

A Gene Regulatory Network Design Of A Synchronous Single-Input Delay Flip-Flop

Imad Hoteit¹, Nawwaf Kharma² and Luc Varin³

^{1,2}Electrical & Computer Engineering Dept., Concordia University, Montreal (QC), Canada

³Biology Department, Concordia University, Montreal (QC), Canada

¹i_hoteit@ece.concordia.ca, ²kharma@ece.concordia.ca, ³varin@alcor.concordia.ca

Abstract

We present a fully detailed design of the very first *synchronous single-input* delay flip-flop (or BioD) implemented as a gene regulatory network in *Escherichia coli* (*E. coli*). The device has one data input (trans-acting RNA), one clock input (far-red light) and an output that reports the state of the device using green fluorescent protein (GFP). The proposed (simulated but not synthesized) device builds on the toggle switch of (Gardner et al., 2000) to provide a more sophisticated device that can be synchronized with other devices in/out of the same cell, and which requires only one input. We provide the first results of a deterministic simulation of a mathematical model of the new device, one which provides evidence that the device is likely to work as required when actually synthesized.

Introduction

The complex processes that take place in a cell are governed by gene expression which is regulated at several levels during the pathway leading from DNA to protein. Apart from the regulation at the DNA level, gene expression may be regulated during transcription, post-transcription, translation, and during post-translational modification of proteins. Notably, much of the control of gene expression is done either by the regulatory proteins or by mRNAs which are essentially the products of other genes. Hence, the interactions between DNA, RNA, proteins, and other molecules, form a gene regulatory network (GRN). While examining these components individually has provided invaluable information, it is essential (a) to thoroughly investigate these components in variable environments and/or performing variable functions, and (b) to integrate this knowledge to generate valuable genetic devices. Here comes the role of synthetic biology that aims at systematically designing, building, combining and testing new biological functions and systems that do not occur in nature. Indeed individual parts such as promoters and transcription factors can be assembled to synthesize GRNs that perform desired functionalities, such as computing machines.

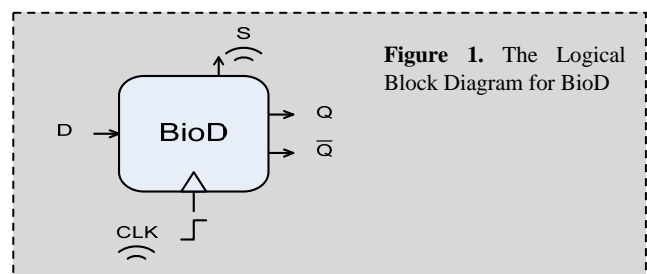
The synthesis of computing machines via the manipulation of DNA (within or without living organisms) started in 1994 when Adleman executed an experimental procedure that used DNA, in vitro, to solve an instance of the directed Hamiltonian path problem (Adleman, 1994). In contrast, in vivo cell-based or cellular computing started in 1998 with the

modification of the genome of prokaryotic cells (typically *E. coli*) to realize one- and two-input combinatorial Boolean logic gates (e.g. NOT, AND and IMPLIES) (Knight, Jr. and Sussman, 1998; Weiss et al., 1998); and a similar feat recently was achieved with eukaryotic cells by Kramer (Kramer et al., 2004). Along another dimension, time-dependant or sequential Boolean logic devices have also been implemented in living cells, starting most notably with a 2-input toggle switch by Gardner (Gardner et al., 2000), and a synthetic oscillator by Elowitz (Elowitz and Leibler, 2000). In fact, in one decade this field has grown to generate many elementary devices (Drubin et al., 2007; Boyle and Silver, 2009; Tigges et al., 2009; Haynes and Silver, 2009), including band-pass filters (Stricker et al., 2008) and counters (Friedland et al., 2009). More complicated devices like engineered multicellular pattern generators (Basu et al., 2005), single cell biosensors (Levskaya et al., 2005; Tecon et al., 2006), tumor-targeting bacteria (Anderson et al., 2006), and cell-based computers (Cox, III et al., 2007; Balagadde et al., 2008) have also been built or proposed.

Despite the numerous works on genetic switches, all proposed designs work *asynchronously*. This means that the switch's operation cannot be synchronized with the operation of other parts, using a single global clock. Henceforth, we call a *synchronous single-input delay* switch a *BioD*; a novel GRN that changes states in response to a clock signal by having its output expression follow its input.

Circuit Design and Modeling

BioD is a synthetic *E. coli* cell that expresses a gene regulatory network acting as a delay switch. By delay switch, we mean a logical device that has an input (D), a clock (CLK), and an output (Q) equal to its state (S); see Figure 1 (\bar{Q} is the



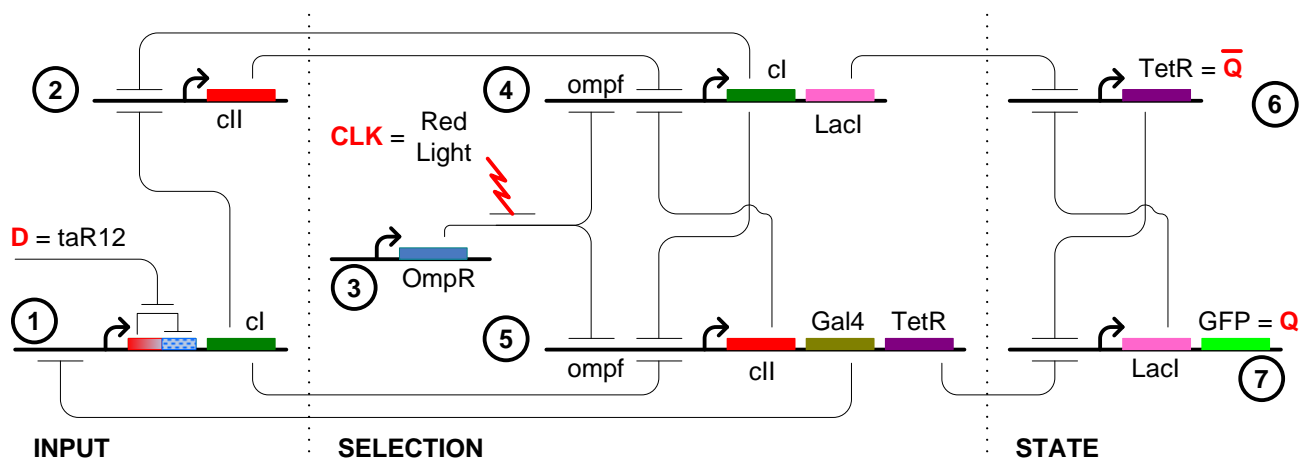


Figure 2. Gene regulatory network for BioD. The network consists of three sections. *STATE* reflects the state of the network. *SELECTION* affects the *state* switch when the far-red light signal is ON. *INPUT* drives the *selection genes*' activation.

second output and is equal to the logical complement of Q). The state of a delay switch is held constant unless and until its input differs from its state, on the rising edge of the clock. In that case, the next state of the delay switch will copy the value of the input (i.e., $Q = D$). Hence, a cell that acts as a delay switch is effectively a 1-bit memory device, controlled by an input and a clock. The BioD also exhibits its state by expressing (or not) a fluorescent protein.

BioD

BioD has two inputs: *trans-activating* RNA or *ta*RNA as input D , and the presence or absence of far-red light as the clock (CLK). It has two complementary outputs (Q and \bar{Q}) defining the state of the flip-flop: the ON state is indicated by the presence of green fluorescence, while the opposite OFF state

is indicated by its absence. As with its electronic counterpart, the output follows the input on the rising edge of the clock. The gene network is comprised of three parts: *input genes*, *state genes* and *selection genes* (as shown in Figure 2).

Input Genes. The *input genes* convey to the *selection genes* whether an input signal is present or not. They do so by tipping the balance of the dual-repression of the *selection genes* – discussed below.

In order to sense input D , gene 1 is designed to be self-repressed, but in such a manner that can only be induced by D . To achieve this, a form of ribo-regulation is used called *cis*-regulation – which means “acting from the same molecule”. The *cis*-regulation or in our case, *cis*-repression prevents the translation of the gene 1 transcripts by causing them to bend and cover the ribosome binding site (RBS) like a

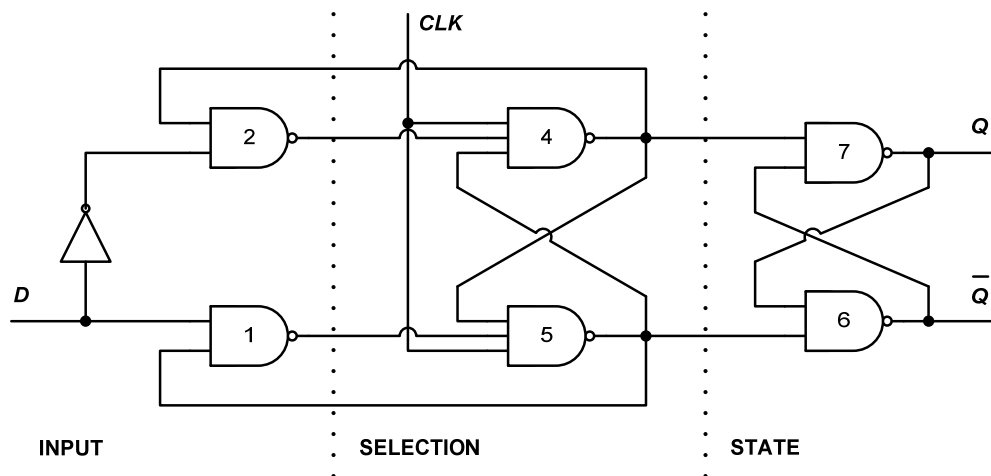


Figure 3. Logic diagram of *BioD* (provided here for simplicity). The above circuit behaves much like the GRN in Figure 2. It is not an exact representation of course, but helps follow the steps the circuit takes to change states. Gene numbers above are matched to gate numbers here. A low CLK signal neutralizes the selection gates 4 and 5, and sends a high signal (or identity for NAND gates) to the state gates 6 and 7; keeping them unchanged. Since the outputs of input gates 1 and 2 are complements, when the CLK signal is turned ON, only one of gates 4 or 5 becomes active and thus (i) affects one of the state gates (6 or 7) and (ii) disables its enabling input gate (gate 1 or 2). The input gates are re-enabled after the CLK goes low, leaving them free to respond to input D .

lock. The key, comes in the form of trans-activating RNA (taRNA) which, when matched with the *cis*-repressed RNA, unlocks the RBS allowing translation (Isaacs et al., 2004). The taRNA chosen for input D is τ_{aR12} which unlocks the *cis*-repression we introduced in gene 1, namely $crR12$.

When the input D is present, the transcripts of gene 1 get translated into cI proteins (from the λ phage). cI in turn represses gene 2. In the absence of input D however, the *cis*-repressed transcripts of gene 1 do not get translated into proteins, lifting the repression of gene 2 and allowing its expression.

The presence of input D results in the production of the cI protein, while its absence results in the production of the cII protein (from the P22 phage).

State Genes. The *state genes* are very similar to Gardner's toggle switch (Gardner et al., 2000). They consist of two co-repressed genes (i.e. only one expressed at a time), and as such define the state of the BioD device. Genes 7 and 6 represent the complementary outputs Q and \bar{Q} respectively. A green fluorescent protein (GFP) signals the output Q , while its absence signals the complementary output \bar{Q} . The co-repressed nature of the toggle switch means that when either gene is active, it enters into a stable state where it represses the other, and insures its own continued expression. In our case, that stable state can only be affected by the *selection genes*.

As can be seen in Figure 2, the *selection genes* can affect the *state genes* independently of the current state of the BioD. As will be discussed below, genes 4 and 5 are mutually exclusive when active; protecting the *state genes* from conflicting signals. Furthermore, they will either reinforce the repression currently in place in the *state genes* (resulting in no state change), or they will repress the presently dominant gene until the balance is tipped, and the other takes over the state of the device. Which of the two genes 4 or 5 is activated depends on the *input genes* at the time the CLK signal is turned ON.

Selection Genes. The *selection genes* are always OFF until turned ON by far-red light (the CLK input). In the absence of far-red light, genes 4 and 5 are always repressed by the phosphorylated version of $OmpR$, i.e. $OmpRP$. Gene 3 is constitutively expressed and produces $OmpR$. $OmpR$ is phosphorylated in the presence of the $EnvZ$ enzyme. $EnvZ$ is connected to $Cph1$, which in the presence of far-red light, induces a conformational change in $EnvZ$ preventing the phosphorylation of $OmpR$. The genes that produce $EnvZ$ and $Cph1$ (and a few others needed for the light response system (Levskaya et al., 2005)) are not shown in Figure 2.

The phosphorylation of $OmpR$ is dominant in the absence of far-red light and negligible in its presence. Therefore, the far-red light signal causes a drop in $OmpRP$ levels and a corresponding rise in $OmpR$ levels. This drop affects genes 4 and 5 using their promoter, as $ompF$ is both activated by $OmpR$ and repressed by $OmpRP$. Both the functionality of $ompF$ and the complementary levels of $OmpR$ and $OmpRP$ result in a system that is quick to start or stop transcription in both genes 4 and 5.

The *selection genes* also respond to and affect the *input genes*. As previously mentioned, BioD is an edge-triggered device, i.e. it responds to the input when the CLK signal turns

on, but not to a change in the input when the CLK signal is on. This is achieved by designing genes 4 and 5 to only be turned off by the CLK signal. When far-red light is introduced, and one of genes 4 or 5 turns on, that gene immediately starts repressing the genes that can potentially repress it; namely, gene 4 represses genes 2 and 5, and gene 5 represses genes 1 and 4. As a result, any change in the input D when the CLK signal is already on, does not translate to the *selection genes* until the CLK signal is turned off, and the repression of the *input genes* is lifted.

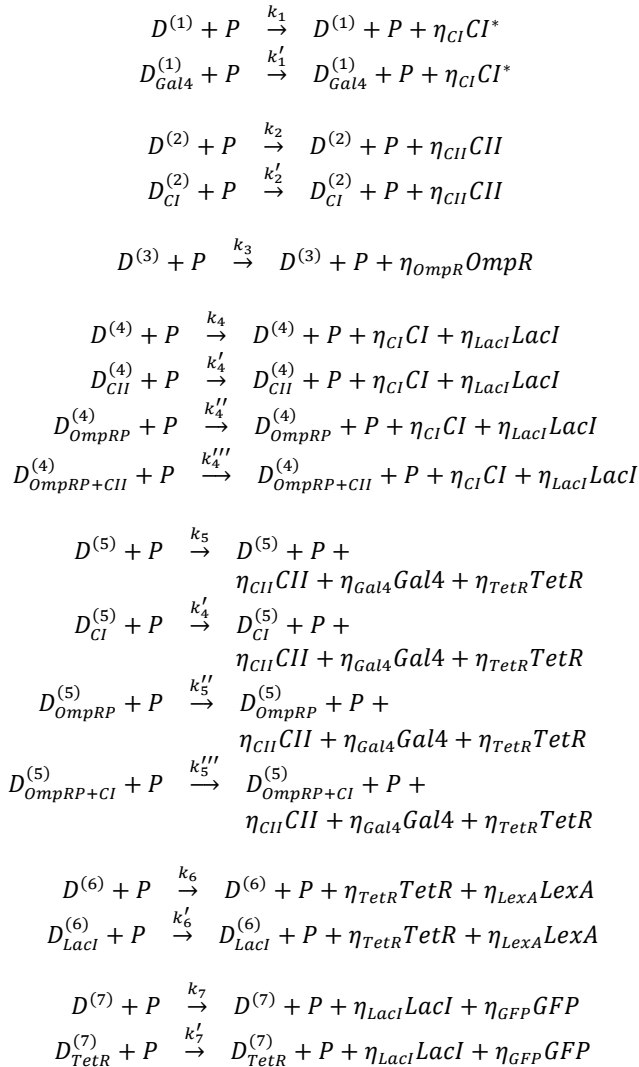
Given that the dynamics of such a gene network are non-trivial, we provide a single fully detailed scenario tracing through one important sequence of transitions. The scenario is that of a change of state, from OFF to ON, in response to a turned ON input (D), whose level must stabilize, prior to the introduction of the CLK signal (far-red light clocking). When the state of the BioD is OFF, gene 6 is ON, expressing two products. Since one of them ($TetR$) is repressing gene 7, gene 7 is considered OFF. In the absence of red light, the constitutively expressed (and subsequently phosphorylated) repressor ($OmpRP$) blocks any production from the *selection genes* (4 and 5). Hence, the status quo of the *state genes* is maintained. Lastly gene 1 is ON, induced by the input (D), while gene 2 is OFF, repressed by the product of gene 1, namely cI . After clocking, the concentration of $OmpRP$ (which was repressing genes 4 and 5) starts falling. The only other repressor of gene 4 (i.e. cII from gene 2) is already OFF. So gene 4 can start producing, and as such, it starts repressing gene 5, which is still repressed by cI from gene 1. At this stage, gene 1 is ON, gene 2 is OFF, gene 4 is ON, gene 5 is OFF, while gene 6 is still ON and gene 7 is still OFF. Turning our attention to gene 4, note that one of the repressors it produces is identical to the one generated by gene 7, namely $LacI$. Its production starts switching off gene 6, resulting in a gradual increase in the expression of gene 7. Once gene 7 is fully expressed, it represses gene 6 (via its own $LacI$ protein), ensuring the continuation of gene 7's new ON state. Hence, we have achieved a network change of state (indicated by GFP) from OFF to ON (following the value of the input (D)). For as long as the CLK signal is ON, the new state is maintained. If a significant change in the input level occurs while the clock is ON, the repressions of genes 2 and 5 would not disappear, since gene 4 is ON and produces cI . Indeed, as long as gene 4 is ON, it has the ability to keep itself from being repressed by other genes, that is, by repressing them. It is only when the CLK signal is removed and both genes 4 and 5 are OFF that the system is free to respond to the input (D) again.

Model

The network in Figure 2 is simulated deterministically. The fast reactions (not shown) involve the binding of proteins to one another and to the DNA. The slow reactions (shown below) involve transcription of mRNA and translation of proteins. The important reactions are presented here as a single combined process.

We define the following terms and chemical species: $D^{(n)}$, the DNA protein-binding site in the promoter of gene n ; $D_X^{(n)}$, the $D^{(n)}$ bound by repressor/activator X ; P , RNA polymerase; k_n , rate of production of gene n (promoter strength); $k_n^{[1]}$,

effective production rate of gene n after repression/activation; η_X , number of proteins molecules per transcript of gene X .



where cI^* depicts the cI protein that is produced by the *cis*-repressed transcripts of gene 1 – therefore is dependent on input D .

We simulate the device using a system of rate equations with the concentrations as the dynamical variables. A timing diagram is displayed in Figure 4.

Results & Discussion

We presented the design of a gene regulatory network (GRN) that, if synthesized and integrated into the genome of an appropriate strain of *E. coli*, will give us a single-cell synchronous single-input toggle switch (we call the *BioD*). The *BioD* accepts as input trans-acting RNA, which allows it be linked to other GRNs; it is clocked using far-red light, which allows external synchronization of its operation; it indicates its state by expressing green fluorescent protein (GFP), which allows easy external monitoring of the state.

The whole design is modular in that it allows alteration of the input sensing and output expression parts without affecting the toggling functionality of the device.

In the sequel, we present the results of simulating the device using a system of rate equations. The results confirm our expectation that the device will toggle when and only when required – though its speed can still be improved.

Simulation

The core functionality of our *BioD* device is illustrated in Figure 4. The highlighted areas indicate the presence of an input. The reddish hue reflects the presence of the clock input (CLK), while the grey diagonal pattern reflects the presence of the data input (D). The examples provided have two different data cycles intersecting (or not) with four different clock cycles. This setting allows us to show that the device can indeed go from one state to the other with nothing more than the introduction of the inputs it was designed to respond to; in other words, the device does not get stuck in any one state.

Ideally, with four separate CLK inputs, the state of the device should follow the D input four times. In this case, the state should turn ON, then OFF, and then OFF again and finally ON. Figure 4a displays those exact state changes in a deterministic run whose initial condition is an OFF state. The normalized GFP expression output follows the input only at the rising edge of the clock. However while the clock *is* ON or *is* OFF, any changes in the input do not propagate to the output. This plot is used to demonstrate the overall input/output relationship. Figure 4b shows the changes in the protein levels – here the levels of $LexA$ and GFP were not displayed because they do not affect the behavior of the device.

In the ON level, the expression of a substance is defined mainly by its rates of synthesis and degradation. As expected, some proteins have multiple stable levels of expression. Since cI , cII , $LacI$ and $TetR$ are not only produced in the *selection genes*, but can also be found in either the *input* or *state genes*, the expression of those proteins is significantly increased with the presence of the CLK signal. $TetR$ has four levels of expression: (i) OFF, (ii) gene 6 is ON, (iii) gene 5 is ON, and (iv) genes 5 and 6 are ON. $LacI$ has similar levels of expression using genes 4 and 7. In the case of cI however, since gene 4 can only turn on if gene 1 is active, it only has three levels of expression (and similarly for cII).

Tracing the various signals in Figure 4b shows that, the simulation starts with two active proteins, $TetR$ (the state of the device is OFF) and cII (unrepressed since input D is OFF). Here is a step-by-step explanation of the changes seen in the timing diagram.

First, input D is introduced, causing the repression of gene 2 (or cII) to start. Since gene 1's transcripts are now translated and gene 2 is OFF, gene 4 becomes on an edge-trigger to be turned ON, while gene 5 is doubly repressed by $OmpRP$ and now by cI . The CLK signal is introduced, stopping the phosphorylation of $OmpR$ and activating gene 4. This raises the levels of cI and $LacI$. The latter represses gene 6 and starts turning the state of the device ON. As $TetR$ fades away, the GFP levels start climbing. Then the CLK signal is turned OFF followed by the input D . These two actions turn OFF gene 4 and disable gene 1 respectively. With both inputs OFF, the cI repressors produced by genes 1 and 4

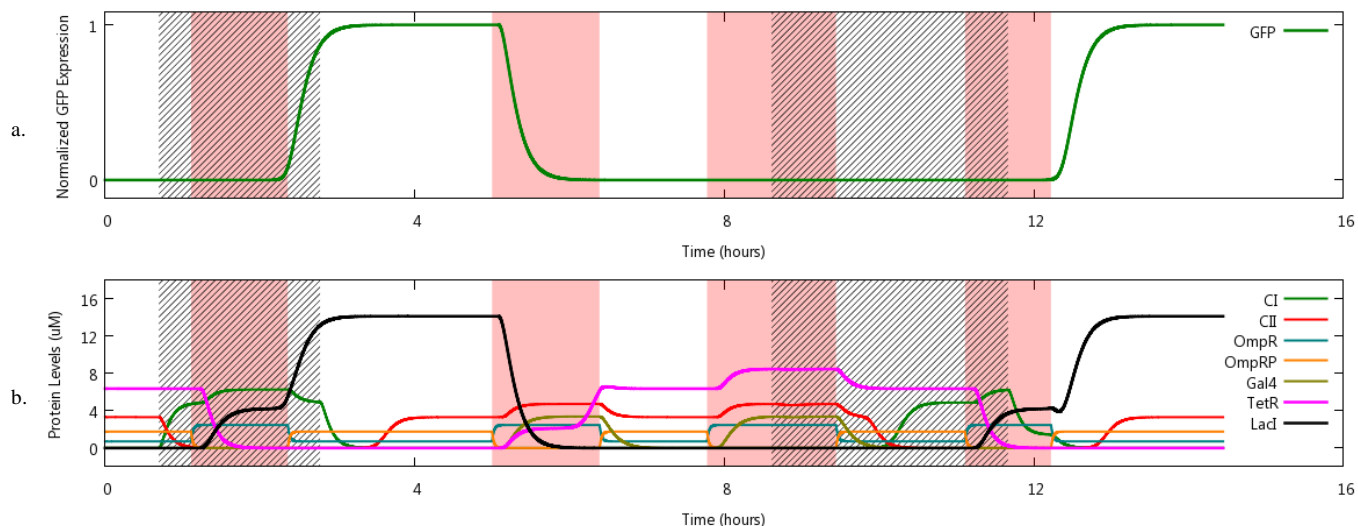


Figure 4. Deterministic simulation of *BioD*. The two timing diagrams are displaying different signals of the same run. The highlighted areas indicate the presence of an input. The red hue indicates the *CLK* signal (FR light). The grey diagonal pattern indicates the presence of the input *D*. **a.** Normalized GFP expression **b.** Protein levels

degrade without replacement allowing *cII* to return to its previous level. *LacI* which is now produced by gene 7 reaches its un-repressed (ON) state equilibrium.

The second state change occurs when the *CLK* signal is turned on again. Since *cII* is expressed at that time (no input *D*), gene 5 turns ON, causing the repression of gene 1 (through *Gal4*) and the repression of gene 7 (through *TetR*), and raising the level of *cII* (as it is produced by both genes 2 and 5). When the *CLK* is removed, gene 5 is turned OFF, but *cII* and *TetR* remain high, while *Gal4* is repressed. Note that the *TetR* levels are now produced by gene 6 (which took over the state of the toggle switch from gene 7), and no longer by gene 5.

The third *CLK* signal starts now. Gene 5 is again turned ON; the levels of *cII*, *Gal4* and *TetR* climb. In the middle of the *CLK* pulse, the input *D* is introduced. This causes no change in the network. Since input *D* only affects gene 1, its effects were muzzled because the clock had already turned on gene 5 which repressed gene 1. It is only after the clock is turned OFF that the gene 1 repression is lifted. At this point, even though the *CLK* signal is removed, the input *D* is still present, and since gene 1 is no longer repressed by gene 5 (or *Gal4*), *cI* is translated and represses gene 2. The state of the device however does not change since the *state genes* are not directly affected by the *input genes*.

The fourth *CLK* signal turns the state of the device back ON. In the presence of input *D*, the *CLK* turns gene 4 ON causing a similar sequence of events witnessed following the first *CLK* signal.

A Note Regarding Frequency

The frequency of operation of the *BioD*, that is to say the frequency at which the device can change its state in response to the input is closely related to the genes used to build the network. Indeed, while the design of the *BioD* allows for the use of other genes than the ones presented in this paper,

different genes do have different properties modeled by different synthesis rates, degradation rates, diffusion rates, and promoter/repressor dissociation constants, to name a few. All of these parameters indirectly control the time it takes for the system to respond to an input change, and the time it takes to finish a state change and reach a steady state.

In our case, and when going from an OFF to ON to OFF state, the *CLK* signal had to be sustained for at least 22 minutes to get a sustained state change, while it had to be removed for at least 64 minutes for the network to regain its steady state. That gave the smallest period (or max. frequency) of approximately 86 minutes (5160 seconds).

A Note Regarding Speed

Speed is a main area of improvement. Indeed, the slowest reactions in a cell are the ones involving repressors and ultimately their transcription and translation. The time it takes to fulfill these operations depends on the promoter strength, the coding sequence that is being transcribed/translated, and the presence of RNA polymerases and/or ribosomes nearby. The impact of repressors is further delayed until the mature protein manages to hit the proper operator site, at the right angle and speed. Using post transcriptional regulation like *taRNA* or RNA interference (RNAi) where possible to effect the state change in *BioD* will make the system significantly faster. The first such place would be where the *selection genes* interact with the *state genes*. Instead of producing repressors for genes 6 or 7, the use of RNAi to prevent one of them from translating repressor proteins would make the entire system significantly faster. Since we already make use of *taR12* to sense the input, we would therefore need another two independent riboregulators that do not interfere with *taR12* or with each other.

Conclusion & Extension

In this paper, we sought a proof of concept for the first synchronous single-input delay flip-flop implemented as a gene regulatory network in *E. coli*. The simulation we present provides evidence that the device can toggle from the ON state to the OFF state and back, according to the intended functionality. The inherent symmetry of the design reduces the number of genes needed for the device, but introduces some complexity (which is palpable when tracing the various changes the device goes through when toggling).

The BioD is effectively a 1-bit memory element that can operate synchronously (on a clock) with any number of other elements. As such, it can be used to hold the state of a finite state machine, but it could also be used to build a memory bank, an event sequence detector/effector, a decision-making system, and numerous other memory-requiring devices.

References

- Aleman, L.M. (1994). Molecular computation of solutions to combinatorial problems. *Science* 266, 1021-1024.
- Anderson, J.C., Clarke, E.J., Arkin, A.P., and Voigt, C.A. (2006). Environmentally controlled invasion of cancer cells by engineered bacteria. *J. Mol. Biol.* 355, 619-627.
- Balagadde, F.K., Song, H., Ozaki, J., Collins, C.H., Barnet, M., Arnold, F.H., Quake, S.R., and You, L. (2008). A synthetic *Escherichia coli* predator-prey ecosystem. *Mol. Syst. Biol.* 4, 187.
- Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H., and Weiss, R. (2005). A synthetic multicellular system for programmed pattern formation. *Nature* 434, 1130-1134.
- Boyle, P.M. and Silver, P.A. (2009). Harnessing nature's toolbox: regulatory elements for synthetic biology. *J. R. Soc. Interface* 6 *Suppl* 4, S535-S546.
- Cox, R.S., III, Surette, M.G., and Elowitz, M.B. (2007). Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* 3, 145.
- Drubin, D.A., Way, J.C., and Silver, P.A. (2007). Designing biological systems. *Genes Dev.* 21, 242-254.
- Elowitz, M.B. and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335-338.
- Friedland, A.E., Lu, T.K., Wang, X., Shi, D., Church, G., and Collins, J.J. (2009). Synthetic gene networks that count. *Science* 324, 1199-1202.
- Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339-342.
- Haynes, K.A. and Silver, P.A. (2009). Eukaryotic systems broaden the scope of synthetic biology. *J. Cell Biol.* 187, 589-596.
- Isaacs, F.J., Dwyer, D.J., Ding, C., Pervouchine, D.D., Cantor, C.R., and Collins, J.J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* 22, 841-847.
- Knight, T. F., Jr. and Sussman, G. J. Cellular gate technology. UMC98: First International Conference On Unconventional Models Of Computation. [1], 257-272. 1-1-1998.
- Kramer, B.P., Fischer, C., and Fussenegger, M. (2004). BioLogic gates enable logical transcription control in mammalian cells. *Biotechnol. Bioeng.* 87, 478-484.
- Levskaya, A., Chevalier, A.A., Tabor, J.J., Simpson, Z.B., Lavery, L.A., Levy, M., Davidson, E.A., Scouras, A., Ellington, A.D., Marcotte, E.M., and Voigt, C.A. (2005). Synthetic biology: engineering *Escherichia coli* to see light. *Nature* 438, 441-442.
- Stricker, J., Cookson, S., Bennett, M.R., Mather, W.H., Tsimring, L.S., and Hasty, J. (2008). A fast, robust and tunable synthetic gene oscillator. *Nature* 456, 516-519.
- Tecon, R., Wells, M., and van der Meer, J.R. (2006). A new green fluorescent protein-based bacterial biosensor for analysing phenanthrene fluxes. *Environ. Microbiol.* 8, 697-708.
- Tigges, M., Marquez-Lago, T.T., Stelling, J., and Fussenegger, M. (2009). A tunable synthetic mammalian oscillator. *Nature* 457, 309-312.
- Weiss, R., Homsy, G., and Nagpal, R. Programming biological cells. Eighth International Conference on Architectural Support for Programming Languages and Operating Systems. Wild and Crazy Ideas Session [8]. 1998.