

Chapter 5 Network Motifs of Single Input Module (SIM) and Toggle Switch

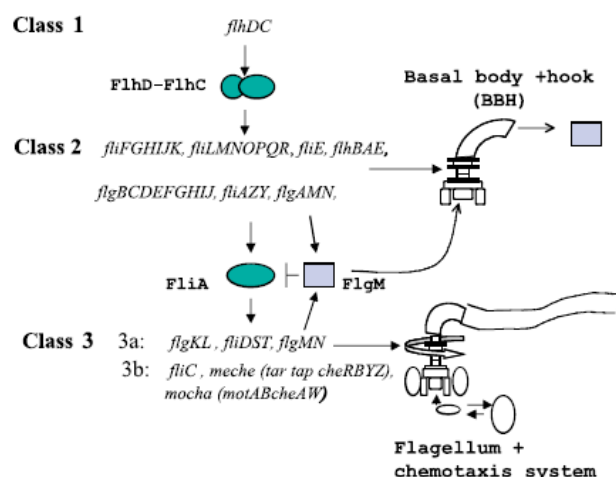
The genetic program of a living cell is determined by a complex web of gene networks. The proper execution of this program relies on faithful signal propagation from one gene to the next. The primary goal of systems biology is to understand the design principles of the transcription networks that govern the timing of gene expression.

In this chapter, we are going to show that by studying the transcriptional timing programs in single cells, the design principles for the assembly of molecular machinery, metabolic pathways and cell fate decision circuitry can be revealed. Because the proper execution of these programs relies on faithful signal propagation from one gene to the next, understanding how noise propagates in gene networks is crucial and will be discussed in the final section.

5.1 Transcriptional Timing Program of Flagella Biosynthesis in *E. coli* (Science 292, 2080 (2001) and Cell 117, 713–720 (2004))

In the first part of this chapter we will discuss the timing program of the flagella pathway of *E. coli*. This timing program of transcription is shown to associate with multiple steps of the flagella assembly.

When growth conditions become mildly unfavorable, *E. coli* produces several rotating flagella and swims away. This system regulates the transcription of 14 operons, which are arranged in a regulatory cascade of three classes. The class 1 operon encodes the transcriptional activator (**FlhDC**) of class 2 operons. Class 2 genes include structural components of a rotary motor called the basal body-hook (BBH) structure, as well as the transcriptional activator (**FlhA**) for class 3 operons. Class 3 includes flagellar filament structural genes and the chemotaxis signal transduction system that directs the cells' motion. A timing program must be implemented to ensure that class 3 genes are not transcribed before functional BBH structures are completed.



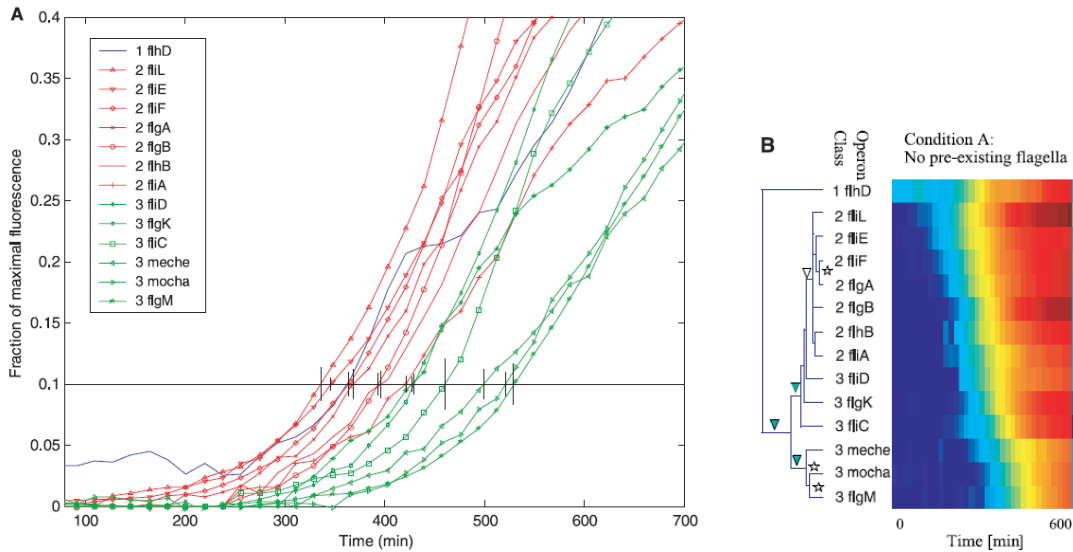
Experimental Design: Many *E. coli* strains are designed and grown in parallel. Each strain carries a plasmid in which one of the class 2 promoters governs expression of a fast-folding green fluorescent protein (GFP). The promoter activity is revealed by $P(t) = dGFP(t)/dOD(t)$, which can be estimated

using the data of GFP fluorescence and optical density (OD) of the cell.

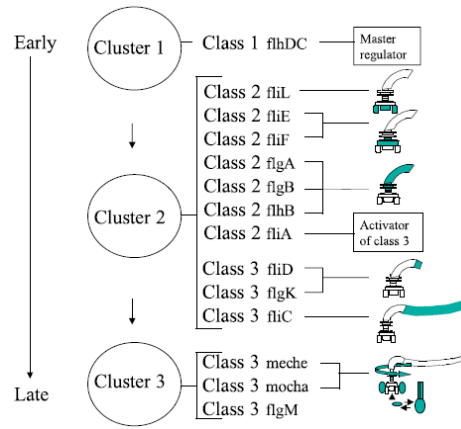
To investigate the input function of the two inputs (*FlhDC* and *FliA*) in a controlled way, a strain in which the *flhD*, *fliA* and *flgM* genes are deleted is prepared. The *fliA* and *flhDC* genes are co-expressed exogenously by using three plasmids. One plasmid containing the **lac** promoter is used to control the expression of the *flhDC* gene and another plasmid containing the **tet** promoter governs the expression of the *fliA* gene. The third plasmid, which contains one of the flagella promoters, is employed to control the GFP expression.

The strain also has a chromosomal cassette including genes **lacI** and **tetR**. The first gene **lacI** codes for the lactose repressor *LacI*, which inhibits the **lac** promoter activity. The gene product of *tetR* is the tetracycline repressor *TetR*, which can inhibit the **tet** promoter activity. Isopropyl- β -D-thiogalactopyranoside (IPTG) and anhydrotetracycline (aTC) can bind to and inhibit the repression of the lactose and tetracycline repressors, respectively. Therefore in the absence of IPTG and aTc, this *E. coli* strain has negligible *fliL* promoter activity.

Results: The flagella system is turned on during the exponential phase of growth. Clustering the fluorescence levels of the operons according to similarity in their expression profiles showed that they fall into clusters that correspond to the genetically defined classes 1 and 2 as the data are split hierarchically into a tree.

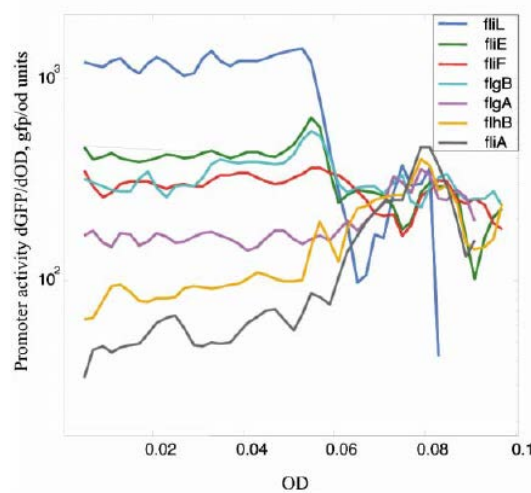


To define the temporal order of expression, first considering each splitting from the top down and computing the average log fluorescence (normalized by the maximal fluorescence) for the two subtrees $\log(f_1)$ and $\log(f_2)$. Next, the response time was computed with $t_i = -\int \log[f_i(t)]dt$. The subtree with the smaller t_i was then positioned to the left. The observed order corresponds to the spatial position of the gene products during flagellar motor assembly.



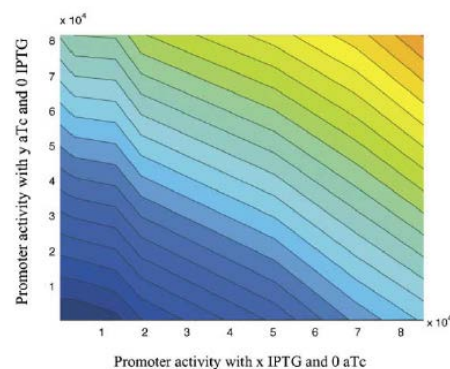
As shown the master regulator *flhDC* is the first gene to turn on and the transcription factor *flhA*, responsible for turning on class 3 genes, is the last class 2 gene to turn on.

The measured promoter activities exhibit two phases. In the first phase, the promoter activity is approximately constant, with different promoter activity for different operon. The order of promoter activity matches with the temporal order in which the genes were expressed.

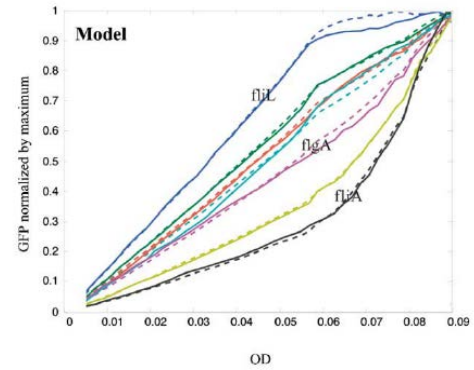
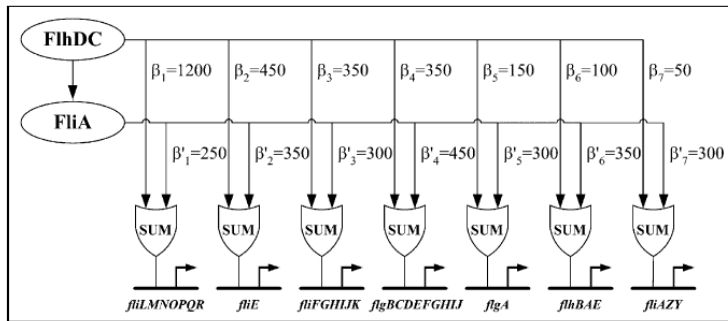


In the second phase, which corresponds to the onset of expression of the class 3 promoters, the promoters exhibit a peak of activity with similar promoter activity.

The class 2 genes in a strain with deleted *flhD*, *flhA* and *flgM* genes can be activated by exogenously expressed *FlhA* in the absence of *FlhDC*, suggesting the *cis*-regulatory input function to be an OR gate. The effects of *FlhA* and *FlhDC* were found to be additive, yielding an input function best described as a SUM gate.



The activity of each promoter of the flagella class 2 genes can be described by $P_i(t) = \beta_i X(t) + \beta'_i Y(t)$, where $X(t)$ and $Y(t)$ are the effective activity profiles of *FlhDC* and *FliA*, respectively. A blueprint of the transcriptional timing program for the entire class 2 operons can therefore be constructed in terms of the two hidden functions $X(t)$ and $Y(t)$. This model well depicts the dynamics of all of the class 2 promoters.



(Left) The blueprint of the flagella class 2 system model and (right) the experiment and model for GFP fluorescence as a function of OD. The model GFP normalized by maximal level versus cell number (dashed lines) is plotted together with experimental data (full lines).

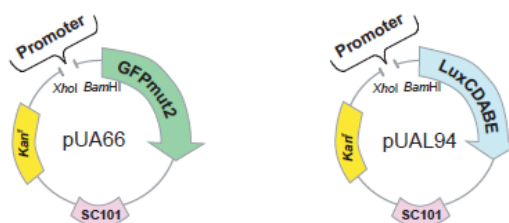
The model for the class 2 flagella system has a multi-output feedforward loop (FFL) architecture. In the transcriptional network, the activation coefficients (β_i) for the master regulator (*FlhDC*) show a hierarchy, whereas the coefficients (β'_i) for the downstream regulator (*FliA*) are quite similar for the different promoters. As the concentration of the transcription factor *FlhDC* gradually increases in the cell, it first binds and activates the operons with the highest affinity sites, and only later does it bind and activate operons with lower affinity sites.

Conclusion: The observed timing order of the flagella motor assembly corresponds to the spatial position of the gene products. The regulatory sites of the operons are ranked in affinity, resulting in a timing program in which the earlier the gene products participate in flagella assembly, the stronger and earlier its promoter is activated.

5.2 Transcriptional Timing Program of Amino Acid Biosynthesis (AAB) in *E. coli* (Nature Genetics 36, 486–491 (2004))

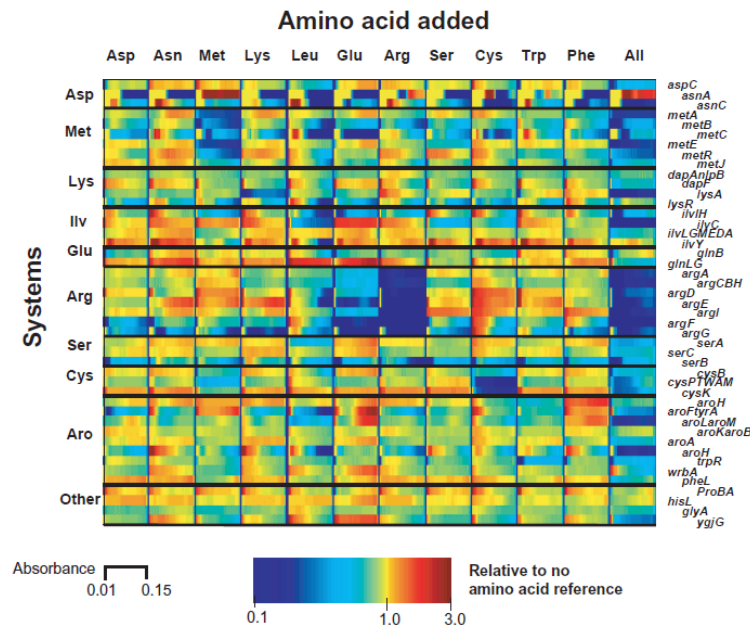
In this section, the transcriptional timing program of amino acid biosynthesis will be given to illustrate the underlying principles of unbranched biosynthesis pathways.

Experimental Design



Experimental design of monitoring promoter activity with high resolution and accuracy. Plasmids contain either the reporter gene *gfpmut2* or the operon *LuxCDABE*. BamHI and XhoI cloning site is included for inserting the promoter for transcribing one of the AAB enzymes, a low-copy origin SC101 and a kanamycin resistance gene.

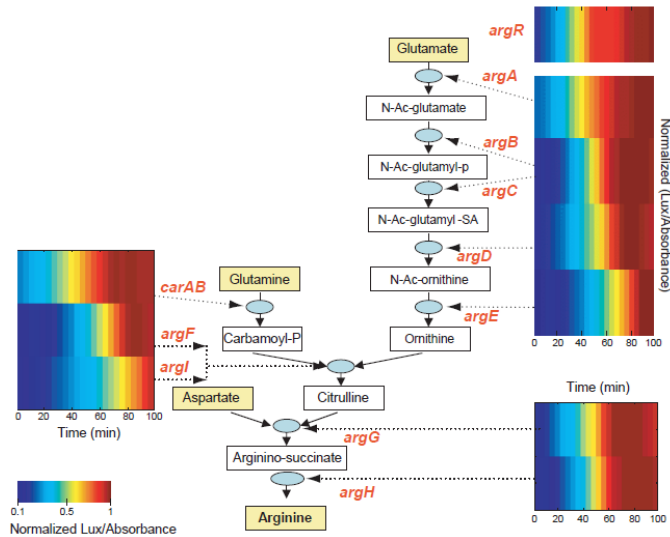
Results: The dynamics of the AAB promoter activity after a shift from medium with no amino acids to defined medium supplemented with one amino acid were first investigated.



Reporter strains were grown on defined medium supplemented with one amino acid (indicated at the top), with all 20 amino acids (last column). Rows represent promoters, arranged by amino acid system, and columns represent experimental conditions, with 31 absorbance points between A = 0.01 and A = 0.15 per condition.

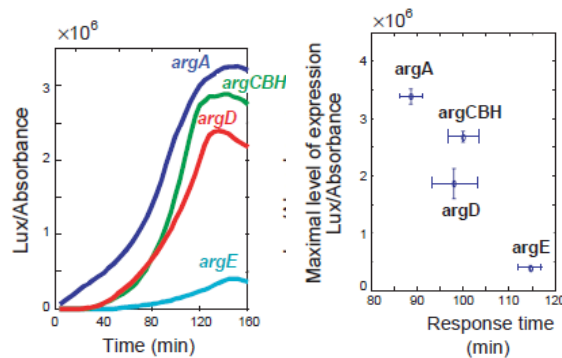
Virtually all promoters were active in the absence of amino acids, whereas promoter activity was repressed in the presence of the complete set of amino acids (the last column marked with All). Transferring the cells to medium that included one amino acid generally resulted in the down regulation of the corresponding biosynthesis pathway. However, cross-activations/repressions, in which a given amino acid affected the promoter activity of genes belonging to other amino acid pathways can also be found.

For arginine synthesis, glutamate is a precursor metabolite. Two other amino acids, glutamine and aspartate, are also needed for arginine production. As shown below, their products converge into the main pathway from two additional linear pathways.



The temporal transcriptional order matches the functional order of enzymes in each unbranched pathway. For example, in the main pathway that converts glutamate to ornithine, *argA* was the first promoter to be upregulated, followed by *argBC*, *argD* and *argE*. Temporal expression also matches gene order in the other two unbranched pathways in the arginine system: *carAB* preceded *argF* and *argI*, in a pathway that converts glutamine to citrulline, and *argG* preceded *argH* in the pathway that converts citrulline to arginine.

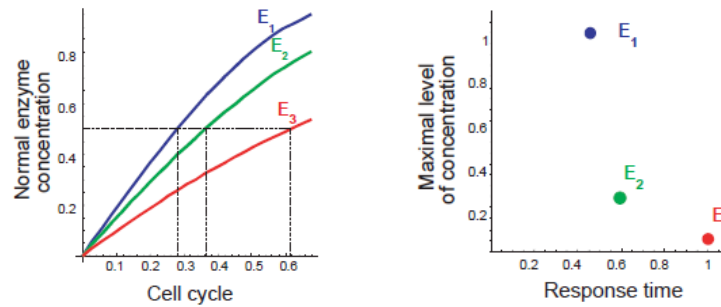
The promoter activity in the arginine biosynthesis pathway from *argA* through *argE* exhibits a hierarchy in the activity levels: *argA* reached the highest promoter activity, followed by *argBC*, *argD* and *argE*. The closer an enzyme is to the beginning of the pathway, the higher its maximal level of promoter activity.



The temporal order and the activity hierarchy can be quantified by the response time, defined as the time it takes to reach half of the maximal promoter activity level. The earlier the enzyme is involved in a pathway, the shorter is the response time and the higher the maximal promoter activity of its gene.

Unbranched biosynthesis pathways can be viewed as production pipelines, in which the end product is produced from an initial substrate S_0 through several intermediate steps. Consider a model

biosynthetic pathway composed of three enzymes (E_1 , E_2 , and E_3): $S_0 \xrightleftharpoons{E_1} S_1 \xrightleftharpoons{E_2} S_2 \xrightleftharpoons{E_3} S_3$ with S_0 denoting the saturated substrate, and S_3 the final product. An optimal transcriptional program can be found by optimizing the model parameters for minimal cost. The cost function used is composed of two factors: the cost to produce the enzymes and the rate and precision at which it approaches its goal. The optimized solution reveals a hierarchy in promoter strengths as well as temporal order similar to that experimentally observed.

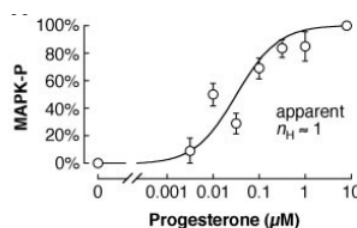


Conclusion: Two design principles can be discovered in the unbranched metabolic pathways: the earlier the enzyme acts in the pathway, the shorter the response time of its promoter activity and the higher its maximal promoter activity. The resulting transcription program is optimal under constraints of rapidly reaching a production goal with minimal total enzyme production. These two design principles characterize gene systems regulated by single-input-module (SIM) motif, in which a single regulator (such as *argR*) controls a set of target genes.

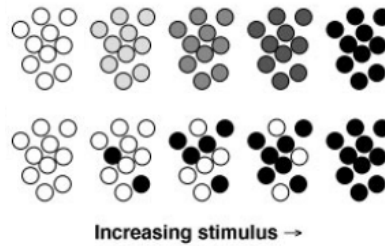
5.3 An All-or-None Cell Fate Switch in *Xenopus* Oocytes (Science 280, 895-898 (1998))

Oocyte maturation is a cell fate switch, which allows oocytes to reside in either the G2 arrest or the metaphase arrest state for extended periods of time.

Progesterone-induced maturation is thought to be triggered by activation of a cascade of mitogen-activated protein kinases (MAPK)—Mos, Mek-1, and Erk2. Progesterone causes the accumulation of Mos, which phosphorylates and activates Mek-1. Active Mek-1 in turn phosphorylates and activates Erk2 (MAPK-P), which then brings about activation of the Cdc2–cyclin B complex. Overall responses of oocytes to progesterone are shown below with each point represents a sample of 11 to 39 oocytes.



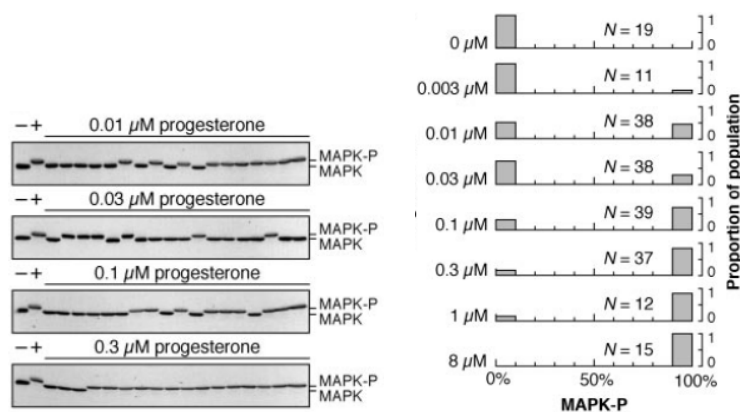
The graded overall response could mean that each of the individual oocytes had a graded response (figures in top row).



Or if the concentration of progesterone required to switch the oocytes on is different for each oocyte, even individual oocytes had perfectly switchlike responses, populations of oocytes would also yield a graded response.

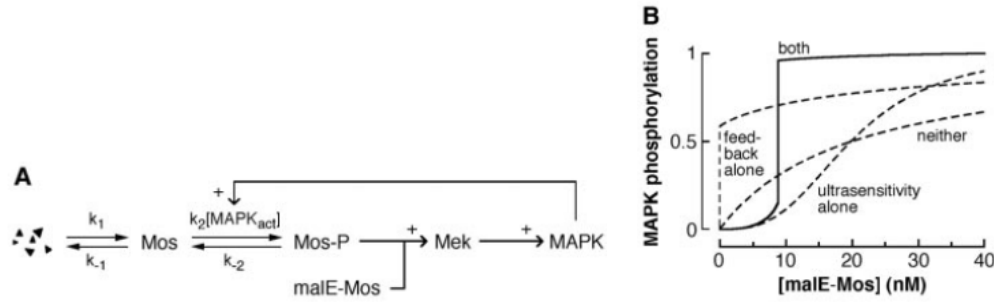
Experimental Design: These two possibilities can be distinguished by examining individual oocytes treated with intermediate concentrations of progesterone. If the individual responses are graded, each oocyte should have an intermediate amount of MAPK phosphorylation (MAPK-P); if they are switchlike, the oocytes should have either very high or very low MAPK-P.

Results: The following figures show the results of the steady-state phosphorylation of ERK2 in 190 individual progesterone-treated oocytes and 19 individual untreated oocytes. Every oocyte had either very high (>90% of maximal) or very low (<10% of maximal) amounts of MAPK-P. Thus, the response of the individual oocytes was essentially all-or-none with a lower bound for the Hill coefficient about 42.



To determine whether the all-or-none character of the response was generated by the MAPK cascade, purified malE-Mos (a recombinative Mos and maltose-binding protein, which is a direct activator of Mek-1) was microinjected into oocytes and assessed the resulting MAPK-P. The Hill coefficient inferred for the oocytes' individual responses was 35.

The responses seen in intact oocytes were much more switchlike than those seen in oocyte extracts, which can be attributed to from a positive feedback loop known to operate in intact oocytes, and known not to operate in extracts. A positive feedback loop would markedly increase the abruptness of the MAPK cascade's response.

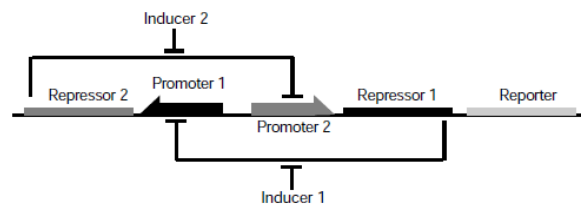


MAPK cascade behaves like a ultrasensitive switch with a negligible response to small stimuli, but once the stimuli reach a threshold the system switches from off to on over a narrower range of stimulus concentrations. Thus the MAPK cascade and the protein synthesis–dependent positive feedback loop can act together with the ultrasensitivity of the MAPK cascade filtering small stimuli out of the feedback loop.

Conclusion: The MAPK cascade is activated in a highly ultrasensitive all-or-none fashion during *Xenopus* oocyte maturation. This behavior is proposed to arise from two known properties of the oocyte’s MAPK cascade: positive feedback, which ensures that the oocyte cannot rest in a state with intermediate MAPK phosphorylation, and the cascade’s intrinsic ultrasensitivity, which establishes a threshold for activation of the positive feedback loop.

5.4 Genetic Toggle Switch in *E. coli* (Nature 403, 339-342 (2000))

The toggle switch is composed of two repressors and two constitutive promoters. Each promoter is inhibited by the repressor that is transcribed by the opposing promoter. This toggle switch design can achieve robust bistable behavior over a wide range of parameter values. Therefore two states are tolerant of the fluctuations in gene expression.



In the absence of inducers, two stable states are possible: one in which promoter 1 transcribes repressor 2, and one in which promoter 2 transcribes repressor 1. Switching is accomplished by transiently introducing an inducer of the currently active repressor. The inducer permits the opposing repressor to be maximally transcribed until it stably represses the originally active promoter.

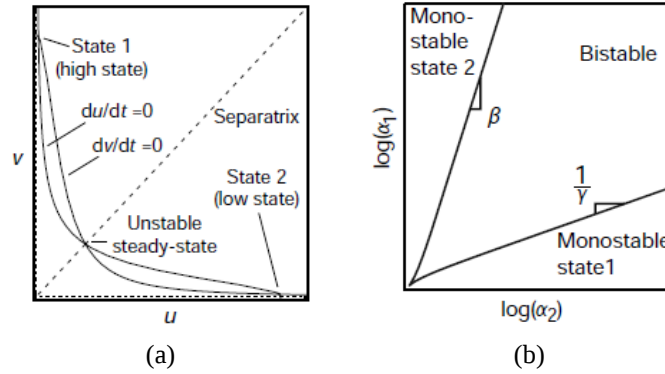
The dynamics of the toggle switch can be described with the following equations

$$\begin{aligned}\frac{dU}{dt} &= \frac{\alpha_1}{1+V^\beta} - U, \\ \frac{dV}{dt} &= \frac{\alpha_2}{1+U^\gamma} - V\end{aligned}$$

where U is the concentration of repressor 1, V denotes the concentration of repressor 2. α_1 is the

effective rate of synthesis of repressor 1, α_2 is the effective rate of synthesis of repressor 2, which are a lumped parameter that includes repressor binding, transcription rate, and translation rate. β is the cooperativity of repression of promoter 2 and γ is the cooperativity of repression of promoter 1.

The geometric structure of the equation system reveals the origin of the bistability: the nullclines ($dU/dt = 0$ and $dV/dt = 0$) can intersect at three points, producing one unstable and two stable steady states.

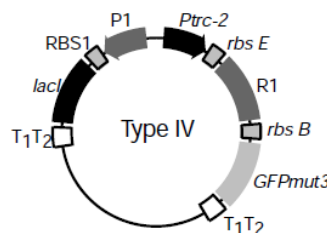


Geometric structure of the equations reveals a bistable toggle switch with balanced promoter strengths. (b) The bistable region. The lines mark the transition between bistability and monostability. The slopes of the bifurcation lines are determined by the exponents β and γ for large α_1 and α_2 .

Three key features of the equation system become apparent. First, the nullclines can intersect three times because of the sigmoidal shape, which arises for $\beta, \gamma > 1$ (i.e., the bistability of the system depends on the *cooperative* repression of transcription). Second, the rates of synthesis of the two repressors must be balanced. If the rates are not balanced, the nullclines will intersect only once, producing a single stable steady state. Third, the structure of the toggle network creates two basins of attraction. Thus, a toggle with an initial condition anywhere *above* the separatrix will ultimately settle to state 1, whereas a toggle starting *below* the separatrix will settle to state 2.

Experimental Design: The promoters used in the toggle can be $P1 = P_{LtetO-1}$, which is repressible by $R1 = TetR$ (or $P1 = P_{LS1con}$, which is repressible by lambda phage repressor CI), and a promoter P_{trc-2} (repressible by *LacI*).

The order of the transcriptional efficiencies of the promoters is $P_{LS1con} > P_{trc-2} > P_{LtetO-1}$. The rates of synthesis of the repressors (α_1 and α_2) can be modified the downstream ribosome binding sites (RBS).

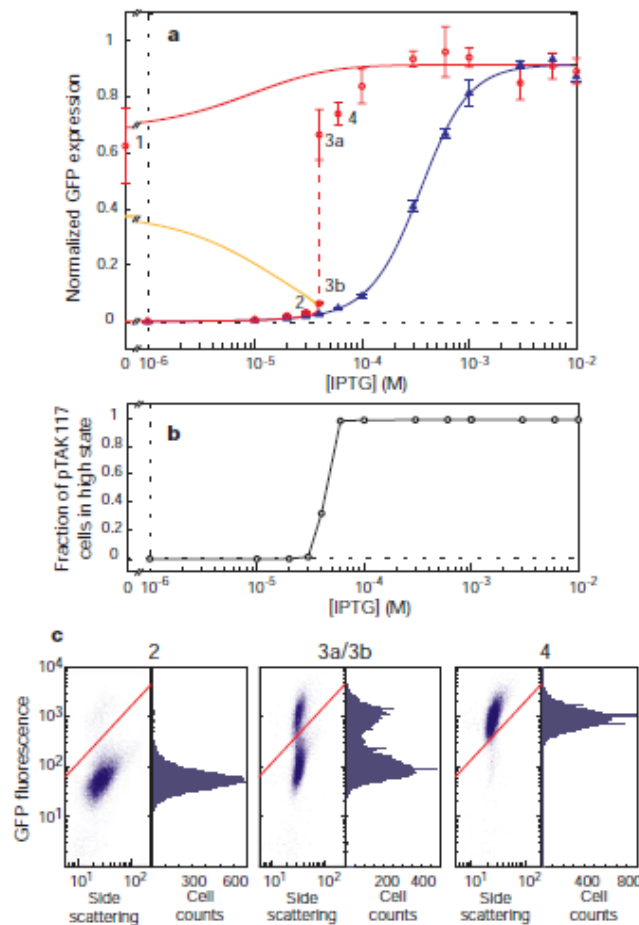


The toggle switch plasmid. Promoters are marked by solid rectangles with arrowheads. Genes are denoted with solid

rectangles. Ribosome binding sites and terminators (T1T2) are denoted by outlined boxes. Different **P1** promoters, RBS1 ribosome binding sites, and/or R1 repressors, can be used to develop various toggle switches.

In the toggle plasmids, the *gfpmut3* gene is arranged as the second cistron downstream of the **P_{trc-2}** promoter. Thus transcription from **P_{trc-2}**, and hence, repression of **P1**, results in the expression of *gfpmut3*. For clarity, this state is termed the *high* state. The opposing state, in which **P1** is transcribed and **P_{trc-2}** is repressed, is termed the *low* state.

Results The threshold, or bifurcation, in the toggle switch is illustrated in the following figures. In this experiment, both the toggle switch (initially in the *low* state) and the control were grown in 13 different concentrations of IPTG for 17 h. Induction of the control has the familiar sigmoidal shape (blue symbols). In contrast, the toggle (red symbols) follows the induction curve of the control up to an IPTG concentration of 40 μ M (at point 3b), at which point it crosses the bifurcation and exhibits a quasi-discontinuous jump to the *high* expression state. Owing to the natural fluctuations in gene expression, the bifurcation is not a perfect discontinuity. Near the bifurcation, this blurriness is realized as a bimodal distribution of cells.



The toggle plasmid (**red** circles) exhibits a quasi-discontinuous jump to the *high* state whereas the control plasmid (**blue** triangles) exhibits a sigmoidal induction curve. Point 1 is taken from separate experiments measuring the *high* state of the toggle plasmid with no inducer. Points 3a and 3b are the *high* and *low* modes of a bimodally distributed cell population. The theoretical curve (**red**) shows the stable steady states and the **orange** curve shows the unstable steady state of the

toggle. The blue curve shows the steady state of the IPTG-inducible control plasmid.

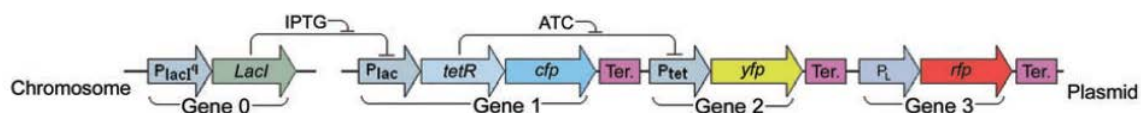
The middle figure shows the fraction of toggle cells in the *high* state at various concentrations of IPTG. The sudden switching to the *high* state is more clearly visible.

The figures in the bottom row show the scatter plots (left plots) and histograms (right plots) illustrating the condition of the toggle cells at points 2, 3a/3b and 4 near the bifurcation point. *High*-state and *low*-state cell populations are divided by the red line in the scatter plots. The two states of the toggle are clearly evident in the bimodally distributed cells (point 3).

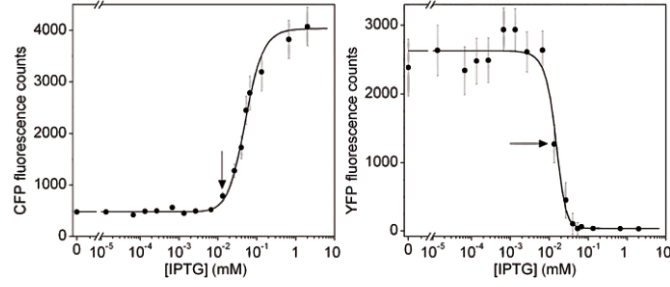
5.5 Noise Propagation in Gene Networks (Science 307, 1965-1969 (2005) and Nature 467, 167-173 (2010))

Studying noise propagation in gene networks is crucial for understanding signal fidelity in natural networks and designing noise-tolerant gene circuits. However, how expression fluctuations propagate from one gene to the next is largely unknown. In this section, we will discuss how to design an experimental strategy and a data analysis scheme to address this issue.

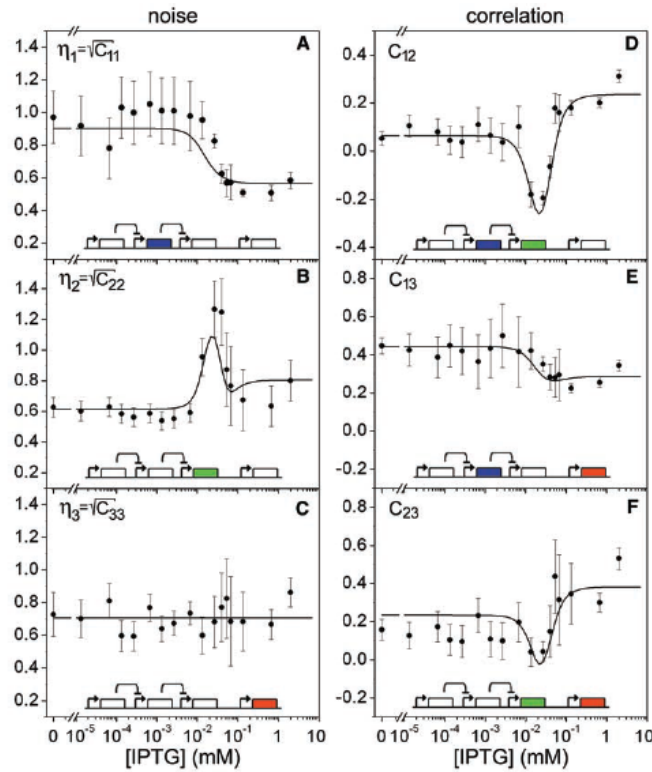
Experimental Design: A genetic network was designed to measure how fluctuations in an upstream gene transmit downstream. The synthetic network consisted of four genes, of which the expression level of three genes were monitored in single *E. coli* cells by cyan, yellow, and red fluorescent proteins (CFP, YFP, and RFP). The first gene *lacI* is constitutively transcribed. The gene product of *lacI* will down-regulate the transcription of the second gene *tetR* that is bicistronically transcribed with *cfp*. The gene product of *tetR* in turn downregulates the transcription of the third gene, reported by YFP. The fourth gene *rfp*, which is not part of the cascade, is under the control of the lambda repressor promoter P_L . This gene was used to monitor the effect of global fluctuations. The inducers IPTG and aTC were used to tune the expression of the upstream gene and the coupling between the two genes.



Results: The response of single cells to various amounts of inducers can be monitored by measuring fluorescence of the three fluorescent reporters for 2000 individual cells. The average signal of the upstream gene (gene 1) displayed a sigmoidal response to changes in the concentration of IPTG in the growth media. In response, the average signal of the downstream gene (gene 2) behaved inversely and decreased sharply at larger IPTG concentrations. However, the average expression does not capture the population behavior, because the expression of most cells is quite different from the average.



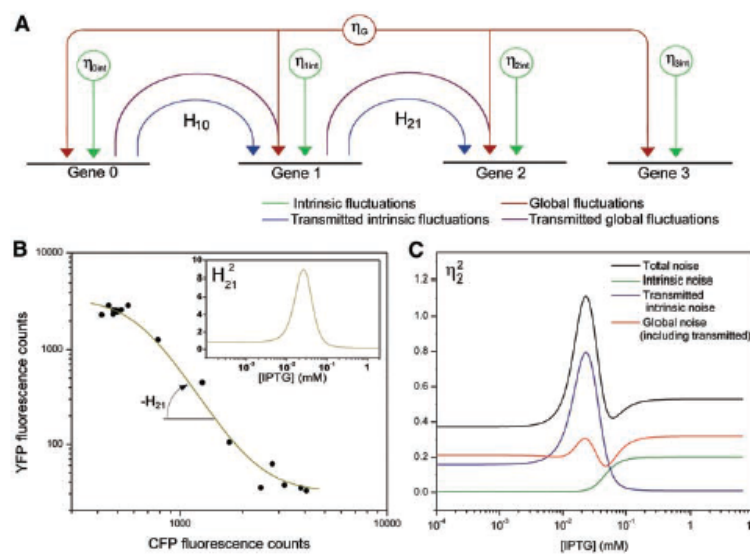
To quantify the expression fluctuations and the degree of correlation between different genes, the correlation $C_{ij} = (\langle F_i F_j \rangle - \langle F_i \rangle \langle F_j \rangle) / \langle F_i \rangle \langle F_j \rangle$ can be defined from the fluorescence levels F_i in individual cells. The brackets $\langle \dots \rangle$ denote averaging over all cells in the population, and the indices i and j refer to the gene number. The self-correlation is identical to the square of the coefficient of variation $\eta_i = \sigma_i / \mu_i$, which is defined as the standard deviation σ_i of the expression distribution normalized to the mean μ_i . The six correlations were measured and plotted in the following figures as a function of the IPTG concentration.



The noise properties of the upstream gene, reflected in $\eta_1 = \sigma_1 / \mu_1$ shown in A, are very different from those of the downstream gene, reflected in $\eta_2 = \sigma_2 / \mu_2$ shown in B, even though both genes are repressed by the same upstream repressor (*LacI*). The correlations C_{13} and C_{23} are dependent on IPTG concentration (E and F) even though RFP is not part of the cascade and one should expect a correlation that is independent of IPTG.

To clarify these issues, a stochastic model based on the Langevin approach can be developed. The differential equations describing the dynamics of the system are modified by adding stochastic terms that reflect the two sources of noise: intrinsic fluctuations due to low numbers of molecules and

global fluctuations in cellular components that change the reaction rates for all genes. Thus, the noise in each gene of the synthetic network can be decomposed into three components: intrinsic noise in that specific gene (green arrows), transmitted intrinsic noise (blue arrows) from the upstream genes, and global noise modulated by the network. The different intrinsic noise sources are usually uncorrelated, whereas the transmitted global noise (purple arrows), which arise from the same sources, does not simply add to the direct global noise (red arrows). We can characterize the transmitted intrinsic noise (blue arrows) with the logarithmic gain H_{ji} , which reflects how the average expression of the downstream gene j changes as the expression of the upstream gene i is varied. As shown in B of the following figures, the logarithmic gain H_{21} can be obtained as the negative of the slope in $\log\text{-}\log$ space of the mean expression of YFP as a function of mean CFP expression. The squared H_{21} exhibits pronounced peak occurring at an IPTG concentration for which the response of the downstream gene is most sensitive to changes in the upstream signal.



Different noise components can be resolved with the model. For example, gene 2 reveals increased intrinsic noise (green line) at higher IPTG concentrations because the noise component varies as the inverse of the square root of the mean. The transmitted intrinsic component (blue line) corresponds to the square of the logarithmic gain H_{21}^2 times the noise in the upstream gene η_1 . The global noise component (red line) reflects the modulation from the transmitted global noise. As shown, the main features of the noise in this gene are determined by the network interactions.

Conclusion: The noise in a gene affects expression fluctuations of its downstream genes. The noise has a correlated global component that is modulated by the network. Measuring the correlation between a constitutive gene and a gene embedded in a network provides a sensitive probe for correlated sources of noise. This would have been difficult to reveal by monitoring single genes.

More ref. on this topics: [Avigdor Eldar and Michael B. Elowitz, Functional roles for noise in genetic circuits, Nature Vol. 467|9, 167-173 \(2010\). doi:10.1038/nature09326](#)