

ARTICLE

Received 11 Oct 2012 | Accepted 15 Apr 2013 | Published 14 May 2013

DOI: 10.1038/ncomms2867

Transcription factor binding kinetics constrain noise suppression via negative feedback

Andreas Grönlund^{1,2}, Per Lötstedt³ & Johan Elf¹

Negative autoregulation, where a transcription factor regulates its own expression by preventing transcription, is commonly used to suppress fluctuations in gene expression. Recent single molecule *in vivo* imaging has shown that it takes significant time for a transcription factor molecule to bind its chromosomal binding site. Given the slow association kinetics, transcription factor mediated feedback cannot at the same time be fast and strong. Here we show that with a limited association rate follows an optimal transcription factor binding strength where noise is maximally suppressed. At the optimal binding strength the binding site is free a fixed fraction of the time independent of the transcription factor concentration. One consequence is that high-copy number transcription factors should bind weakly to their operators, which is observed for transcription factors in *Escherichia coli*. The results demonstrate that a binding site's strength may be uncorrelated to its functional importance.

¹Department for Cell and Molecular Biology, Science For Life Laboratory, Uppsala University, SE-751 24 Uppsala, Sweden. ²Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, SE-901 87 Umeå, Sweden. ³Division of Scientific Computing, Department of Information Technology, Uppsala University, SE-751 05 Uppsala, Sweden. Correspondence and requests for materials should be addressed to J.E. (email: johan.elf@icm.uu.se).

The number of molecules in a biological cell is always fluctuating because of the probabilistic nature of chemical reactions^{1,2}. Negative feedback is therefore commonly used to maintain homeostasis and to reduce the fluctuations in intracellular processes such as gene expression^{3–5}. For example, nearly half of the ~300 transcription factors (TFs) in *Escherichia coli* make use of negative feedback on their own expression^{6,7}. However, the chemical noise that makes regulation necessary also limits the accuracy of feedback regulation⁸. For this reason, we need to consider the physical limitations in how feedback systems can be implemented by chemical components when analysing optimal strategies for suppressing noise in gene expression⁹.

Recent single molecule *in vivo* measurements show that it takes much longer time for a TF molecule to find its chromosomal operator site than what is expected from *in vitro* measurements. The discrepancy is due to slow intracellular diffusion, significant binding to non-specific DNA and that chromosomal DNA is crowded by other proteins¹⁰. As a consequence, it takes about 4 min for a single LacI repressor molecule to find and bind a single *lac* operator in the *E. coli* chromosome¹¹. The slow association kinetics, as quantified by a low association rate constant k_a , implies that the dissociation rate from the operator has to be low (low k_d) to accomplish strong feedback with a high binding constant $K = k_a/k_d$. However, k_d also sets the time scale for how often the feedback system samples the cellular concentration that should be regulated. We should therefore ask if transcriptional noise is better suppressed with fast or strong negative feedback.

Results

Modelling slow TF binding. The consequences of limited TF-binding rates can be modelled by the basic steps in expressing a TF gene and the TF binding to the DNA^{12–19}, illustrated in Fig. 1. In summary, an active gene is transcribed into mRNA, the mRNA is translated into a repressor protein several times or degraded, and the repressor binds the operator site mediating the negative feedback or is degraded. When the operator is bound the gene is inactive until the repressor dissociates. We now ask what dissociation rate leads to optimal noise suppression when the average expression level is maintained constant by compensating stronger repression with higher transcription rate. By posing the optimization problem in this way, we implicitly assume that the purpose of the negative feedback system is to maintain a certain concentration of the regulated protein, and that the parameters that evolution can work with are the strength of the promoter and the strength of repressor binding.

It is straightforward to perform stochastic simulations

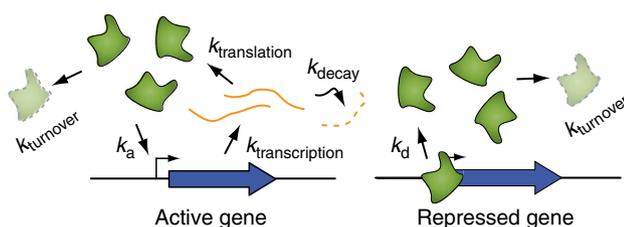


Figure 1 | The negative feedback motif. The TF protein binds the regulatory region of its own gene and suppresses expression at high TF concentration. The rate constant for binding k_a is limited by the complex process of finding a specific site in the genome, whereas the dissociation rate k_d can be very high at the expense of poor binding strength. At low k_d , the TF binds strongly and the mean response in the regulation is sensitive. However, when considering the discrete nature of chemical reaction, a low k_d implies that the TF dissociates rarely and that all mRNAs are made in bursts at these events.

corresponding to the reaction system in Fig. 1 (ref. 20). The results of such simulations are displayed in Supplementary Fig. S1 and will serve as reference point for the mathematical analysis. The fluctuations are quantified using the variance, σ^2 , to mean, $\langle x \rangle$, ratio $\text{VMR} = \sigma^2 / \langle x \rangle$. As we fix the average number of molecules $\langle x \rangle$ for varying feedback, finding an optimal k_d in terms of minimizing the VMR is thus equivalent to optimize noise in terms of the coefficient of variation $\sigma^2 / \langle x \rangle^2$ or the s.d. over the mean $\sigma / \langle x \rangle$. In Fig. 2a, we show that the VMR has a minimum with respect to the dissociation rate k_d . Higher dissociation rates imply weak binding, insensitive regulation and small noise suppression. Lower dissociation rates imply that mRNAs have to be made in noisy bursts to maintain the average protein expression level despite rare dissociation events. This inevitable property of negative transcriptional autoregulation is lost in the quasi steady state (QSS) assumption where it is assumed that the TF binding equilibrates rapidly in its binding to operator sites. The QSS assumption leads to the misconception that stronger binding, compensated by a strong promoter to maintain constant copy number, results in more sensitive regulation and better noise suppression. The reason for the error is that the slow TF kinetics that follows with strong feedback is in conflict with the QSS assumption. In Supplementary Fig. S2, the discrepancy for strong feedback is shown, an effect that is previously demonstrated in comparisons with simulations¹⁶.

Optimal binding strength. To understand how the noise relates to the binding strength, we need to solve the chemical master equation²¹ for the well-stirred system corresponding to the reaction network in Fig. 1. The molecules are considered to be well-mixed in the cell since the search time of finding the binding site is of the order of minutes and proteins diffuse across the cell on the order of seconds. Possible effects of diffusion noise are neglected²². We derive an expression for the variance in the number of repressor molecules, see Methods section and Supplementary Note 1, where a key step is to approximate third order moments in terms of lower order moments. In Fig. 2a we display the analytical solution for the VMR, solid lines and simulation results for the complete system. As a reference, we display the VMR of unregulated protein synthesis (dashed line), which is approached in the limit of no feedback and given by^{13,14}

$$\text{VMR}_0 = 1 + b \quad (1)$$

where b is the average number of proteins made per mRNA molecule. Noise is suppressed below the unregulated level, VMR_0 , when the dissociation rate is higher than $k_{d,\text{min}}$.

$$k_{d,\text{min}} \approx c_p m_p / (1 + b) \quad (2)$$

where c_p is the rate of protein turnover, which can be a combination of dilution and degradation, m_p is the mean number of repressor protein molecules and $c_p m_p$ is therefore equal to the average rate of protein synthesis. See Supplementary Note 2 for details. When several proteins are made per mRNA ($b \gg 1$), Equation 2 implies that the repressor has to dissociate at a rate corresponding to the protein synthesis rate divided by the number of proteins made per mRNA. This result makes sense, as the mRNAs otherwise would have to be made in bursts. The denominator is $1 + b$, and not b , as it is sufficient that the repressor dissociates at the rate of protein synthesis, $c_p m_p$, to maintain the uncorrelated protein synthesis events, even if $b < 1$. High-copy number *E. coli* TF exists in numbers up to several thousands^{23–25}. If we consider as an example a high-copy number TF to be made at a rate of 1,000 molecules per generation and with $b = 8$, the median for *E. coli*²⁶, the TF needs to dissociate more than 100 times per generation or negative feedback adds more noise than the unregulated situation.

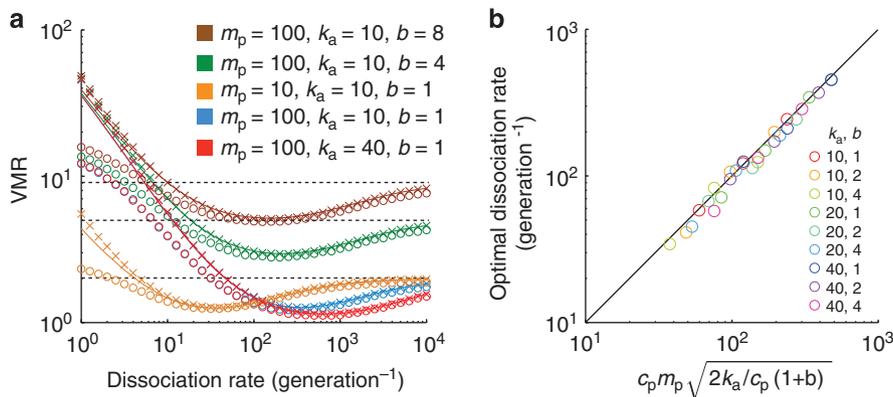


Figure 2 | Optimal dissociation rate of negative autoregulation. The left panel (a) displays the VMR as a function of the dissociation rates (k_d), for various average amount (m_p) of the regulated TF protein, TF association rate per operator (k_a) and average numbers of TFs made per mRNA ($b = k_{\text{translation}}/k_{\text{decay}}$). The solid lines are the analytical expressions given in the Supplementary Note 1. The crosses are measurements from stochastic simulations of fixed size cells where cell growth is approximated with protein decay. Circles are measurements from stochastic simulations of cells that are growing, replicating and dividing with measurements taken in the middle of the cell cycle. Rates are defined in terms of the generation time and TFs are only diluted by cell growth. In addition to the parameters given in the figure, the mRNA lifetime is 1/10 of the generation time. The dashed lines are VMR of unregulated synthesis. In the right panel (b), we display how the expression in Equation 3 for the optimal dissociation rate compares with that obtained from simulations of exponentially growing cells for specific parameter combinations (inset), each with three different mean protein copy numbers (20, 40 and 80). Note that the VMR is approximated in the middle of the cell cycle (see Supplementary Note 1).

In order to gain noise suppression by the regulation, the transcription events should not only be uniformly distributed in time, which is the case for unregulated synthesis, but respond to deviations from the average value. For this reason, the operator binding state has to be given some time to equilibrate to free protein concentration between expression events. We find that the optimal dissociation rate is

$$k_{d,\text{opt}} \approx c_p m_p \sqrt{2k_a/c_p(1+b)} \tag{3}$$

where k_a is the association rate for the repressor to a single operator site. See Supplementary Note 2 for details. In Fig. 2b, we show that equation 3 describes the optimal dissociation rate for all relevant parameters by comparing the equation with the optimal dissociation rate obtained from fully stochastic simulations of growing cells. The result can also be described in terms of the fraction of time the operator should be free, that is,

$$\frac{\text{Free}}{\text{Free} + \text{Bound}} = \frac{1/k_a m_p}{1/k_a m_p + 1/k_{d,\text{opt}}} \tag{4}$$

$$\approx \frac{1}{1 + \sqrt{(1+b)\tau_G/2\tau_s}}$$

where the search time of the TF, $\tau_s = 1/k_a$, and the protein turnover time, $\tau_G = 1/c_p$. For example, if the search time is 4 min, which is close to the physical limit¹¹, the generation time is 22 min and $b = 8$, the operator should be free 17% of the time independent of the number of regulated protein molecules. A highly abundant autorepressor TF should therefore bind weakly to its operator. With the previous example of 1,000 molecules, Equation 3 gives that a TF should (roughly) be bound on average one second before dissociating for optimal noise suppression.

Global TF-binding sequence statistics. In *E. coli*, the expression of 51% of all genes are under control of seven highly expressed global transcriptional regulators (CRP, FNR, IHF, FIS, ArcA, NarL and Lrp). Among these all but ArcA are employing negative feedback on their own synthesis^{6,7}. However, all the six TFs bind weakly to the operators mediating the feedback compared with their binding to other specific sites (Fig. 3), where the binding strengths are estimated using the heterology index^{27,28}. The

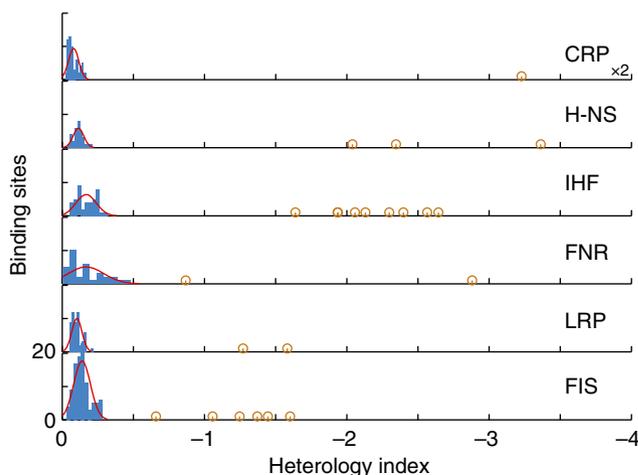


Figure 3 | Global regulators show weak autorepressor binding. The figure displays the histograms of the binding site heterology index for six of the seven global regulators controlling 51% of the genes in *E. coli*. The blue bars display the histogram of the heterology index of the forward repressor binding sites, that is, sites in the promoter of other genes. The orange markers correspond to the heterology index of the repressor binding sites in the promoter of the regulators' own gene. The absolute heterology index is proportional to the reduction in binding energy compared with consensus binding sequence. The red lines are fitted normal distributions of the histograms of the forward repressor bindings.

heterology index is for a given TF-binding site an estimation of the total reduction of binding free energy compared with if each base pair would be occupied by the most representative base pair in a set of recognition sites for the TF. The observed weak autorepressor binding sites cannot be explained from the need of having a high expression of the TFs, as the same expression level could be achieved by a combination of stronger promoters and stronger feedback. Instead, the results from our calculations demonstrate another explanation; the binding sites need to be weak to keep them free a significant fraction of the time despite the high concentration of these TFs.

Discussion

Recently Lestas *et al.*⁸ derived a fundamental limit for noise suppression in a feedback control system. When applied to the case where the regulated quantity itself is mediating the feedback, there is no information loss between making the species and regulating its production and the noise can therefore theoretically be suppressed to zero. Such feedback would be implemented by having infinite synthesis rate when the copy number is below the set point and zero synthesis when the copy number is at the set point or higher. However, in practice the physical implementation of the control systems implies that gene expression cannot be controlled this accurately. For example, although the repressor is bound to the operator, the copy number can fall below the set point without the control system being able to respond. In order to determine and respond to the copy number deviations, the TF has to dissociate such that the rebinding time can be used as a measure of the free concentration. The rebinding time will, however, be exponentially distributed and information is lost by making a decision based on a single-rebinding event. To estimate the concentration accurately in the system, the repressor has to bind and dissociate several times for each synthesis event such that the average occupancy at the time scale of the gene expression accurately reflects the concentration in the cells. Related constraints for optimal information transmission in gene regulatory systems where a signal molecule is regulated by upstream regulators have previously been analysed²⁹ and how feedback may improve the fidelity in transmission of time-varying signals³⁰.

In summary, we have derived the optimal dissociation rate for the case of a single auto-regulatory repressor protein and in the Supplementary Fig. S2 we demonstrate that the same principles apply for cooperative repression schemes. Based on these results, we conclude that there is no reason to consider weak-binding sites as unimportant. On the contrary, for regulatory fidelity by abundant TFs it is critical that binding is weak.

Methods

Solving the moment equations. The first and second moments at stationary conditions have been calculated from the chemical master equation by performing two moment closure steps, where third order moments are approximated with lower order moments (see Supplementary Note 1). After the approximation steps are performed, the second moment M_{pp} of the protein copy number is shown to satisfy the cubic equation

$$M_{pp}(M_{pp} - bm_p)^2 + \kappa m_p^4 (M_{pp} - bm_p) = \lambda m_p^6 \quad (5)$$

where

$$\kappa = \frac{k_d + c_p}{k_a m_p} - 1 - \frac{c_p}{k_d} - \frac{1}{m_p} \quad (6)$$

and

$$\lambda = \frac{k_d + c_p}{k_a m_p} \left(1 + \frac{1}{m_p} \right) \quad (7)$$

The cubic equation can be solved using Cardano's method and the second moment M_{pp} will depend upon the mean protein number m_p , the protein turnover rate c_p , the number of proteins translated per mRNA b , the binding association rate k_a and the dissociation rate k_d . The VMR of the protein copy number can be calculated from the second moment as

$$\text{VMR} = (M_{pp} - m_p^2) / m_p \quad (8)$$

Finding the optimal dissociation rate. The dissociation rate $k_{d,\text{opt}}$ that minimize the VMR can be obtained by differentiating the cubic equation (5) with respect to k_d and setting the derivative of M_{pp} to zero (see Supplementary Note 2).

References

1. Berg, O. G. A model for the statistical fluctuations of protein numbers in a microbial population. *J. Theor. Biol.* **71**, 587–603 (1978).

2. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).
3. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
4. Savageau, M. A. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature* **252**, 546–549 (1974).
5. Dublanche, Y., Michalodimitrakis, K., Kümmerer, N., Foglierini, M. & Serrano, L. Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol. Syst. Biol.* **2**, 41 (2006).
6. Martinez-Antonio, A. & Collado-Vides, J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.* **6**, 482–489 (2003).
7. Gama-Castro, S. *et al.* RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K12 integrated within genetic sensory response units (sensor units). *Nucleic Acids Res.* **39**, D98–D105 (2011).
8. Lestas, I., Vinnicombe, G. & Paulsson, J. Fundamental limits on the suppression of molecular fluctuations. *Nature* **467**, 174–178 (2010).
9. Grönlund, A., Löfstedt, P. & Elf, J. Delay-induced anomalous fluctuations in intracellular regulation. *Nat. Commun.* **2**, 419 (2011).
10. Li, G. -W., Berg, O. G. & Elf, J. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.* **5**, 294–297 (2009).
11. Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E. G., Berg, O. G. & Elf, J. The lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598 (2012).
12. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl Acad. Sci. USA* **98**, 8614–8619 (2001).
13. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69–73 (2002).
14. Paulsson, J. Models of stochastic gene expression. *Phys. Life Rev.* **2**, 157–175 (2005).
15. Singh, A. & Hespanha, J. P. Optimal feedback strength for noise suppression in autoregulatory gene networks. *Biophys. J.* **96**, 4013–4023 (2009).
16. Stekel, D. J. & Jenkins, D. J. Strong negative self regulation of prokaryotic transcription factors increases the intrinsic noise of protein expression. *BMC Syst. Biol.* **2**, 6 (2008).
17. Tkačik, G., Gregor, T. & Bialek, W. The role of input noise in transcriptional regulation. *PLoS One* **3**, e2774 (2008).
18. Walczak, A. M., Onuchic, J. N. & Wolynes, P. G. Absolute rate theories of epigenetic stability. *Proc. Natl Acad. Sci. USA* **102**, 18926–18931 (2005).
19. Kim, K. -Y. & Wang, J. Potential energy landscape and robustness of a gene regulatory network: toggle switch. *PLoS Comput. Biol.* **3**, e60 (2007).
20. Lu, T., Volfson, D., Tsimring, L. & Hasty, J. Cellular growth and division in the Gillespie algorithm. *Syst. Biol.* **1**, 121–128 (2004).
21. van Kampen, N. G. *Stochastic Processes in Physics and Chemistry*, 5th edn (Elsevier, Amsterdam, 2004).
22. Bialek, W. & Setayeshgar, S. Physical limits to biochemical signaling. *Proc. Natl Acad. Sci. USA* **102**, 10040–10045 (2005).
23. Cai, S. J. & Inouye, M. EnvZ-OmpR interaction and osmoregulation in *Escherichia coli*. *J. Biol. Chem.* **277**, 24155–24161 (2002).
24. Overgaard, M., Borch, J. & Gerdes, K. RelB and RelE of *Escherichia coli* form a tight complex that represses transcription via the ribbon–helix–helix motif in RelB. *J. Mol. Biol.* **394**, 183–196 (2009).
25. Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. & Ishihama, A. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.* **181**, 6361–6370 (1999).
26. Taniguchi, Y. *et al.* Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533 (2010).
27. Berg, O. G. & von Hippel, P. H. Selection of DNA binding sites by regulatory proteins: II. The binding specificity of cyclic AMP receptor protein to recognition sites. *J. Mol. Biol.* **200**, 709–723 (1988).
28. Berg, O. G. Selection of DNA binding sites by regulatory proteins: the LexA protein and the arginine repressor use different strategies for functional specificity. *Nucleic Acids Res.* **16**, 5089–5105 (1988).
29. Tkačik, G., Walczak, A. M. & Bialek, W. Optimizing information flow in small genetic networks. III. A self-interacting gene. *Phys. Rev. E* **85**, 041903 (2012).
30. de Ronde, W. H., Tostevin, F. & ten Wolde, P. R. The effect of feedback on the fidelity of information transmission of time-varying signals. *Phys. Rev. E* **82**, 031914 (2010).

Acknowledgements

This work was supported by the European Research Council (203083), the Swedish Foundation for Strategic Research, the Swedish Research Council, Göran Gustafssons stiftelse, eSENCE, and the Knut and Alice Wallenberg Foundation.

Author contributions

A.G., P.L. and J.E. jointly conceived the study and wrote the paper.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Grönlund, A. *et al.* Transcription factor binding kinetics constrain noise suppression via negative feedback. *Nat. Commun.* 4:1864 doi: 10.1038/ncomms2867 (2013).