], have a significant overlap with transcription factor binding sites [47, 48]. Our results could be useful in the design of synthetic control systems that either generate oscillations [41, 42] or tightly control fluctuations around the mean number of molecules [5, 6].

Acknowledgements. We thank L. Bruno, D. J. Jörg, A. C. Oates and K. Uriu for valuable comments on the manuscript. LGM acknowledges support from ANPCyT PICT 2012 1954 and PICT 2013 1301.

References

[1] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, Molecular Biology of the Cell 3E (Garland Science, 1994) p. 1408.

[2] U. Alon, An Introduction to Systems Biology: Design Principles of Biological Circuits (CRC press, Boca Raton, Florida, 2006).

[3] D. Thieffry, A. M. Huerta, E. P´erez-Rueda, and J. Collado-Vides, BioEssays 20, 433 (1998).

[4] N. Rosenfeld, M. B. Elowitz, and U. Alon, J. Mol. Biol. 323, 785 (2002).

[5] A. Becskei and L. Serrano, Nature 405, 590 (2000).

[6] Y. Dublanche, K. Michalodimitrakis, N. Kümmener, M. Foglierini, and L. Serrano, Mol. Sys. Biol. 2, 1 (2006).

[7] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, Nat. Rev. Genet. 6, 451 (2005).
[8] J. M. Raser and E. K. O’Shea, Science 309, 2010 (2005).

[9] L. S. Tsimring, Rep. Prog. Phys. 77, 026601 (2014).

[10] T. M. Norman, N. D. Lord, J. Paulsson, and R. Losick, Annu. Rev. Microbiol. 69, 381 (2015).

[11] D. J. Stekel and D. J. Jenkins, BMC Syst. Biol. 2, 6 (2008).

[12] A. Singh and J. P. Hespanha, Biophys. J. 96, 4013 (2009).

[13] A. Grönlund, P. Löstedt, and J. Elf, Nat. Commun. 4, 1864 (2013).

[14] I. Lestas, G. Vinnicombe, and J. Paulsson, Nature 467, 174 (2010).

[15] L. G. Morelli and F. Jülicher, Phys. Rev. Lett. 98, 228101 (2007).

[16] P. Morant, Q. Thommen, F. Lemaire, C. Vandermøëre, B. Parent, and M. Lefranc, Phys. Rev. Lett. 102, 068104 (2009).

[17] J. Wang, M. Lefranc, and Q. Thommen, Biophys. J. 107, 2403 (2014).

[18] K. Takebayashi, Y. Sasaki, Y. Sakaib, T. Watanabe, S. Nakanishi, and R. Kageyama, J. Biol. Chem. 269, 5150 (1994).

[19] M. Kazemian, H. Pham, S. A. Wolfe, M. H. Brodsky, and S. Sinha, Nucleic Acids Res. 41, 8237 (2013).

[20] M. Rydenfelt, H. G. Garcia, R. S. Cox, and R. Phillips, PLoS ONE 9, e113437 (2014).

[21] J. P. Bothma, H. G. Garcia, S. Ng, M. W. Perry, T. Gregor, and M. Levine, eLife 4, 1 (2015).

[22] A. Wagner, Nucleic Acids Res. 25, 3594 (1997).

[23] A. Wagner, Bioinformatics 15, 776 (1999).

[24] A. P. Lifanov, V. J. Makeev, A. G. Nazina, and D. A. Papatsenko, Genome Res. 13, 579 (2003).

[25] V. Gotea, A. Visel, J. M. Westlund, M. A. Nobrega, L. A. Pennacchio, and I. Ovcharenko, Genome Res. 20, 565 (2010).

[26] E. E. Thomas, Curr. Opin. Genet. Dev. 15, 640 (2005).

[27] I. M. Lengyel, D. Soroldoni, A. C. Oates, and L. G. Morelli, Papers in Physics 6, 060012 (2014).

[28] N. G. Van Kampen, Stochastic processes in physics and chemistry, Vol. 11 (North Holland, 1992) p. 465.

[29] R. Grima, D. R. Schmidt, and T. J. Newman, J. Chem. Phys 137, 035104 (2012).
Supplementary notes

Supplementary Note 1: Interpeak distribution

We use the time interval between two consecutive maxima of the burst to define the burst duration $T$. Varying $N$ values we compute the distribution of the burst duration from stochastic simulations, FIG 5. As long as the total number of binding sites $N$ increase the mean value ($\langle T \rangle$) increases and the coefficient of variation of the peaks $CV = \sigma_T/\langle T \rangle$ decreases.

Supplementary Note 2: Spectrum analysis and fitting

With $Ns = 100$ stochastic simulations for every set of parameters $\{N, M, r, d_f, k_b, k_u, d_k\}$ we compute the frequency spectrum. For each frequency of the spectrum we add the values of the power spectrum of all the realizations. We add the Frequency spectrum of all the time series for the same set of parameters because for our time series this procedure reduces the fluctuations of the spectrum. Then we fit the Spectrum with a Lorentzian distribution, with an extra scale $A$ coefficient since the sum of all the power spectrum is not normalized.

The Lorentzian function has a location parameter $f_c$ which specifies the peak of the distribution, and a dispersion parameter $\gamma$ which specifies the half-width at half-maximum. $2\gamma$ specifies the full width at half maximum.

$$A \frac{\pi \gamma}{\pi \gamma (x - f_c)^2 + \gamma^2}. \quad (3)$$

Once we perform the fit using least squares, we use the information of the fit to compute the Quality factor $Q$ defined as:

$$Q = \frac{\Delta f}{f_c} \quad (4)$$

where $f_c$ is the central frequency and $\Delta f$ the half power bandwidth, $\Delta f = 2\gamma$. Higher values of $Q$ indicate a narrower frequency spectrum and a more defined characteristic time. In oscillatory time series the Quality factor measures the number of oscillations over which period fluctuations are small and thus serves as a measure for oscillator precision.

Supplementary Note 3: Amplitude distribution, Burst duration distribution, and decay time for $M = 0$

When only one protein inhibit the synthesis, this is when $M = 0$, the oscillatory dynamic is defined mainly by two types of events, the bursty expression, and the decay of the proteins, FIG 2 main text. We can try to estimate both the amplitude and the duration of the cycles from the parameters of the system. Expression events, the bursts, occur only when $m = 0$.

For illustration we can start in the initial state $(n, m) = (0, 0)$, production is the only possible reaction of the system with a rate $r$ and a characteristic time $\tau = 1/r$. Once we have a production event the system is in the state $(n, m) = (1, 0)$, and now degradation and binding events are possible in the system. After $\tilde{n}$ production events the state of the system is $(n, m) = (\tilde{n}, m)$. In this situation the probabilities of degradation, synthesis, and binding are:

$$p_{deg, \tilde{n}} = \frac{\tilde{n}d}{r + \tilde{n}d + k_bN\tilde{n}} \quad (5)$$
$$p_{syn, \tilde{n}} = \frac{r}{r + \tilde{n}d + k_bN\tilde{n}} \quad (6)$$
$$p_{bin, \tilde{n}} = \frac{k_bN\tilde{n}}{r + \tilde{n}d + k_bN\tilde{n}} \quad (7)$$

and are the only possible reactions during a burst event. For typical values of this work when $M = 0$ and $\tilde{n} > 1$, $r \gg d\tilde{n}$ and $k_bN\tilde{n} \gg d\tilde{n}$, so the probability of the degradation reaction $p_{deg, \tilde{n}} \sim 0$. We can neglect degradation events and only analyse synthesis and binding events during the burst.

From probabilities (6) and (7), and the fact that there is no degradation, can compute the probability distribution of a binding event after $\tilde{n}$ production...
Figure 7: (a) Distribution of peaks maxima from the simulations (blue histogram) and from the theoretical estimation (red line). From the stochastic and the estimation distributions we can compute the mean values of the maximums values, bottom panel. (b) Time series of total protein number (blue) and the maximum values of the peaks (green dots). From the maximum values we can estimate the decay of the oscillations with a deterministic approach (red line). (c) Burst duration distribution of the cycles with the estimation. From the maximum values of total protein number (blue) and the maximum values of the maximum values, bottom panel. (b) Time series estimation distributions we can compute the mean values of the periods, bottom panel.

After a binding event $m \neq 0$ and no more synthesis events are possible, setting the maximum value of the total protein of number $n_T = \tilde{n}$ for the cycle. With the probability distribution (8) we can compute the distribution and the mean value of the burst maxima, FIG 7(a).

After the initial burst and a binding event, since $k_bN\tilde{n} \gg d\tilde{n}$ the bound protein often fully occupies all available sites, $m \sim N$. Under this condition fluctuations of $m$ that cross the threshold $M$ can be neglected and decay time can be analysised in a more deterministic fashion. When there is at least one protein bound to the DNA, only events of degradation, binding and unbinding are possible. In terms of the total amount of protein $n_T = m + n$, events of binding and unbind do not change the total amount of proteins and only degradation events change the number of proteins. The deterministic differential equations when $m > 0$ for the the free, bound and total proteins are

$$\begin{cases} \frac{dn}{dt} = -k_b(N-m)n + k_u m - dn \\ \frac{dm}{dt} = k_b(N-m)n - k_u m - dm \end{cases}$$

We can add both equations to analyse the evolution of the total number of proteins $n_T = n + m$

$$\frac{dn_T}{dt} = -dn_T$$

So in absence of production, when $m \neq 0$, the deterministic evolution of the total number of proteins until a new burst synthesis event is

$$n_T(t) = Ae^{-dt},$$

where $A$ maximum value of the burst. In FIG 7(b) it is shown a time series with the decay prediction. The cycle stops when the last bound protein $m = 1$ unbounds from the DNA or is degraded and a new synthesis burst occurs.

We want to estimate is how many proteins there are in the system before the regulatory region in the DNA is free again, this is the value of $n_T$ when $m = 1$. From the stationary probability matrix, we can compute the probability distribution for $n_T$ total proteins in the system and only one protein bound to the DNA, this is $P(n_t | m = 1)$. This probability distribution let us compute the last number of total proteins or the minimum value of total protein number in the cycle $n_{T,min}$ before a unbinding or degradation of a bound protein event

$$n_{T,min} = \sum_{n_T \geq 1} P(n_t | m = 1)n_T$$

Using the maximum value of the burst $n_{T,burst}$, the minimum $n_{T,min}$, and the decay evolution (11) we can compute the duration of the cycle as

$$\tau = \frac{-1}{d(\ln n_{T,min} - \ln n_{T,burst})}$$

with probability distribution (8). In figure 7 the distribution of burst duration is shown with an estimation of the mean duration and the standard deviation. This characteristic time is set by the burst size, the decay of the total number of proteins, and the number of binding sites $N$ and regulatory architecture $M$ of the DNA.

Supplementary Note 4: Noise suppression and bound proteins statistics

We computed the mean value and the statistics of the bound proteins $m$ for different parameters in order to give an explanation to the behaviour of the CV when $M$ varies, FIG 3 Main Text.

We computed the mean value of the bound proteins $\langle m \rangle$ and the confidence intervals in a situation in which there is no degradation and no synthesis for a fixed value of total number of proteins, $d = r = 0$ and $n_T = 20$, FIG 8(a). This situation provides a reference of the fluctuations of the bound proteins because the binding and unbinding dynamics and not because of fluctuations in the number of proteins due to synthesis or degradation. The mean value $\langle m \rangle$ and the confidence interval under this situation are independent of the regulatory threshold and only depends on the binding and unbinding rates and the number of binding sites.

When there is degradation and synthesis, for low values of $M$ bound protein $m$ fluctuations around the
Figure 8: (a) Mean value and confidence interval for bound proteins. Green line indicates $\langle m \rangle$ when there is no degradation or synthesis $d = r = 0$. Green shaded regions from dark to light indicates 90%, 98% and 99.99% confidence intervals respectively. Purple dots indicates the value of the mean value for the different regulatory thresholds $M$. Shaded purple region indicates the 90% confidence interval for the different regulatory thresholds $M$. Orange line indicates the reference value of the regulatory threshold $M$. (b) Probability of finding the DNA not inhibited, $\text{Prob}(m \leq M)$. Parameters: $k_b = k_u = 335.5$, $N = 14$.

Mean $\langle m \rangle$ are not enough to cross the threshold, FIG 8(a). Degradation of the total number of proteins is needed in order to obtain synthesis events. This degradation down-deviates the mean value of the total number proteins from the mean value $\langle n_T \rangle$. Since probability of finding the gene active is very low, FIG 8(b), synthesis events occur in a burst leading to deviation over the mean, in order to fix the mean value $\langle n_T \rangle = 20$. The need of degradation to obtain synthesis events, and the very short synthesis activity events leads to a bursty dynamic and high and long fluctuations from the mean. Because of the need of degradation, mean value of bound proteins is lower than in the no production no synthesis case $\langle m \rangle < \langle \tilde{m} \rangle$, and the fluctuations of bound proteins are bigger.

For high values of $M$ bound protein $m$ fluctuations are enough to cross the threshold. Further more, for the highest values of $M$ the mean value is under the threshold $\langle m \rangle \leq M$, FIG 8(a). In this situation the probability of finding the gene active is big and, FIG 8(b), and in order to keep $\langle n_T \rangle = 20$ the synthesis rate is low. This leads to a slow response when the total number of proteins is under the mean down-deviating the total number of proteins. Since the number of free binding sites is typically low, binding events that overcome the threshold are not so frequent. This leads to an over synthesis even when the number of proteins is over the mean value $n_t \geq \langle n_t \rangle$. The slow response in both situations results in a lazy control and increase the deviations from the mean.