Designing Genetic Feedback Controllers

Andreas W. K. Harris, James A. Dolan, Ciarán L. Kelly, James Anderson
and Antonis Papachristodoulou

***Abstract*—By incorporating feedback around systems we wish to manipulate, it is possible to improve their performance and robustness properties to meet pre-specified design objectives. For decades control engineers have been successfully implementing feedback controllers for complex mechanical and electrical systems such as aircraft and sports cars. Natural biological systems use feedback extensively for regulation and adaptation but apart from the most basic designs, there is no systematic framework for designing feedback controllers in Synthetic Biology. In this paper we describe how classical approaches from linear control theory can be used to close the loop. This includes the design of genetic circuits using feedback control and the presentation of a biological phase lag controller.**

***Index Terms*—Feedback Control, Robust Control, Systems and Synthetic Biology.**

I. Introduction

Synthetic Biology is a relatively new field, aspiring to use Engineering principles (such as standardisation and modularity) to design biological circuits and organisms [1], [2]. It aims to design new biological parts, devices and systems or redesign existing ones for a specific application. Several application areas have been identified to date, from bioenergy to healthcare, to agriculture and the environment. Synthetic Biology therefore has the potential to create new industries and economies and holds great promise. The early vision of Synthetic Biology set very ambitious goals, as described in [3] but the many challenges that it currently faces [2], [4] mean that several of the promised advances will be very difficult to achieve. Some of these challenges are addressed in natural biological systems through the extensive use of feedback control at all levels of biological organisation [5], [6]. Based on this observation and insights from classical control engineering, this paper aims to develop a framework to design feedback loops in genetic circuits.

Control Engineering has developed a number of methods for the analysis and design of feedback control systems. Some of these ideas have had extensive application in Systems Biology [7], [8] but also more recently in Synthetic Biology [9]–[11]. One application of Control Engineering to Synthetic Biology is in the development of tools to support an 'Engineering Design Cycle' that involves modelling, system identification and parameter estimation and a 'design' approach to modify biological systems systematically to achieve a particular objective, by tuning 'dials' [12].

Perhaps the most intriguing application of the ideas of Control Engineering to Biology is in the design and implementation of *feedback controllers*, in much the same way these are designed for a standard engineering system. Previous work in this area includes the design of controllers using DNA [13], [14]. By closing the loop, the advantages of feedback can be harvested, such as improved robustness and reduced sensitivity to noise. Other design objectives can also be achieved, through careful controller design [15].

The question that arises is what types of feedback can be designed within a biological context and what such controllers look like. Previous work has implemented feedback at the genetic level (autorepressor [16], toggle switch [17] and repressilator [18]), but also at the protein level [19]. However, to date very few have actually designed specific controllers and the size of these synthetic systems is restricted to just a few genes. In fact, the design of more complicated controllers such as the ones found in Engineering still remains a challenge [20], as well as the implementation of simple components such as a summing junction [21]. The use of accurate mechanistic models [22] in conjunction with the reliable design of feedback and controllers are key elements that will allow the size of synthetic systems to increase.

In this paper we aim to design bio-controllers using ideas borrowed from Control Engineering, building on our previous work in [23]. The approach we take is to investigate what configurations would result in dynamics resembling traditional engineering controllers and use them to achieve a desired objective through optimisation.

The paper is organised as follows. In Section II we provide some background; we discuss the problem we wish to solve and review different modelling approaches. Section III discusses frequency representations of the gene circuits while in Section IV we discuss how the controller design can be performed, which leads us to demonstrate our approach in an example in Section V. Section VI concludes the paper.

II. Problem Setup and Background

This section introduces the concepts of gene expression and expression regulation as well as standard modelling techniques used to study these processes. This will

A.W.K. Harris, J. Anderson and A. Papachristodoulou are with the Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, U.K. e-mail: andreas.harris, james.anderson, antonis@eng.ox.ac.uk. J.A. Dolan is with the Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PX. e-mail: jad67@cam.ac.uk. C.L. Kelly is with the Department of Bioengineering, Imperial College London, South Kensington Campus, London SW7 2AZ. e-mail: ciaran.kelly@imperial.ac.uk.

allow us to develop a design framework for biological feedback control.

Gene expression is the method by which information is transferred from the passive DNA to the active proteins that perform the vast majority of cellular tasks required for life. This process can be broadly split into two stages. The first is transcription, where a section of DNA is read out into RNA. The second is translation, where the short strand of transcribed RNA is turned into protein. One of the ways these processes are regulated in nature, but also the most common method of regulation in synthetic systems, is through Transcription Factors (TF). These control the initiation rate of the transcription of a gene and thereby its total expression. A specific TF will have a specific target promoter region in front of the regulated gene.

TFs are proteins themselves and come in two different types: they either increase the transcription rate, in which case they are called *activators*, or they inhibit transcription, in which case they are called *repressors*. In natural systems, genes are commonly regulated by multiple TFs at once. These bind to separate promoter regions in tandem in front of the gene, which can be used when designing synthetic systems.

A. Gene Expression Regulation

We begin by modelling the interaction between a TF, X, and the gene it regulates, G

$$\mathrm{G} + n\mathrm{X} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \mathrm{G{:}X}_n,$$

where n is the number of TF molecules that bind to the gene and $\mathrm{G{:}X_n}$ is the gene-TF complex. The forward and reverse reaction rates are given by the positive constants k_f and k_r respectively. Using generalised mass action kinetics, ODE models of the above biochemical reaction can be developed [8]. The following derivation is based on that presented in [7]; ignoring intermediate binding stages, the dynamics are given by:

$$\frac{\mathrm{d[G{:}X}_n]}{\mathrm{d}t} = k_f[\mathrm{G}][\mathrm{X}]^n - k_r[\mathrm{G{:}X}_n], \quad (1a)$$

$$\frac{\mathrm{d[G]}}{\mathrm{d}t} = k_r[\mathrm{G{:}X}_n] - k_f[\mathrm{G}][\mathrm{X}]^n, \quad (1b)$$

where the concentration of a species is represented by square brackets around its label.

Proceeding from here to the full model requires two assumptions, that are as follows:

> A1: TF-gene interactions are assumed to be at steady state when modelling gene expression.
> A2: In the case of an activator, the transcription initiation rate is only increased when the active TF is bound to the promoter region. In the case of a repressor, the transcription initiation rate is only decreased when the active TF is bound to the promoter region.

The total concentration of the gene is

$$[\mathrm{G}]_T = [\mathrm{G}] + [\mathrm{G{:}X}_n].$$

This conservation relation is derived from (1). Invoking A1, the steady state fractional occupations of the free and bound gene can be calculated:

$$\frac{[\mathrm{G{:}X}_n]}{[\mathrm{G}]_T} = \frac{[\mathrm{X}]^n}{K^n + [\mathrm{X}]^n}$$

$$\frac{[\mathrm{G}]}{[\mathrm{G}]_T} = \frac{K^n}{K^n + [\mathrm{X}]^n}$$

where $K^n := \frac{k_r}{k_f}$. The above are Hill functions approximating the fast binding and un-binding of two species. In this context, K is referred to as the apparent dissociation constant and n is the Hill coefficient. These fractions can be thought of as the time average of the state of the gene and can therefore be used to model expression rates by invoking A2. The contribution to expression rate from a promoter is therefore:

$$\Gamma_{\mathrm{A}} = \frac{\beta[\mathrm{A}]^n}{K^n + [\mathrm{A}]^n}, \quad (2a)$$

$$\Gamma_{\mathrm{R}} = \frac{\beta K^n}{K^n + [\mathrm{R}]^n}, \quad (2b)$$

where Γ_{A} is the contribution from an activated promoter region and Γ_{R} is the contribution from a repressed promoter region, A is an activating TF, R is a repressing TF and the parameter β describes the strength of the promoter.

For later modelling and design is it worth noting the derivatives of these functions with respect to the TF concentration and how these are related, as it simplifies notation.

$$\frac{\mathrm{d}\Gamma_{\mathrm{A}}}{\mathrm{d[A]}} = \frac{n_{\mathrm{A}}\beta K_{\mathrm{A}}^{n_{\mathrm{A}}}[\mathrm{A}]^{n_{\mathrm{A}}-1}}{(K_{\mathrm{A}}^{n_{\mathrm{A}}} + [\mathrm{A}]^{n_{\mathrm{A}}})^2} =: \gamma_{\mathrm{A}}, \quad (3a)$$

$$\frac{\mathrm{d}\Gamma_{\mathrm{R}}}{\mathrm{d[R]}} = -\frac{n_{\mathrm{R}}\beta K_{\mathrm{R}}^{n_{\mathrm{R}}}[\mathrm{R}]^{n_{\mathrm{R}}-1}}{(K_{\mathrm{R}}^{n_{\mathrm{R}}} + [\mathrm{R}]^{n_{\mathrm{R}}})^2} =: -\gamma_{\mathrm{R}}, \quad (3b)$$

where

$$\gamma_{\mathrm{X}} := \frac{n_{\mathrm{X}}\beta K_{\mathrm{X}}^{n_{\mathrm{X}}}[\mathrm{X}]^{n_{\mathrm{X}}-1}}{(K_{\mathrm{X}}^{n_{\mathrm{X}}} + [\mathrm{X}]^{n_{\mathrm{X}}})^2},$$

where X can either be an activating or repressing TF and we have taken advantage of the fact that the derivatives of Γ (2) have the same form excepting the sign. The function γ_{X} describes the derivative of the expression rate contribution from a promoter which is positive in the case of an activating TF and negative in the case of a repressing TF.

Even though a promoter is regulated, there is a probability that the promoter attracts the transcriptional machinery anyway. This effect is modelled by including a constant basal level of expression is denoted β_0.

The expression of a protein is balanced by its decay which occurs through a number of processes. Among these are degradation by the cellular machinery, but also dilution due to cell growth. These processes are lumped

into a single parameter in a linear term dependent on the concentration of the protein being degraded or diluted and is denoted α.

It is worth noting that the parameters β, β_0, K, n and α and all the concentrations are non-negative, thus $\Gamma_{\mathrm{A}}, \Gamma_{\mathrm{R}}, \gamma_{\mathrm{X}} \geq 0$.

We will now see how the mechanisms discussed above can be used to model the two simplest cases of regulation: a protein whose expression is activated and a protein whose expression is repressed.

B. Activated Expression

Following the discussions above, a protein Y, whose expression is activated by the TF A, has the following rate of change of concentration over time

$$\frac{\mathrm{d}[\mathrm{Y}]}{\mathrm{d}t} = \beta_0 + \Gamma_{\mathrm{A}} - \alpha[\mathrm{Y}] \tag{4}$$

where β_0 is the basal expression rate, the function Γ_{A} yields the contribution to expression by the activator A, and α is the degradation/dilution rate. A diagram of this genetic circuit can be seen in Fig. 1(a), where the incoming activating TF A interacts with its promoter and regulates the expression of gene Y producing protein Y.

The letter Y is used to denote the output protein of the system following the standard engineering convention. This is what would be measured in a real biological system through e.g., fluorescent tagging.

C. Repressed Expression

Analogous to the case above of activated expression, the rate of change of the concentration of a protein, whose expression is repressed by the TF R is

$$\frac{\mathrm{d}[\mathrm{Y}]}{\mathrm{d}t} = \beta_0 + \Gamma_{\mathrm{R}} - \alpha[\mathrm{Y}] \tag{5}$$

where now the function Γ_{R} yields the contribution to expression by the repressor R. A diagram of this genetic circuit can be seen in Fig. 1(b), where the incoming repressing TF R interacts with its promoter and regulates the expression of gene Y producing protein Y.

D. Expression Regulation by Multiple TFs

As previously mentioned, it is common for the expression of a gene to be regulated by several TFs. In this case, it is assumed these can be modelled through a product of the expression rates (2) as in [7], [8]. In such a way, a gene's expression can be regulated by an arbitrary number of TFs:

$$\frac{\mathrm{d}[\mathrm{Y}]}{\mathrm{d}t} = \beta_0 + \prod_{i=1}^{N} \Gamma_{\mathrm{A}_i} \prod_{j=1}^{M} \Gamma_{\mathrm{R}_j} - \alpha[\mathrm{Y}], \tag{6}$$

where [Y] is the concentration of output protein and the products are over the N activating TFs, $\mathrm{A}_1, \ldots, \mathrm{A}_N$, and the M repressing TFs, $\mathrm{R}_1, \ldots, \mathrm{R}_M$, with their contributions to expression supplied by the activation and repression functions Γ_{A_i} and Γ_{R_j} respectively. Using the above equation, the regulation of gene expression by any finite number of TFs can be described. A genetic circuit diagram of this system can be seen in Figure 2(a), where the TFs interact with their specific promoters and regulate the expression of the protein Y.

III. Frequency Domain Representation

In order to apply classical linear control theoretical analysis and design methodologies to genetic circuits, the dynamics of the activator and repressor must first be expressed in the frequency domain. To do so requires linearisation and Laplace transformation of the differential equations (4), (5), and the more general case (6). Detailed derivations of these and introductions to linearisation and Laplace transforms can be found in the Supplementary Information [24].

A. Single Input Systems

Consider first the activated expression, Equation (4). Let y^* and a^* denote, respectively, the equilibrium concentrations of the expressed protein and activating TF about which the linearisation is performed. The linearised dynamics are given by

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\alpha y + \gamma_{\mathrm{A}}^* u, \tag{7}$$

where $y = [\mathrm{Y}] - y^*$ and $u = [\mathrm{A}] - a^*$ are small perturbations about the equilibrium of the system and γ_{A}^* indicates the function γ_{A} evaluated at the equilibrium. The change in notation for the TF concentration to u is because this is treated as an input to the system, for which the standard engineering notation is u.

The linearised dynamics of the repressed system are similar, taking the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\alpha y - \gamma_{\mathrm{R}}^* u, \tag{8}$$

where again $y = [\mathrm{Y}] - y^*$ and $u = [\mathrm{R}] - r^*$ are small perturbations about the equilibrium of the system, r^* is the equilibrium repressor concentration and γ_{R}^* is the function γ_{R} evaluated at the equilibrium.

Taking the Laplace transform of the systems above yields the transfer function

$$H_{\mathrm{X}}(s) = \frac{Y(s)}{U(s)} = \frac{\pm\gamma_{\mathrm{X}}^*}{s + \alpha} \tag{9}$$

where activated expression ($\mathrm{X} = \mathrm{A}$) takes the positive ($+$) and repressed expression ($\mathrm{X} = \mathrm{R}$) takes the negative ($-$) sign.

This system can be represented by the block diagram seen in Figure 1(c). A framework for the use of these diagrams will be developed. Note how the protein expression process is split across two blocks. The left block quantifies the interaction between a TF, the input, and a promoter. This can be thought of as a gain block. The second block represents the expression of the protein.

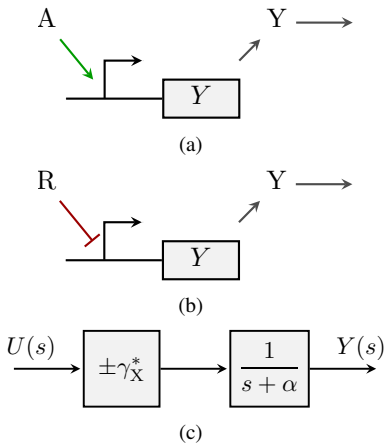

Fig. 1. (a) A genetic circuit diagram of activated expression. (b) A genetic circuit diagram of repressed expression. (c) Block diagram representation for a single gene with expression either activated, where X = A and the gain block takes the positive (+), or repressed, where X = R and the gain block takes the negative (−).

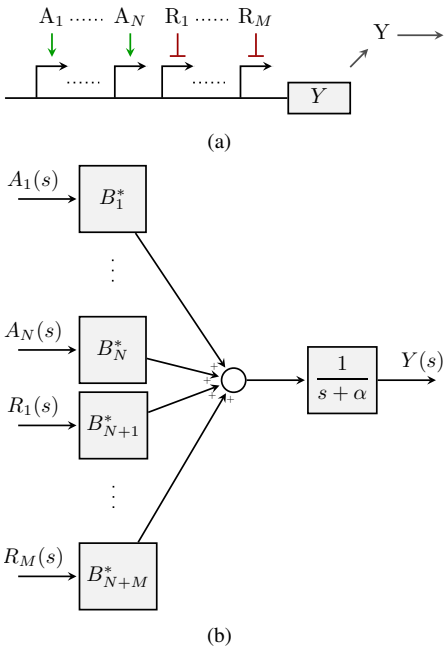

Fig. 2. (a) A genetic circuit diagram and (b) a block diagram of protein expression regulated by multiple transcription factors.

These are the two fundamental building blocks that exist in these systems and throughout the paper these will be used and combined in different ways exhibiting various behaviours.

B. Multiple Input System

Taking Equation (6) as a starting point, the linearisation of the system is of a similar form to the single input cases, though there are some important differences. The linearisation takes the form

$$\dot{y} = -\alpha y + \mathbf{B}^* u, \tag{10}$$

where now the input $u \in \mathbb{R}^{N+M}$, is a small perturbation about the equilibrium inputs $a_1^*, \ldots, a_N^*, r_1^*, \ldots, r_M^*$, and the input matrix

$$\mathbf{B}^* = \left[\frac{\mathrm{d}[\dot{\mathrm{Y}}]}{\mathrm{d}[\mathrm{A}_1]}, \frac{\mathrm{d}[\dot{\mathrm{Y}}]}{\mathrm{d}[\mathrm{A}_2]}, \ldots, \frac{\mathrm{d}[\dot{\mathrm{Y}}]}{\mathrm{d}[\mathrm{A}_N]}, \ldots \right. \left. \frac{\mathrm{d}[\dot{\mathrm{Y}}]}{\mathrm{d}[\mathrm{R}_1]}, \frac{\mathrm{d}[\dot{\mathrm{Y}}]}{\mathrm{d}[\mathrm{R}_2]}, \ldots, \frac{\mathrm{d}[\dot{\mathrm{Y}}]}{\mathrm{d}[\mathrm{R}_M]} \right] \Bigg|_{y^*, u^*} \tag{11}$$

is a row vector evaluated at equilibrium of same dimension as u. It is worth noting that the derivatives w.r.t. the repressing TFs $\mathrm{R}_1, \ldots, \mathrm{R}_M$ are negative. For a full derivation, see the Supplementary Information [24].

Laplace transforming the linearisation yields the following equation:

$$Y(s) = \frac{1}{s+\alpha} \mathbf{B}^* U(s), \tag{12}$$

where $U(s)$ is the Laplace transform of the vector of inputs. The term $\mathbf{B}^* U(s)$ can be thought of as a weighted sum of the inputs as is demonstrated by the block diagram in Figure 2(b). The inputs are weighted by their corresponding term in the input matrix $\mathbf{B}^*$ and are collected at the summing junction before being fed to the gene regulating expression. In this way the promoters of a gene process the inputs before expression occurs.

C. Block Diagram Framework

The analysis of the above system motivates a block diagram framework for gene circuit analysis that facilitates the extraction of the transfer function of a system and allows the application of linear design principles. This section explains the basic rules of converting a standard genetic circuit diagram such as the ones seen in Figures 1(a),(b) and 2(a) into a block diagram and extracting the relevant transfer function.

1) First, the number of genes involved in the circuit is counted. An expression block with the transfer function $\frac{1}{s+\alpha_i}$ pertains to each gene, where α_i is the degradation/dilution rate of the resulting protein.
2) The inputs to each gene are then processed through a promoter gain block whose form depends on the types (activating and/or repressing) and number of input TFs going to a specific gene and can be calculated using Equations (3) in the case of a single input to a gene and (11) in the case of multiple inputs.
3) Once processed through the promoter gain blocks the inputs are then added together and fed to the gene expression block, which then outputs another protein.
4) The final step is to link up the genes using the topology provided by the genetic circuit diagram.

This methodology allows one to construct the block diagram from which the transfer function can be extracted

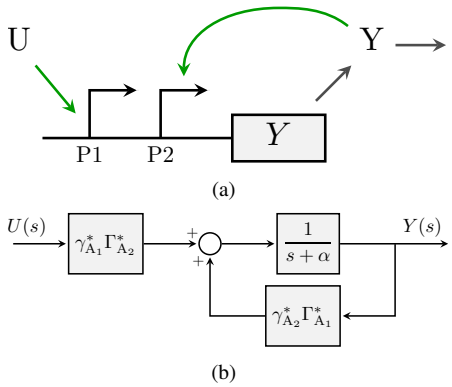

Fig. 3. (a) A genetic circuit diagram and (b) a block diagram of protein expression with autoactivation.

using standard techniques.

D. Example: Autoactivation

Autoactivation occurs when a gene activates its own expression. This example demonstrates the use of the framework and is interesting as it gives insight into the possibility of a genetic integrator. The genetic circuit diagram can be seen in Figure 3(a). An example of a biological system that would behave as such would be a gene encoding the protein 'XylS' with the P_m promoter in front of it. The block diagram can be produced using the methods discussed and is exhibited in Figure 3(b). The input to output transfer function is

$$H(s) = \frac{\gamma^*_{A_1}\Gamma^*_{A_2}}{s + \alpha - \gamma^*_{A_2}\Gamma^*_{A_1}}. \tag{13}$$

The position of the pole of this transfer function is at $s = -\alpha + \gamma^*_{A_2}\Gamma^*_{A_1}$. If $\alpha < \gamma^*_{A_2}\Gamma^*_{A_1}$, this pole is in the right half plane and would cause instability. If $\alpha = \gamma^*_{A_2}\Gamma^*_{A_1}$, the system would have a pole at the origin providing an integrator.

The ODE model of the system provides further insight to its behaviour and takes the form

$$\frac{\mathrm{d}[\mathrm{Y}]}{\mathrm{d}t} = \beta_0 + \Gamma_{A_1}\Gamma_{A_2} - \alpha[\mathrm{Y}],$$

where Γ_{A_1} provides the dependence on the input TF and Γ_{A_2} supplies the feedback of the output protein. This system has up to 3 equilibria and only one of these is stable. It is around the stable equilibrium that the linearisation is performed. In the stable case $\alpha > \gamma^*_{A_2}\Gamma^*_{A_1}$ and the response time of the system is increased as the 3dB bandwidth has now been reduced to $\alpha - \gamma^*_{A_2}\Gamma^*_{A_1}$. This is in agreement with the literature [7], [8] and also with simulation of the ODE model.

The implementation of a biological integrator would be a powerful tool as it would provide a means to track a step reference with no steady state error. Given the form of (13), one might think that the autoactivator would provide a way of achieving this by setting $\alpha = \gamma^*_{A_2}\Gamma^*_{A_1}$, but this is not possible. The parameters in biological systems are far too uncertain to be able to tune to the accuracy necessary to behave as an integrator reliably. This uncertainty would cause instability.

IV. Designing a Feedback Controller

Using the modelling framework set up in Section II-A and the frequency domain analysis and block diagram framework laid out in Section III, we can now turn to the task of designing a genetic feedback controller. The standard implementation of such controllers is through unit negative feedback and comparison with a reference producing an error, which is fed to the controller. First let us review the possible controllers.

The *ideal* PID controller takes the form

$$K_{PID}(s) = K_P\left(1 + \frac{1}{T_I s} + T_D s\right)$$

where K_P, T_I and T_D are parameters defining the proportional, integral and derivative gains of the controller. Whilst PID controllers offer many advantages we have shown in the previous section that building a genetic integrator is problematic, furthermore, in order to build a PID controller (for any system not necessarily a biological one) with a non-zero K_D term a filter would need to be implemented in order to make the transfer function proper. This added complexity makes constructing an ideal genetic PID controller difficult at present.

Alternatively, phase lead and lag controllers can be considered. These are described by the transfer function

$$K_{PL}(s) = K\frac{s+z}{s+p}, \tag{14}$$

where K is the gain, $z > 0$ defines the position of the zero and $p > 0$ defines the position of the pole. When $|z| < |p|$ (14) is said to be a phase lead controller, and when $|z| > |p|$ (14) is a phase lag controller.

By tuning the parameters K, z and p of the phase lead or lag controller it is possible to change the frequency response of the system. Phase lead controllers typically

1) improve the phase margin (provide robustness to un-modelled time delays),
2) damp out transient oscillations and reduce settling time.

Phase lag controllers typically

1) improve disturbance rejection (provide robustness to noise at the input),
2) reduce steady state error.

An advantage of the lead-lag architecture is that the cascade connection of two such controllers allows the designer to shape the frequency response as desired i.e. over certain frequency ranges the benefits of a phase lead controller can be achieved, whilst in another frequency range the phase lag controller dominates. Such a controller is referred to as a lead-lag controller and takes the form:

$$K_{PLL}(s) = K\frac{(s+z_{lead})(s+z_{lag})}{(s+p_{lead})(s+p_{lag})},$$

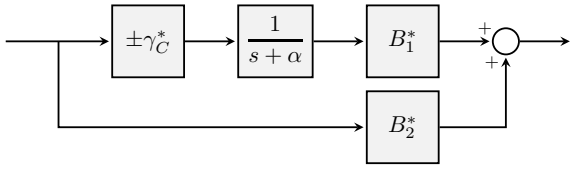

Fig. 4. Block diagram of a genetic phase lead (+) and phase lag (-) controller.

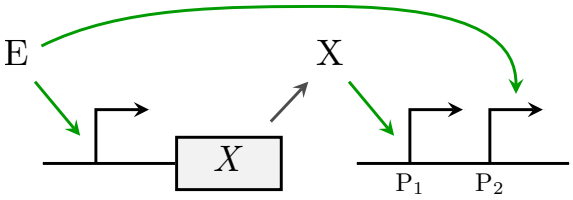

Fig. 5. A genetic circuit diagram of a genetic phase lag controller.

where K is the gain. Of course there are conflicting design trade-offs: Phase lead controllers increase bandwidth making the system vulnerable to noise and phase lag controllers can deteriorate the transient response.

The block diagram framework laid out above allows the combination of different blocks in an attempt to achieve a certain transfer function. With this in mind the block diagram in Figure 4 was developed which has the transfer function:

$$H(s) = B_2^* \frac{s + \alpha \pm \frac{\gamma_C^* B_1^*}{B_2^*}}{s + \alpha}, \tag{15}$$

where γ_C^* is the controller promoter gain of the controller, the associated gains of the downstream promoters are elements of the downstream input matrix $\mathbf{B}^* = [B_1^*, B_2^*]$ derived from (11). This transfer function does indeed have the form of a phase lag controller when the error activates expression of the state protein X and γ_C^* takes the positive sign. A genetic circuit diagram of the phase lag controller is simple and takes the form seen in Figure 5, where the error fed into the controller is denoted E.

When the error represses expression of the state protein X, the transfer function represents a phase lead controller. This corresponds to the negative sign in Equation (15).

V. Example

This section describes the detailed design of a genetic phase lag controller in feedback with a single gene acting as plant. The biological topology is achieved using a protease to supply the desired negative feedback. The parameter values found during the design process can be obtained by engineering ribosome binding sites, promoter strengths, plasmid copy number, as well as degradation rates of the species involved. It must be underlined that the engineering process of tuning the implemented system is not trivial with current available techniques.

The above described mathematical framework is used to complete this task, as well as the MATLAB optimisation toolbox, and the design objectives are to:

- Minimise the rise time of the closed loop system.
- Minimise steady state error.
- Retain a reasonable phase margin to ensure robustness.

A. Modelling

The plant, i.e. the system we wish to control, is the activated expression of a gene modelled as

$$\frac{\mathrm{d}[\mathrm{Y}]}{\mathrm{d}t} = \beta_{P0} + \Gamma_1([\mathrm{X}]) - \alpha_P[\mathrm{Y}].$$

The controller is a phase lag compensator, which is put in cascade with the plant. The genetic phase lag controller requires a second promoter in front of the downstream gene to transfer information from the controller to the plant so that both the error and state protein can regulate it (see Fig. 5). This results in the following model for the full system:

$$\frac{\mathrm{d}[\mathrm{Y}]}{\mathrm{d}t} = \beta_{P0} + \Gamma_1([\mathrm{X}])\Gamma_2([\mathrm{E}]) - \alpha_P[\mathrm{Y}], \tag{16a}$$

$$\frac{\mathrm{d}[\mathrm{X}]}{\mathrm{d}t} = \beta_{C0} + \Gamma_3([\mathrm{E}]) - \alpha_C[\mathrm{X}], \tag{16b}$$

where β_{P0} and β_{C0} are basal expression rates of the plant and controller respectively and α_P and α_C are the degradation rates of the plant and controller respectively. All TFs are activators so the subscript A has been dropped from the Γ functions for clarity. In addition to this, their dependency has been stated explicitly. It is through the error that unit feedback is supplied:

$$[\mathrm{E}] = \begin{cases} [\mathrm{Y}_{ref}] - [\mathrm{Y}] & \text{if } [\mathrm{E}] \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where the external reference to the system is $[\mathrm{Y}_{ref}]$.

A full block diagram of this model can be seen in Figure 6(a), where the functions in the promoter gain blocks have been calculated using Equation (11). This diagram allows us to calculate the transfer function from the error signal $E(s)$ to the output $Y(s)$, which we can then use to design our system. The full form of this open loop transfer function is

$$H(s) = \frac{Y(s)}{E(s)} = \Gamma_1^*\gamma_2^* \frac{s + \alpha_C + \frac{\gamma_1^*\Gamma_2^*\gamma_3^*}{\Gamma_1^*\gamma_2^*}}{(s+\alpha_C)(s+\alpha_P)}. \tag{17}$$

For later reference, it is worth noting that the zero of transfer function (17) is

$$z = -\alpha_C - \frac{\gamma_1^*\Gamma_2^*\gamma_3^*}{\Gamma_1^*\gamma_2^*}. \tag{18}$$

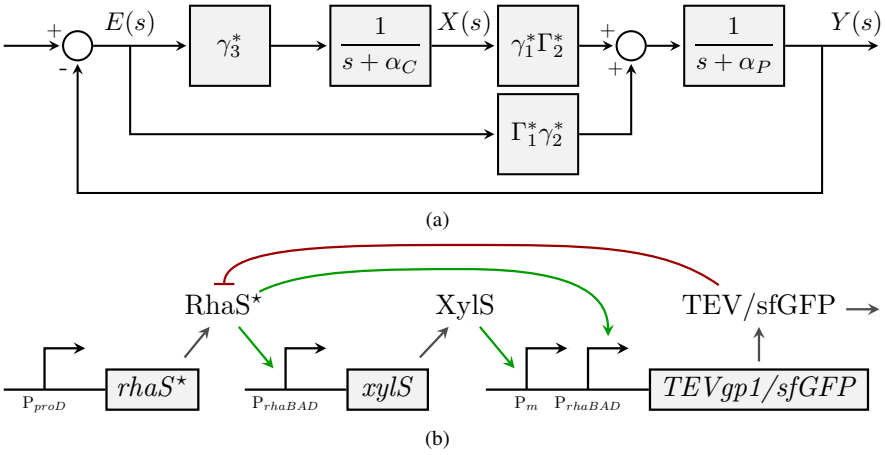

Fig. 6. (a): Block diagram showing the theoretical design example with the phase lag controller in cascade with the plant around which unit negative feedback is applied. (b): Biological implementation of the designed circuit using a protease to supply negative feedback.

B. Optimisation Framework

This section details the design process used to achieve the design objectives for the system. This process derives the parameters that define the controller and a certain input-output response. The optimisation that will be performed is formulated as follows:

$$\begin{aligned} \max_{\theta} \quad & H_\theta(0) \\ \text{s.t.} \quad & \mathbf{c}_{eq}(\theta) = \mathbf{0} \\ & \mathbf{c}(\theta) \leq \mathbf{0} \\ & \underline{\theta}_i \leq \theta_i \leq \overline{\theta}_i, \quad i = 1, \ldots, m \end{aligned}$$

The cost function used is the open loop low frequency gain of the open loop system given by:

$$H_\theta(0) = \frac{Y(0)}{E(0)} = \frac{\Gamma_1^*\gamma_2^*\alpha_C + \gamma_1^*\Gamma_2^*\gamma_3^*}{\alpha_C\alpha_P}. \tag{20}$$

The equality constraints are given by:

$$\mathbf{c}_{eq}(\theta) = \begin{bmatrix} \beta_{P0} + \Gamma_1^*\Gamma_2^* - \alpha_P y^* \\ \beta_{0C} + \beta_{C0} + \Gamma_3^* - \alpha_C x^* \end{bmatrix}.$$

These ensure that the optimisation is performed around the equilibrium of the ODE model of the system (16). As the Laplace transform and linearisation are performed about the equilibrium and the equilibrium is dependent on the parameters, this needs to be recalculated at every iteration of the optimisation.

The inequality constraints on the system aid the placement of the zero of the transfer function (18) with the aim of pushing it to the left, far away from the origin, without degrading the phase too much. The constraint is given by

$$\mathbf{c} = \lambda\alpha_P - \alpha_C - \frac{\gamma_1^*\Gamma_2^*\gamma_3^*}{\Gamma_1^*\gamma_2^*},$$

where $\lambda\alpha_P$ fixes an upper bound on the position of the zero in the transfer function.

The decision variables are

$$\theta = \begin{bmatrix} y^* & x^* & \beta_2 & \beta_3 & \alpha_C \end{bmatrix}^T,$$

where the first two y^* and x^* are the equilibrium, β_2 is the strength of the promoter at the plant with which the error signal interacts, β_3 is the strength of the promoter at the controller with which the error signal interacts and α_C is the degradation rate of the controller protein X. The bounds on these decision variables are given in Table I, and have been chosen so as to be biologically achievable.

The optimisation is then performed with the parameter values given in Table II held constant. The results produced a system with the equilibrium at $(y^*, x^*) = (0.89\text{nM}, 0.46\text{nM})$ with the parameters $\beta_2 = \beta_3 = 15\text{nMh}^{-1}$ and $\alpha_C = 1.85\text{h}^{-1}$. The uncontrolled system that is used as a benchmark for comparison is similar to the controlled system excepting the removal of the controller gene. The error is simply fed straight to the plant without being processed. The low frequency gain of the controlled system is 21.5dB higher than in uncontrolled case and the controlled system has a phase margin of 84.6°.

The bode plots of the controlled and uncontrolled systems can be seen in Figure 7 and step responses to a 10% increase in reference can be seen in Figure 8. The reduced steady state error can clearly be seen in the step response, and the other obvious effect of the phase lag controller is the slight overshoot and the underdamped response. The system has a reasonable phase margin and the rise time of the system has also been reduced.

An obvious next line of inquiry would be to investigate how the linearisation compares to the nonlinear system. The comparison is omitted due to lack of space, however, using Lyapunov theory [25], [26] and sum-of-squares programming [27] the worst case difference (over all initial conditions) between the two approaches was shown to be small. Figure 8 compares simulations of the linear and nonlinear models and shows that they are in good agreement. For specific instances, we refer the reader to the Section 3 of the Supplementary Information [24], where the nonlinear system has been subjected to

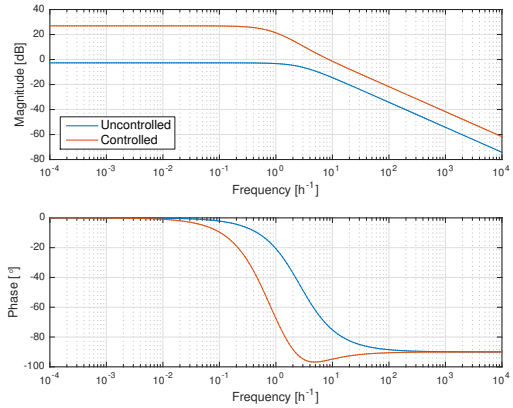

Fig. 7. Bode magnitude and phase plots showing both the controlled and uncontrolled system.

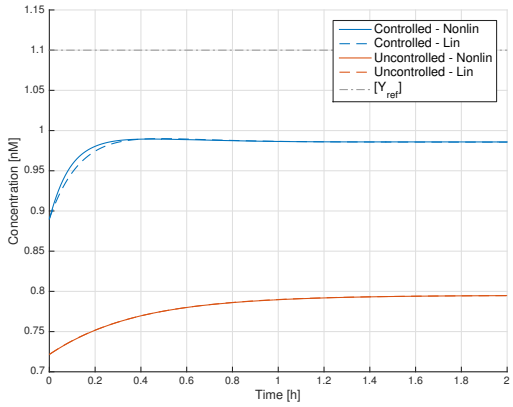

Fig. 8. Step responses of the controlled and uncontrolled systems both linear and nonlinear to a 10% increase in reference, y_{ref}.

perturbations of an order of magnitude greater than those valid for the linear analysis.

C. Biological Implementation

A biological implementation of this theoretical system, consisting of three genetic modules is proposed; see Figure 6(b). The *TEVgp1/sfGFP* gene encodes a fusion polypeptide between the *Tobacco etch virus* P1 protease (TEV protease; the plant) [28] and superfolder Green Fluorescent Protein (sfGFP; allows quantification of output) [29]. Expression of this gene is induced through two promoters: the meta-cleavage pathway operon promoter

TABLE I
THE BOUNDS ON THE DECISION VARIABLES EMPLOYED IN THE OPTIMISATION IN SECTION V-B.

Variable	Lower bound	Upper bound
y^*	0nM	100nM
x^*	0nM	100nM
β_2	0nMh^{-1}	15nMh^{-1}
β_3	0nMh^{-1}	15nMh^{-1}
α_C	0.7h^{-1}	13h^{-1}

TABLE II
THE THE VALUES OF THE CONSTANT PARAMETERS IN THE OPTIMISATION IN SECTION V-B.

Parameter	Value	Parameter	Value
β_{P0},β_{C0}	0.15nMh^{-1}	α_P	0.7h^{-1}
β_1	15nMh^{-1}	y_{ref}	1nM
K_1,K_2,K_3	0.5	λ	5
n_1,n_2,n_3	2		

(P$_m$) from *Pseudomonas putida*, which is activated by the XylS transcription factor in the presence of benzoate or m-toluate [30]; and the L-rhamnose inducible promoter (P$_{rhaBAD}$), which is activated by the RhaS transcription factor [31]–[33].

TEV protease is a cysteine protease with a molecular mass of 49 kDa [34], [35]. Its native role in the life cycle of the virus is to cleave the large 346 kDa polyprotein at specific sites into its mature protein products [28], [36]. The consensus recognition site for this enzyme is the sequence: Glu-Asn-Leu-Tyr-Phe-Gln-(Gly/Ser) (ENLYFQ(G/S)), with proteolysis occurring between the Gln and Gly/Ser residues [34]. It is also regularly used to direct proteolysis of various targeted proteins [37], [38]. To provide the desired negative feedback, the TEV-protease recognition site is inserted into a region of the positive transcription factor RhaS. Specifically, the cleavage site is inserted at residue 167 of RhaS, which lies between the N-terminal domain responsible for L-rhamnose and dimerisation and the C-terminal domain responsible for DNA and σ_{70} (sigma 70) binding binding [39]. This should ensure minimal disruption to the function of RhaS and accessibility of the recognition site to TEV protease.

The *rhaS** gene is the source of this engineered version of the activator RhaS (denoted RhaS* to indicate that it has been engineered) and is expressed constitutively at a high level using the engineered *proD* promoter (P$_{proD}$) [40]. The *xylS* gene is the controller module and expresses the activator XylS, through induction of P$_{rhaBAD}$ by the unprocessed RhaS* variant. The reference signal input to the system, mathematically denoted y_{ref}, can be thought of the activator RhaS* prior to digestion.

Once constructed and tested, this system would be optimised through parameter estimating, changing and tuning parts, etc, until the desired output is reached.

VI. CONCLUSION

In this paper we have presented a framework for designing genetic feedback controllers for improving the performance and robustness of genetic networks. Biological realisations of the classical phase lead and lag controllers have been presented. We have illustrated the 'design' approach in detail using an example, and discussed a starting point as to how this could be implemented in practice. Future work will consider other possible types of realisable genetic controllers and further develop the linear design framework to this end.

ACKNOWLEDGMENT

The authors would like to thank Dr Ali Raza Awan for discussions on biological implementation, as well as the Engineering and Physical Sciences Research Council for funding this research under grants EP/I031944/1 and EP/M002454/1. James Anderson acknowledges St. John's College Oxford for funding through a Junior Research Fellowship and Andreas Harris acknowledges funding through the EPSRC Systems Biology Doctoral Training Centre at the University of Oxford.

REFERENCES

[1] P. E. Purnick and R. Weiss, "The second wave of synthetic biology: from modules to systems," *Nature reviews Molecular cell biology*, vol. 10, no. 6, pp. 410–422, 2009.

[2] J. Anderson, N. Strelkowa, G.-B. Stan, T. Douglas, J. Savulescu, M. Barahona, and A. Papachristodoulou, "Engineering and ethical perspectives in synthetic biology," *EMBO reports*, vol. 13, no. 7, pp. 584–590, 2012.

[3] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, "Synthetic biology: new engineering rules for an emerging discipline," *Molecular Systems Biology*, vol. 2, no. 1, 2006.

[4] R. Kwok, "Five hard truths for synthetic biology," *Nature*, vol. 463, pp. 288–290, 2010.

[5] T.-M. Yi, Y. Huang, M. I. Simon, and J. Doyle, "Robust perfect adaptation in bacterial chemotaxis through integral feedback control," *Proceedings of the National Academy of Sciences*, vol. 97, no. 9, pp. 4649–4653, 2000.

[6] H. Kurata, H. El-Samad, R. Iwasaki, H. Ohtake, J. C. Doyle, I. Grigorova, C. A. Gross, and M. Khammash, "Module-based analysis of robustness tradeoffs in the heat shock response system," *PLoS Computational Biology*, vol. 2, no. 7, p. e59, 2006.

[7] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, 1st ed. Chapman & Hall/CRC, 2007.

[8] B. Ingalls, *Mathematical Modeling in Systems Biology*. The MIT Press, 2013.

[9] T. P. Prescott and A. Papachristodoulou, "Synthetic biology: A control engineering perspective," in *European Control Conference*. IEEE, 2014, pp. 1182–1186.

[10] T. Afroz and C. L. Beisel, "Understanding and exploiting feedback in synthetic biology," *Chemical Engineering Science*, vol. 103, pp. 1–12, Nov. 2013.

[11] Y.-J. Shin and L. Bleris, "Linear control theory for gene network modeling." *PloS one*, vol. 5, no. 9, pp. 1–16, Jan. 2010.

[12] J. A. J. Arpino, E. J. Hancock, , J. Anderson, M. Barahona, G.-B. V. Stan, A. Papachristodoulou, and K. Polizzi, "Tuning the dials of synthetic biology," *Microbiology*, vol. 159, no. Pt 7, pp. 1236–1253, 2013.

[13] B. Yordanov, J. Kim, R. L. Petersen, A. Shudy, V. V. Kulkarni, and A. Phillips, "Computational design of nucleic acid feedback control circuits," *ACS Synthetic Biology*, July 2014.

[14] Y.-J. Chen, N. Dalchau, N. Srinivas, A. Phillips, L. Cardelli, D. Soloveichik, and G. Seelig, "Programmable chemical controllers made from DNA." *Nature nanotechnology*, vol. 8, no. 10, pp. 755–62, Oct. 2013.

[15] J. C. Doyle, B. A. Francis, and A. Tannenbaum, *Feedback control theory*. Macmillan Publishing Company New York, 1992, vol. 1.

[16] N. Rosenfeld, M. B. Elowitz, and U. Alon, "Negative autoregulation speeds the response times of transcription networks," *Journal of molecular biology*, vol. 323, no. 5, pp. 785–793, 2002.

[17] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in *Escherichia coli*," *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.

[18] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 6767, pp. 335–338, 2000.

[19] Y.-C. Chang, J. P. Armitage, A. Papachristodoulou, and G. H. Wadhams, "A single phosphatase can convert a robust step response into a graded, tunable or adaptive response," *Microbiology*, vol. 159, no. Pt 7, pp. 1276–1285, 2013.

[20] A. Milias-Argeitis, S. Summers, J. Stewart-Ornstein, I. Zuleta, D. Pincus, H. El-Samad, M. Khammash, and J. Lygeros, "*In silico* feedback for *in vivo* regulation of a gene expression circuit," *Nature biotechnology*, vol. 29, no. 12, pp. 1114–1116, 2011.

[21] K. Oishi and E. Klavins, "Biomolecular implementation of linear I/O systems," *IET Systems Biology*, vol. 5, no. 4, pp. 252–260, 2011.

[22] E. J. Hancock, G.-B. Stan, J. A. J. Arpino, and A. Papachristodoulou, "Simplified mechanistic models of gene regulation for analysis and design," *J. R. Soc. Interface*, 2015.

[23] J. Dolan, J. Anderson, and A. Papachristodoulou, "A loop shaping approach for designing biological circuits." in *IEEE Conference on Decision and Control*, 2012, pp. 3614–3619.

[24] A. W. K. Harris, J. A. Dolan, C. L. Kelly, J. Anderson, and A. Papachristodoulou, "Supplementary information," 2015. [Online]. Available: http://sysos.eng.ox.ac.uk/HarDKAP15_supplement.pdf

[25] A. Papachristodoulou and S. Prajna, "A tutorial on sum of squares techniques for systems analysis," in *American Control Conference, 2005. Proceedings of the 2005*, 2005, pp. 2686 – 2700.

[26] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Prentice Hall, Inc., 2001.

[27] A. Papachristodoulou, J. Anderson, G. Valmorbida, P. Seiler, and P. A. Parrilo, "SOSTOOLS – Sum of Squares Optimization Toolbox, User's Guide," http://www.eng.ox.ac.uk/control/sostools, 2013.

[28] R. F. Allison, J. C. Sorenson, M. E. Kelly, F. B. Armstrong, and W. G. Dougherty, "Sequence determination of the capsid protein gene and flanking regions of *Tobacco Etch Virus*: Evidence for synthesis and processing of a polyprotein in potyvirus genome expression," *Proceedings of the National Academy of Sciences*, vol. 82, no. 12, pp. 3969–3972, 1985.

[29] J.-D. Pédelacq, S. Cabantous, T. Tran, T. C. Terwilliger, and G. S. Waldo, "Engineering and characterization of a superfolder green fluorescent protein," *Nature biotechnology*, vol. 24, no. 1, pp. 79–88, 2006.

[30] N. Mermod and K. N. Timmis, "Construction of a broad host range vector for the regulated expression of cloned genes in a range of gram-negative bacteria," *Cellular and Molecular Life Sciences*, vol. 42, no. 1, pp. 105–105, 1986.

[31] J. F. Tobin and R. F. Schleif, "Positive regulation of the *Escherichia coli* L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes," *Journal of molecular biology*, vol. 196, no. 4, pp. 789–799, 1987.

[32] ——, "Transcription from the *rha* operon p_{sr} promoter," *Journal of Molecular Biology*, vol. 211, pp. 1 – 4, 1990.

[33] M. J. Giacalone, A. M. Gentile, B. T. Lovitt, N. L. Berkley, C. W. Gunderson, and M. W. Surber, "Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system," *Biotechniques*, vol. 40, no. 3, p. 355, 2006.

[34] J. C. Carrington and W. G. Dougherty, "A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing," *Proceedings of the National Academy of Sciences*, vol. 85, no. 10, pp. 3391–3395, 1988.

[35] ——, "Processing of the tobacco etch virus 49k protease requires autoproteolysis," *Virology*, vol. 160, no. 2, pp. 355–362, 1987.

[36] T. A. Smith and B. D. Kohorn, "Direct selection for sequences encoding proteases of known specificity." *Proceedings of the National Academy of Sciences*, vol. 88, no. 12, pp. 5159–5162, 1991.

[37] M. Mondigler and M. Ehrmann, "Site-specific proteolysis of the *Escherichia coli* SecA protein *in vivo*." *Journal of bacteriology*, vol. 178, no. 10, pp. 2986–2988, 1996.

[38] R. B. Kapust and D. S. Waugh, "Controlled intracellular processing of fusion proteins by TEV protease," *Protein expression and purification*, vol. 19, no. 2, pp. 312–318, 2000.

[39] J. R. Wickstrum and S. M. Egan, "Amino acid contacts between sigma 70 domain 4 and the transcription activators RhaS and RhaR," *Journal of bacteriology*, vol. 186, no. 18, pp. 6277–6285, 2004.

[40] J. H. Davis, A. J. Rubin, and R. T. Sauer, "Design, construction and characterization of a set of insulated bacterial promoters," *Nucleic acids research*, p. gkq810, 2010.