

RESEARCH ARTICLE

Rational design of a biosensor circuit with semi-log dose-response function in *Escherichia coli*

Haoqian Zhang^{1,2,3}, Ying Sheng¹, Qianzhu Wu¹, Ao Liu¹, Yuheng Lu¹, Zhenzhen Yin¹, Yuansheng Cao^{3,4}, Weiqian Zeng^{3,4} and Qi Ouyang^{3,4,*}

¹ Peking University Team for the International Genetic Engineering Machine Competition (iGEM), Peking University, Beijing 100871, China

² Peking-Tsinghua Joint Center for Life Sciences, School of Life Sciences, Peking University, Beijing 100871, China

³ Center for Quantitative Biology, Peking University, Beijing 100871, China

⁴ State Key Laboratory for Artificial Microstructures and Mesoscopic Physics, School of Physics, Peking University, Beijing 100871, China

* Correspondence: qi@pku.edu.cn

Received September 20, 2013; Accepted October 9, 2013

A central goal of synthetic biology is to apply successful principles that have been developed in electronic and chemical engineering to construct basic biological functional modules, and through rational design, to build synthetic biological systems with predetermined functions. Here, we apply the reverse engineering design principle of biological networks to synthesize a gene circuit that executes semi-log dose-response, a logarithmically linear sensing function, in *Escherichia coli* cells. We first mathematically define the object function semi-log dose-response, and then search for tri-node network topologies that can most robustly execute the object function. The simplest topology, transcriptional coherent feed-forward loop (TCFL), among the searching results is mathematically analyzed; we find that, in TCFL topology, the semi-log dose-response function arises from the additive effect of logarithmical linearity intervals of Hill functions. TCFL is then genetically implemented in *E. coli* as a logarithmically linear sensing biosensor for heavy metal ions [mercury (II)]. Functional characterization shows that this rationally designed biosensor circuit works as expected. Through this study we demonstrated the potential application of biological network reverse engineering to broaden the computational power of synthetic biology.

Keywords: synthetic biology; gene circuit design; reverse engineering; logarithmically linear sensing

INTRODUCTION

Significant advances have been made in synthetic biology. For instance, the genetic toggle switch [1], the genetic oscillator [2,3], the push-on push-off genetic switch [4], and the genetic program performing edge detection [5] have been experimentally synthesized in different laboratories. These achievements demonstrated that artificially designed biological circuits can perform predicted reliable functions and are of great potential for applications in biological engineering. However, due to the lack of effective rational design approach, most of these circuits have been built in a fairly *ad hoc* manner, in which iterative design and following troubleshooting are

often required to achieve object functions [6,7]. Therefore, how to develop rational approaches that effectively support the design and building of genetic circuit is becoming a critical problem in synthetic biology [8,9].

On the other hand, systems biology aims to quantitatively study biological regulatory network, focusing on the interactions between biological components and under what design principles those interactions happen [10,11]. Obviously, investigation, understanding and designing are integrally linked. Therefore, it is reasonable that exploiting natural design principles depicted by systems biology to guide the rational design of artificial biological network could be a potential solution to the problem of synthetic biology [8,9,11]. It has been recently revealed in

systems biology that network topology plays a critical role in robustly producing specific functions—the functional repertoire of a network is limited by its network topology [12–14]. Although the implementation of biological networks dramatically diversifies, a particular biological function is permitted by only a limited number of network topologies among which some may be more favored due to fewer parameter constraints [12,13].

Here we apply the method of biological network analysis developed by Ma et al. [13] to design transcriptional gene circuits with predetermined functions. If the network topologies with the highest robustness for an object function are biologically implemented, the object function will be achieved with less laborious trial-and-error fine-tuning or iterative design. As a proof of concept, we *in silico* designed and experimentally constructed a gene circuit that enables *E. coli* cells to report the presence of mercury (II) ions in water with a semi-log dose response curve (SLDRC), in which the output value (the fluorescence of green fluorescence protein of the cell) is proportional to the logarithmical value of the input (mercury (II) concentration in water) (Figure 1A). The results showed that the biosensor gene circuit performed logarithmically linear sensing function

as expected. Through this study we addressed the possibility of applying biological network reverse engineering approach to broaden the capacity of synthetic biology to rationally design biological systems.

RESULTS

The procedure can be outlined as the following: (i) we enumerate all possible tri-node network topologies, define the network topologies capable of performing the object function (SLDRC), and quantify their robustness against internal and external fluctuations (Figure 1A); (ii) the topologies with the highest robustness for the SLDRC function were selected for further analysis. The workflow is depicted in Figure 1B. In the following we discuss each process in detail.

Quantitatively defining object function

The object function was first defined according to our expectations. Figure 1A provides the main feature of SLDRC function, logarithmical linearity: The horizontal axis in Figure 1A denotes the input, the logarithmic value of divalent mercury ion [mercury (II)] concentration, and

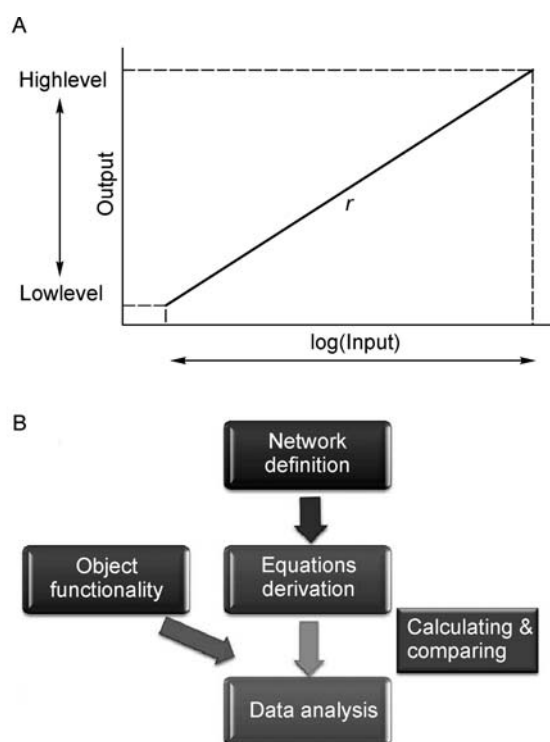


Figure 1. The definition of SLDRC function and the accompanying scheme of biological network reverse engineering. (A) The mathematical definition of semi-log dose-response function (SLDRC function). r is Pearson correlation coefficient and output range was defined as HIGHLEVEL minus LOWLEVEL. (B) The flow chart briefly describing the network reverse engineering approach

vertical axis stands for the output, the fluorescence of green fluorescent protein (GFP) as the reporter of gene expression, which increases with input in a linear relationship. The output range was defined as HIGH-LEVEL minus LOWLEVEL. In addition, we also expected a wide input range; it was predetermined to be wide enough, in our case 10^{-9} mol/L to 10^{-5} mol/L. For network topology searching, a numerical standard, Pearson correlation coefficient r , was set to quantify the Input-Output linear relationship in SLDRC function (Figure 1A). If r is larger than a critical value r_c (r_c was set as 0.98 in this study), we regarded the network topology as “positive”, which means that it is capable of SLDRC function; else as “negative”, in this case, the network motif will be excluded.

Network definition, equations derivation & calculation

Subsequently, we define the scale of networks (the number of nodes) and the interaction function between nodes for network enumeration. As proposed in Ma's study [13], here we consider only a minimal framework of tri-node network: one node for receiving the input (Node A), one node for transmitting output (Node C), and a third node that can play diverse regulatory roles (Node B) (Figure 2A). There are nine direct links between nodes, each with three possible types: positive, negative, or no regulation. Therefore, there are altogether $3^9 = 19683$ tri-node topologies. Because among them there are 3645 topologies that have no direct or indirect links from the input to the output (occlusive Node C), there remain a total of 16038 possible three-node network topologies that contain at least one direct or indirect causal link from the input node to the output node. We set to evaluate all these possible network topologies to get the ensemble of positive network topologies.

To quantitatively describe the dynamics behavior of each network topology, they were all translated into sets of dynamics equations. For this purpose we need to determine the interactions between nodes in a given network. In this study, the derivation of the equation was on the basis of the following assumptions:

(i) The nodes represent transcriptional factors (TF) or reporter gene (GFP). The links between the Nodes stand for activation or inhibition of gene expression control through transcription and translation process. The transcription was quantitatively described using the equilibrium binding probability P for binding of TF to DNA operator and the constant β for maximum transcription rate. Then we adopted a constant λ to adjust P . For the interactions between multiple TFs, the multiplication of λP or $1 - \lambda P$ was exploited to describe [12].

(ii) Our mathematical model focused on protein (TFs or

GFP) production (transcription-translation process), degradation and dilution along with cell growth. This is because other reactions, such as mercury (II)-protein binding process, happen much faster, thus to be regarded to be approximately at steady state, compared with the timescale of transcription-translation process [12,15–18].

(iii) It is well known that gene expression has basal level even without TF regulation [19]. A repressor will down-regulate the initial expression level and an activator will up-regulate it. Accordingly, we assume each gene have a basal level expression β_0 . Furthermore, we assume that every TF species contributes to the final expression level of protein.

As an example, we translated the network topology presented in Figure 2C as the following: it is widely accepted that the possibility of TF binding to the DNA operator in promoter is

$$P = \frac{(X^*/K_d)^n}{1 + (X^*/K_d)^n}. \quad (1)$$

Where X^* and K_d represents the effective concentration of specific TF and the dissociation constant, respectively. According to the assumptions (i) and (iii), the link from Node A to Node C can be derived as:

$$\begin{aligned} \frac{dX_3}{dt} &= \beta_0 + (\beta_m - \beta_0)\lambda P - \alpha X_3 \\ &= \beta_0 + (\beta_m - \beta_0)\lambda \frac{(X_1^*/K_d)^n}{1 + (X_1^*/K_d)^n} - \alpha X_3 \\ &= \beta_0 \left(1 - \lambda \frac{(X_1^*/K_d)^n}{1 + (X_1^*/K_d)^n} \right) + \beta_m \lambda \frac{(X_1^*/K_d)^n}{1 + (X_1^*/K_d)^n} - \alpha X_3. \end{aligned} \quad (2)$$

Where β_0 is the basal transcription rate, β_1 the effective transcription rate contributed by Node A, and α the decay rate of Node C itself.

For the process of mercury (II) sensing, TF at Node A forms a dimer to activate transcription. With the assumption (ii), we could deduce the effective concentration of a specific TF (X_1^*) and the total concentration of the TF (X_1):

$$X_1^* = [SSI] = kI[S]^2 = kIX_1^2. \quad (3)$$

Where I is the concentration of Hg (II). Substituting the above in (2), we obtained the dynamics equation for the network topology shown in Figure 2:

$$\begin{aligned} \frac{dX_3}{dt} &= \beta_0 \left(1 - \lambda \frac{(kIX_1/K_d)^n}{1 + (kIX_1/K_d)^n} \right) \\ &\quad + \beta_m \lambda \frac{(kIX_1/K_d)^n}{1 + (kIX_1/K_d)^n} - \alpha X_3. \end{aligned} \quad (4)$$

The first term in the right side of the equation is the basal

expression level of the circuit in which $1 - \lambda P$ is the possibility that specific TF is off the DNA operator; the second term is the contribution of on protein expression level, in which λP is the possibility that TF is on DNA operator, the third term is the decay rate of Node C.

The above process of translating a network topology into mathematical equations can be generalized to obtain the dynamics equation for any tri-node network topologies. The mathematics formulas are presented in the following:

$$P_{li} = \frac{(kIX_1^2)^n}{K_{li}^n + (kIX_1^2)^n}, X_i = A(i=1), B(i=2),$$

$$C(i=3), P_{ji} = \frac{X_j^{*n}}{K_{ji}^n + X_j^{*n}}, j=2,3.$$

$$\frac{dX_i}{dt} = \beta_0(1 - \lambda_{1i}P_{1i})(1 - \lambda_{2i}P_{2i})(1 - \lambda_{3i}P_{3i})$$

$$+ \beta_m[1 - (1 - \lambda_{1i}P_{1i})(1 - \lambda_{2i}P_{2i})(1 - \lambda_{3i}P_{3i})] - \alpha_i X_i. \quad (5)$$

For each network topology, we randomly sampled 10000 sets of network parameters using Latin Hypercube Sampling (LHS) method (Figure 2B). In order to compare Pearson correlation coefficient r of each network (with randomly selected parameters) with estimated standard

($r_c = 0.98$), the derived ODE equations was numerically integrated to acquire steady-state expression level of output Node C under various input level received by Node A, and then a linear fitting of logarithmic input to output was conducted. As the input range was set to be 10^{-9} mol/L to 10^{-5} mol/L, we selected points in identical logarithmic distance intervals within this range. Considering the possibility of bistability resulted from certain network topology motif, we conducted calculation from high input level to low input level and the reverse, respectively.

In this way, we exhaustively analyzed a total of 16038×10000 different ODEs. The values of their main feature for SLDRC function (Pearson correlation coefficient r) were thus obtained, of which all were compared with numerical standard ($r_c = 0.98$).

Analysis on network topologies

As shown in Figure 3A, for each network topology we defined Q value as the number of parameter sets that enables the network topology to meet the functional criteria ($r > 0.98$). The Q value reflects the robustness of a network topology for SLDRC function, the higher the Q value is, the more robust the network topology executes SLDRC function. All the network topologies were sorted

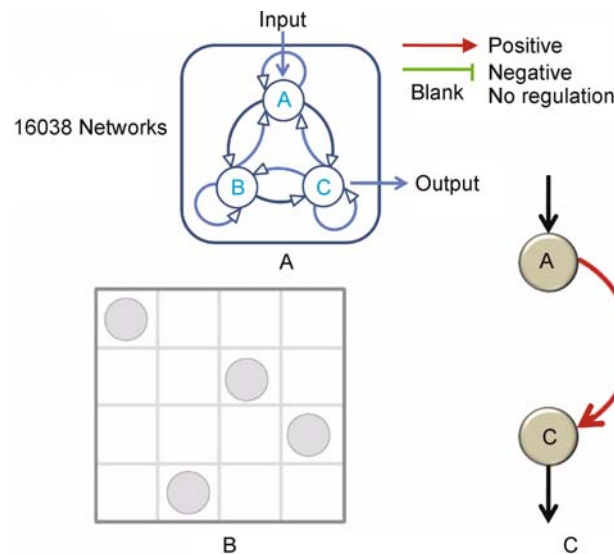


Figure 2. The framework of tri-node transcriptional network. (A) All possible tri-node network topologies. Node A receives input signal, Node C transmits output, and Node B plays regulatory roles. The links between nodes stand for TF interaction through transcription and translation process with three possible types: positive, negative, or no regulation. There are 16038 possible network motifs in total. (B) For each topology, 10000 sets of network parameters were sampled using LHS method. When sampling a function containing N variables, the range of each variable was divided into M equally probable intervals, and M sample points were then sampled to meet the requirements of Latin Hypercube Sampling method: in each axis-aligned hyper-plane, only one cube was filled with sample point. (C) An example network topology that has only one link from node A to node C. Whether this link is positive, negative or no regulation was determined by λ , a constant described in context.

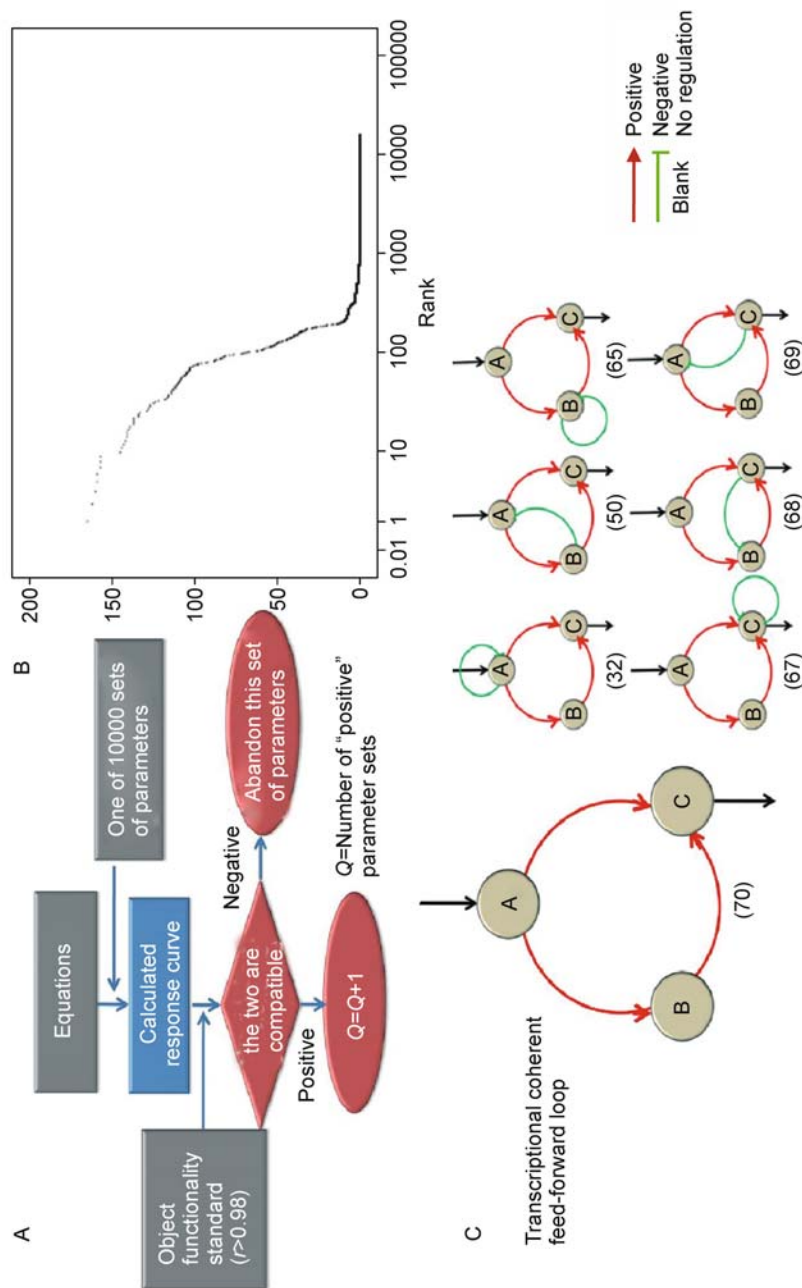


Figure 3. Calculation and analysis on network topologies. (A) The schematic of calculating process. For each network topology, we sampled 10000 sets of parameters, calculated the character (r) of response curve under each set of parameters to assess whether it fits the criteria ($r > 0.98$). Q value was defined as the number of functional parameter sets for each network topology. (B) Ranking of all tri-node networks according to their Q values. The top 74 topologies with Q values higher than 100 were extracted for further analysis in Figure 4. (C) The simplest network topology is a transcriptional coherent feed-forward loop (TCFL). 6 other simplest topologies with Q values higher than 100 were also listed. Their ranks were indicated in brackets. All of them contain TCFL motif implying that it is a common feature.

and ranked according to their Q values (Figure 3B). We found that most network topologies have Q or low Q values, while there are a small fraction of the network topologies with high Q values.

Listed in Figure 3C are all of the simplest topologies with 4 or less direct links between the three nodes with high Q values. Their ranks were also indicated in Figure 3C. Notably, there is only one 3-link topology out of the entire 7 simplest topologies, and it seems to be necessary among all: two positive links from Node A to Node B and Node C, respectively, and one positive link from Node B to Node C (Figure 3C). We named this network motif as Transcriptional Coherent Feed-forward Loop (TCFL).

In order to explore whether TCFL motif is necessary for SLDRC, we analyzed all of the top 74 SLDRC topologies listed in Figure 3B. Results shown in Figure 4A demonstrate that TCFL is indeed necessary for SLDRC function. Furthermore, to investigate what additional features contribute to SLDRC except minimal TCFL network motif, we clustered top 74 SLDRC network topologies. The results clearly indicated that, apart from TCFL and self-activation at Node C, there should be no positive link anywhere else (Figure 4B).

To fully understand how TCFL network motif exhibits SLDRC function and to provide more information for gene circuit implementation, we analyzed the TCFL topology and its parameter preference.

When a specific gene circuit taking TCFL topology reaches its steady state,

$$\frac{dX_1}{dt} = \beta_0 - \alpha_1 X_1 = 0. \quad (6)$$

$$\begin{aligned} \frac{dX_2}{dt} = & \beta_0 \left(1 - \lambda_{12} \frac{kIX_1^2}{K_{12} + kIX_1^2} \right) \\ & + \beta_m \left[1 - \left(1 - \lambda_{12} \frac{kIX_1^2}{K_{12} + kIX_1^2} \right) \right] - \alpha_2 X_2 = 0. \end{aligned} \quad (7)$$

$$\begin{aligned} \frac{dX_3}{dt} = & \beta_0 \left(1 - \lambda_{13} \frac{kIX_1^2}{K_{13} + kIX_1^2} \right) \left(1 - \lambda_{23} \frac{X_2}{K_{23} + X_2} \right) \\ & + \beta_m \left[1 - \left(1 - \lambda_{13} \frac{kIX_1^2}{K_{13} + kIX_1^2} \right) \left(1 - \lambda_{23} \frac{X_2}{K_{23} + X_2} \right) \right] \\ & - \alpha_3 X_3 = 0. \end{aligned} \quad (8)$$

which give:

$$X_1 = \frac{\beta_0}{\alpha_1}. \quad (9)$$

$$X_2 = \frac{(\beta_m - \beta_0) \frac{\lambda_{12} kIX_1^2}{K_{12} + kIX_1^2} + \beta_0}{\alpha_2}. \quad (10)$$

$$\begin{aligned} X_3 = & \frac{1}{\alpha_3} \left\{ \beta_0 + (\beta_m - \beta_0) \left[\frac{\lambda_{13} kIX_1^2}{K_{13} + kIX_1^2} + \frac{\lambda_{23} X_2}{K_{23} + X_2} \right. \right. \\ & \left. \left. - \frac{\lambda_{13} \lambda_{23} kIX_1^2 X_2}{(K_{13} + kIX_1^2)(K_{23} + X_2)} \right] \right\}. \end{aligned} \quad (11)$$

Taking X_3 as the function of I :

$$\begin{aligned} X_3 = & \alpha + \frac{v_1 I}{K_1 + I} + \frac{v_2 I}{K_2 + I} \\ & + \frac{\gamma}{K_2 + I} - \frac{v_1 v_2 I^2}{(K_1 + I)(K_2 + I)}. \end{aligned} \quad (12)$$

in which

$$\begin{aligned} \alpha = & [\beta_0 + (\beta_m - \beta_0)\beta_0/\alpha_2]/\alpha_3, \\ \beta = & [\beta_0 + (\beta_m - \beta_0)\beta_0/\alpha_2]/\alpha_3, \\ v_1 = & (\beta_m - \beta_0)\lambda_{13}, \\ K_1 = & \frac{K_{13}}{kX_1^2} = \frac{K_{13}\alpha_1^2}{k\beta_0^2}, \\ v_2 = & \lambda_{23}k\alpha_1^2(\beta_0 + \lambda_{12}\beta_m - \lambda_{12}\beta_0)/\beta_0^2, \\ K_2 = & K_{12}(\beta_0 + \alpha_2 K_{23})/v_2. \end{aligned}$$

Here we omitted the term of I^2 and $\frac{\gamma}{K_2 + I}$ because $I \ll K_2$. This gives

$$X_3 = \hat{\alpha} + \frac{v_1 I}{K_1 + I} + \frac{v_2 I}{K_2 + I}. \quad (13)$$

in which γ and $\hat{\alpha}$ are integrated constants.

For any $y = \frac{vx}{K+x}$, there always exists a logarithmically linear range: if x is transformed to $\log x$ (linear scale to logarithmical scale), we obtained

$$y = \frac{vx}{K+x} = \frac{v}{1 + e^{\log K - \log x}}. \quad (14)$$

This can be viewed as the result of a horizontal shift of $y = \frac{vx}{1+x} = \frac{v}{1 + e^{-\log x}}$ along the $\log x$ axis by $\log K$. If the curve of $y = \frac{vx}{1+x}$ taking a logarithmic abscissa coordinate could be regressed to a line $\hat{y} = a + b \log \hat{x}$ with coefficient r within an interval $[x_1, x_2]$, for any $y = \frac{vx}{K+x} = \frac{v}{1 + e^{\log K - \log x}}$ there is a regression $\hat{y} = v[a + b(\log \hat{x} - \log K)]$ with the same r within an interval that is approximately $\frac{x_2 - x_1}{K}$.

For $X_3 = \hat{\alpha} + \frac{v_1 I}{K_1 + I} + \frac{v_2 I}{K_2 + I}$, a wider logarithmically linear range could be reached under the restriction of $K_1 \leq K_2 \leq K_1 e^R$. This is due to the additive combining of

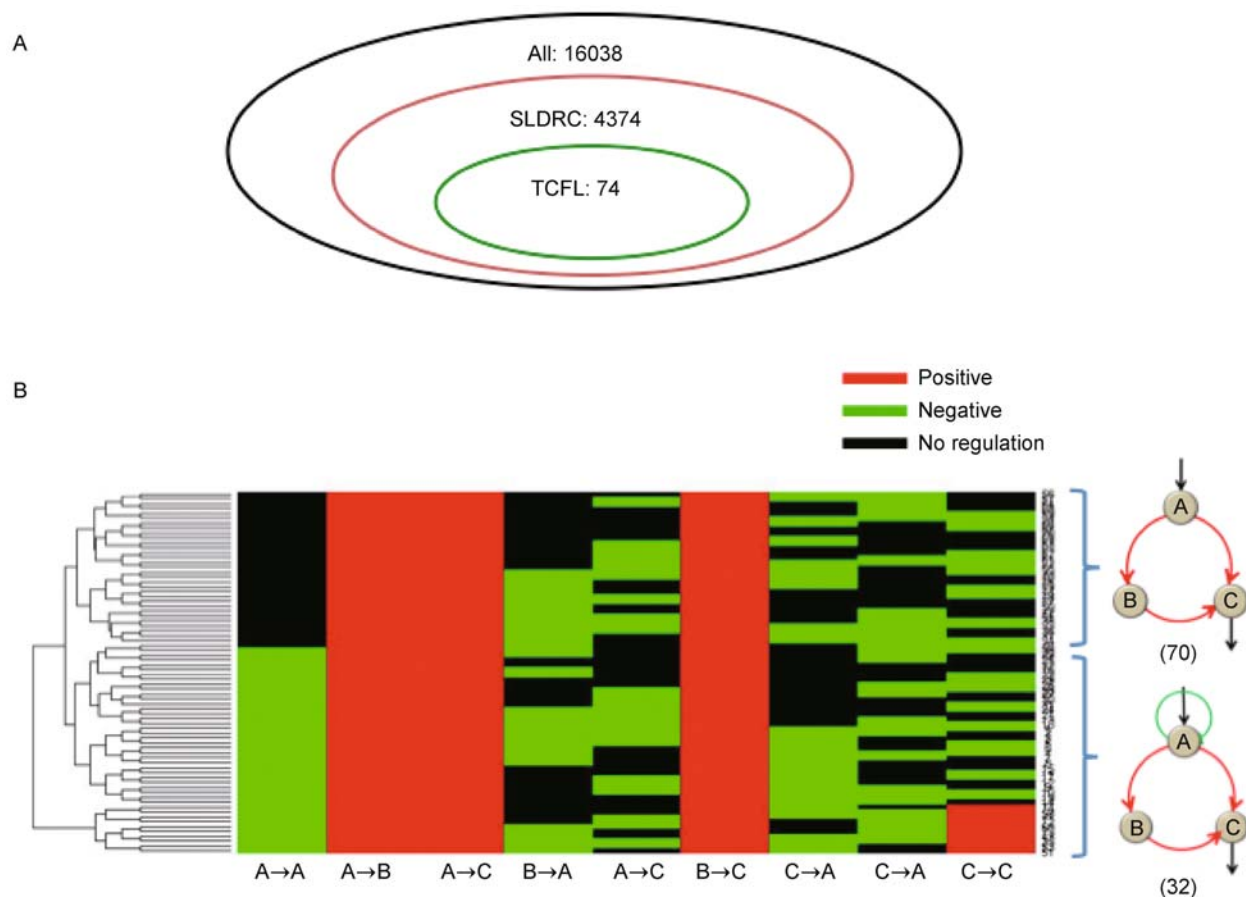


Figure 4. Transcriptional coherent feed-forward loop (TCFL) is necessary for SLDRC function. (A) Venn diagram shows that all of the network topologies that can achieve SLDRC function contain the TCFL motif. (B) Clustergram of top 74 SLDRC network topologies. Nine vertical rectangle bars stand for nine links in Figure 2A, which correspond to Node A self-regulation, Node A to Node B, Node A to Node C, Node B to Node A, Node B self-regulation, Node B to Node C, Node C to Node A, Node C to Node B, and Node C self-regulation, respectively. Red color stands for activation, green for repression and black for no regulation. The network topologies on the right are the minimal motifs revealed by clustering.

logarithmically linear intervals of the second and third terms, thus to form an extended logarithmically linear range that could be formulized as $R + \log \frac{K_2}{K_1}$, with R representing the width of logarithmically linear interval in a single Hill function. For instance, a combination of two logarithmically linear intervals with 3.5 orders of magnitude could give a logarithmically linear range with 6–7 orders of magnitude. To briefly summarize, in TCFL network topology, the network utilizes a redundant pathway (Node A to Node B to Node C) with different sensitivity to compensate for the saturation of single Hill function (Node A to Node C).

The final step of analysis on the network topologies is to conduct parameter sensitivity analysis. This process is to find sensitive parameters for the SLDRC function. It will provide theoretical guidance for the gene circuit

implementation. The analysis showed that the parameters that significantly contribute to SLDRC function are λ_{13} ; λ_{12} ; λ_{23} ; α_1 ; α_2 ; α_3 ; K_{13} ; K_{12} ; K_{23} . It also suggests that, in practical gene circuit construction, smaller α_3 , larger λ_{12} and larger λ_{23} should be adopted.

Genetic implementation of TCFL network topology for SLDRC function

After defining the network topologies that can perform the SLDRC function, we set to genetically implemented the TCFL network topology into a gene circuit. We chose mercury binding transcription activator, MerR, as the input-receiving node [20,21]. In this study, MerR gene comes from Tn21 in *Shigella flexneri* R100 plasmid [22,23]. MerR protein forms dimer and tightly sequesters mercury (II) at 10^{-8} mol/L concentration even in the

presence of 1000-fold higher concentration of molecular thiol-competing ligands [16]. In the absence of mercury (II), MerR protein dimers lightly binds to its DNA operator at promoter *pmerT*, resulting in slight repression of the basal transcription [16,17,21]. In the presence of mercury (II), tight binding of mercury (II) causes a conformational change to activate MerR dimer, leading into tight binding of MerR dimer to *pmerT* promoter and recruitment of RNA polymerase [16,17]. As a consequence, transcription is initiated. Notably, as the binding of MerR dimer to *pmerT* promoter in the absence of mercury (II) (inactivated form) is weak compared with the case of mercury (II) presence (activated form), we reasonably neglected the competitive binding of inactivated MerR dimer.

A conventional bi-node biosensor circuit consisting of only Node A (MerR) and Node C (green fluorescence protein, GFP) was constructed at first (Figure 5A). In details, a DNA fragment carrying *pmerT* promoter was cloned preceding BBa_E0840, a GFP generator from Registry of Standard Biological Parts (partsregistry.org) [24,25]. The expression of MerR was driven by a constitutive promoter, BBa_J23103, from Registry of Standard Biological Parts as well. The *pmerT*-E0840

construct was then cloned into pSB3K3 backbone and BBa_J23103-*merR* into pSB1A3 backbone. All the constructions here and after were performed using standard assembly [24,25].

When induced with gradient concentrations of mercury (II), bacteria carrying two plasmids, *pmerT*-E0840-pSB3K3 and BBa_J23103-*merR*-pSB1A3, were grown in LB broth with appropriate ampicillin and kanamycin at 37°C. Then the overnight culture was 1000-fold diluted into fresh LB broth. When OD₆₀₀ (observing density at wavelength 600 nm) reached 0.4–0.6, the cell culture was disposed into 96-deep-well plate wells, each owning 500 µL; different doses of mercuric chloride solution were supplemented accordingly, 3 replicates for each dose. The final concentration varied from 0 to 10⁻⁶ mol/L. In-plate culture fluorescence and OD₆₀₀ was recorded after overnight growth at 37°C. As shown in Figure 6, the output GFP fluorescence fits well to Hill function. As discussed before, only a narrow logarithmically linear range can be used for the accurate sensing of input level.

According to the results depicted by our *in silico* computation, we next genetically implemented tri-node TCFL by adding Node B, which was expected to transform the dose response curve from Hill function to

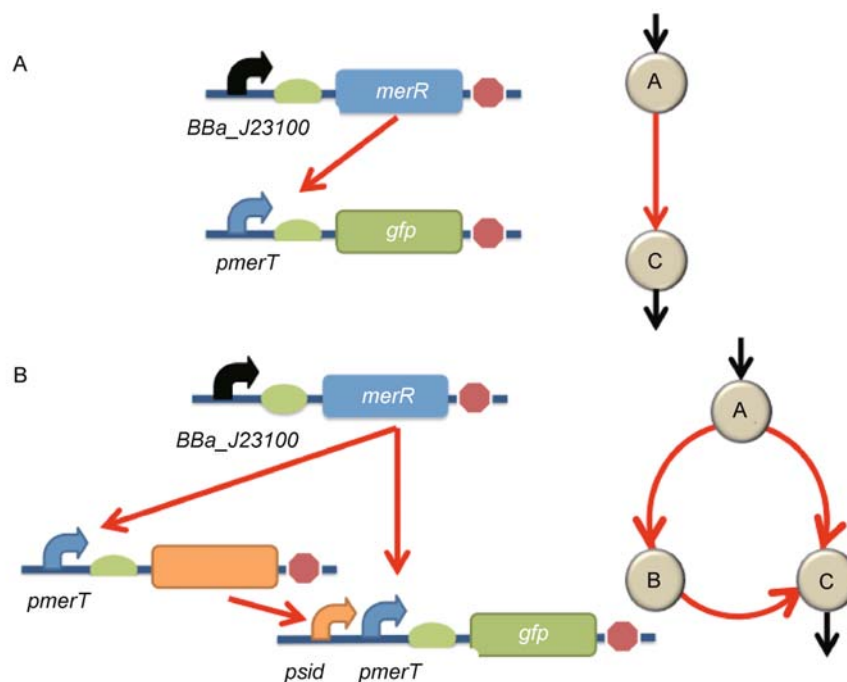


Figure 5. Genetic implementation of bi-node and tri-node TCFL biosensor gene circuits. (A) The detailed schematics of conventional bi-node biosensor circuit. The mercury binding transcription activator MerR was chosen as Node A. The input is mercury (II), since it can bind to the MerR dimer and activate transcription at *pmerT* promoter. The output is GFP expression level. Activation of MerR dimer by mercury (II) binding will initiate transcription of *gfp* gene at *pmerT* promoter. (B) On the basis of biosensor circuit in (A), TCFL, the simplest network topology motif for SLDR function, was genetically implemented. Node A is MerR. For Node B, the gene is transcription activator *ogr* that can activate transcription at *psid* promoter. Node C is GFP reporter gene whose expression was driven by both *psid* (activated by *ogr* activator) and *pmerT* (activated by MerR and mercury (II) complex).

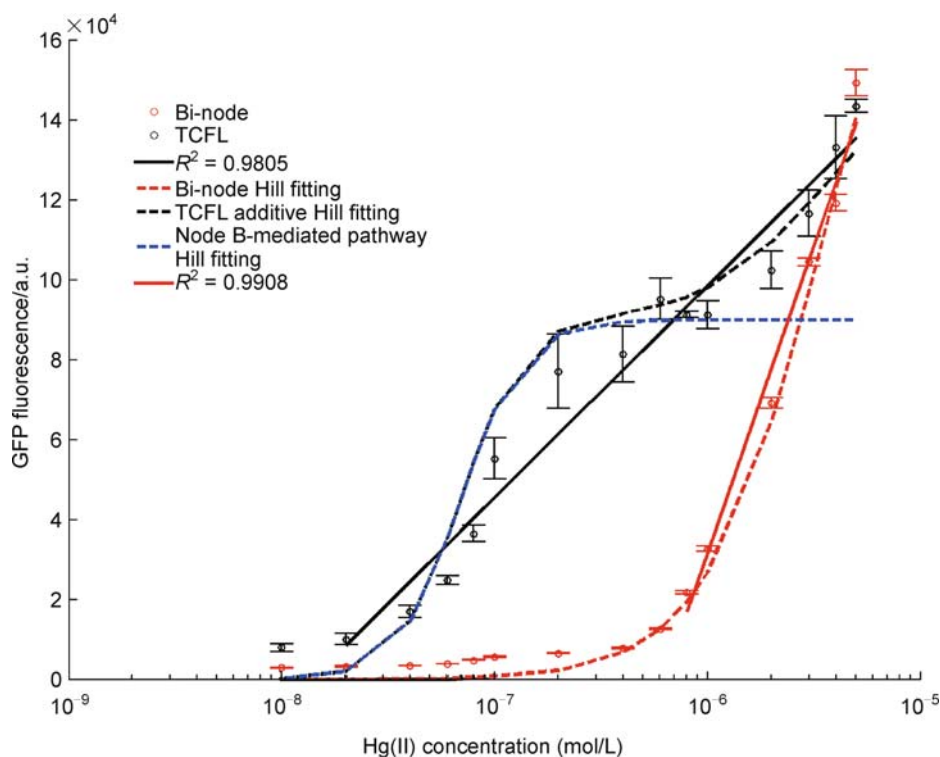


Figure 6. The dose response curves of bi-node and tri-node TCFL biosensor circuits. The response of conventional bi-node biosensor circuit to mercury (II) doses presents a typical Hill function. The TCFL biosensor circuit, however, exhibits semi-log dose-response function as *in silico* analysis predicted. The logarithmically linear sensing range for mercury concentration was significantly extended from 7 fold to nearly 250 fold. The solid lines stand for the logarithmically linear fitting on the dose response curves of bi-node and TCFL biosensor circuit, respectively. The red dashed curve denotes the fitting on the dose response curves of bi-node biosensor circuit using Hill function $G = 2.2 \times 10^5 \frac{I^{1.59}}{K_2^{1.59} + I^{1.59}}$. The value of K_2 was then taken as the apparent constant of Hill function term denoting the contribution of Node A-to-Node C direct pathway in TCFL, thus to fit the dose response curve of TCFL biosensor circuit, as black dashed curve shows. The formulas used for fitting is two additively combined Hill function terms, $G = 9.0 \times 10^4 \frac{I^3}{K_1^3 + I^3} + \alpha \times 2.2 \times 10^5 \frac{I^{1.59}}{K_2^{1.59} + I^{1.59}}$, with $K_1 = 6.91 \times 10^{-8}$ and $K_2 = 3.45 \times 10^{-6}$. $\alpha = 0.3$ describes the attenuation of Node A-to-Node C direct pathway due to the competition between the direct and indirect pathways, given the consideration of saturation effect in biochemical reactions. The putative “dose response curve” of the Node B-mediated Node A-to-Node C pathway is also presented in the dashed blue curve, showing the contribution of this indirect pathway to the SLDR function of TCFL biosensor circuit; it was numerically calculated using the dose response curves of bi-node and tri-node biosensor circuits.

SLDR. As suggested in theoretical analysis, smaller α_3 (degradation rate of GFP), larger λ_{12} (transcription activation of Node B by MerR-Hg (II) complex) and larger λ_{23} (transcription activation of GFP by Node B) should be adopted. Therefore, for Node B, we chose a strong transcription activator *ogr* from P2 phage [26,27]. Once expressed, it could recruit RNA polymerase constitutively and activate transcription at its cognate *psid* promoter [27]. When mercury (II) activates Node A (MerR), *ogr* will be expressed and activate the expression of GFP at Node C together with activated Node A (MerR) (Figure 5B). Additionally, the ribosome binding site (RBS, the major determinant of translation rate) of GFP is

BBa_B0034, the most consensus RBS sequence Registry of Standard Biological Parts; GFP protein has no degradation tag to guarantee low degradation rate (smaller α_3). Reassuringly, without further parameter fine-tuning in bench work, the dose response curve exhibits SLDR function. The logarithmically linear sensing range for mercury concentration was significantly expanded from 7 fold to nearly 250 fold (Figure 6).

The experimental result is consistent with our theoretical analysis. In Figure 6 both bi-node and TCFL tri-node biosensor circuits could be regressed to lines within appropriate logarithmically linear intervals. When the Hill function was exploited to fit the dose response curve of

bio-node biosensor circuit, we found the dissociation constant to be $K_2 = 3.45 \times 10^{-6}$ (Figure 6). We took this constant as the apparent constant of the Hill term describing the contribution of the direct pathway (from Node A directly to Node C) in the TCFL biosensor circuit. Then using the experimental data of TCFL tri-node biosensor circuit we found the Hill term describing the contribution of indirect pathway (Node B-mediated Node A Node C) to be $K_1 = 6.91 \times 10^{-8}$ (Figure 6). According to the analysis in the section of Analysis on Network Topologies, the estimation of the logarithmically linear sensing range was $\log(7) + \log(K_2/K_1) = 2.53$, which is consistent with the experimentally measured results, $\log(250) = 2.39$, of the TCFL tri-node biosensor circuit. Notably, the slopes of bi-node and tri-node biosensor circuits are different (Figure 6). This is probably because in genetic implementation the Hill coefficients of the two pathways (the direct and Node B-mediated indirect pathways) in TCFL have different values.

It is possible to fine-tune the sensitive parameters revealed by *in silico* analysis to expand the logarithmically linear sensing range. Our analysis indicates that any change of K_{12} and K_{13} (corresponding to the binding constant of MerR to *pmerT* promoter preceding TFs of Node B and Node C, respectively) will attenuate SLDR function character. Therefore, fine-tuning of the binding constants of MerR to *pmerT* promoter could probably improve SLDR function character. Previous studies suggested that mutations at semi-conserved region of MerR binding site would change the binding constant of MerR-DNA interactions [28]. To construct a spectrum of such parameter, we conducted mutagenesis at semi-conserved region of MerR binding site. A saturated mutagenesis library was built, followed by measurement of 100 mutants. If this parameter spectrum was applied to fine-tune TCFL motif gene circuit, it is probable that SLDR function character will be improved.

DISCUSSION

In summary, we applied the approach of biological network reverse engineering to synthesize a gene circuit that executes semi-log dose response in *Escherichia coli* cells. The major advantage of biosensor circuit with SLDR function is the expanded working range compared with conventional bi-node biosensor whose dose response curve is typically a Hill function. This gene circuit could work as a biosensor with logarithmically linear sensing function; it implements Weber's law behavior, which states the detection of input fold changes rather than the absolute levels: if input varies as a geometric progression, the output will change in an arithmetic progression. Very recently Lu et al. reported that SLDR function could also be implemented using

forward engineering design [29]. They exploited feedbacks and copy number control to dramatically expand the logarithmically linear sensing range of bi-node sensor circuit, which is quite different from our reverse engineering approach. Considering the mechanisms to achieve SLDR function, these two approaches could compensate each other.

With the continuous endeavor in system biology, more and more advances in our understanding of design principles of natural biological networks have been achieved. It is expected that these advances, including the design principles of how biological components interact with each other and the computation methods developed in this process, will provide fundamental basis for designing synthetic biological systems in the area of synthetic biology [7–9]. In this context, it has been recently reported that the reverse engineering approach can be taken to design artificial signal transduction regulatory networks that can robustly achieve polarization in yeast cells [30]. This study (biochemical networks of signal transduction) together with ours (transcriptional regulatory network) can serve as good examples to demonstrate the promising power of biological network reverse engineering approach.

However, there are still problems hidden behind. The most critical issue of biological network reverse engineering approach is the limited complexity of network for exhaustive topology searching. The topology of tri-node network can be exhausted, but those of quarto-node network cannot due to currently limited computation ability. It is very difficult to exhaust quarto-node network topologies and to investigate smaller “positive” fraction out of 3^{16} motifs. New mathematical tools are needed to push forward this line.

MATERIALS & METHODS

Strains and growth media

E. coli strain Trans1T1 and Luria-Bertani (LB) medium (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl) were used throughout the studies. Kanamycin (25 µg/mL) and ampicillin (50 µg/mL), were supplemented as needed. The inducer mercuric chloride was obtained from Sigma Aldrich (# M1136).

Plasmid construction

BioBrick standard assembly [25] ENREF 1 was used for plasmid construction throughout the study unless otherwise specified. Site-directed mutagenesis was conducted using the MutanBEST Kit (Takara) following the manufacturer's protocol.

Fluorescence and OD₆₀₀ recording

Cells that harbour the appropriate plasmids were incubated in LB broth at 37°C overnight in culture tubes and then diluted 1000-fold into fresh LB broth in beaker flasks. Once the OD₆₀₀ reached 0.3–0.4 (Spectrumbiol 721 N), each culture was transferred into 96-deep-well plates (500 µL culture per well), supplemented with appropriate inducers, and incubated overnight (250 r/m, 37°C). Finally, the inducer was washed away and the cells were resuspended with PBS solution; 20 µL of the resuspension was diluted 10-fold into 200 µL of PBS solution for a microplate reader recording of the fluorescence and OD₆₀₀ values (Thermo Scientific Varioskan Flash) with appropriate settings.

ACKNOWLEDGEMENTS

This work is part of the project for the 2010 team of Peking University in the international genetically engineered machine (iGEM) competition. H. Zhang designed the project, performed the experiments and modeling simulation, and wrote the manuscript. Y. Sheng, A. Liu, and Q. Wu performed the experiments. Y. Lu and Z. Yin performed the modeling simulation. Y. Cao and W. Zeng performed the modeling simulation and wrote the manuscript. Q. Ouyang designed the project and wrote the manuscript.

We would like to thank F. Hao, X. He, W. Wei, C. Xu, and L. Ji for their technical assistance; the BioBrick Foundation for providing DNA materials; and Anne O. Summers for supplying the plasmid carrying MerR gene. We thank Peking University for its financial support. This work is also partially supported by the National Nature Science Foundation of China (Nos. 10721463, 11074009), the National Basic Research Program of China (Nos. 2009CB918500, 2012AA02A702), and the National Science Fund for Talent Training in Basic Science of China (Nos. J1030310, J1103205).

COMPLIANCE WITH ETHICS GUIDELINES

The authors Haoqian Zhang, Ying Sheng, Qianzhu Wu, Ao Liu, Yuheng Lu, Zhenzhen Yin, Weiqian Zeng, Yuan Sheng Cao and Qi Ouyang declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

- Gardner, T. S., Cantor, C. R. and Collins, J. J. (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403, 339–342
- Elowitz, M. B. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, 403, 335–338
- Danino, T., Mondragón-Palomino, O., Tsimring, L. and Hasty, J. (2010) A synchronized quorum of genetic clocks. *Nature*, 463, 326–330
- Lou, C., Liu, X., Ni, M., Huang, Y., Huang, Q., Huang, L., Jiang, L., Lu, D., Wang, M., Liu, C., et al. (2010) Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch. *Mol. Syst. Biol.*, 6, 350
- Tabor, J. J., Salis, H. M., Simpson, Z. B., Chevalier, A. A., Levskaya, A., Marcotte, E. M., Voigt, C. A. and Ellington, A. D. (2009) A synthetic genetic edge detection program. *Cell*, 137, 1272–1281
- Kwok, R. (2010) Five hard truths for synthetic biology. *Nature*, 463, 288–290
- Lu, T. K., Khalil, A. S. and Collins, J. J. (2009) Next-generation synthetic gene networks. *Nat. Biotechnol.*, 27, 1139–1150
- Nandagopal, N. and Elowitz, M. B. (2011) Synthetic biology: integrated gene circuits. *Science*, 333, 1244–1248
- Randall, A., Guye, P., Gupta, S., Duportet, X. and Weiss, R. (2011) Design and connection of robust genetic circuits. *Meth. Enzymol.*, 497, 159–186
- Bruggeman, F.J., Hornberg, J.J., Boogerd, F.C., and Westerhoff, H.V. (2007). Introduction to systems biology. EXS97, 1–19.
- Smolke, C. D. and Silver, P. A. (2011) Informing biological design by integration of systems and synthetic biology. *Cell*, 144, 855–859
- Alon, U. (2007). An introduction to systems biology : design principles of biological circuits (Boca Raton, FL, Chapman & Hall/CRC).
- Ma, W., Trusina, A., El-Samad, H., Lim, W. A. and Tang, C. (2009) Defining network topologies that can achieve biochemical adaptation. *Cell*, 138, 760–773
- Artyukhin, A. B., Wu, L. F. and Altschuler, S. J. (2009) Only two ways to achieve perfection. *Cell*, 138, 619–621
- Buchler, N. E., Gerland, U. and Hwa, T. (2005) Nonlinear protein degradation and the function of genetic circuits. *Proc. Natl. Acad. Sci. U.S.A.*, 102, 9559–9564
- Helmann, J. D., Ballard, B. T. and Walsh, C. T. (1990) The MerR metalloregulatory protein binds mercuric ion as a tricoordinate, metal-bridged dimer. *Science*, 247, 946–948
- Ralston, D. M. and O'Halloran, T. V. (1990) Ultrasensitivity and heavy-metal selectivity of the allosterically modulated MerR transcription complex. *Proc. Natl. Acad. Sci. U.S.A.*, 87, 3846–3850
- Wickner, S., Maurizi, M. R. and Gottesman, S. (1999) Posttranslational quality control: folding, refolding, and degrading proteins. *Science*, 286, 1888–1893
- J., Xiao, J., Ren, X., Lao, K., and Xie, X.S. (2006). Probing gene expression in live cells, one protein molecule at a time. *Science* 311, 1600–1603.
- Hobman, J. L., Wilkie, J. and Brown, N. L. (2005) A design for life: prokaryotic metal-binding MerR family regulators. *Biometals*, 18, 429–436
- Nascimento, A. M. and Chartone-Souza, E. (2003) Operon mer: bacterial resistance to mercury and potential for bioremediation of contaminated environments. *Genet. Mol. Res.*, 2, 92–101
- Liebert, C. A., Hall, R. M. and Summers, A. O. (1999) Transposon Tn21, flagship of the floating genome. *Microbiol. Mol. Biol. Rev.*, 63, 507–522
- Nakaya, R., Nakamura, A. and Murata, Y. (1960) Resistance transfer agents in *Shigella*. *Biochem. Biophys. Res. Commun.*, 3, 654–659
- Bower, A. G., McClintock, M. K. and Fong, S. S. (2010) Synthetic biology: a foundation for multi-scale molecular biology. *Bioeng Bugs*, 1, 309–312

25. Canton, B., Labno, A. and Endy, D. (2008) Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.*, 26, 787–793
26. Christie, G. E. and Calendar, R. (1985) Bacteriophage P2 late promoters. II. Comparison of the four late promoter sequences. *J. Mol. Biol.*, 181, 373–382
27. Julien, B. and Calendar, R. (1996) Bacteriophage PSP3 and phiR73 activator proteins: analysis of promoter specificities. *J. Bacteriol.*, 178, 5668–5675
28. Park, S. J., Wireman, J. and Summers, A. O. (1992) Genetic analysis of the Tn21mer operator-promoter. *J. Bacteriol.*, 174, 2160–2171
29. Daniel, R., Rubens, J. R., Sarpeshkar, R. and Lu, T. K. (2013) Synthetic analog computation in living cells. *Nature*, 497, 619–623
30. Chau, A. H., Walter, J. M., Gerardin, J., Tang, C. and Lim, W. A. (2012) Designing synthetic regulatory networks capable of self-organizing cell polarization. *Cell*, 151, 320–332