# Enhancement of the Stability of Genetic Switches by Overlapping Upstream Regulatory Domains

Patrick B. Warren[1,2] and Pieter Rein ten Wolde[1,3]

[1]*FOM Institute for Atomic and Molecular Physics, Kruislaan 407, 1098 SJ Amsterdam, The Netherlands*
[2]*Unilever R&D Port Sunlight, Bebington, Wirral CH63 3JW, United Kingdom*
[3]*Division of Physics and Astronomy, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands*
(Received 4 November 2003; published 25 March 2004)

We study genetic switches formed from pairs of mutually repressing operons. The switch stability is characterized by a well-defined lifetime, which grows very rapidly, albeit subexponentially, with the number of copies of the most-expressed transcription factor. The switch stability can be drastically enhanced by overlapping the upstream regulatory domains such that competing regulatory molecules mutually exclude each other. Our results suggest that robustness against biochemical noise can provide a selection pressure that drives operons together in the course of evolution.

Biochemical networks are the analog computers of life. They allow living cells to detect, transmit, and amplify environmental signals, as well as integrate different signals in order to recognize patterns in, say, the food supply. Indeed, biochemical networks can perform a variety of computational tasks analogous to electronic circuits. However, their design principles are markedly different. In a biochemical network, computations are performed by molecules that chemically and physically interact with each other. These interactions are stochastic in nature. This becomes particularly important when the concentrations are low. In gene regulatory networks, this is generally the case: not only the DNA, but also the proteins that regulate gene expression are often present in very small numbers, which can be as low as ten, or even fewer. Hence, one would expect that gene regulatory networks, in contrast to electronic circuits, are highly stochastic and error prone [1–5]. An important question, therefore, is how the ability to resist biochemical noise constrains the design of the network [2,3].

In prokaryotes, the expression of operons—groups of contiguous genes that are transcribed into single mRNA molecules—is regulated by the binding of transcription factors (TFs) to upstream regulatory domains on the DNA. A spatial arrangement in which two operons are transcribed in diverging directions (i.e., from opposite strands of the DNA) allows the upstream regulatory domains to interfere with each other. This affords additional regulatory control. In particular, biochemical noise in the expression of operons can become correlated or anticorrelated. Just as the existence of operons provides for correlated gene expression, interference between the regulatory domains of two diverging operons allows for a correlated or an anticorrelated expression of operons. Here, we show that this can have a dramatic influence on the stability of gene regulatory networks.

Recently, we performed a statistical analysis of the spatial distribution of operons on the genome of

*Escherichia coli* [6]. The analysis identified a large number of motifs in which the regulatory domains of two operons overlap and interfere [6]. Among them are well-known examples such as the lysA-lysR and the araBAD-araC operon pairs [7]. But perhaps the best known and arguably the most studied example of such a motif is provided by the λ-phage switch that consists of two adjacent operons that mutually repress each other [8]. Here, we study a minimalist model of such a switch, as shown in Fig. 1. In particular, we compare the stability of an "exclusive" (XOR) switch, for which the simultaneous binding of the repressive TFs for both operons is inhibited, to that of a general switch. We find that the exclusive switch is much more stable than the general switch. This demonstrates the potential importance of such motifs in making gene regulatory networks robust against biochemical noise. And although we focus here on prokaryotes, this mechanism could also apply to eukaryotes, for which, in fact, evidence for correlated and anticorrelated gene expression has been reported [9,10].
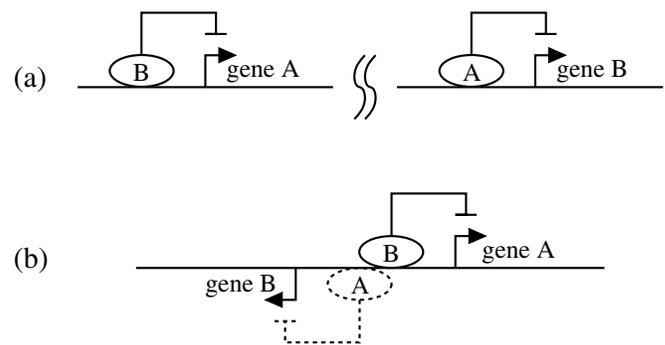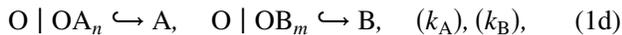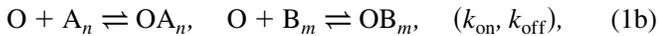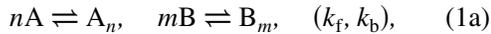


FIG. 1. A toggle switch consisting of two operons that mutually repress each other (i.e., operon A codes for TF A that represses the expression of operon B, and vice versa). (a) General switch. (b) Exclusive switch, in which the operons are transcribed in diverging directions; the regulatory domains overlap and only one TF can bind at the time.

The starting point of our analysis is a set of chemical reactions that constitute the switches shown in Fig. 1. As chemical species, we introduce a pair of TFs which can exist as monomers, A and B, or multimers, $A_n$ and $B_m$. The state of the genome is represented by O, $OA_n$, etc., depending on the binding of the TF multimers. Adopting a condensed notation in which "|" indicates alternative sets of reactants and "$\hookrightarrow$" indicates that the reactants are not destroyed by the reaction (e.g., O $\hookrightarrow$ A means O $\rightarrow$ O + A), the set of chemical reactions are

$$nA \rightleftharpoons A_n, \quad mB \rightleftharpoons B_m, \quad (k_f, k_b), \quad (1a)$$

$$O + A_n \rightleftharpoons OA_n, \quad O + B_m \rightleftharpoons OB_m, \quad (k_{on}, k_{off}), \quad (1b)$$

$$OA_n + B_m \mid OB_m + A_n \rightleftharpoons OA_nB_m, \quad (k_{on}, k_{off}), \quad (1c)$$

$$O \mid OA_n \hookrightarrow A, \quad O \mid OB_m \hookrightarrow B, \quad (k_A), (k_B), \quad (1d)$$

$$A \mid B \rightarrow \varnothing. \quad (\mu_A), (\mu_B). \quad (1e)$$

These reactions account for, respectively, the formation of multimers, the binding of TF multimers to the genome [Eqs. (1b) and (1c)], the expression of TF monomers, and the degradation of TF monomers. Repression of gene expression is implicit in Eqs. (1d), thus A is expressed if and only if $B_m$ is not bound, etc. Reaction rates are as indicated, and we define equilibrium constants for multimerization, $K_d = k_f/k_b$, and binding to the genome, $K_b = k_{on}/k_{off}$.

While detailed and biologically faithful models can be constructed as has been done for the $\lambda$-phage switch [11,12], the above model is intentionally "as simple as possible." However, as Cherry and Adler have shown [13], the TFs must bind cooperatively to the DNA in order to make a working switch. In the present model, cooperativity is introduced through the binding of TF multimers rather than monomers. Binding of TFs as (homo) dimers or tetramers is commonly observed in nature [7].

In our model the genome is in one of four states $\{O, OA_n, OB_m, OA_nB_m\}$. We now include the effect of interference between the upstream regulatory domains by disallowing some of these states. This leads to four distinct cases, which are shown in Table I. They are implemented by excluding some of the reactions in Eqs. (1b) and (1c). For example, the exclusive switch is obtained by discarding the reactions in Eqs. (1c) thereby removing the state $OA_nB_m$.

TABLE I. Distinct possibilities for the subsets of allowed genome states for our switch model.

| Case/genome states | O | $OA_n$ | $OB_m$ | $OA_nB_m$ |
|---|---|---|---|---|
| General | √ | √ | √ | √ |
| Exclusive | √ | √ | √ | × |
| Partially cooperative | √ | √ | × | √ |
| Totally cooperative | √ | × | × | √ |

We first use mean-field theory to analyze the behavior of Eqs. (1). Switching behavior corresponds to the appearance of two distinct stable states in the space of TF molecule numbers. Previously, general switches were studied by Cherry and Adler [13], and a specific example of the exclusive switch was studied by Kepler and Elston [14]. We extend the analysis of Cherry and Adler to determine where switching behavior can occur, for all the cases in Table I. First, for $n = m = 1$, no switching behavior can be found for any case. This confirms that some form of cooperative binding is required. For the totally cooperative switch though, switching behavior cannot be found for any values of $n$ and $m$. For the remaining cases, we have analyzed in detail the situation for $n = m = 2$ where both TFs bind as dimers. Figure 2 shows the regions in the $(\mu_A/k_A, \mu_B/k_B)$ plane where switching behavior is found. Clearly, switching behavior is more extensive for the exclusive switch than for the general switch and is strongly suppressed for the partially cooperative switch. Thus we conclude that, at least in mean-field theory, the structure of the switch has a strong influence on the extent of switching behavior.

To go beyond mean-field theory, we have simulated the reactions in Eqs. (1) using Gillespie's kinetic Monte Carlo scheme which generates trajectories appropriate to the chemical master equation [15]. We focus on dimerizing ($n = m = 2$) general and exclusive switches, and on the symmetry line $k_A = k_B = k$ and $\mu_A = \mu_B = \mu$. We will use the expression rate $k \approx 0.1-1$ s$^{-1}$ [12] as a unit of
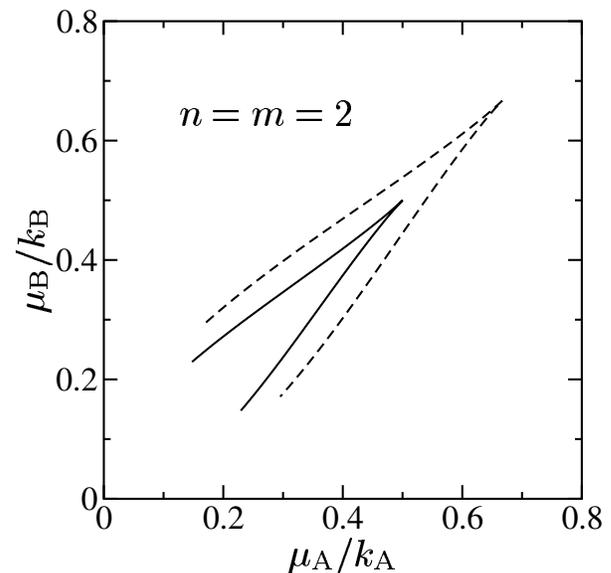


FIG. 2. In mean field theory, switching behavior is confined to a wedge in the $(\mu_A/k_A, \mu_B/k_B)$ plane. Results are shown for dimerizing ($n = m = 2$) switches. It is seen that the region of bistability is larger for the exclusive switch (dashed line) than for the general switch (solid line). For the partially cooperative dimerizing switch ($OB_2$ disallowed), the wedge moves to $\mu_A/k_A \lesssim 0.10$ and $\mu_B/k_B \lesssim 0.019$.

(inverse) time and the degradation rate $\mu$ as the main control parameter. The choice of the rate constants is biologically motivated, in particular, we expect the expression to be a slow step and the binding equilibrium to be biased in favor of bound states [12]. For a baseline set we use $k_f/V = k_b = k_{on}/V = 5k_{off} = 5k$ ($K_d/V = 1$ and $K_b/V = 5$), where $V \approx 2 \, \mu m^3$ is the cell volume [12]; we assume one copy of the genome is present.

We monitor the total numbers of the TFs, $N_A$ and $N_B$, including those in dimers and those bound to the genome. If the system is behaving as a switch then we typically see that one of the TFs is strongly repressed compared to the other one. A switching event occurs when the roles of the two TFs flip spontaneously, as shown in Fig. 3.

We can obtain more insight into the switching behavior by sampling the probability distribution $P(N_A, N_B)$ for states in the $(N_A, N_B)$ plane, as shown in Fig. 4. Switching behavior appears as a double maximum in probability in this representation, and the transition state is seen to lie at low numbers of both TFs. Three points are worthy of note. First of all, it is seen that the positions of the two stable steady states do not depend much on the architecture of the switch. This is not surprising, because if one species dominates, both switches will behave similarly. What is perhaps more surprising is that the pathways for switching are different. The transition paths of the exclusive switch cross the transition state surface at higher

values of $N_A = N_B$, as compared to the general switch. The reason is that in the general switch both genes can be repressed simultaneously, while in the exclusive switch only one gene can be turned off at a time. More importantly, however, the barrier for flipping the switch is higher for the exclusive switch than for the general switch, as can be seen in the insets in Fig. 4. This is because for a switch to flip, two events have to happen. First of all, the system has to wait for a rare fluctuation by which the concentration of the dominant species decreases; this allows for the synthesis of the other component. Subsequently, the latter component has to bind to its operator site in order to toggle the switch. In the general switch, the latter event is more probable, because the minor component can bind to its site as soon as it is synthesized, while in the exclusive switch the dominant species first has to dissociate from the DNA. This is the main reason why the exclusive switch is more stable than the general switch.

We have also characterized the switching dynamics by constructing the cumulative distribution function $F(\Delta t)$ for the time intervals $\Delta t$ between zero crossings of the order parameter $N_A - N_B$. About 50% of $F$ arises from noise on a time scale $\Delta t \sim k^{-1}$ as the system jitters around the transition state, but for $\Delta t \gg k^{-1}$, and provided we are well into the switching regime, we invariably see Poisson statistics with $F \to 1 - \exp[-\Delta t/\tau]$ [see Fig. 3(b)]. This first confirms that the switch states have a well-defined lifetime $\tau$, and second allows us to extract an accurate estimate of the value of $\tau$.

Bialek has suggested that the switch lifetime may grow exponentially with the number of molecules involved in switching between states [16]. Motivated by this, we monitor the mean number $\overline{N}$ of the most-expressed TF, defined to be the time average of $\max(N_A, N_B)$. We can also calculate $\overline{N}$ from mean-field theory, and we find good agreement between this and the value measured in the simulations, as $\mu$ varies.
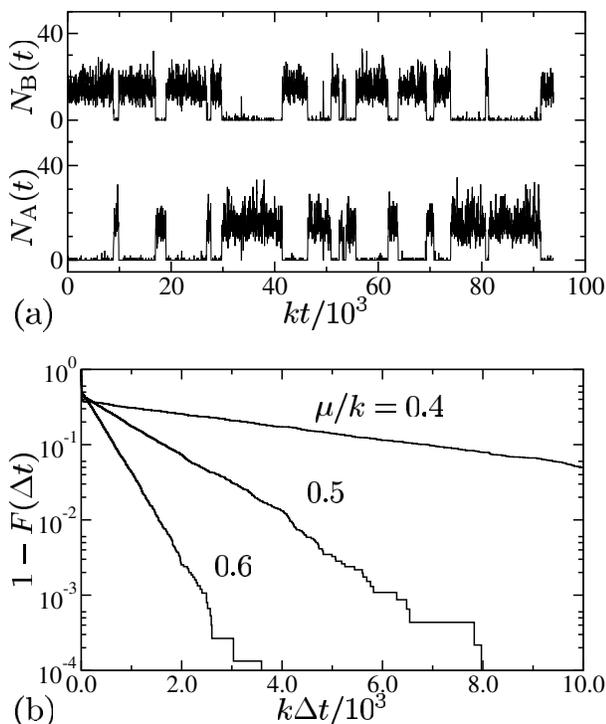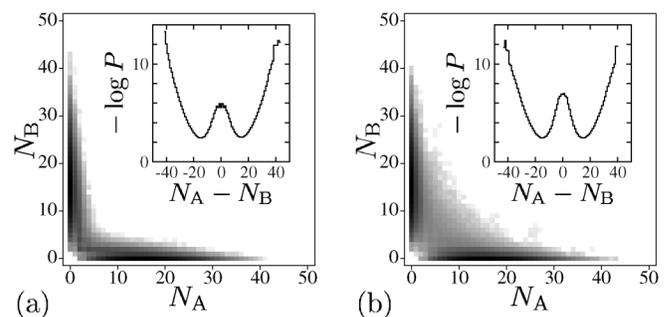


(a)



(b)

FIG. 3.   (a) Typical numbers of TFs as a function of time. (b) Cumulative distribution functions for the time intervals between zero crossings of $N_A - N_B$. Results are for the dimerizing exclusive switch at $\mu/k = 0.45$ unless stated otherwise.



FIG. 4.   Probability density in $(N_A, N_B)$ plane constructed from $2.5 \times 10^6$ samples (total duration of $kt \approx 5 \times 10^6$), for (a) general and (b) exclusive dimerizing switches at $\mu/k = 0.45$. Grey scale indicates bin count, logarithmically, from $\leq 1$ (white) to $\geq 10^5$ (black). Insets show probability density collapsed onto the $N_A - N_B$ line, plotted as a dimensionless "free energy" $-\log[P(N_A - N_B)]$ (the ordinate zero is arbitrary).
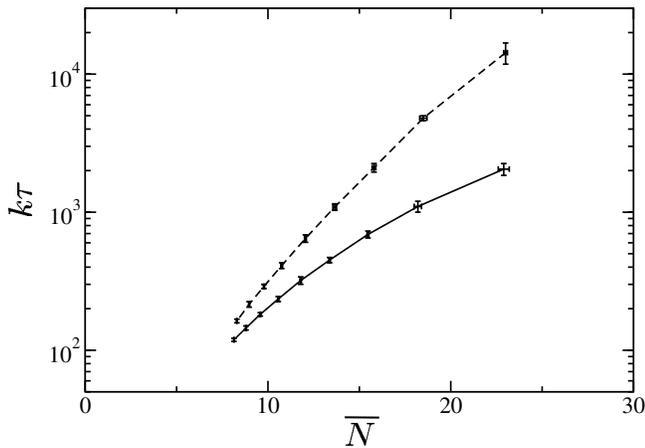
FIG. 5.   Switch lifetime as a function of the mean number of the most-expressed TF, for the general (solid line) and exclusive (dashed line) cases. The exclusive switch becomes orders of magnitude more stable than the general switch at high numbers of the expressed TF.

other on the DNA and are transcribed in diverging directions—a network motif that has been identified by our statistical analysis of the gene regulatory network of *E. coli* [6]. Hence, our simulations suggest that robustness against biochemical noise can provide a selection pressure that drives pairs of operons, that either regulate each other or are controlled by a common transcription factor, towards each other in the course of evolution.

Qualitative support for Bialek's conjecture comes from Fig. 5. It shows that $\tau$ grows very rapidly with $\overline{N}$, which is the basic reason why extremely stable switches can be built with at most a few hundred expressed proteins. In contrast to Bialek's conjecture, however, $\tau(\overline{N})$ appears to be subexponential in $\overline{N}$. Interestingly, we can fit $\tau(\overline{N})$ to a form that is suggested by an analysis of a related problem, namely, that of switching between broken-symmetry phases in a driven diffusive model [17,18]. This suggests that the ultimate scaling is $\tau \sim \overline{N}^{\alpha} \exp[b\overline{N}]$, where $\alpha$ and $b$ are constants. Note that this corresponds to Bialek's conjecture, but with a logarithmic correction in $\overline{N}$.

More importantly, however, Fig. 5 clearly demonstrates that the switch construction has a marked influence on the stability of the switch. It shows that the lifetime of the exclusive switch grows much more rapidly with mean copy number than that of the general switch. Our simulations cover $10 \lesssim \overline{N} \lesssim 30$, but if we extrapolate our results to $\overline{N} \approx 100$, then $k\tau \approx 10^4$–$10^6$ for the general switch but $k\tau \approx 10^8$–$10^{10}$ for the exclusive switch. In the latter case, this corresponds to lifetimes measured in tens of years. Such extremely long lifetimes have been reported for the $\lambda$ phage [12].

In summary, a genetic switch is intrinsically stochastic, because of the molecular character of its components. However, our simulations demonstrate that the stability of a genetic switch can be strongly enhanced by spatially arranging the operons such that competing regulatory molecules mutually exclude each other at the operator regions (i.e., regulatory domains). Such a spatial arrangement can be achieved if the two operons lie next to each

[1] H. H. McAdams and A. Arkin, Proc. Natl. Acad. Sci. U.S.A. **94**, 814 (1997).
[2] M. B. Elowitz and S. Leibler, Nature (London) **403**, 335 (2000).
[3] N. Barkai and S. Leibler, Nature (London) **403**, 267 (2000).
[4] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002).
[5] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[6] P. R. ten Wolde and P. B. Warren, q-bio/0310029.
[7] R. Wagner, *Transcription Regulation in Prokaryotes* (OUP, Oxford, 2000).
[8] M. Ptashne, *A Genetic Switch: Phage Lambda and Higher Organisms* (Blackwell, Oxford, 1992), 2nd ed.
[9] P. J. Willy, R. Kobayashi, and J. T. Kadonaga, Science **290**, 982 (2000).
[10] Z. Szallasi, TRENDS Pharmacolog. Sci. **22**, 110 (2001).
[11] A. Arkin, J. Ross, and H. H. McAdams, Genetics **149**, 1633 (1998).
[12] E. Aurell, S. Brown, J. Johanson, and K. Sneppen, Phys. Rev. E **65**, 051914 (2002).
[13] J. L. Cherry and F. R. Adler, J. Theor. Biol. **203**, 117 (2000).
[14] T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116 (2001).
[15] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[16] W. Bialek, in *Advances in Neural Information Processing*, edited by T. K. Leen, T. G. Dietterich, and V. Tresp (MIT Press, Cambridge, 2001), Vol. 13; see also cond-mat/0005235.
[17] M. R. Evans, D. P. Foster, C. Godrèche, and D. Mukamel, Phys. Rev. Lett. **74**, 208 (1995).
[18] C. Godrèche *et al.*, J. Phys. A **28**, 6039 (1995).