BIOINFORMATICS ORIGINAL PAPER Vol. 27 no. 19 2011, pages 2700–2706
doi:10.1093/bioinformatics/btr465

Systems biology

Advance Access publication August 17, 2011

Robust synthetic gene network design via library-based search method

Chih-Hung Wu[1,*], Hsiao-Ching Lee[2] and Bor-Sen Chen[1,*]

[1]Lab of Systems Biology, Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan 30013 and [2]Department of Biological Science and Technology, National Chiao Tung University, Hsinchu 30068, Taiwan, R.O.C.

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Synthetic biology aims to develop the artificial gene networks with desirable behaviors using systematic method. These networks with desired behaviors could be constructed using diverse biological parts, which may limit the development to complex synthetic gene networks. Fortunately, some well-characterized promoter libraries for engineering gene networks are widely available. Thus, a synthetic gene network can be constructed by selecting adequate promoters from promoter libraries to achieve the desired behaviors. However, the present promoter libraries cannot be directly applied to engineer a synthetic gene network. In order to efficiently select adequate promoters from promoter libraries for a synthetic gene network, promoter libraries are needed to be redefined based on the dynamic gene regulation.

Results: Based on four design specifications, a library-based search method is proposed to efficiently select the most adequate promoter set from the redefined promoter libraries by a genetic algorithm (GA) to achieve optimal reference tracking design. As the number and size of promoter libraries increase, the proposed method can play an important role in the systematic design of synthetic biology.

Contact: g883743@alumni.nthu.edu.tw; bschen@ee.nthu.edu.tw

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received on May 17, 2011; revised on July 19, 2011; accepted on July 29, 2011

1 INTRODUCTION

Synthetic biology seeks the improvement of rational design and reprogramming of biological systems based on design principles. During the past decade, synthetic biology had significant progress in designing biological parts. Synthetic biologists have assembled these parts into genetic circuits to achieve basic functions such as toggle switches (Gardner *et al.*, 2000; Kramer *et al.*, 2004), oscillators (Atkinson *et al.*, 2003; Goh *et al.*, 2008; Stricker *et al.*, 2008; Tigges *et al.*, 2009), pulse generators (Basu *et al.*, 2004), genetic counters (Friedland *et al.*, 2009), logic evaluators (Rinaudo *et al.*, 2007; Win and Smolke, 2008), filters (Hooshangi *et al.*, 2005; Sohka *et al.*, 2009a, b), sensors (Kobayashi *et al.*, 2004; Win and Smolke, 2007) and cell–cell communicators (Basu *et al.*, 2005; Kobayashi *et al.*, 2004; Pai *et al.*, 2009). Recently, the design concepts of next-generation gene networks, such as tunable filters, analog-to-digital and digital-to-analog converters, adaptive learning networks and protein-based computational circuits, have been proposed to enable the construction of more complex biological systems (Lu *et al.*, 2009). With the full understanding of biological parts and modules, synthetic biologists can construct useful next-generation synthetic gene networks with real-world applications in medicine, biotechnology, bioremediation and bioenergy (Lu *et al.*, 2009).

At present, engineered organisms contain only very simple genetic circuits since there are problems for assembling biological parts into more complex genetic circuits. For example, although standard biological parts can be obtained from the Registry of Standard Biological Parts (http://partsregistry.org), only a few parts have measured characteristics describing quantitative behavior in the composed devices (Kelly *et al.*, 2009). There is still lack of an efficient method to select adequate biological parts from usable libraries to engineer a genetic circuit to track the desired reference trajectory (Wu *et al.*, 2011).

In this article, we first redefine the presented promoter libraries based on promoter activity, and then a proposed library-based search method is used to efficiently select adequate promoters through their promoter activities in order to engineer a synthetic gene network with desired behaviors despite intrinsic fluctuations and environmental disturbances on the host cell. The proposed library-based search method can solve the global optimization problem by employing genetic algorithm (GA) (Chen and Chen, 2010; Goldberg, 1989; Grefenstette, 1986; Holland, 1992; Katayama and Narihisa, 2000; Renders and Flasse, 1996). Hence, the most adequate promoters from redefined promoter libraries can be efficiently selected to achieve a desired reference trajectory. Finally, two design examples are given to illustrate the design procedure of the proposed method *in silico*. In the future, synthetic biologists can employ the proposed library-based search method to select the most adequate promoter set from redefined promoter libraries to construct a predictable synthetic gene network.

2 METHODS

In order to efficiently select adequate promoters from different promoter libraries for synthetic gene networks to achieve desired behaviors, the promoter libraries should be redefined based on the promoter activities according to the promoter characteristics and experimental data in the present promoter libraries. The design procedure for a synthetic gene network is shown in Figure 1, including following steps: (i) characterize the redefined promoter libraries based on the promoter activities; (ii) construct a stochastic

*To whom correspondence should be addressed.

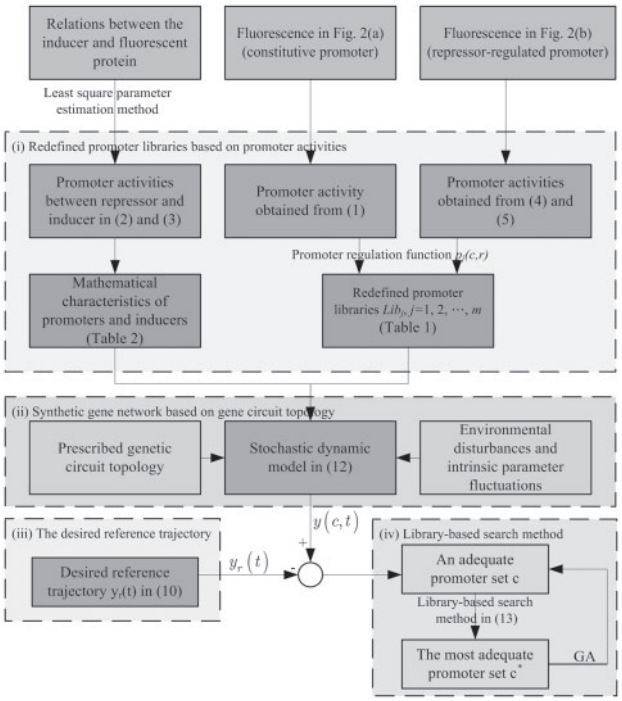

Fig. 1. The design procedure for synthetic gene networks. The design procedure can divide four steps: (i) construct redefined promoter libraries based on dynamic gene regulation; (ii) construct a synthetic gene network from prescribed genetic circuit topology; (iii) provide a desired reference trajectory to be tracked; (iv) the most adequate promoter set c^* can be obtained using library-based search method through GA.

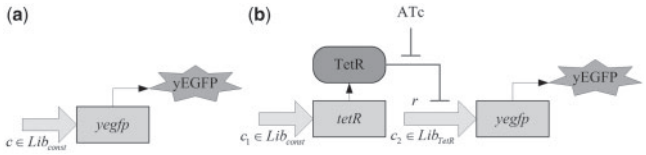

Fig. 2. The schematic diagram for measuring the promoter activities of promoters including (**a**) constitutive promoter library $\text{Lib}_{\text{const}}$ and (**b**) repressor-regulated promoter library Lib_{TetR}.

dynamic model based on the prescribed genetic circuit topology; (iii) provide the desired reference trajectory to be tracked; and (iv) employ the library-based search method to select the most adequate promoter set c^* by employing GA.

2.1 Redefinition of promoter libraries based on promoter activities

In general, two types of promoters are used to regulate gene expression in gene networks, i.e. constitutive- and repressor-regulated promoters. Their promoter activities of them are both indirectly measured by the fluorescent protein (Canton *et al.*, 2008; Murphy *et al.*, 2007), and a dynamic model can be used to indirectly evaluate the promoter activities (Ellis *et al.*, 2009; Kelly *et al.*, 2009; Leveau and Lindow, 2001). The scheme of indirectly measuring the promoter activities of promoters is shown in Figure 2. It is worth to note that activatable promoters can also be used in synthetic gene networks. Since activatable promoters, however, are not used in this study, their promoter libraries are not mentioned here. The related discussion of promoter libraries for activatable promoters is mentioned in the Supplementary Material.

Table 1. The redefined constitutive, TetR- and LacI-regulated promoter libraries (i.e. Lib_{const}, Lib_{TetR} and Lib_{LacI}) comprise different promoters (i.e. J_k, $k = 0, \ldots, 18$, T_k and L_k, $k = 0, \ldots, 20$) with their corresponding promoter activities of c_s and c_r obtained from experimental data of previous promoter libraries (Ellis *et al.*, 2009) (See Supplementary Material)

Lib_{TetR}			Lib_{LacI}			Lib_{const}	
	c_s	c_r		c_s	c_r		c_s
T_0	2121	0.1724	L_0	1657.5	0.3018	J_0	5901.40
T_1	1604	0.7576	L_1	923.97	0.2567	J_1	5048.74
T_2	1376.6	0.1936	L_2	860.87	0.2244	J_2	4242.43
T_3	1169.8	0.4672	L_3	674.92	1.9189	J_3	4149.75
T_4	974.52	0.0753	L_4	651.58	1.1680	J_4	3445.38
T_5	942.77	0.2281	L_5	570.07	3.5062	J_5	3310.99
T_6	967.17	0.1493	L_6	527.83	0.5497	J_6	3019.05
T_7	738.57	0.0702	L_7	323.45	0.1248	J_7	2745.64
T_8	641.74	0.7135	L_8	327.77	0.1772	J_8	2103.84
T_9	564.24	0.2620	L_9	309.74	0.5439	J_9	1955.55
T_{10}	501.35	0.0756	L_{10}	298.35	0.1146	J_{10}	1443.49
T_{11}	469.35	0.0788	L_{11}	250.16	0.1326	J_{11}	917.53
T_{12}	466.16	0.1636	L_{12}	248.03	0.1171	J_{12}	896.68
T_{13}	356.88	0.0927	L_{13}	239.32	0.1010	J_{13}	593.15
T_{14}	348.95	0.1483	L_{14}	190.2	0.0959	J_{14}	375.35
T_{15}	274.79	0.1067	L_{15}	163.84	0.4813	J_{15}	245.60
T_{16}	250.04	0.0857	L_{16}	166.42	0.0989	J_{16}	48.66
T_{17}	188.77	0.1366	L_{17}	131.63	0.1190	J_{17}	39.39
T_{18}	119.57	0.0753	L_{18}	108.96	0.0903	J_{18}	2.32
T_{19}	111.57	0.1185	L_{19}	101.89	0.0982		
T_{20}	70.909	0.1606	L_{20}	85.673	0.2174		

In order to redefine the constitutive promoter library, $\text{Lib}_{\text{const}}$, a dynamic model in Figure 2a for a promoter c in $\text{Lib}_{\text{const}}$ can be built. The promoter activity of c can be determined by assuming the dynamic model at a steady state (Kelly *et al.*, 2009)

$$c = \frac{\beta}{\eta}\frac{\gamma_{\text{yEGFP}}}{\alpha} y_{ss}, \tag{1}$$

where y_{ss} denotes fluorescence of the gene network at the steady state; β and γ_{yEGFP} denote the degradation rates of mRNA *yegfp* and protein yEGFP, respectively. α denotes the translation rate; η is the ratio of fluorescence to the concentration of yEGFP (Hooshangi *et al.*, 2005; Leveau and Lindow, 2001).

In this study, we redefine the constitutive promoters BBa_J23100-BBa_J23118 in the Registry of Standard Biological Parts (http://partsregistry.org) to build a constitutive promoter library $\text{Lib}_{\text{const}}$ and denote as J_0–J_{18} listed in Table 1. The detailed procedure of evaluating the promoter activity is shown in Supplementary Material.

Since a repressor-regulated promoter is regulated by the repressor, it has always the minimum and maximum expression with and without the repressor binding, respectively. Therefore, a promoter in repressor-regulated promoter libraries has two promoter activities, with and without saturating concentration of repressor binding.

In Figure 2b, for example, a constitutive promoter expresses the TetR repressor, which is regulated by anhydrotetracycline (ATc) and represses the TetR-regulated promoter. In order to simplify the illustration, we denote the repressor activity r of TetR as the following form (Alon, 2007; Nakanishi *et al.*, 2008):

$$r = \frac{\text{TetR}}{1 + (\text{ATc}/K_{\text{ATc}})^{n_{\text{ATc}}}}, \tag{2}$$

Table 2. Mathematical characteristics of promoters and their regulation parameters

Parameter	Description	Value	Units	Ref.
K_{TetR}	TetR binding affinity	7.3093	M	+
K_{LacI}	LacI binding affinity	60.1405	M	+
K_{ATc}	ATc-TetR dissociation rate	26.3236	ng/ml	+
K_{IPTG}	IPTG-LacI dissociation rate	0.0598	mM	+
n_{TetR}	Binding co-operativity between TetR and DNA	2	–	Braun *et al.* (2005)
n_{LacI}	Binding co-operativity between LacI and DNA	2	–	Iadevaia and Mantzariz (2006)
n_{ATc}	Binding co-operativity between ATc and TetR	4	–	Braun *et al.* (2005)
n_{IPTG}	Binding co-operativity between IPTG and LacI	1	–	Iadevaia and Mantzariz (2006)
γ_{yEGFP}	yEGFP degradation rate	0.001 925	1/min	Ellis *et al.* (2009)
γ_{TetR}	TetR degradation rate	0.1386	1/min	Tuttle *et al.* (2005)
γ_{LacI}	LacI degradation rate	0.1386	1/min	Tuttle *et al.* (2005)
γ_{CI}	CI degradation rate	0.042	1/min	Arkin *et al.* (1998)
β	mRNA degradation rate	0.288	1/min	Canton *et al.* (2008) and Kelly *et al.* (2009)
α	Translation rate	24	1/min	Canton *et al.* (2008) and Kelly *et al.* (2009)

+ All parameter values are obtained from empirical studies in the literature or estimated via experimental data from Ellis *et al.* (2009).

where K_{ATc} denotes the ATc-TetR dissociation rate, and n_{ATc} denotes the binding co-operativity between ATc and TetR. Then we denote the promoter regulation function $p_{\mathrm{TetR}}(c,r)$ as (Ellis *et al.*, 2009; Nakanishi *et al.*, 2008; Semsey *et al.*, 2009)

$$\begin{aligned} &p_{\mathrm{TetR}}(c,r)=c_r+\tfrac{c_s-c_r}{1+(r/K_{\mathrm{TetR}})^{n_{\mathrm{TetR}}}},\\ &c=\{c_r,c_s\}\in \mathrm{Lib}_{\mathrm{TetR}} \end{aligned} \tag{3}$$

where K_{TetR} and n_{TetR} denote the TetR-DNA binding affinity and the binding co-operativity between regulatory protein TetR and DNA, respectively. The promoter c in $\mathrm{Lib}_{\mathrm{TetR}}$ has two promoter activities c_r and c_s, i.e. $c=\{c_r,c_s\}\in \mathrm{Lib}_{\mathrm{TetR}}$. The promoter regulation function $p_{\mathrm{TetR}}(c,r)$ in (3) has the minimum value c_r and maximum value c_s with and without the saturating concentration of the repressor activity in (2). Finally, the promoter activities c_r and c_s of c can be determined as

$$c_r=p^{\min}_{\mathrm{TetR}}(c,r)=\frac{\beta}{\eta}\frac{\gamma_{\mathrm{yEGFP}}}{\alpha}y_{\min}(c,t) \tag{4}$$

$$c_s=p^{\max}_{\mathrm{TetR}}(c,r)=\frac{\beta}{\eta}\frac{\gamma_{\mathrm{yEGFP}}}{\alpha}y_{\max}(c,t), \tag{5}$$

where $y_{\min}(c,t)$ and $y_{\max}(c,t)$ denote the minimum and maximum fluorescence at the steady state, and γ_{yEGFP} denotes the degradation rate of yEGFP. The promoter activities c_r and c_s of each promoter c in $\mathrm{Lib}_{\mathrm{TetR}}$ are shown in Table 1.

Furthermore, the parameters K_{ATc} and n_{ATc} from the repressor activity in (2) and K_{TetR} and n_{TetR} from the promoter regulation function in (3) are identified and listed in Table 2.

The promoter activity $c=\{c_r,c_s\}$ of each promoter in the LacI-regulated promoter library $\mathrm{Lib}_{\mathrm{LacI}}$ can be also determined by a similar process, as listed in Table 1. The detailed procedure for redefining the promoter libraries is shown in Supplementary Material.

Finally, the promoter regulation function can be extended to illustrate the promoter activity in the constitutive promoter. The promoter regulation function for a constitutive promoter can be rewritten and described as

$$p_{\mathrm{const}}(c,0)=c_r+\frac{c_s-c_r}{1+(0/K_{\mathrm{const}})^{n_{\mathrm{const}}}}=c_s \tag{6}$$

2.2 Construction of synthetic gene network based on circuit topology

Consider the synthetic two-stage transcriptional cascade in Figure 3 (Hooshangi *et al.*, 2005). The *yegfp* expression is controlled by LacI protein, which can be repressed by the TetR repressor. TetR is expressed constitutively

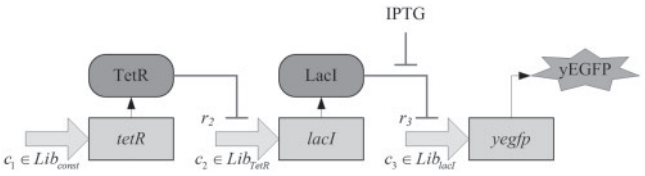

Fig. 3. A simple synthetic transcriptional cascade. The protein TetR, expressed constitutively from the constitutive promoter c_1, inhibits the production of LacI by binding the TetR-regulated promoter c_2, and LacI inhibits the production of yEGFP by binding the LacI-regulated promoter c_3.

from the promoter selected from $\mathrm{Lib}_{\mathrm{const}}$. When IPTG is added to the growth medium, IPTG binds to the LacI repressor and then induces the *yegfp* expression. Suppose $x_1(c_1,t), x_2(c_2,t), x_3(c_3,t), X_1(c_1,t), X_2(c_2,t)$ and $X_3(c_3,t)$ as the concentrations of mRNA *tetR*, *lacI*, *yegfp* and proteins TetR, LacI, and yEGFP, respectively. The dynamic model in Figure 3 with promoters c_1, c_2 and c_3, selected from the promoter library $\mathrm{Lib}_{\mathrm{const}}$, $\mathrm{Lib}_{\mathrm{TetR}}\mathrm{Lib}_{\mathrm{LacI}}$, respectively, is modeled as follows (Gardner *et al.*, 2000; Kobayashi *et al.*, 2003)

$$\begin{aligned} &\begin{cases} \dot{x}_i(c_i,t)=p_j(c_i,r_i)-\beta x_i(c_i,t)\\ \dot{X}_i(c_i,t)=\alpha x_i(c_i,t)-\gamma_i X_i(c_i,t) \end{cases}\\ &y(c,t)=X_3(c_3,t),\ i=1,2,3\\ &p_j(c_i,r_i)\in\{p_{\mathrm{const}}(c_1,0),p_{\mathrm{TetR}}(c_2,X_1),p_{\mathrm{LacI}}(c_3,X_2)\}\\ &c=\{c_1,c_2,c_3\}\in\{\mathrm{Lib}_{\mathrm{const}},\mathrm{Lib}_{\mathrm{TetR}},\mathrm{Lib}_{\mathrm{LacI}}\} \end{aligned} \tag{7}$$

where $y(c,t)$ denotes the desired concentration of yEGFP; $\{\gamma_1,\gamma_2,\gamma_3\}=\{\gamma_{\mathrm{TetR}},\gamma_{\mathrm{LacI}},\gamma_{\mathrm{yEGFP}}\}$; and the three promoters c_1, c_2 and c_3 are considered as a promoter set, $c=\{c_1,c_2,c_3\}$.

In general, biological parts are inherently uncertain in this network. We assume that the promoter activities, degradation rates of mRNAs and proteins, translation rates are stochastically uncertain due to gene expression noises in transcriptional and translational processes, thermal fluctuations, DNA mutations and evolutions (Alon, 2007) as follows:

$$\begin{aligned} &c_i=\{c_{r_i},c_{s_i}\}\to c_i+\Delta c_i n_i(t)=\{c_{r_i},c_{s_i}\}+\{\Delta c_{r_i},\Delta c_{s_i}\}n_i(t)\\ &\beta\to\beta+\Delta\beta n_i(t),\alpha\to\alpha+\Delta\alpha n_i(t),\gamma_i\to\gamma_i+\Delta\gamma_i n_i(t),i=1,2,3 \end{aligned} \tag{8}$$

where Δc_{r_i}, Δc_{s_i}, $\Delta\gamma_i$, $\Delta\beta$ and $\Delta\alpha$ denote the amplitudes of stochastic parameter variations, and $n_i(t)$ is a Gaussian noise with zero mean and unit variance.

Synthetic gene networks often also suffer from environmental disturbances on the host cell, as follows:

$$\begin{cases} \dot{x}_i(c_i,t) = p_j(c_i,r_i) - \beta x_i(c_i,t) + (p_j(\Delta c_i,r_i) - \Delta\beta x_i(c_i,t))n_i(t) + v_i(t) \\ \dot{X}_i(c_i,t) = \alpha x_i(c_i,t) - \gamma_i X_i(c_i,t) + (\Delta\alpha x_i(c_i,t) - \Delta\gamma_i X_i(c_i,t))n_i(t) + v_i(t) \end{cases}$$
$$y(c,t) = X_3(c_3,t),\ i = 1, 2, 3$$
$$p_j(c_i,r_i) \in \{p_{\text{const}}(c_1,0), p_{\text{TetR}}(c_2,X_1), p_{\text{LacI}}(c_3,X_2)\}$$
$$c = \{c_1,c_2,c_3\} \in \{\text{Lib}_{\text{const}}, \text{Lib}_{\text{TetR}}, \text{Lib}_{\text{LacI}}\} \tag{9}$$

The design purpose is to select the most adequate promoter set $c^* = \{c_1^*, c_2^*, c_3^*\}$ from corresponding promoter libraries so that the observed output $y(c^*,t)$ can track the following desired reference trajectory.

$$y_r(t) = f_r(t) \tag{10}$$

Hence, in order to engineer a stochastic synthetic gene network in (9), which can robustly and optimally track a desired trajectory (10), a library-based search method is provided to efficiently select the most adequate promoter set $c^* = \{c_1^*, c_2^*, c_3^*\}$ from promoter libraries.

2.3 Synthetic gene network design: library-based search method

Consider a more general design for a synthetic gene network. We extend the above gene network in Figure 3 with promoters from constitutive, TetR- and LacI-regulated promoter libraries to an n-gene network with promoters from m promoter libraries $\text{Lib}_j, j = 1,2,\ldots m$ based on a prescribed genetic circuit topology. In the n-gene–gene network, the promoter regulation function for the promoter can be represented as follows:

$$\begin{aligned} p_j(c_i,r_i) &= c_{r_i} + \frac{c_{s_i} - c_{r_i}}{1 + (r_i/K_j)^{n_j}} \\ c_i &= \{c_{r_i}, c_{s_i}\} \in \text{Lib}_j, j = 1, 2, \ldots, m \end{aligned}, \tag{11}$$

where K_j and n_j are the corresponding parameters to be identified or estimated. For example, the parameters K_{TetR} and n_{TetR} are for the TetR-regulated promoter library estimated and listed in Table 2. Hence, a dynamic model of the n-gene circuit topology with n promoters selected from m promoter libraries with intrinsic parameter fluctuations and environmental disturbances is represented as

$$\begin{aligned} \dot{X}(c,t) &= f(X,c,t) + \Delta f(X,c,t)n(t) + v(t) \\ y(c,t) &= \text{H}X(c,t) \\ c &= \{c_1,c_2,\ldots,c_n\} \in \text{Lib}_j, j = 1, 2, \ldots, m \end{aligned}, \tag{12}$$

where $x_i(c_i,t)$ and $X_i(c_i,t)$ denote the concentrations of mRNA and proteins in the synthetic gene network, respectively. $X(c,t) = [x_1(c,t), X_1(c,t), \ldots, x_n(c,t), X_n(c,t)]^T$ denotes the concentrations of mRNAs and proteins in the synthetic gene network; $f(X,c,t)$ denotes the non-linear gene regulation of transcription and translation based on the circuit topology; $\Delta f(X,c,t)$ denotes the parameter fluctuations of the non-linear gene regulation $f(X,c,t)$; and $y(c,t) = [y_1(c,t), \ldots, y_p(c,t)]^T$ denotes the observed proteins we are interested in. H is a $p \times n$ output matrix determining what proteins we are interested in.

The stochastic dynamic system in (12) can be established from the prescribed genetic circuit topology, and the most adequate promoter set $c^* = \{c_1^*, c_2^*, \ldots, c_n^*\}$ selected from corresponding promoter libraries Lib_j, $j = 1, 2, \ldots, m$ can make the synthetic gene network track the prescribed desired trajectories via the library-based search method. For convenience in illustrating the proposed library-based search method, four design specifications are given as follows:

(1) Given the desired reference trajectory $y_r(t) = f_r(t)$, which dimension is consistent with the numbers of fluorescent proteins $y(c,t)$ in (12).

(2) Given the promoter libraries $\text{Lib}_j, j = 1, \ldots, m$ as in Table 1.

(3) Given the environmental disturbances $v_i(t)$ and standard deviations of promoter activities, $\Delta c_i = \{\Delta c_{r_i}, \Delta c_{s_i}\}$, of degradation rates of mRNAs and proteins, $\Delta\beta$ and $\Delta\gamma_i$, and of translation rates, $\Delta\alpha$, to be tolerated in the host cell. These are quantitative specifications of system robustness of synthetic gene networks in the host cell.

(4) Select a promoter set $c = \{c_1, c_2, \ldots, c_n\}$ from the promoter libraries $\text{Lib}_j, j = 1, \ldots, m$ to minimize the following cost function, i.e.

$$J(c) = E\int_0^{t_f} (y(c,t) - y_r(t))^T (y(c,t) - y_r(t))dt \tag{13}$$

If the cost function in (13) can be minimized by selecting the most adequate promoter set $c^* = \{c_1^*, c_2^*, \ldots, c_n^*\}$ under design specifications (1)–(4), then the observed output $y(c^*,t)$of the synthetic gene network can optimally and robustly track the prescribed reference trajectories $y_r(t)$ under the intrinsic parameter fluctuations and environmental disturbances on the host cells.

Using conventional optimal design methods, the many combinations of promoter sets required to satisfy four design specifications generally will waste a large amount of computation time and trail-and-error experiments. In this study, a library-based search method using a genetic algorithm (GA) is employed to select the most adequate promoter set c^* from corresponding promoter libraries. GA is a stochastic optimization algorithm, originally motivated by the mechanisms of natural selection and evolutionary genetics. GA has been proven to be efficient for solving constrained optimization problems in many areas (Grefenstette, 1986) and further details on GA can are available elsewhere (Goldberg, 1989). The analysis of computational complexity is shown in the Supplementary Material.

By the library-based search method and four design specifications, the most adequate promoter set c^* can be selected in design specification (2). Despite intrinsic parameter fluctuations and environmental disturbances in design specifications (3), the gene network can achieve the desired reference trajectory in design specification (1) by minimizing the cost function $J(c)$ in design specification (4). In fact, this library-based optimal reference tracking problem in (13) is a highly non-linear optimization problem. In conventional search algorithms, the optimal solution can be obtained, but it may not be the global optimization. The library-based search method using GA is an iterative procedure to select the most adequate promoter set $c^* = \{c_1^*, c_2^*, \ldots, c_n^*\}$ that satisfies design specifications (1)–(4). When the most adequate promoter set c^* is selected, design specifications (1)–(4) for the synthetic gene network can be satisfied, and then the synthetic gene network can track the desired reference trajectory robustly and optimally.

A design procedure for this library-based search method using GA is given as follows.

2.4 Design procedure

(1) Build redefined promoter libraries from the experimental data of maximum and minimum outputs and input–output relationship between the fluorescence, repressor and inducer.

(2) Construct a genetic circuit topology such as Figure 3 or 4 with promoters, and then build the stochastic dynamic model in (12) for the synthetic gene network.

(3) Provide the design specifications (i)–(iv) for the synthetic gene network in (12).

(4) Generate initial promoter sets from the redefined promoter libraries.

(5) Calculate the cost function $J(c)$ in (13) for each promoter set c in the population.

(6) Create offspring by GA operator (i.e. reproduction, cross-over and mutation).

(7) Evaluate the new promoter sets and calculate the cost function of each promoter set obtained by natural selection.

(8) Stop if the search goal is achieved, or an allowable generation is attained; else go to step 6.

In each iteration or generation of GA, these genetic operators are performed to generate new populations (i.e. promoter sets), and these new populations are evaluated via the cost function in (13). On the basis of

these genetic operators and evaluations, a better new population of candidate solutions is formed in each genetic generation.

3 RESULTS

In this section, two *in silico* design examples are given to illustrate the design procedure of the proposed library-based search method. First of all, synthetic biologists need to engineer a genetic circuit topology, and then a dynamic model could be constructed. Finally, the proposed library-based search method is used to obtain the most adequate promoter set $c^*=\{c_1^*, c_2^*, \ldots, c_n^*\}$.

3.1 Design of synthetic transcription cascade

Consider the dynamic system (7) or (9) of the synthetic transcriptional cascade shown in Figure 3. The design specifications are provided and listed as follows:

(i) The desired reference trajectory to be tracked is

$$y_r(t)=\begin{cases}10^6, & t\le 40, t>80\\ 2\times 10^7, & 40<t\le 80\end{cases} \tag{14}$$

i.e, the reference trajectory $y_r(t)$ is high from 40 to 80 h and low at other times. In this design example, 10 mM IPTG is added to induce the network transition from 40 h and is removed at 80 h. The concentration of yEGFP will track the reference trajectory $y_r(t)$ in (14).

(ii) Since $c_1\in \mathrm{Lib}_{\mathrm{const}}$, $c_2\in \mathrm{Lib}_{\mathrm{TetR}}$ and $c_3\in \mathrm{Lib}_{\mathrm{LacI}}$, $c_1\in J_m$, $c_2\in T_k$ and $c_3\in L_k$, $m=0, 1, \ldots, 18$, $k=0, 1, \ldots, 20$ are looked up in Table 1.

(iii) Suppose the standard deviations of parameter fluctuations $\Delta c_i=\{\Delta c_{r_i}, \Delta c_{s_i}\}=\{0.1c_{r_i}, 0.1c_{s_i}\}$, $i=1, 2, 3$, $\Delta d=0.1\times d$, $\Delta\alpha=0.1\times\alpha$, $\Delta\gamma_1=0.1\times\gamma_{\mathrm{TetR}}$, $\Delta\gamma_2=0.1\times\gamma_{\mathrm{LacI}}$, $\Delta\gamma_3=0.1\times\gamma_{\mathrm{yEGFP}}$ and environmental disturbances $v_i(t)$ are independent Gaussian white noises with zero means and unit variances.

(iv) The following mean-square tracking error needs to be minimized

$$J(c)=\min_{c=\{c_1,c_2,c_3\}} E\int_0^{t_f}(y(c,t)-y_r(t))^2\mathrm{d}t \tag{15}$$

In order to solve the constrained optimal tracking design problem of synthetic gene network via the proposed library-based search method, GA operators are set as follows: (i) a roulette wheel selection is used to increase the selection efficiency of the population with a lower cost function score; (ii) the cross-over rate is 0.8; (iii) the chromosome mutates uniformly with the mutation rate 0.1. Then the most adequate promoter set $c^*=\{c_1^*, c_2^*, c_3^*\}=\{J_4, T_2, L_1\}$ can be obtained. The simulation result shows that the proposed library-based search method can efficiently find the most adequate promoter set c^* (See Supplementary Figure S2).

3.2 Design of synthetic genetic oscillator

Consider a synthetic gene oscillator with negative feedback loops shown in Figure 4 (Elowitz and Leibler, 2000). The repressor protein LacI inhibits the expression of the gene *cI*, whose protein product in turn inhibits the expression of the gene *tetR*. The repressor protein TetR inhibits the transcription of *lacI* and *yegfp*. The negative feedback loops lead to temporal oscillations if the adequate promoter set is selected. For convenience, the concentrations of mRNA: *lacI*, *cI*, *tetR*, *yegfp* and proteins: LacI, CI, TetR, yEGFP are denoted by $x_1(c_1,t)$, $x_2(c_2,t)$, $x_3(c_3,t)$, $x_4(c_4,t)$, and $X_1(c_1,t)$, $X_2(c_2,t)$, $X_3(c_3,t)$, $X_4(c_4,t)$, respectively. The dynamic system for circuit topology in Figure 4 is given as

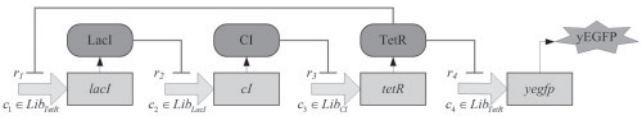

Fig. 4. The synthetic oscillator constructed by four repressor-regulated promoters c_1, c_2, c_3 and c_4, where $c_1, c_4\in T_k$ and $c_2\in L_k$, $k=0, 1, \ldots, 20$ in Table 1. $c_3\in Lib_{CI}$ is a CI-regulated promoter that has no library at present, so we suppose the promoter regulation function has a fixed form as (17).

$$\begin{aligned}&\begin{cases}\dot{x}_i(c_i,t)=p_j(c_i,r_i)-\beta x_i(c_i,t)\\ \dot{X}_i(c_i,t)=\alpha x_i(c_i,t)-\gamma_i X_i(c_i,t)\end{cases}\\ &y(c,t)=X_4(c_4,t),\ i=1, 2, 3, 4\\ &p_j(c_i,r_i)\in\left\{p_{\mathrm{TetR}}(c_1,X_3), p_{\mathrm{LacI}}(c_2,X_1), p_{\mathrm{CI}}(c_3,X_2), p_{\mathrm{TetR}}(c_4,X_3)\right\}\\ &c=\{c_1,c_2,c_3\}\in\{\mathrm{Lib}_{\mathrm{TetR}},\mathrm{Lib}_{\mathrm{LacI}},\mathrm{Lib}_{\mathrm{TetR}}\}\end{aligned} \tag{16}$$

Since there is no CI-regulated promoter library at present, we suppose that the promoter regulation function for CI has a fixed form as

$$p_{CI}(c_3,X_2)=\frac{150}{1+(X_2/K_{CI})^{n_{CI}}}, \tag{17}$$

where K_{CI} and n_{CI} are binding affinity and binding co-operativity between CI and DNA, respectively.

Based on the synthetic gene network in (3) with intrinsic fluctuations and environmental disturbances, the design specifications are provided and listed as follows:

(i) The desired reference trajectory to be tracked is

$$y_r(t)=4000\sin(0.016\pi t)+22000 \tag{18}$$

(ii) The indexes of TetR- and LacI-regulated promoter libraries in Table 1 are both from 0 to 20. The promoter regulation function for the CI-regulated promoter library is shown in (4) with $K_{CI}=20$ and $n_{CI}=2$.

(iii) According to the values of system parameters, we assume the parameter fluctuations with the standard deviations $\Delta c_i=\{\Delta c_{r_i}, \Delta c_{s_i}\}=\{0.1c_{r_i}, 0.1c_{s_i}\}$, $i=1,\ldots,4$, $\Delta d=0.1d=0.0288$, $\Delta\alpha=0.1\alpha=2.4$, $\Delta\gamma_1=0.1\gamma_{\mathrm{LacI}}$, $\Delta\gamma_2=0.1\gamma_{\mathrm{CI}}$, $\Delta\gamma_3=0.1\gamma_{\mathrm{TetR}}$ and $\Delta\gamma_4=0.1\gamma_{\mathrm{yEGFP}}$.

(iv) The following mean-square tracking error is to be minimized.

$$J(c)=\min_{c=\{c_1,c_2,c_4\}} E\int_0^{t_f}(y(c,t)-y_r(t))^2\mathrm{d}t \tag{19}$$

The proposed library-based search method is used to solve the constrained optimal tracking design problem through GA. GA operators are set as follows: (i) a roulette wheel selection is used to increase the selecting efficiency of the population with a lower cost function score; (ii) the cross-over rate is 0.8; (iii) the chromosome mutates uniformly with the

mutation rate 0.1. Then the most adequate promoter set $c^* = \{c_1^*, c_2^*, c_4^*\} = \{T_{17}, L_{19}, T_8\}$ can be obtained (Supplementary Figure S3).

4 DISCUSSION

The main challenge in genetic circuit design lies in selecting well-matched genetic parts that combine and produce the desired behavior reliably. Some studies have proposed to engineer a synthetic gene network to achieve a desired steady state (Chen and Wu, 2009) or an H_2 optimal tracking of a desired oscillator (Chen and Wu, 2010). However, the above design methods need to tune some kinetic parameters to achieve the desired steady or oscillatory state. In fact, tuning intrinsic parameters of biological devices to fit the designed parameter is currently quite a difficult or even unfeasible task for biotechnology. Further, it is still hard to select adequate biological parts to implement a desired cellular function with quantitative value. To overcome this problem, synthetic biologists usually create many versions of synthetic circuits with diverse characteristic by directed evolution, point mutation or random combinational of DNA components, and the functions of these versions are investigated to engineer the gene circuit with the desired behavior (Chatterjee and Yuan, 2006; Guet *et al.*, 2002; Yokobayashi *et al.*, 2002). But when the design of the genetic circuit is complex, the number of mutated versions needed to be created and tested is dramatically increased. Hence, these experimental steps become tedious and time consuming due to the significant amount of trail-and-error experiments. Our method is to select the existent promoters from promoter libraries and will be more appealing to synthetic biologists in the future.

In this study, the present promoter libraries are first redefined based on the dynamic gene regulation from experimental data, and a library-based search method is then introduced to engineer a synthetic gene network with desired behaviors by satisfying four design specifications. Through re-characteristic biological part datasheets, we can design a synthetic gene network with a new function in the existing libraries without having to perform a large number of trail-and-error experiments. Using GA operators, the library-based search method is used to mimic the realistic gene network evolution under natural selection to fit four design specifications but with much fast evolution needed for synthetic gene networks to achieve their desired behaviors. Hence, the proposed method can efficiently engineer a synthetic gene network to perform its desired behaviors despite intrinsic parameter fluctuations and environmental disturbances.

In design examples of the synthetic gene network, the most adequate promoter set c^* is selected to achieve the minimum tracking error between the desired reference trajectory and the observed output (i.e. fluorescence) despite intrinsic parameter fluctuations and environmental disturbances *in vivo*, as shown in Supplementary Figure S2. It is worth to notice that this method is also useful for constructing a synthetic oscillator (Fig. 4). After constructing the synthetic gene network and providing four design specifications, the library-based search method can efficiently select the most adequate promoter set to track the desired trajectory, as shown in Supplementary Figure S3. For the oscillator design in Fig. 4, only 71.4% of the promoter sets selected from Table 1 have oscillatory behavior, and only a few set of these can track the reference trajectory. Furthermore, by the proposed method, the design procedure of a synthetic gene network can be simplified for *in vivo* experiments.

5 CONCLUSION

In this study, promoter libraries are redefined based on their promoter activities so that they are more suitable for the gene circuit design of synthetic biology. The library-based search method can produce biological insights for how to design a synthetic gene network with prescribed functions. A GA-based search method is proposed to mimic the gene network evolution under natural selection to select adequate promoters to fit four design specifications with more fast evolution speed. The proposed design method to select the most adequate promoter set from promoter libraries will help synthetic biologists to simplify the design procedure of synthetic gene networks and accelerate the progress of synthetic biology. With advances in modern DNA synthesis technologies, together with well-defined characteristics of promoters and other standard biological parts, in the future our systematic method could be widely applied for the faster design cycle of synthetic gene networks.

Funding: This work was supported by the National Science Council of Taiwan under grant No. NSC 99-2745-E-007-001-ASP and No. 100-2745-E-007-001-ASP.

Conflict of Interest: none declared.

REFERENCES

Alon,U. (2007) *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, London.

Arkin,A. *et al.* (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. *Genetics*, **149**, 1633–1648.

Atkinson,M.R. *et al.* (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. *Cell*, **113**, 597–607.

Basu,S. *et al.* (2005) A synthetic multicellular system for programmed pattern formation. *Nature*, **434**, 1130–1134.

Basu,S. *et al.* (2004) Spatiotemporal control of gene expression with pulse-generating networks. *Proc. Natl Acad. Sci. USA*, **101**, 6355–6360.

Braun,D. *et al.* (2005) Parameter estimation for two synthetic gene networks: a case study. *IEEE Int. Conf. Acoustics Speech Signal Process.*, **5**, 769–772.

Canton,B. *et al.* (2008) Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.*, **26**, 787–794.

Chatterjee,R. and Yuan,L. (2006) Directed evolution of metabolic pathways. *Trends Biotechnol.*, **24**, 28–38.

Chen,B.S. and Chen,P.W. (2010) GA-based Design Algorithms for the Robust Synthetic Genetic Oscillators with Prescribed Amplitude, Period and Phase. *Gene Regul. Syst. Biol*, **4**, 35–52.

Chen,B.S. and Wu,C.H. (2009) A systematic design method for robust synthetic biology to satisfy design specifications. *BMC Syst. Biol.*, **3**, 66.

Chen,B.S. and Wu,C.H. (2010) Robust Optimal Reference-Tracking Design Method for Stochastic Synthetic Biology Systems: T-S Fuzzy Approach. *IEEE T Fuzzy Syst.*, **18**, 1144–1159.

Ellis,T. *et al.* (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.*, **27**, 465–471.

Elowitz,M.B. and Leibler,S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**, 335–338.

Friedland,A.E. *et al.* (2009) Synthetic gene networks that count. *Science*, **324**, 1199–1202.

Gardner,T.S. *et al.* (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature*, **403**, 339–342.

Goh,K.I. *et al.* (2008) Sustained oscillations in extended genetic oscillatory systems. *Biophys. J.*, **94**, 4270–4276.

Goldberg,D. (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning. 1989*. Addision Wesley, MA, USA.

Grefenstette,J.J. (1986) Optimization of control parameters for genetic algorithms. *IEEE Trans. Syst. Man Cybernet.*, **16**, 122–128.

Guet,C.C. *et al.* (2002) Combinatorial synthesis of genetic networks. *Science*, **296**, 1466–1470.

Holland,J.H. (1992) *Adaptation in Natural and Artificial Systems: an Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Bradford Book, Cambridge.

Hooshangi,S. *et al.* (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl Acad. Sci. USA*, **102**, 3581–3586.

Iadevaia,S. and Mantzaris,N.V. (2006) Genetic network driven control of PHBV copolymer composition. *J. Biotechnol.*, **122**, 99–121.

Katayama,K. and Narihisa,H. (2000) An efficient hybrid genetic algorithm for the traveling salesman problem. *Electron. Commun. Jpn*, **84**, 76–83.

Kelly,J.R. *et al.* (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.*, **3**, 4.

Kobayashi,T. *et al.* (2003) Modeling genetic switches with positive feedback loops. *J. Theor. Biol.*, **221**, 379–399.

Kobayashi,H. *et al.* (2004) Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl Acad. Sci. USA*, **101**, 8414–8419.

Kramer,B.P. *et al.* (2004) An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.*, **22**, 867–870.

Leveau,J.H. and Lindow,S.E. (2001) Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J. Bacteriol.*, **183**, 6752–6762.

Lu,T.K. *et al.* (2009) Next-generation synthetic gene networks. *Nat. Biotechnol.*, **27**, 1139–1150.

Murphy,K.F. *et al.* (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl Acad. Sci. USA*, **104**, 12726–12731.

Nakanishi,H. *et al.* (2008) Dynamical analysis on gene activity in the presence of repressors and an interfering promoter. *Biophys. J.*, **95**, 4228–4240.

Pai,A. *et al.* (2009) Engineering multicellular systems by cell-cell communication. *Curr. Opin. Biotechnol.*, **20**, 461–470.

Renders,J.M. and Flasse,S.P. (1996) Hybrid methods using genetic algorithms for global optimization. *IEEE Trans. Syst. Man Cybernet.*, **26**, 243–258.

Rinaudo,K. *et al.* (2007) A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat. Biotechnol.*, **25**, 795–801.

Semsey,S. *et al.* (2009) Dominant negative autoregulation limits steady-state repression levels in gene networks. *J. Bacteriol.*, **191**, 4487–4491.

Sohka,T. *et al.* (2009a) Morphogen-defined patterning of Escherichia coli enabled by an externally tunable band-pass filter. *J. Biol. Eng.*, **3**, 10.

Sohka,T. *et al.* (2009b) An externally tunable bacterial band-pass filter. *Proc. Natl Acad. Sci. USA*, **106**, 10135–10140.

Stricker,J. *et al.* (2008) A fast, robust and tunable synthetic gene oscillator. *Nature*, **456**, 516–519.

Tigges,M. *et al.* (2009) A tunable synthetic mammalian oscillator. *Nature*, **457**, 309–312.

Tuttle,L.M. *et al.* (2005) Model-Driven Designs of an Oscillating Gene Network. *Biophy. J.*, **89**, 3873–3883.

Win,M.N. and Smolke,C.D. (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc. Natl Acad. Sci. USA*, **104**, 14283–14288.

Win,M.N. and Smolke,C.D. (2008) Higher-order cellular information processing with synthetic RNA devices. *Science*, **322**, 456–460.

Wu,C.H. *et al.* (2011) Multiobjective H2/H∞ synthetic gene network design based on promoter libraries. *Math. Biosci.* [Epub ahead of print, doi.org/10.1016/j.mbs.2011.07.001].

Yokobayashi,Y. *et al.* (2002) Directed evolution of a genetic circuit. *Proc. Natl Acad. Sci. USA*, **99**, 16587–16591.