

Characterization of Performance and Dynamics of Genetic Networks

Mahendra Kumar Prajapat*, Kirti Jain**
Supreet Saini***

Indian Institute of Technology Bombay, Mumbai, Maharashtra 400 076
India (Tel: 91-22-2576-7216; e-mail: *mahendra.prajapat@iitb.ac.in, **kirtijn1@gmail.com, ***saini@che.iitb.ac.in).

Abstract: Living cells are continuously monitoring their surroundings and making appropriate decisions to enhance chances of survival. Information regarding environmental conditions is processed in biochemical and genetic networks, eventually leading to a cellular phenotype. However, for all transfer of information, a large number of possible network topologies are available to the cell to choose from. In the common bacterium *Escherichia coli*, frequency of certain topologies far outnumbers others. What decides the choice of a particular network over the others? In this work, we work with the simplest transcriptional network - an interaction between a regulator, R and its target gene, T - and develop a computational and experimental framework to characterize performance of different possible networks in terms of a list of objective functions. Our results suggest and we speculate that a group of proteins evolve into a particular network topology so as to optimize the most crucial objective function for their particular role in cellular physiology.

Keywords: Genetic networks, dynamics, evolution, physiology

1. INTRODUCTION

Networks are information processing centres of the cell (Fraser et al., 2013). Developments in the last decade or so have demonstrated that living cells exhibit a distinct preference for certain network topologies over others (Dekel et al., 2005; Mangan and Alon, 2003). This results in an over-representation of these topologies among a large number of available options to the cell. Mathematical and experimental analysis of these topologies has clearly demonstrated that these over-represented topologies outperform their competitors when compared against a defined objective function (Goentoro et al., 2009; Kaplan et al., 2004). This understanding of networks is crucial for design of industrially-relevant synthetic circuits in living cells.

The analysis of genetic networks and their relative advantages has, however, not included the significance of the biochemical parameters associated with the network. Varying the values of the biochemical parameters, a network can be designed to exhibit vastly different dynamic and steady-state behavior. Hence, the role of network topology must be analyzed in conjunction with the available biochemical parameter space. In this work, we analyze the most simple of the genetic interactions in a living cell - a transcriptional regulator, R controlling the production of its target gene, T. This interaction can be any of the six possible topologies as shown in Fig. 1. Why cells chose one topology over the other remains an open question.

In addition, what is the reason for the choice of under-represented topologies in a certain networks? Do these under-

represented topologies also confer an advantage to the cell in certain physiological conditions?

To answer this question, we employed a computational and experimental approach using the common bacterium *Escherichia coli* (*E. coli*). We analyze the network in terms of a series of objective functions that could be physiologically relevant for a cell. Our results show that different topologies (each analyzed over its entire biochemical space) offer specific features/properties to the cell. These properties could be advantageous /disadvantageous to the cell depending on its precise function in the cellular working. Our results indicate that it is this factor that results in choice of a specific topology over the other. In addition, this approach provides a framework for studying evolution of topology of genetic networks. We validate our computational work with experimental studies by studying the characteristic features from 5 promoters from each of the 6 classes as shown in Figure 1. Our experimental results demonstrate that, in accordance, with the computational work, the distribution of time of activation and cost of response is spread in agreement with the computational work. The precise objective function which dictates evolution of a individual network will need more closer inspection.

2. METHODS

2.1 Mathematical Modeling

All six models as represented in Fig. 1 were modeled using Gillespie algorithm to account for stochasticity (Gillespie, 1976; Gillespie, 1977).

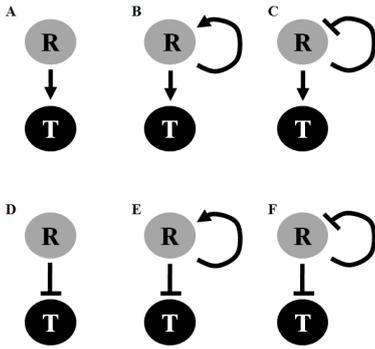


Fig. 1. Transcription Factor and Target Gene Interactions: A) R (regulator) activates T (target); B) R activates R and T; C) R activates T and represses R; D) R represses T; E) R represses T and activates R; and F) R represses R and T. Arrows indicate transcriptional activation of the promoter, and blunt-ends indicate transcriptional repression of the promoter.

For each model, the dynamics of transition from the target T "Off" to "On" and from "On" to "Off" were recorded.

The assumptions involved in our modeling work are as follows:

- i. The regulator R, acts as a dimer to control gene expression. This assumption is true for a large number of transcription factors in bacteria (Sadamoto and Muto, 2013; Yu et al., 2010),
- ii. 500 cells were simulated for each transition in each network. This number was considered large enough to represent behavior of larger sets of data,
- iii. Gene expression control were assumed to be via Hill equation,
- iv. Degradation of R and T was proportional to their amounts in the cell,
- v. The parameter values were varied in accordance with the observed values from a number of previous reports (Rosenfeld et al., 2005; Süel et al., 2006; Rosenfeld et al., 2007). It was assumed that under these values, the entire physiologically relevant parameter space is being covered, and
- vi. "Off" conditions were modeled as absence of a signal variable, s . Under "On" conditions, the signal variable, s , took the value 1, and allowed for R-independent expression of the regulator at a low basal rate. In our simulations, the signal parameter represents the environmental conditions that induce expression of R.

The governing equations used for the network A are given as follows. All other networks were mathematically framed similarly.

$$\frac{d(R / kdR)}{dt} = c \times s - R \quad (1) \quad \frac{d(T / kdT)}{dt} = \frac{a \times R^2}{b^2 + R^2} - T \quad (2)$$

Where, kdR and kdT are the degradation constants for R and T respectively; c is the activation constant for transcription of R in response to environmental conditions; and a and b are the Michealis-Menten constants associated with regulation of expression of target T by regulator R.

Networks B, C, E, and F had 5 parameters each, and networks A and D were defined by 3 parameters only (Appendix A). Each of the networks B, C, E, and F was simulated for 5 parameter values each, resulting in 3125 network simulations. Networks A and D were simulated for 15 values of each parameter. The parameter value range was taken to be representative of typical values found in bacterial transcription control (Appendix B). Hence, a total of 15^3 networks (3375) were simulated for topologies A and D. The model simulation data was then used to characterize the performance of each network in terms of the following parameters:

- i. Steady state value of the target gene, T and regulator R (on transition from "OFF" to "ON" state),
- ii. Steady state value of T and R on transition from "ON" to "OFF" state,
- iii. Time of activation, time taken to reach 50% of the total change in the value of T (on transition from "OFF" to "ON",
- iv. Time of deactivation, time taken to reach 50% of the total change in the value of T on removal of the activating signal, and
- v. Cost of response, amount of R required to produce one unit of target, T, in response to changing environmental conditions.

Averages for each of the above (v) criteria from all networks were reported below.

2.2 Experimental Procedures

All experiments were conducted in *E. coli* strain RP437 in Luria-Bertani (LB) media at 37°C. Amplification of promoters was done using Taq Polymerase DNA Polymerase. Kanamycin was used to make a final concentration of 40µg/ml in all experiments. All reporter plasmids were constructed with the fusion of promoter of interest with green fluorescent protein (*gfp*) gene on a medium copy number plasmid, pPROBE *gfp*[tagless] (Miller et al., 2000). Fluorescent and absorbance after growth of 14 hours were measured on a TECAN Safire2. All experiments were conducted in triplicate. The average values and standard deviations are reported.

3. RESULTS

3.1 Natural Frequency of R and T motif in *E. coli*

E. coli has over a hundred transcription factors encoded in its genome. These transcription factors further regulate expression of a large number of genes. In fact, some of the transcription factors are known to have more than 100 target

promoters. Overall, this results in a large number of transcription factor-target promoter interactions. In addition, a number of transcription factors are known to regulate their own-expression. Thus to understand the natural frequencies of the six possible R-T interactions, we counted all such interactions in *E. coli*. The frequencies are as represented in Fig. 2 below.

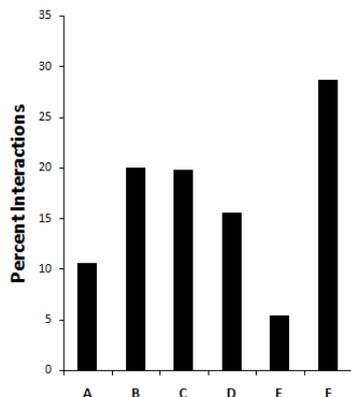


Fig. 2. Frequencies of regulator, R and target, T interactions in *E. coli*. Total transcription interactions is 1773. Target gene information was taken from RegulonDB. A, B, C, D, E, and F network structure is as defined in Fig. 1 (Salgado et al., 2013).

These results suggest that certain network topologies are preferred by *E. coli* over others. However, there are still some networks which remain in the network topologies with lower frequencies. Why has evolutionary pressure not forced these networks to move towards the more frequent topologies? To characterize each of these networks, we summarize our simulation results below.

3.2 Steady State Analysis

The transition of the network from an “OFF” to an “ON” position was simulated by a signal variable, s , which, in inducing conditions took the value 1, and in repressing conditions remained at zero. Our steady state analysis demonstrates the various qualitative features associated with each of the six topologies under consideration (Fig. 3). For network A, transition from OFF to ON results in a wide-spectrum of response on the R-T plane. However, large fractions of the response is expensive (high R, and low T). However, on transition to OFF state, the system response collapses, and at steady state the system exhibits a wide variety of T amounts over a narrow R range (at very small R values). This suggests that network A is an ineffective mechanism to switch off networks, while it exhibits a large range of steady state ON values. Network B also exhibits a poor transition to OFF, owing to the inbuilt positive feedback in the network. In addition, networks C, E, and F exhibit a large dynamic range of steady state R and T values. Network D is able to demonstrate a qualitatively unique response than others as it excludes a large fraction of the R-T space in the transition to the ON state. On transition to the OFF state, however, the entire R-T space is exhibited by the network in its response. These variations are likely important in dictating the qualitative response of the system, and more importantly,

decide the fate of the network behaviour in case of mutations in the promoter region, which alter the biochemical constants associated with the interactions. For instance, the analysis suggests that any mutation in network 1 is not likely to lead to a system response (in the OFF state) outside the narrow band as shown in the Figure above.

3.3 Cost Analysis

In addition to the steady state response of the system, an important characteristic in quantification of network performance is the cost of response. We quantify this by the number of R molecules needed to bring about a unit change in amounts of T in the cell. A large number of R required for a unit change in T is likely to be detrimental to the overall cellular behaviour.

Our results indicate that some network designs are inherently “more expensive” than the others. As shown in Fig. 4, we note that networks A, B, and C are more expensive than D, E, and F. In addition to the differences in the average costs associated with the production of T, there is also a great variation in the spread of the cost of T over different values of the biochemical parameters. For instance, we note that there is minimal variation in the cost in networks E and F, when they transition from the OFF to ON state. On the other hand, network D exhibits a narrow range of cost when transitioning from OFF to ON state, but a much wider range of the cost values when the cells transition from the ON to OFF state.

The cells are likely to have evolved network designs keeping in consideration the most critical parameter for their functioning. For example, if low cost of transition was the sole criteria for choice of a network – the likely networks will be E and F, since any change in the biochemical parameters in these networks does not result in change in the cost of production of T.

Overall, our results show that the networks E and F are able to produce T with the least number of R molecules. However, this comes with the disadvantage that E and F networks are able to exhibit poor control over expression of T, as cells transition from one environmental condition to another.

3.4 Dynamics of Activation and Deactivation

In response to the changing environmental conditions, cells need to quickly adapt to not only enhance their chances of survival but also accomplish complex and critical tasks in their life-cycle (Saini and Rao, 2010; Kumar et al., 2013). A faster dynamic response from a bacterium would not only enable it to adapt better to its new surroundings, but also provide a competitive advantage over other competitors. In our context, in each of our simulations, we quantified the time of response as the time taken by the system to reach 50% of its final response upon transition from one state to another. For example, if on transition from OFF to ON state, the value of T changes from 30 to 100, the time of response corresponds to the time taken to reach $T = 65$, half of the total response from 30 to 100. Upon calculation of the response time for both the transitions, we plotted the frequencies of particular response times as shown in Fig. 5 below.

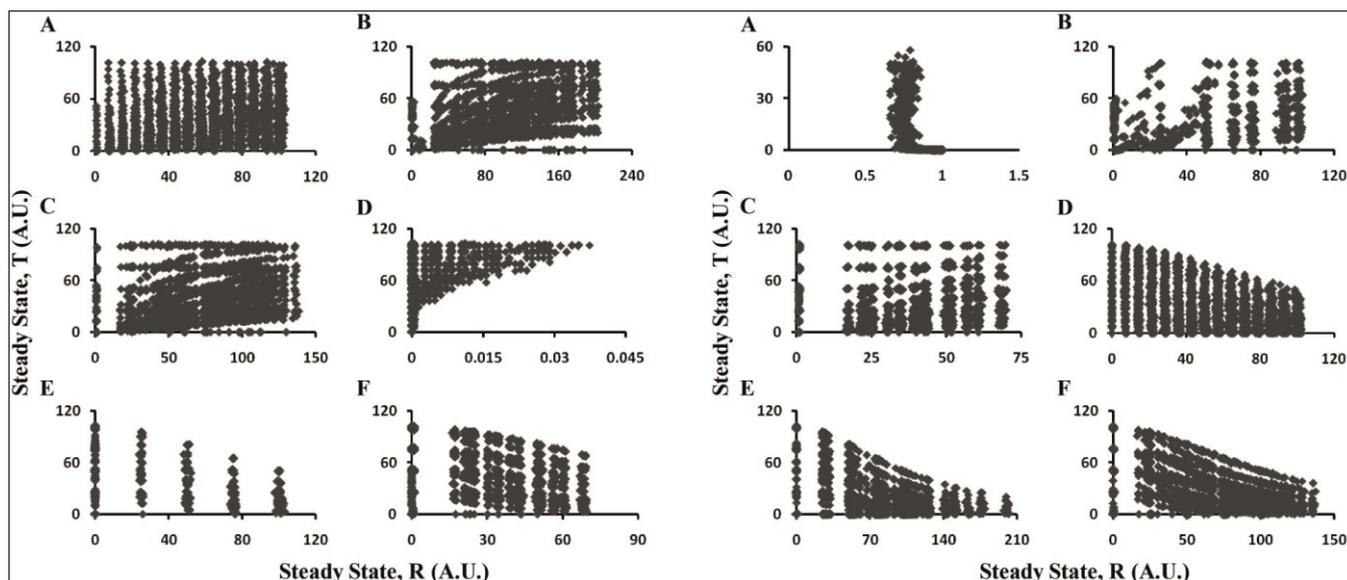


Fig. 3. Steady state regulator (R) vs. target (T) profiles as cells transition from "OFF" to "ON" state (left panel), and as cells transition from "ON" to "OFF" state (right panel). Labels A-F correspond to the networks as shown in Fig. 1.

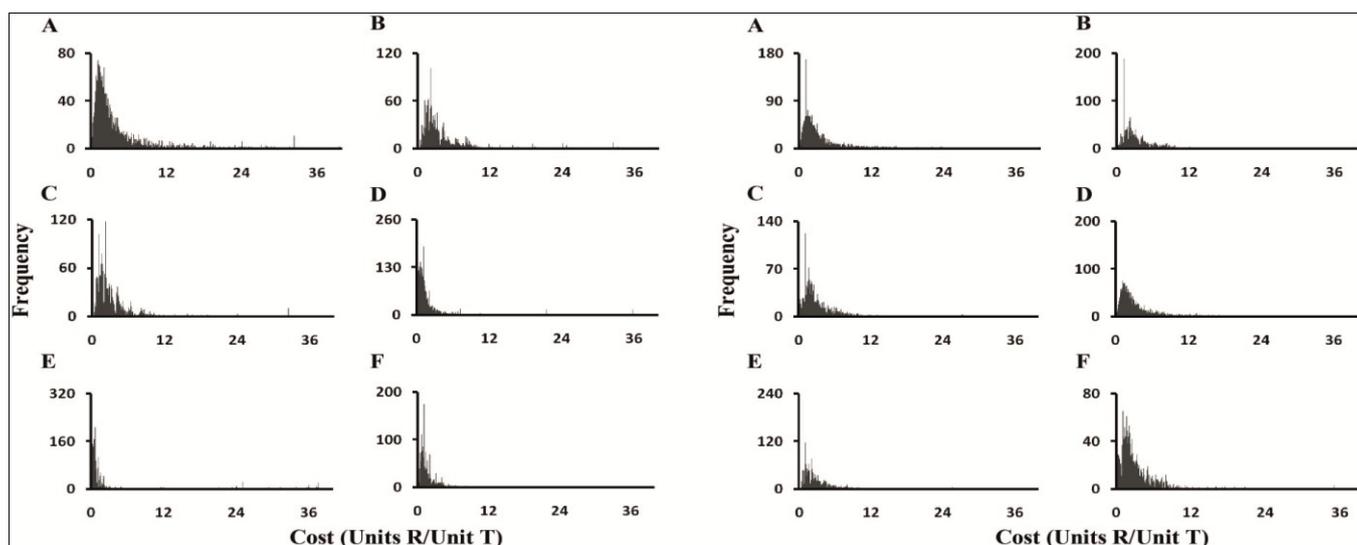


Fig. 4. Cost of production of target protein T in various circuits as cells transition from OFF to ON conditions (left panel) and from ON to OFF conditions (right panel). Cost is calculated in terms of number of R molecules needed to produce a unit T molecule. Labels A-F correspond to the networks as shown in Fig. 1.

Our results show that in networks A, B, and C – where the target T is being activated by regulator R, the spread of time of response is greater in the OFF to ON transition, as compared to the ON to OFF transition. On the other hand, in networks D, E, and F, the spread is greater for the OFF to ON transition, as compared to the ON to OFF transition. This result is perhaps to be expected as negative regulation is known to equilibrate cellular behaviour.

Variation in the time of response was maximum for the activation time in network C. This is contrary to the expectation that negative feedback speeds up and reduces the cell-to-cell variability in responses in cellular networks. However, our results show that negative feedback perhaps

makes the cell more susceptible to changes in the response time, should it acquire mutations.

3.5 Experimental Validation

To test results of our computational work experimentally, we developed reporter fusions for five promoters (all from *E. coli*) in each of the six networks designed as shown in Fig.1.

The following promoters were used for this study: Network A: *arcA*, *cbl*, *cra*, *flhDC*, *malT*, and *modE*; Network B: *marA*, *gadE*, *gadX*, *fucR*, *hyfR*, and *idnR*; Network C: *cynR*, *cysB*, *oxyR*, *fnr*, *fis*, and *gadW*; Network D: *allR*, *arcA*, *cra*,

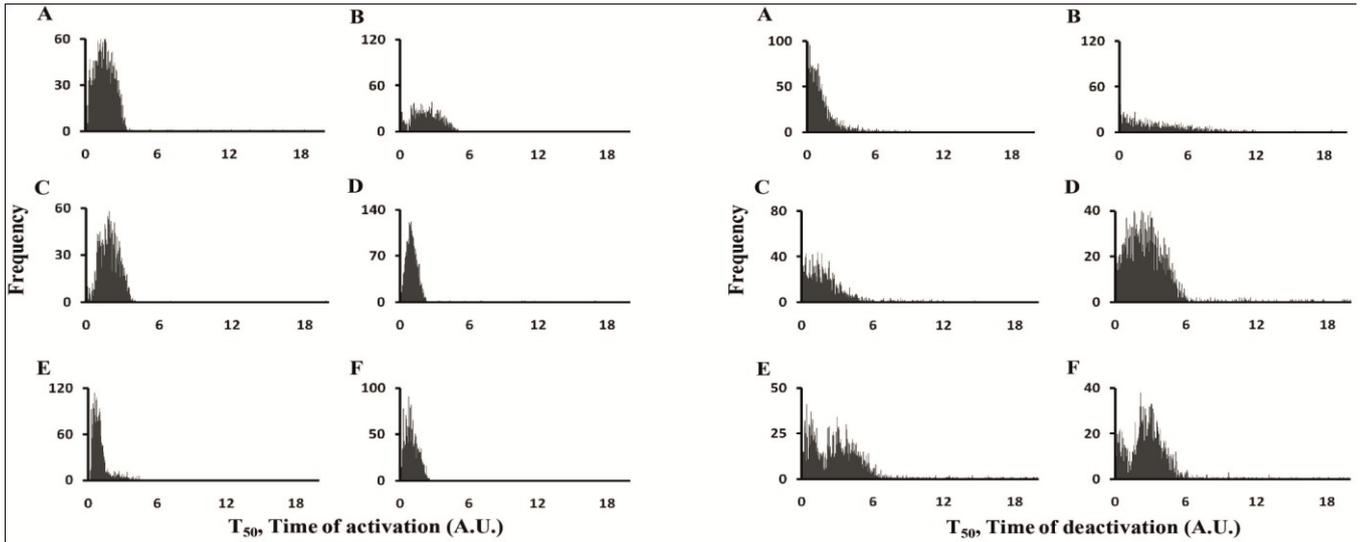


Fig. 5. Time of activation (left panel) and deactivation (right panel) for the networks considered in the present study. Time of activation/deactivation is quantified by the time (in arbitrary units, A.U.) taken to reach the 50% of the total response of the network from one steady state (OFF/ON) to another (ON/OFF). Labels A-F correspond to networks as shown in Fig. 1.

dinJ, *fadR*, and *gntR*; Network E: *ada*, *araC*, *idnR*, *marA*, *cpxR*, and *csgD*; and Network F: *fis*, *galR*, *ihf*, *lexA*, *nac*, and *nagC*. Cells carrying each reporter fusion were grown overnight in LB media at 37°C and the time of activation, time of deactivation, cost of production measured. The following formula was used to measure the cost of production experimentally:

$$Cost = \frac{\Delta RFU / OD(R)}{\Delta RFU / OD(T)} \quad (3)$$

Where RFU corresponds to the fluorescence in a particular samples as compared to a reference. This fluorescence is assumed to be proportional to the protein being controlled by the promoter fusion. OD is a measure of number of cells present in the sample. Hence, RFU/OD is a measure of protein amounts per unit cell. As shown in Fig. 6A, we note that there is considerable variation in the average time of activation for the six networks under consideration. In addition, for each network, the standard deviation across the average of five promoters is different for the six networks.

As shown in the Figure, both these facets are in agreement with the computational predictions as shown in Fig. 5.

In a similar manner, the cost of production of T when the system switches from the OFF to the ON state was calculated. To calculate the cost, the fluorescence from the R reporter and the T reporter were calculated in both ON and OFF state. The cost was then calculated as per equation 3. Again, consistent our simulations, we note that the cost for networks A, B, and C was greater than the cost for D, E, and F. The average cost values and the standard deviation for the five promoters in each of the six networks is as reported in Fig. 6B.

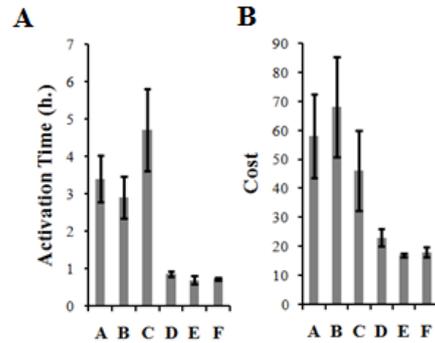


Fig. 6. Average time of activation (A) and cost (B) for the six network designs. The error bars represent the standard deviation across the five promoters in each network group.

4. CONCLUSIONS

In this work, we hypothesize that the choice of a network to accomplish a simple regulatory task (i.e. regulation of a target gene T by a regulator R) is based on several considerations. These include the range of dynamic response, time of response, cost of response, and the susceptibility to response in case of a mutation. In addition to the factors considered above, additional factors like cellular heterogeneity (both transient and steady-state), cell-to-cell variability also likely impact the choice of a particular network topology in a regulatory network. To fully understand the complexity of choice of a network design in a regulatory interaction, we speculate that each network will have to be studied in conjunction with the specific physiological role that it plays. Future experimental work in this direction is likely to develop an enhanced understanding of evolution of network structures.

REFERENCES

- Dekel, E., Mangan, S., and Alon, U. (2005). Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Physical Biology*, 2 (2), 81-8.
- Fraser, J.S., Gross, J.D., and Krogan, N.J. (2013). From Systems to Structure: bridging networks and mechanism. *Molecular Cell*, 49 (2), 222-31.
- Gillespie, D.T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22, 403-34.
- Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81, 2340-61.
- Goentoro, L., Shovai, O., Kirschner, M.W., and Alon, U. (2009). The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular Cell*, 36 (5), 894-9.
- Kaplan, S., Bren, A., Dekel, E., and Alon, U. (2008). The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Molecular Systems Biology*, Article number 203; doi:10.1038/msb.2008.43.
- Kumar, M., Gayen, K., and Saini, S. (2013). Role of extracellular cues to trigger the metabolic phase shifting from acidogenesis to solventogenesis in *Clostridium acetobutylicum*. *Bioresource Technology*, 138, 55-62.
- Mangan, S. and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America*. 100 (21), 11980-5.
- Miller, W.G., Leveau, J.H., and Lindow, S.E. (2000). Improved gfp and inaZ broad-host-range promoter-probe vectors. *Molecular Plant-Microbe Interactions: MMPI*, 13 (11), 1243-50.
- Mitrophanov, A., Y., Groisman, E., A. (2008). Positive feedback in cellular control systems. *Bioessays*, 30 (6), 542-555.
- Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. (2005). Gene regulation at the single-cell level. *Science*, 307(5717), 1962-5.
- Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. (2007). Accurate prediction of gene feedback circuit behavior from component properties. *Molecular Systems Biology*, 3, 143.
- Saini, S., Ellermeier, J.R., Slauch, J.M., and Rao, C.V. (2010). The role of coupled positive feedback in expression of the SPI1 type three secretion system in *Salmonella*. *PLoS Pathogens*, 6 (7): e1001025. doi:10.1371/journal.ppat.1001025
- Sadamoto, H., and Muto, H. (2013). Fluorescence Cross-correlation Spectroscopy (FCCS) to Observe Dimerization of Transcription Factors in Living Cells. *Methods in Molecular Biology*, 977, 229-41.
- Salgado, H., Peralta-Gil, M., Gama-Castro, S., et al. (2013). RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acid Research*, 41 (Database issue), D203-13.
- Sneppen, K., Krishna, S., Semsey, S. (2010). Simplified models of biological networks. *Annu. Rev. Biophys.*, 39, 43-59.
- Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M., and Elowitz, M. B. (2006). An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440 (7083), 545-50.
- Yu, Z., Reichheld, S. E., Savchenko, A., Parkinson, J., and Davidson, A. R. (2010). A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *Journal of Molecular Biology*, 400(4), 847-64.

Appendix A

Network-B

$$\frac{d(R/kdR)}{dt} = c \times s + \frac{p \times R^2}{q^2 + R^2} - R \quad \frac{d(T/kdT)}{dt} = \frac{a \times R^2}{b^2 + R^2} - T$$

Network-C

$$\frac{d(R/kdR)}{dt} = c \times s + \frac{p}{1 + (R/q)^2} - R \quad \frac{d(T/kdT)}{dt} = \frac{a \times R^2}{b^2 + R^2} - T$$

Network-D

$$\frac{d(R/kdR)}{dt} = c \times s - R \quad \frac{d(T/kdT)}{dt} = \frac{a}{1 + (R/b)^2} - T$$

Network-E

$$\frac{d(R/kdR)}{dt} = c \times s + \frac{p \times R^2}{q^2 + R^2} - R \quad \frac{d(T/kdT)}{dt} = \frac{a}{1 + (R/b)^2} - T$$

Network-F

$$\frac{d(R/kdR)}{dt} = c \times s + \frac{p}{1 + (R/q)^2} - R \quad \frac{d(T/kdT)}{dt} = \frac{a}{1 + (R/b)^2} - T$$

Appendix B

The variables a , b , p , q and c are chosen from the uniformly distributed range of [0,100].

The variables a , b , p , q and $c \in [0.0005, 25, 50, 75, 100]$ for networks B, C, E, F.

The variables a , b and $c \in [0.0005, 7.1405, 14.2805, 21.4205, 28.5605, 35.7005, 42.8405, 49.9805, 57.1205, 64.2605, 71.4005, 78.5405, 85.6805, 92.8205, 99.9605]$ for networks A and D.

The parameter values of the variables are based on the representative physiological range observed in a number of studies conducted with bacterial promoters and transcription regulation (Mitrophanov et al., 2008; Rosenfeld et al., 2005; Rosenfeld et al., 2007; Sneppen et al., 2010; Süel et al., 2006).