

Modeling transcriptional factor cross-talk to understand parabolic kinetics, bimodal gene expression and retroactivity in biosensor design

Hannah Aris^{1†}, Shayan Borhani^{1†}, Devorah Cahn^{1†}, Colleen O'Donnell^{1†}, Elizabeth
Tan^{1†}, and Peng Xu^{1*}

¹Department of Chemical, Biochemical and Environmental Engineering, University of Maryland
Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250

[†]These authors contributed equally to this article and their names are listed in
alphabetical order.

*Corresponding author email: pengxu@umbc.edu; phone: +1-410-455-2474

Abstract

Transcriptional factor-based biosensor has been widely used to reprogram and adjust cellular activity to the changing environment. These biosensors translate an internal cellular signal to a transcriptional reporter output. Previous examples have demonstrated transcriptional factor cross-talk could lead to complex gene expression dynamics. In this report, we formulated a mechanism-based kinetic model to simulate and predict how TF cross-talk reshape the transcriptional dynamics of an engineered biosensor. Our model comprises TF cross-talk and Hill-type equation to quantify the degree of gene repression and de-repression; it also accounts for protein degradation, cell growth (dilution) effect and the carrying capacity of the system. Our simulation fits well with the biphasic parabolic gene expression pattern of an experimentally validated malonyl-CoA sensor. We discovered that exponential growth could lead to hysteresis in the TF cross-talk model. With Logistic growth to limit the carrying capacity, we find that bimodal gene expression pattern could arise, and this phenomenon is rooted in the hysteresis characteristics of the TF cross-talk model. The computational insights obtained in this study will guide us to design accurate, sensitive TF-based biosensors and may serve as a diagnostic platform to troubleshoot the complex transcriptional dynamics in biosensor design. The computational framework developed here will also provide an educational toolbox for synthetic biologist and biochemical/biomolecular engineering students to understand biosystem design and transcriptional regulation.

Keywords: malonyl-CoA sensor, transcriptional factor cross-talk, FapR, kinetic model, hysteresis, bimodal gene expression

Introduction

Biosensors are indispensable tools to detect or respond to a specific biochemical signal [1-3]. These sensors have been used to report environmental pollutants or diagnose disease markers. Most of the biosensors are constructed on the basis of biomolecular interactions, including protein-ligand binding (i.e. enzyme-substrate) [4, 5], protein-protein binding (i.e. immunological interaction or GPCR receptor) [6], protein-DNA-RNAP (i.e. transcriptional regulation) [7], RNA-RNA (i.e. riboregulators and toehold switches) [8-11] and DNA/RNA-ligand (i.e. aptamers) interactions [12, 13].

Transcriptional factor (TF) based biosensors consist of both repressor or activator proteins regulating the transcriptional activity of a specific promoter. These TF-based sensors contain a N-terminal DNA-binding domain (DBD) and a C-terminal ligand-binding domain (LBD). The DBD of a transcriptional factory typically interacts with a cis-regulatory DNA sequence (generally called operator or enhancer) to restrict or enhance the formation of RNA polymerase (RNAP)-promoter complex. For instance, a repressor binds to the operator and prevents RNAP proceeding forward to decrease transcription; on the opposite, an activator binds to the enhancer and promotes the formation of more stable RNAP-promoter complex to increase transcription [14]. Apart from the DNA-binding domain (DBD), the C-terminal ligand-binding domain forms the sensing parts that can respond to metabolite (small molecule) concentration change and environmental stress signals (salt, osmotic pressure, pH, oxygen, redox, light or radiation *et al*). This unique structure allows the TF-based biosensor to transduce a C-terminal ligand-binding activity to the N-terminal DNA-binding activity. Upon interaction with a small molecule or environmental stress signal, TFs will undergo a conformational change to decrease or increase its DNA-binding affinity. The occupancy of TF binding states will instead affect the transcriptional activity of the RNAP-promoter complex. To detect the transcriptional activity, the RNAP is designed to drive (actuate) the expression of a reporter protein that outputs an easily measured optical signal (absorbance,

fluorescence or luminescence) [15]. In principle, small molecule or environmental stimuli input will form a quantitative dose-response curve with the reporter output. From a structural perspective, the C-terminal ligand-binding domain of TFs determines the specificity of the input-output relationship, while the N-terminal DNA-binding domain of TFs determines the leakiness (background noise) and sensitivity of the input-output relationship, the promoter strength determines the dynamic response range of the dose-response curve.

Malonyl-CoA is an essential building block molecule involved in the biosynthesis of a range of important molecules, including fatty acids [16-19], phenylpropanoids [20], flavonoids [21], and non-ribosomal polyketides [22-24]. As a result, several genetically-encoded malonyl-CoA sensors have been developed in *E. coli* [15, 16, 25], *S. cerevisiae* [26-28] and mammalian cells [29] for sensing or regulating the level of malonyl-CoA inside the cell for various biotechnological applications. These malonyl-CoA sensors were built with the transcriptional repressor *FapR*, the primary regulator protein controlling the gene expression of fatty acid biosynthetic pathway in gram-positive bacteria [30, 31]. Taking the mostly characterized FapR as an example, *B. subtilis* FapR contains two domains: the C-terminal ligand-binding domain specifically recognizes malonyl-CoA, and the N-terminal DNA-binding domain specifically binds with the DNA sequence *fapO* (27bp, TAGAATTAGTACCTGATACTAATAATT) [32]. Binding of FapR with *fapO* will block the access of RNAP to transcribe the downstream gene, the process so called repression; malonyl-CoA binds with the C-terminal of FapR and causes FapR dissociate from *fapO* to restore transcription, the process so called de-repression.

Previously, we have developed a malonyl-CoA sensor and observed cross-talk effect between FapR and LacI. Unexpectedly, the engineered malonyl-CoA sensors displayed biphasic gene expression dynamics [15]. In this study, we formulated a mechanistic biophysical model to simulate the cross-talk between FapR and LacI repressor, and predict the transcriptional output of the engineered malonyl-CoA sensors. Our model comprises transcriptional factor competitive

inhibition, Hill-type equation to quantify the degree of gene repression and de-repression, it also incorporates protein degradation and cell growth (dilution) effect and the carrying capacity of the system. We modeled the unintended interactions of transcriptional factor cross-talk and our simulation is in good agreement with the biphasic parabolic gene expression pattern of an experimentally validated malonyl-CoA sensor. We expanded this model and elucidated that the bimodal (or double-hump) gene expression pattern is deeply rooted in the hysteresis characteristic of the TF cross-talk model under Logistic growth constraint. We formulated a few design principles underlying transcriptional factor-based sensor design. The computational insights obtained in this study will guide us to design more accurate, sensitive TF-based biosensors and will serve as a diagnostic platform to understand the complex transcriptional dynamics of the engineered sensor. The computational framework developed here will also provide an educational toolbox for synthetic biologist and biochemical/biomolecular engineering students to understand biosystem design and transcriptional regulation.

Results and discussion

A rational explanation of the parabolic kinetics of the malonyl-CoA sensor

One critical factor to control the dynamic response range of FapR-based biosensor is the amount of repressor protein FapR and the saturation constant between FapR and malonyl-CoA. To control the amount of FapR, we previously constructed a genetically encoded malonyl-CoA sensor [15, 25] using a titratable promoter (T7) to drive the expression of FapR, which can be induced by the amount of inducer IPTG present to the cell (Figure 1a). We then inserted the FapR binding site *fapO* (27bp) downstream of the T7 promoter to drive the reporter gene *eGFP* (Figure 1a). Interestingly, instead of observing a linearly increasing or monotonically increasing dynamic response (Figure 1b), we observed a biphasic parabolic type response kinetics (Figure 1c). Our first explanation is that there might be saturation effect due to the limited number of T7 RNAP to

occupy the T7 promoter. It perfectly explains the first phase of the kinetics: increasing the amount of IPTG will rapidly increase eGFP output and the green fluorescence signal reaches a plateau at some timepoint (from 250 min to 300 min). However, this RNAP occupancy and promoter saturation couldn't support the kinetics of the second half of the parabolic curve: the plateau doesn't persist and the eGFP output plummets off with increasing amount of IPTG after 300 min.

The 27bp *fapO* binding site is a pseudo-palindromic (inverted repeat) sequence, which could form an intra-strand self-complementary structure that can specifically bound with FapR. Due to the sequence synergy between *fapO* and *lacO*, we hypothesize that there might be transcription factor cross-talk between the repressor protein FapR and the non-cognate DNA-binding site *lacO*. That said malonyl-CoA repressor protein FapR and lactose repressor protein LacI will compete for the same binding site of *lacO*, in addition to the specific interaction of FapR with *fapO*. This is not counter-intuitive as *fapO* has the DNA sequence 5'TAGAATTAGTACCTGATACTAATAATT3', which shares a large portion of sequence synergy with *lacO* 5'AATTGTGAGCGGATAACAATT3'. Indeed, recent studies have demonstrated the leakiness and stringency of T7-*lacO* system is determined by the exact sequence of the LacI binding site (lactose operator *lacO*)[33].

A model captures the parabolic gene expression dynamics of malonyl-CoA sensor

Synthetic biology community has developed Pigeon as a design visualization tool for complex genetic circuits design [34]. Pigeon uses simple syntax to translate DNA/RNA/protein interaction into a genetic circuit image. Connected to a web-server (<http://pigeoncad.org/>), we input the text descriptions of T7 promoter, DNA-binding site (*fapO* and *lacO*), transcriptional repressor (FapR and LacI) and their interactions, Pigeon automatically output a hieroglyphic image that describes LacI and FapR's competition for the *lacO/fapO* site upstream of the reporter gene eGFP (Figure 2a). Then we deconvolute the interactions and rearrange the image layout to better describe TF cross-talk between FapR and LacI (Figure 2b), for simple display, we omitted the *lacO* site in Figure 2b. This figure could be further simplified using systems biology graphic language [35] for

easy mathematical description and modeling (Figure 2c). This procedure serves an example to teach biochemical/biomolecular engineering students how to abstract the essential information from a specific genetic circuitry and facilitate the understanding of biosystem design with simplified systems/synthetic biology language.

To better understand how transcriptional factor cross-talk reshapes the transcriptional dynamics, we built a biophysical model that accounts for TF competitive binding, gene expression repression and derepression, protein degradation and cell dilution effect. A detailed explanation of the equation parameters can be found in Table 1.

Without considering cell growth (i.e. in cell-free systems), we use a first-order exponential decay to describe the production and degradation of the repressor protein LacI (y1), as described in Equation 1. This is consistent with the fact that LacI is coupled with cell growth and its expression is controlled by a constitutive promoter in most *E. coli* expression system. Similarly, a Hill-type inhibition kinetics was used to describe the net production rate of the repressor FapR (y2), which is repressed by the amount of repressor LacI (y1) and this repression could be removed (or antagonized) by the amount of the inducer IPTG (y4), as described in Equation 2. A competitive Hill-type inhibition kinetics was used to describe the net production rate of the reporter protein GFP (y3), which is repressed by both the repressor protein FapR (y2) and LacI (y1); and the repression could be removed by the effector molecule IPTG (y4) and malonyl-CoA (y5), as described in Equation 3. Equation 4 and Equation 5 simply describe the exponential decay (degradation) of the two effector molecules IPTG (y4) and malonyl-CoA (y5), respectively.

MATLAB was used to simulate the time-dependent species concentration change. A set of protein synthesis rate (α) and degradation rate (γ), repressor protein saturation constant (K) and Hill coefficient (m, n, p, q, r and s) were assigned to frame the biophysical parameter of the proposed model (Eq1 to Eq 5). Then initial species concentration ($y1_0$ for LacI initial concentration, $y2_0$ for FapR initial concentration, $y3_0$ for GFP initial concentration and $y5_0$ for

malonyl-CoA initial concentration) with varying concentration of the inducer IPTG ($y4_0$) was set as the starting point to model the gene expression dynamics. Detailed biophysical parameters have been presented in the Methods and Computational Framework section.

Based on the hypothesis of TF cross-talk model (Eq 1 to Eq5), our simulation results clearly demonstrate the parabolic biphasic transcriptional output (Figure 3b) as we vary the level of the inducer IPTG ($y4$). The net rate of the repressor protein LacI ($y1$) displays an exponential decay (plot not shown here), which is not surprising as we define our system here is a cell-free system without LacI re-synthesis. In addition, the net accumulation of the repressor protein FapR ($y2$) displays a sigmoidal increase pattern as we increase the level of IPTG ($y4$) in the system (Figure 3a), which is reasonable as the expression of FapR is repressed by LacI and IPTG abolishes this repression. At the first phase (i.e. $t < 25$ in Figure 3), the accumulation of FapR is negligible compared to the amount of LacI existing in the system; as a result, LacI ($y1$) will dominate the repression of fluorescence output ($y3$). Indeed, as we increased the level of IPTG, we obtained rapidly increased fluorescence output ($y3$). At the second phase (i.e. $t > 25$ in Figure 3), the expression of FapR ($y2$) reaches a critical amount which is comparable to the level of LacI existing in the system, this corresponds to a flipped gene expression pattern: a transition from an increasing GFP output to a decreasing GFP output in Figure 3b. As time persists, FapR outcompetes LacI to repress expression of the fluorescence reporter, which could explain the fact that fluorescence output undergoes a rapidly decreasing GFP signal. Taken together, our TF competitive binding model perfectly matched our experiment data and explained the parabolic biphasic kinetics: more IPTG leads to stronger GFP expression in the first phase due to LacI dominance, and this pattern is flipped into the opposite scenario that more IPTG leads to weaker GFP expression in the second phase due to FapR dominance. The weakness of our model is that reporter output is always proportionally increased with the level of malonyl-CoA in the system

(Figure 3c), as we didn't specify the carrying capacity of this cell-free system. This model will be refined by our cell-dilution model in the following section.

Stretch and flip the input-output relationship of malonyl-CoA sensor

We then explored what the GFP output would change if we tweak the kinetic parameters of the proposed model (Equation 1 to Equation 5). Our previous model doesn't account for the replenishing rate of malonyl-CoA (α_5 is 0 in Eq5). We added an additional constraint to specify the replenishing rate of malonyl-CoA (α_5 is 0.15 in Eq 5). Due to the persistent supply of malonyl-CoA, GFP output kinetic curve (Figure 4a) in the second phase ($t > 50$) is stretched, indicating that supply of malonyl-CoA is critical to remove the FapR repression in the second phase. This stretched kinetic curve also confirms our previous explanation that FapR dominates the GFP output at the second phase.

We also explored how the GFP output would change if we vary the Hill coefficients of the system. In the previous models (Figure 3), the Hill coefficient $m=n=1$, which means IPTG (y_4) has the same de-repression effect on FapR (y_2) and GFP (y_3). We then changed the Hill coefficients $m=1$, but $n=0.5$, which means IPTG (y_4) will pose relatively weaker de-repression effect on GFP (y_3) compared to FapR (y_2). Interestingly, our model predicts flipped GFP expression kinetics (Figure 4b) compared to Figure 3b: more IPTG leads to decreased GFP expression plateau in Figure 4b *versus* more IPTG leads to increased GFP expression plateau in Figure 3b. It could be simply interpreted as relative expression of FapR (y_2) is larger than the relative expression of GFP (y_3) at same level of inducer IPTG (y_4). Due to the fact that FapR (y_2) also represses the expression of GFP, the reporter output (y_2) will be remarkably reduced when we raise the concentration of IPTG (y_4). As the result, the gene expression kinetic curve is flipped as illustrated in Figure 4b. Not surprisingly, posing further constraints on the system (malonyl-CoA replenishing rate α_5 is set as 0.15) lead to a stretched and flipped gene expression pattern, as shown in Figure 4c.

This part concludes that gene expression kinetics could be completely different from the desired gene expression pattern in biosensor design. It also highlights the importance to choose the appropriate operators and transcriptional factor pairs to make the system behave as it is designed. Should the TF-based biosensor not functional or unexpected reporter output arising, it is always an effective way to retrofit the kinetic parameters and repeat the design-build-test cycle.

Exponential growth leads to hysteresis in malonyl-CoA sensor

Unlike electrochemical biosensors that are normally invasive, most of transcriptional factor-based biosensor is designed to be functional inside the cell. Fluorescence reporter output is always a function of a multitude of factors: cell growth, temperature, pH conditions and cellular redox state. In the following section, we will discuss what the sensor output would change if we couple cell dilution effects with the transcriptional factor cross-talk model; specifically, we will explore the sensor output dynamics under exponential cell growth and Logistic cell growth conditions, using malonyl-CoA sensor model as a prototype.

To investigate how cell growth impacts the reporter output, we re-defined the cell-free malonyl-CoA sensor models (Equation 1 to Equation 5). Specifically, we included one more equation (Eq. 11) to describe how biomass (y_6) evolves with time. We then coupled protein and metabolites synthesis rate α_i ($i=1, 2, 3, 4$ and 5) with the specific growth rate μ . This modification leads to a set of six new equations (Equation 6 to Equation 11) accounting for cell growth effects. Since IPTG (y_4) is an exogenously-added inducer, it is following the logic that we don't need to couple IPTG (y_4) with cell growth in Equation 9. Here we used a generalized Logistic growth model as this is the most common cell-growth pattern for most organisms in a batch-culture.

We first considered the exponential growth, supposed that there are unlimited nutrients and no growth inhibitors exist in our malonyl-CoA sensor system (Figure 5a). For exponential growth, we

simply assign an infinity positive number to the maximal biomass ($y_{6,max}$) to eliminate the inhibitor factor $(1 - y_6/y_{6,max})$ in the set of differential equations from Eq. 6 to Eq. 11. One change on the kinetic parameters is that we set the same synthesis rate and degradation rate for the two repressor protein LacI (y_1) and FapR (y_2), namely $\alpha_1 = \alpha_2 = 0.4$ and $\gamma_1 = \gamma_2 = 0.05$. Because of the relatively large degradation rate, our model predicts that both LacI (y_1) and FapR (y_2) will undergo exponential decay in the first phase ($t < 25$). As a result, expression of the reporter gene GFP (y_3) will undergo exponential increase at this phase ($t < 25$ in Figure 5a). Total cellular malonyl-CoA (y_5) will remain linearly increase with time due to the coupling with cell growth. Counterintuitively, expression of FapR later undergoes an increase (Figure 5a), this is due to the depletion of repressor LacI (y_1) or at least the term y_1^p is negligible compared to the saturation constant K_1 in Eq. 7. The expression of FapR (y_2) is lagging behind the effector IPTG (y_4) induction. This phenomenon is so called hysteresis. In another word, the expression of FapR reflects the past history of the system: FapR expression has a persistent memory of the past that will continue to respond to IPTG even the repressor protein LacI depletes. Due to the persistent increase of FapR (y_2), the reporter output (y_3) will undergo a decay at the second phase (roughly $60 < t < 150$ in Figure 5a). Eventually expression of FapR (y_2) will undergo steady decay due to protein degradation, as a result, reporter output GFP (y_3) will undergo steady increase because the synthesis rate outcompetes the degradation rate (roughly $t > 150$ in Figure 5a).

Logistic growth leads to hysteresis and bimodal gene expression in malonyl-CoA sensor

To model system dynamics, one should always consider the carrying capacity of the system, this is especially true for understanding complex biological behavior. When resources are limited, cell growth exhibits a logistic sigmoidal-shaped pattern, which stipulates that population expansion keeps decreasing as resource becomes scarce, leveling off when the carrying capacity of the system is reached [36, 37]. For this reason, we then examined the effect of Logistic growth on the

dynamic output of our engineered malonyl-CoA sensor. To model Logistic effect, we simply assign a finite positive number to the maximal biomass ($y_{6,max} = 20$) in the set of differential equations from Eq. 6 to Eq. 11. We coupled all the five biological variables (y_1 for LacI, y_2 for FapR, y_3 for GFP, y_5 for malonyl-CoA and y_6 for biomass) with the specific growth rate. IPTG (y_4) is an exogenously-added inducer irrelevant to the biology itself, we just let this species exponentially decay. It should be noted that the specific growth rate contains an inhibitor factor $\mu_{max}(1 - y_6/y_{6,max})$, which specifies that specific growth rate is transitioning from the maximal rate μ_{max} at the beginning ($t = 0, y_6 = 0$) to zero rate at the end (t approaches $+\infty, y_6 = y_{6,max}$) when the system reaches its capacity.

With this, we will be able to account for Logic growth effect and predict the dynamic behavior of the engineered malonyl-CoA sensor. Like the exponential growth effect, our model predicts that both LacI (y_1) and FapR (y_2) will undergo exponential decay in the first phase ($t < 25$) when the model incorporates Logistic growth (Figure 5b). Reporter output GFP (y_3) will rapidly increase at this period ($t < 25$) in response to the decay of FapR and LacI. Due to the hysteresis effect as discussed previously, expression of FapR “remember” the past history of IPTG (y_4) induction; instead of continuing decay, the amount of FapR will later undergo steady increase (roughly $40 < t < 80$, Figure 5b). As FapR (y_2) binds *fapO* to shut down GFP (y_3) expression, reporter GFP will be leveling off as FapR increases. It is obvious that the decay of reporter GFP expression is much lagging behind the rising of the repressor FapR (roughly $75 < t < 100$, Figure 5b). This indicates the hysteresis of FapR expression propagates to the delay of the reporter GFP expression. Due to coupling of the state variables FapR (y_2) and malonyl-CoA (y_5) with cell growth and the tradeoff between synthesis rate (α_i) and degradation rate (γ_i), expression of FapR and synthesis of malonyl-CoA displays a continuing decay as time evolves (roughly $t > 100$, Figure 5b), albeit malonyl-CoA experiences a much weaker decay (due to the very small degradation constant $\gamma_5 = 0.0006$). Interestingly, reporter output GFP (y_3) exhibits a second peak, which is generally called

a bimodal expression pattern (Figure 5b, Figure 6b and Supplementary Figures S1, S2, S3, S4 and S5). This bimodal gene expression pattern is deeply rooted in the hysteresis characteristics of the TF cross-talk model posed by the carrying capacity of the system (where we have used both exponential growth and Logistic growth to constraint our system).

Then we varied the level of inducer IPTG (y_4) and cellular malonyl-CoA (y_5) to examine how strong the hysteresis behavior could lead to. With a range of the inducer IPTG level from 2.5 to 50, the expression of FapR (y_2) displays hysteresis behavior (Figure 6a): a rapid decay ($t < 25$) and a rapid increase ($25 < t < 80$) followed by another decay ($t > 100$). The first decay of FapR expression ($t < 25$) is associated with the relatively large degradation constant of FapR; the rising of FapR expression ($25 < t < 80$) simply represents a memory response to the IPTG induction that is governed by transcriptional factor cross-talk; and the second decay ($t > 100$) of FapR expression is rooted in the Logistic-coupling effects (Eq. 7): the systems displays continuing decreased rate as it reaches the carrying capacity of the system. Consequently, the reporter GFP output displays a bimodal gene expression pattern (Figure 6b). The first rising of GFP expression (roughly $t < 25$) is due to the decay of FapR and LacI expression; the first decay of GFP expression (roughly $75 < t < 120$) represents a memory response to the past FapR rising history ($25 < t < 80$, Figure 6a). The second rising of GFP expression (roughly $150 < t < 200$ in Figure 6b) is simply a response to the past memory of the decay of the FapR (roughly $t > 100$ in Figure 6a); the second decay of GFP expression (roughly $t > 200$ in Figure 6b) is rooted in the Logistic-coupling effect (Eq. 8): the system rate eventually decreases to 0 as it reaches the carrying capacity.

As we increase the level of IPTG, we could induce stronger hysteresis effects: as exemplified by a pronounced FapR peak (Figure 6a) and a widen bimodal GFP expression (Figure 6b) when IPTG level is increased to 50. The delay in expression of FapR and reporter GFP is reinforced as the IPTG level is increased from 2.5 to 50 (Figure 6a and 6b). For example, the reporter output doesn't experience the second delay and only displays a singular mode gene expression pattern

at low IPTG level (2.5), but transitions to bimodal gene expression when IPTG is increased to above 5. More importantly, the peak reporter output is further shifted afterward on the right, it is a characteristic of the hysteresis behavior. Under all bimodal gene expression scenarios, the reporter output (y_3) is increased as we increase the cellular malonyl-CoA level (y_5 , Figure 6c), which tells us that the complicated system still functions as a malonyl-CoA sensor, albeit with transcriptional factor cross-talk and cell dilution effect. Phase plot between FapR (y_2) and GFP expression (y_3) clearly demonstrates the hysteresis nature of transcriptional factor cross-talk and cell dilution effect (Figure 6d), as predicted by our model (Eq. 6 to Eq. 11). The trajectory of FapR-reporter output (GFP) expression transverses back and forth on the phase plane, and eventually bifurcates at two distinct stable states (Figure 6d).

It should be noted that hysteresis and bimodal gene expression is a common phenomenon in biological systems. Positive feedback control and noncooperative regulation has been the hallmark for generating multistability, hysteresis and bifurcation in synthetic gene networks [38-42]. In the cell dilution (exponential or Logistic growth) model, cell dilution actually creates an additional negative control to all the system variables. Due to incoherent nature of the TF cross-talk model, the overall effect of the cell dilution created a positive feedback control and the noncooperative regulation, which is the source of hysteresis and bimodal gene expression in the malonyl-CoA sensor.

Conclusions

A grand challenge in synthetic biology is to design and engineer predictable transcriptional output driving different level of cellular/metabolic activity. Metabolite-responsive transcriptional factors (MRTFs) have emerged as promising tools to address this challenge. These biosensors translate an internal cellular signal to a transcriptional reporter output. Reductionist tends to simplify protein-DNA-RNA interactions and overlook cellular-context effects, leading to undesirable gene expression dynamics with perplexing signal outputs. In this work, we formulated an ordinary

differential equation based- mathematical models to understand the design constraints of transcriptional factor-based biosensors, with the aim to improve the robustness and predictability of the engineered system. Specifically, we proposed a mechanism-based kinetic model to simulate and predict how transcriptional factor (TF) cross-talk, autoregulation and feedback reshape the transcriptional dynamics of an engineered sensor-reporter system. Our computational framework allows us to investigate the effects of protein cooperativity, DNA binding affinity, non-cognate DNA cross-talk, cell growth dilution and autoregulation *et al.* The computational insights obtained in this study will guide us to design more accurate, sensitive TF-based biosensors and may serve as a diagnostic platform to troubleshoot the complex transcriptional dynamics in biosensor design. The computational framework developed here will also provide an educational tutorial for biochemical/biomolecular engineering students to understand biosystem design and transcriptional regulation.

Methods and computational framework

All the simulation is generated by Matlab R2016a and the numerical solutions of differential equations are solved by ode45 solver. MATLAB code will be shared upon request for educational use only. The genetic circuit glyphs are generated by the online servers of pigeoncad (<http://pigeoncad.org/>) [34]. The plots are designed with OriginPro 2017.

For Figures 3a and 3b, physical parameters are $\alpha_1 = 0.5$; $\gamma_1 = 0.1$; $\alpha_2 = 0.2$; $m = 1$; $K_1 = 1$; $p = 2$; $\gamma_2 = 0.01$; $\alpha_3 = 0.1$; $n = 1$; $s = 1$; $K_2 = 0.1$; $q = 2$; $K_3 = 1$; $r = 2$; $\gamma_3 = 0.05$; $\gamma_4 = 0.0005$; $\alpha_5 = 0$ and $\gamma_5 = 0.0006$ in MATLAB programming. And initial conditions are $y_1_0=10$; $y_2_0=0.5$; $y_3_0=0$; $y_5_0=4$.

For Figure 3c, physical parameters are $\alpha_1 = 0.5$; $\gamma_1 = 0.1$; $\alpha_2 = 0.2$; $m = 1$; $K_1 = 1$; $p = 2$; $\gamma_2 = 0.01$; $\alpha_3 = 0.1$; $n = 1$; $s = 1$; $K_2 = 0.1$; $q = 2$; $K_3 = 1$; $r = 2$; $\gamma_3 = 0.05$; $\gamma_4 = 0.0005$; $\alpha_5 = 0$ and $\gamma_5 = 0.0006$ in MATLAB programming. And initial conditions are $y_1_0=10$; $y_2_0=0.5$; $y_3_0=0$; $y_4_0=25$.

For Figure 4a, physical parameters are $\alpha_1 = 0.5$; $\gamma_1 = 0.1$; $\alpha_2 = 0.4$; $m = 1$; $K_1 = 1$; $p=2$; $\gamma_2 = 0.001$; $\alpha_3 = 1$; $n = 1$; $s = 1$; $K_2 = 0.1$; $q = 2$; $K_3 = 1$; $r = 2$; $\gamma_3 = 0.05$; $\gamma_4 = 0.0005$; $\alpha_5 = 0.15$ and $\gamma_5 = 0.0006$ in MATLAB programming. And initial conditions are $y_1_0=10$; $y_2_0=4$; $y_3_0=0$; $y_5_0=4$.

For Figure 4b, physical parameters are $\alpha_1 = 0.5$; $\gamma_1 = 0.1$; $\alpha_2 = 0.2$; $m = 1$; $K_1 = 1$; $p = 2$; $\gamma_2 = 0.01$; $\alpha_3 = 0.1$; $n = 0.5$; $s = 1$; $K_2 = 0.1$; $q = 2$; $K_3 = 1$; $r = 2$; $\gamma_3 = 0.05$;

gamma4 = 0.0005; alpha5 = 0 and gamma5 = 0.0006 in MATLAB programming. And initial conditions are y1_0=10; y2_0=0.5; y3_0=0; y5_0=4.

For Figure 4c, physical parameters are alpha1 = 0.5; gamma1 = 0.1; alpha2 = 0.4; m = 1; K1 = 1; p=2; gamma2 = 0.001; alpha3 = 1; n = 0.5; s = 1; K2 = 0.1; q = 2; K3 = 1; r = 2; gamma3 = 0.05; gamma4 = 0.0005; alpha5 = 0.15 and gamma5 = 0.0006 in MATLAB programming. And initial conditions are y1_0=10; y2_0=4; y3_0=0; y5_0=4.

For Figure 5a, physical parameters are alpha1 = 0.4; gamma1 = 0.05; alpha2 = 0.4; m = 1; K1 = 1; p = 2; gamma2 = 0.05; alpha3 = 1.4; n = 1; s = 1; K2 = 0.1; q = 2; K3 = 1; r = 2; gamma3 = 0.05; gamma4 = 0.0005; alpha5 = 0.15; gamma5 = 0.0006; miu = 0.06 and Ym = $+\infty$ in MATLAB programming. And initial conditions are y1_0=10; y2_0=4; y3_0=0; y4_0=10; y5_0=4; y6_0=0.1.

For Figure 5b and supplementary Figure S1, S2, S3 and S4, physical parameters are alpha1 = 0.4; gamma1 = 0.05; alpha2 = 0.4; m = 1; K1 = 1; p = 2; gamma2 = 0.05; alpha3 = 1.4; n = 1; s = 1; K2 = 0.1; q = 2; K3 = 1; r = 2; gamma3 = 0.05; gamma4 = 0.0005; alpha5 = 0.15; gamma5 = 0.0006; miu = 0.06 and Ym = 20 in MATLAB programming. And initial conditions are y1_0=10; y2_0=4; y3_0=0; y5_0=4; y6_0=0.1.

For Figure 6a and 6b, physical parameters are alpha1 = 0.4; gamma1 = 0.05; alpha2 = 0.4; m = 1; K1 = 1; p = 2; gamma2 = 0.05; alpha3 = 1.4; n = 1; s = 1; K2 = 0.1; q = 2; K3 = 1; r = 2; gamma3 = 0.05; gamma4 = 0.0005; alpha5 = 0.15; gamma5 = 0.0006; miu = 0.06 and Ym = 20 in MATLAB programming. And initial conditions are y1_0=10; y2_0=4; y3_0=0; y5_0=4 and y6_0=0.1.

For Figure 6c, physical parameters are alpha1 = 0.4; gamma1 = 0.05; alpha2 = 0.4; m = 1; K1 = 1; p = 2; gamma2 = 0.05; alpha3 = 1.4; n = 1; s = 1; K2 = 0.1; q = 2; K3 = 1; r = 2; gamma3 = 0.05; gamma4 = 0.0005; alpha5 = 0.15; gamma5 = 0.0006; miu = 0.06 and Ym = 20 in MATLAB programming. And initial conditions are y1_0=10; y2_0=4; y3_0=0; y4_0=25; y6_0=0.1.

Acknowledgements

This work was supported by the Cellular & Biochem Engineering Program of the National Science Foundation under grant no.1805139. The authors would also like to acknowledge the Department of Chemical, Biochemical and Environmental Engineering at University of Maryland Baltimore County for funding support.

Contributions

This work is derived from ENCH 482/682 (Biochemical Engineering) project at UMBC. A detailed project report is uploaded as supplementary information. PX designed and conceived the study. Authors HA, SB, DC, CO and ET analyzed the data, wrote the Matlab code, run the simulation and complete the project report with discussion with PX. PX wrote the manuscript with input from other authors.

Competing interests

The author(s) declares no competing interests.

References

1. Michener, J.K., et al., *Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways*. Metabolic Engineering, 2012. **14**(3): p. 212-222.
2. Liu, D., T. Evans, and F. Zhang, *Applications and advances of metabolite biosensors for metabolic engineering*. Metabolic Engineering, 2015. **31**: p. 35-43.
3. Xu, P., N. Bhan, and M.A.G. Koffas, *Engineering plant metabolism into microbes: from systems biology to synthetic biology*. Current Opinion in Biotechnology, 2013. **24**(2): p. 291-299.
4. Tiangco, C., et al., *Fiber optic biosensor for transdermal glucose based on the glucose binding protein*. Sensors and Actuators B: Chemical, 2017. **242**: p. 569-576.
5. DeLoache, W.C., et al., *An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose*. Nat Chem Biol, 2015. **11**(7): p. 465-471.
6. Mukherjee, K., S. Bhattacharyya, and P. Peralta-Yahya, *GPCR-Based Chemical Biosensors for Medium-Chain Fatty Acids*. ACS Synthetic Biology, 2015. **4**(12): p. 1261-1269.
7. Zhang, F., J.M. Carothers, and J.D. Keasling, *Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids*. Nat Biotech, 2012. **30**(4): p. 354-359.
8. Callura, J.M., et al., *Tracking, tuning, and terminating microbial physiology using synthetic riboregulators*. Proceedings of the National Academy of Sciences, 2010.
9. Green, Alexander A., et al., *Toehold Switches: De-Novo-Designed Regulators of Gene Expression*. Cell, 2014. **159**(4): p. 925-939.
10. Jang, S., et al., *Development of Artificial Riboswitches for Monitoring of Naringenin In Vivo*. ACS Synthetic Biology, 2017. **6**(11): p. 2077-2085.
11. Xiu, Y., et al., *Naringenin-responsive riboswitch-based fluorescent biosensor module for Escherichia coli co-cultures*. Biotechnology and Bioengineering, 2017. **114**(10): p. 2235-2244.
12. Liang, Joe C., Ryan J. Bloom, and Christina D. Smolke, *Engineering Biological Systems with Synthetic RNA Molecules*. Molecular Cell. **43**(6): p. 915-926.
13. McKeague, M., R.S. Wong, and C.D. Smolke, *Opportunities in the design and application of RNA for gene expression control*. Nucleic Acids Research, 2016. **44**(7): p. 2987-2999.
14. Xu, P., *Production of chemicals using dynamic control of metabolic fluxes*. Current Opinion in Biotechnology, 2018. **53**: p. 12-19.
15. Xu, P., et al., *Design and Kinetic Analysis of a Hybrid Promoter-Regulator System for Malonyl-CoA Sensing in Escherichia coli*. Acs Chemical Biology, 2014. **9**(2): p. 451-458.
16. Liu, D., et al., *Negative Feedback Regulation of Fatty Acid Production Based on a Malonyl-CoA Sensor-Actuator*. ACS Synthetic Biology, 2015. **4**(2): p. 132-140.
17. Xu, P., et al., *Engineering Yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals*. Proceedings of the National Academy of Sciences, 2016. **113**(39): p. 10848-10853.
18. Yu, X., et al., *In vitro reconstitution and steady-state analysis of the fatty acid synthase from Escherichia coli*. Proc Natl Acad Sci U S A, 2011. **108**(46): p. 18643-8.
19. Zha, W., et al., *Improving cellular malonyl-CoA level in Escherichia coli via metabolic engineering*. Metab Eng, 2009. **11**(3): p. 192-8.
20. Bhan, N., et al., *Redirecting carbon flux into malonyl-CoA to improve resveratrol titers: Proof of concept for genetic interventions predicted by OptForce computational framework*. Chemical Engineering Science, 2013. **103**: p. 109-114.
21. Xu, P., et al., *Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA*. Metabolic Engineering, 2011. **13**(5): p. 578-587.

22. Tang, S. and P. Cirino, *Design and Application of a Mevalonate-Responsive Regulatory Protein*. Angewandte Chemie-International Edition, 2011. **50**(5): p. 1084-1086.
23. Tang, S.-Y., et al., *Screening for Enhanced Triacetic Acid Lactone Production by Recombinant Escherichia coli Expressing a Designed Triacetic Acid Lactone Reporter*. Journal of the American Chemical Society, 2013. **135**(27): p. 10099-10103.
24. Johnson, A.O., et al., *Design and application of genetically-encoded malonyl-CoA biosensors for metabolic engineering of microbial cell factories*. Metabolic Engineering, 2017. **44**: p. 253-264.
25. Xu, P., et al., *Improving fatty acids production by engineering dynamic pathway regulation and metabolic control*. Proc Natl Acad Sci U S A, 2014. **111**(31): p. 11299-304.
26. Li, S., et al., *Development of a Synthetic Malonyl-CoA Sensor in Saccharomyces cerevisiae for Intracellular Metabolite Monitoring and Genetic Screening*. ACS Synthetic Biology, 2015. **4**(12): p. 1308-1315.
27. Wang, M., S. Li, and H. Zhao, *Design and engineering of intracellular-metabolite-sensing/regulation gene circuits in Saccharomyces cerevisiae*. Biotechnology and Bioengineering, 2016. **113**(1): p. 206-215.
28. David, F., J. Nielsen, and V. Siewers, *Flux Control at the Malonyl-CoA Node through Hierarchical Dynamic Pathway Regulation in Saccharomyces cerevisiae*. ACS Synthetic Biology, 2016. **5**(3): p. 224-233.
29. Ellis, Jessica M. and Michael J. Wolfgang, *A Genetically Encoded Metabolite Sensor for Malonyl-CoA*. Chemistry & Biology, 2012. **19**(10): p. 1333-1339.
30. Schujman, G., et al., *FapR, a bacterial transcription factor involved in global regulation of membrane lipid biosynthesis*. Developmental Cell, 2003: p. 663-672.
31. Fujita, Y., H. Matsuoka, and K. Hirooka, *Regulation of fatty acid metabolism in bacteria*. Molecular Microbiology, 2007. **66**(4): p. 829-839.
32. Schujman, G., et al., *Structural basis of lipid biosynthesis regulation in Gram-positive bacteria*. Embo Journal, 2006: p. 4074-4083.
33. Jin, E., et al., *Rapid evolution of regulatory element libraries for tunable transcriptional and translational control of gene expression*. Synthetic and Systems Biotechnology, 2017. **2**(4): p. 295-301.
34. Bhatia, S. and D. Densmore, *Pigeon: A Design Visualizer for Synthetic Biology*. ACS Synthetic Biology, 2013. **2**(6): p. 348-350.
35. Ma, W., et al., *Defining Network Topologies that Can Achieve Biochemical Adaptation*. Cell, 2009. **138**(4): p. 760-773.
36. Wong, W.W., L.M. Tran, and J.C. Liao, *A hidden square-root boundary between growth rate and biomass yield*. Biotechnology and Bioengineering, 2009. **102**(1): p. 73-80.
37. Tsoularis, A. and J. Wallace, *Analysis of logistic growth models*. Mathematical Biosciences, 2002. **179**(1): p. 21-55.
38. May, T., et al., *Bimodal and Hysteretic Expression in Mammalian Cells from a Synthetic Gene Circuit*. PLOS ONE, 2008. **3**(6): p. e2372.
39. Ochab-Marcinek, A. and M. Tabaka, *Bimodal gene expression in noncooperative regulatory systems*. Proceedings of the National Academy of Sciences, 2010. **107**(51): p. 22096-22101.
40. Kramer, B.P. and M. Fussenegger, *Hysteresis in a synthetic mammalian gene network*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(27): p. 9517-9522.
41. Angeli, D., J.E. Ferrell, and E.D. Sontag, *Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems*. Proceedings of the National Academy of Sciences, 2004. **101**(7): p. 1822-1827.
42. Riccione, K.A., et al., *A Synthetic Biology Approach to Understanding Cellular Information Processing*. ACS Synthetic Biology, 2012. **1**(9): p. 389-402.

Tables

Table 1. Physical explanation of the equation parameters in this study.

Symbols	Physical explanation
y_1	Repressor LacI concentration
y_2	Repressor FapR concentration
y_3	eGFP readout
y_4	IPTG concentration
y_5	Malonyl-CoA concentration
y_6	Biomass concentration
$\alpha_i (i=1,2,3,4,5)$	Protein or metabolite synthesis rate
$\gamma_i (i=1,2,3,4,5, 6)$	Protein or metabolite degradation constant
$K_i (i=1,2,3)$	Saturation constant of ligand (IPTG or malonyl-CoA) to repressor protein
m, n, p, q, r and s	Hill coefficient
μ_{max}	Maximal specific growth rate
$y_{6,max}$	Maximal biomass reached in batch culture

Equations

$$\frac{dy_1}{dt} = \alpha_1 - \gamma_1 y_1 \quad (\text{Eq.1})$$

$$\frac{dy_2}{dt} = \frac{\alpha_2 y_4^m}{K_1 + y_1^p} - \gamma_2 y_2 \quad (\text{Eq.2})$$

$$\frac{dy_3}{dt} = \frac{\alpha_3 y_4^n y_5^s}{(K_2 + y_2^q)(K_3 + y_1^r)} - \gamma_3 y_3 \quad (\text{Eq.3})$$

$$\frac{dy_4}{dt} = -\gamma_4 y_4 \quad (\text{Eq.4})$$

$$\frac{dy_5}{dt} = \alpha_5 - \gamma_5 y_5 \quad (\text{Eq.5})$$

$$\frac{dy_1}{dt} = \alpha_1 \mu_{\max} \left(1 - \frac{y_6}{y_{6,\max}}\right) - \gamma_1 y_1 \quad (\text{Eq.6})$$

$$\frac{dy_2}{dt} = \frac{\alpha_2 y_4^m}{K_1 + y_1^p} \mu_{\max} \left(1 - \frac{y_6}{y_{6,\max}}\right) - \gamma_2 y_2 \quad (\text{Eq.7})$$

$$\frac{dy_3}{dt} = \frac{\alpha_3 y_4^n y_5^s}{(K_2 + y_2^q)(K_3 + y_1^r)} \mu_{\max} \left(1 - \frac{y_6}{y_{6,\max}}\right) - \gamma_3 y_3 \quad (\text{Eq.8})$$

$$\frac{dy_4}{dt} = -\gamma_4 y_4 \quad (\text{Eq.9})$$

$$\frac{dy_5}{dt} = \alpha_5 \mu_{\max} \left(1 - \frac{y_6}{y_{6,\max}}\right) - \gamma_5 y_5 \quad (\text{Eq.10})$$

$$\frac{dy_6}{dt} = \mu_{\max} \left(1 - \frac{y_6}{y_{6,\max}}\right) y_6 \quad (\text{Eq.11})$$

Figure captions

Figure 1. Schematic representation and gene expression dynamics of malonyl-CoA sensor with eGFP as readout. **(a)** A T7 promoter based genetic circuit to sense malonyl-CoA in *E. coli*. T7: bacteriophage T7 promoter; lacO: lacI repressor binding site; lacI: *E. coli* lactose repressor protein; fapO: fapR repressor binding site; fapR, *B. subtilis* fatty acid biosynthetic pathway repressor; T7-fapO, fapO sequence was inserted downstream of T7 promoter; eGFP: enhanced green fluorescence protein. **(b)** Dependence of eGFP readout on the inducer IPTG: IPTG deactivates LacI and turns on FapR expression to repress the transcription of eGFP, assuming constant level of malonyl-CoA inside the cell. **(c)** Dependence of eGFP readout on the inducer IPTG: LacI competes with FapR for binding with fapO, leading to biphasic gene expression dynamics. This data has been experimentally validated [15].

Figure 2. Simplified representation of constructed malonyl-CoA sensor. **(a)** Pigeoncad glyph generated from Pigeoncad online server [34]. **(b)** Deconvolution of the pigeon glyph and representation of LacI and FapR competitively bind with fapO. **(c)** Symbol representation of the inter-dependence of LacI (y1), FapR (y2), GFP (y3), IPTG (y4) and malonyl-CoA (y5).

Figure 3. Time-dependent species concentration change predicted by equation 1 to equation 5. **(a)** FapR concentration is increased with increased concentration of inducer IPTG (y4) and increased cultivation time t. **(b)** eGFP output exhibits biphasic parabolic kinetics. Note that the plateau GFP output is increased with increasing concentration of IPTG (y4). **(c)** GFP output is proportionally increasing with increased concentration of malonyl-CoA (y5).

Figure 4. Time-dependent GFP output simulated by equation 1 to equation 5. **(a)** GFP expression with malonyl-CoA (y5) replenishing. This GFP expression curve is stretched compared to Figure 3b. **(b)** GFP expression with Hill coefficients $m=1$ and $n=0.5$. The pattern of GFP expression is flipped compared to Figure 3b. **(c)** GFP expression with malonyl-CoA (y5) replenishing and Hill

coefficients $m=1$ and $n=0.5$. This GFP expression curve is stretched and the pattern is flipped compared to Figure 3b. “Flipped” means that the plateau GFP output is decreased with increasing concentration of IPTG (y_4).

Figure 5. Time-dependent species concentration change with cell dilution predicted by Equation 6 to Equation 11. **(a)** Cell maintains exponential growth ($y_{6,max}$ in Equations 6-11 is infinity large $+\infty$ and μ_{max} is 0.06). **(b)** Cell maintains logistic growth ($y_{6,max}$ in Equations 6-11 is 20 and μ_{max} is 0.06). Exact physical parameters and initial conditions can be found in the Methods and Computational Framework section.

Figure 6. Time-dependent species concentration change with logistic-type dilution effect predicted by equation 6 to equation 11. **(a)** FapR expression exhibits hysteresis characteristics, though FapR concentration is increased with increased concentration of inducer IPTG (y_4). **(b)** eGFP output exhibits bimodal gene expression as we increase the concentration of IPTG (y_4). **(c)** Time-dependent bimodal eGFP expression varies at different level of intracellular malonyl-CoA (y_5). **(d)** Hysteresis characteristic of repressor protein FapR and reporter output (GFP expression) in transcriptional factor cross-talk model constrained by logistic growth.

Figures

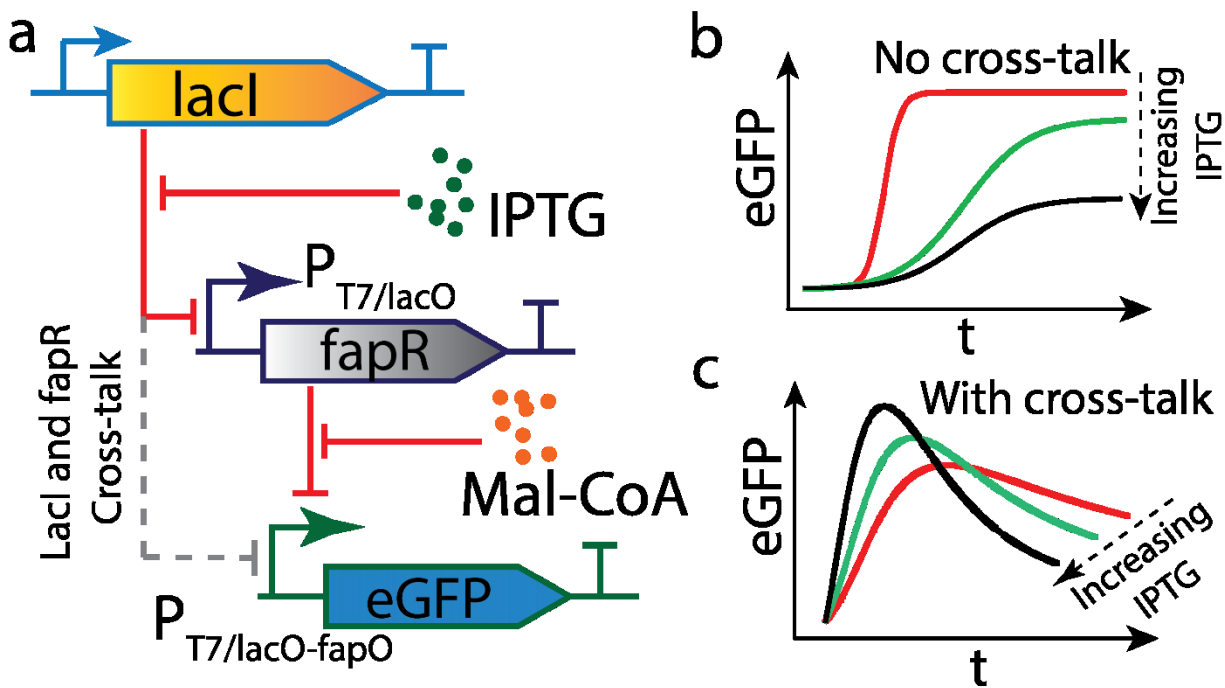


Figure 1

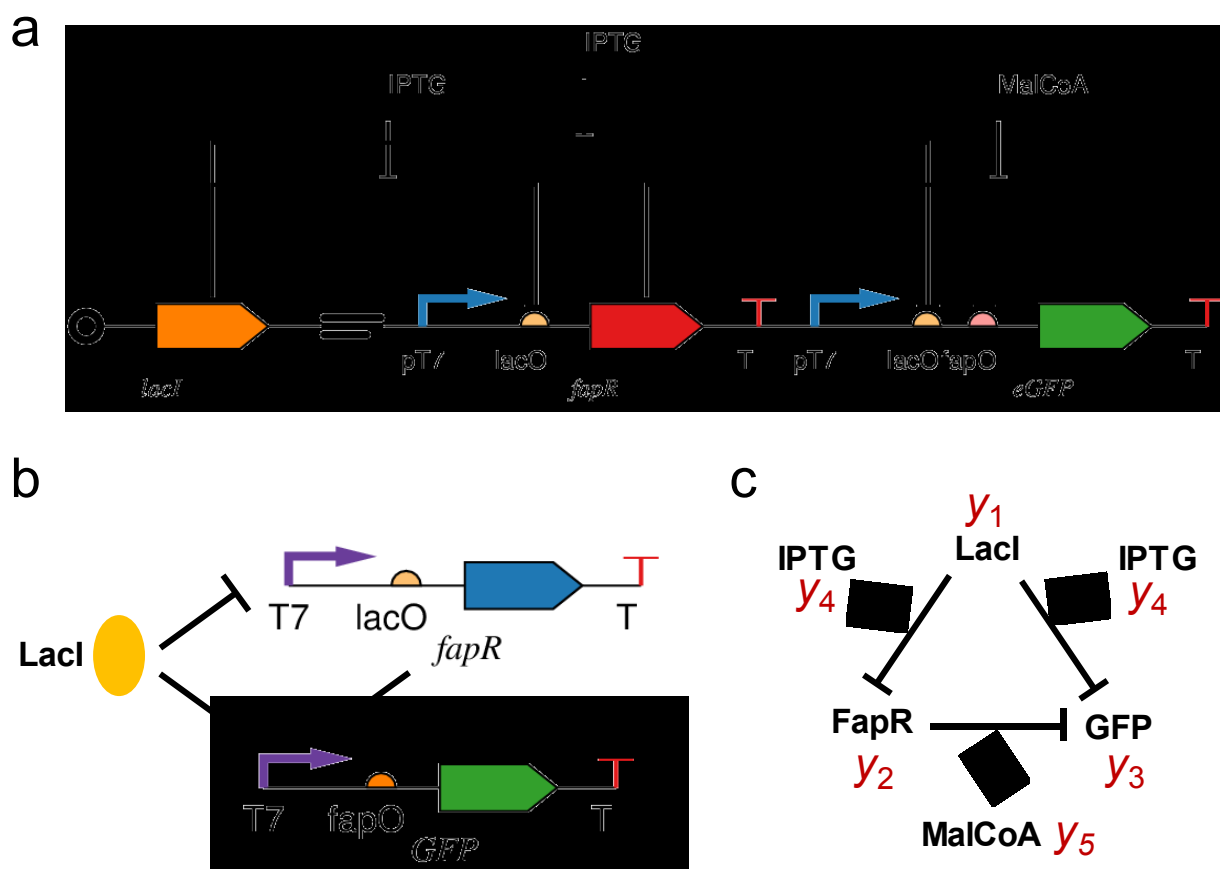


Figure 2

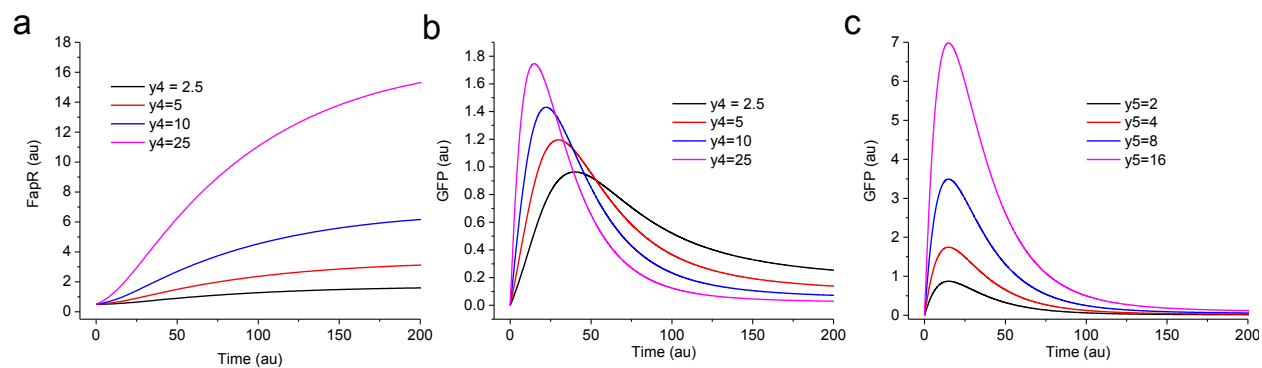


Figure 3

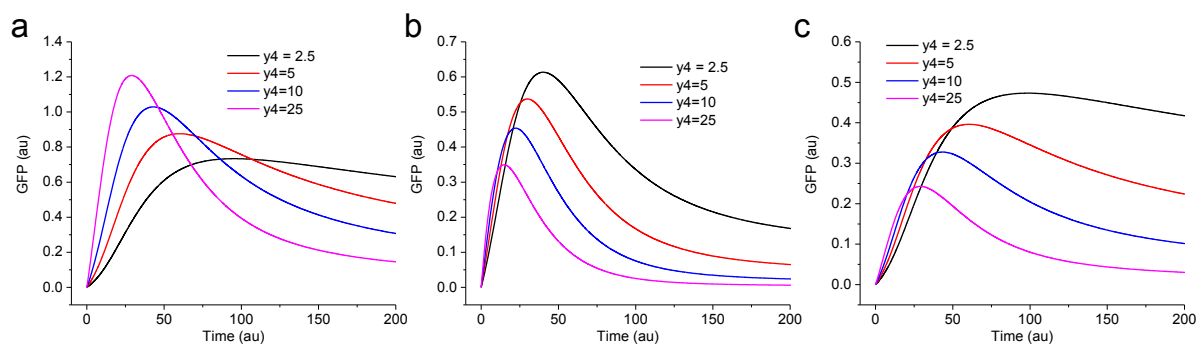


Figure 4

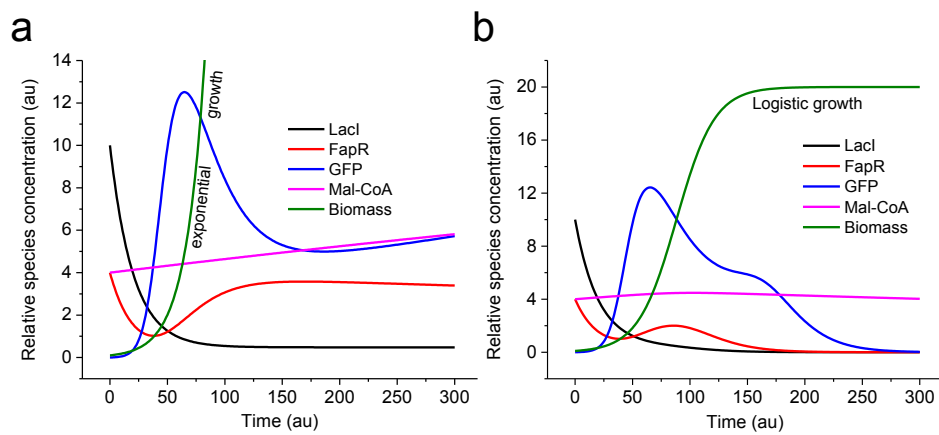


Figure 5

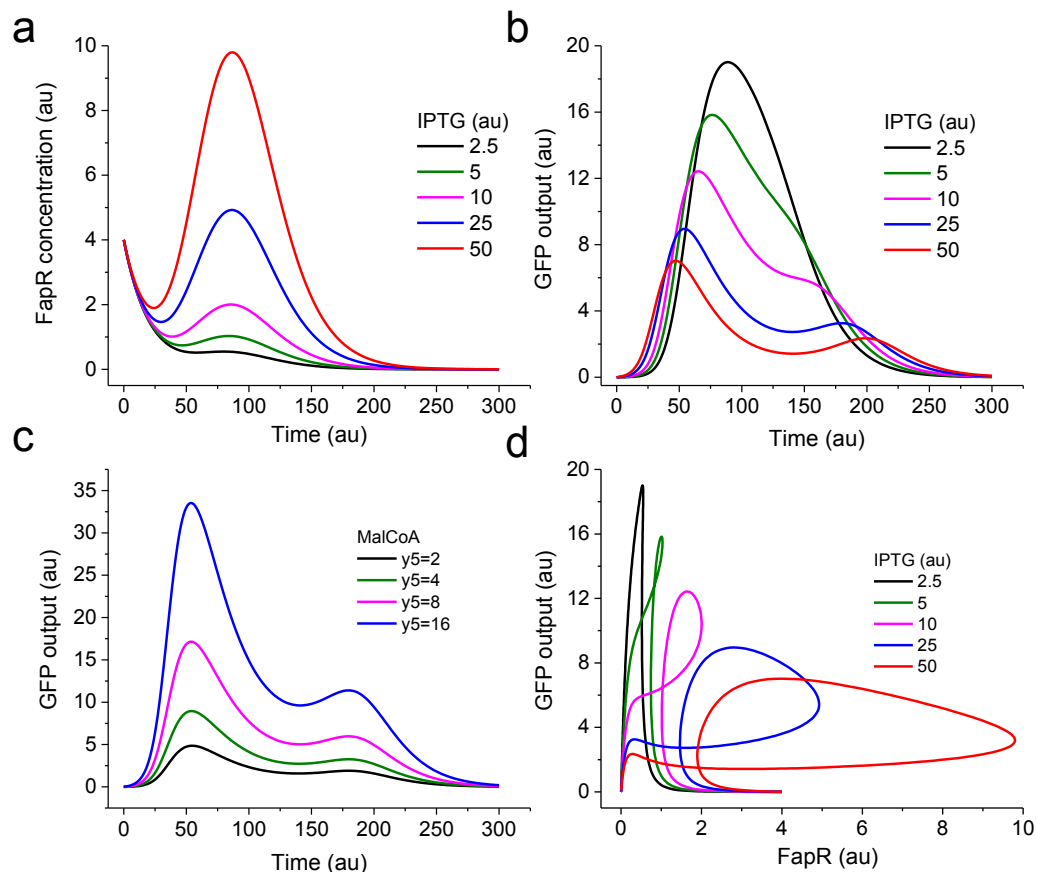


Figure 6