Supplementary Material

Fundamental design principles for transcription-factor-based metabolite biosensors

Ahmad A. Mannan[1], Di Liu[2], Fuzhong Zhang[2*] and Diego A. Oyarzún[1*].

[1] Department of Mathamatics, Imperial College London, UK.
[2] Department of Energy, Environmental & Chemical Engineering, Washington University in St Louis, USA.
[*] Corresponding authors: fzhang@seas.wustl.edu, d.oyarzun@imperial.ac.uk

Summary. Here we present the phenomenological models used in the paper, together with the calculations of the dose-response parameters and further details on the parameter fitting.

1 Phenomenological model for metabolite biosensors

As shown in Figure 1A of the main text, we model metabolite biosensors as a cascade of two modules: a sensing module, which describes the interaction between the metabolite (M) and transcription factor (TF), and a regulation module, which describes the interaction between the transcription factor and protein expression (P). We describe the dose-response curve $P(M)$ as the composition of two Hill functions, $P = f_2(f_1(M))$, where $TF = f_1(M)$ is the activity level of the transcription factor for a given amount of metabolite M, and $P = f_2(TF)$ is the output expression for a given concentration of transcription factor. Depending on whether the interactions are inhibitory or activatory, we use the following Hill functions

$$\text{Inhibitory} \quad : \quad f_i(x) = b_i + \frac{a_i}{1 + (K_i \cdot x)^{n_i}}, \text{ for } i = \{1, 2\}, \tag{1}$$

$$\text{Acivatory} \quad : \quad f_i(x) = b_i + \frac{a_i \cdot (K_i \cdot x)^{n_i}}{1 + (K_i \cdot x)^{n_i}}, \text{ for } i = \{1, 2\}, \tag{2}$$

where the parameters describe different aspects of the sensing and regulation modules (see main text). The dose-response curves for the four architectures in Figure 1A are thus

- Activated-Repressor (AR)

$$f_2(f_1(M)) = b_2 + \frac{a_2}{1 + \left(K_2 \cdot \left(b_1 + \frac{a_1 \cdot (K_1 \cdot M)^{n_1}}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}, \tag{3}$$

- Repressed-Activator (RA)

$$f_2(f_1(M)) = b_2 + \frac{a_2 \cdot \left(K_2 \cdot \left(b_1 + \frac{a_1}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}{1 + \left(K_2 \cdot \left(b_1 + \frac{a_1}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}, \tag{4}$$

- Repressed-Repressor (RR)

$$f_2(f_1(M)) = b_2 + \frac{a_2}{1 + \left(K_2 \cdot \left(b_1 + \frac{a_1}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}, \tag{5}$$

- Activated-Activator (AA)

$$f_2(f_1(M)) = b_2 + \frac{a_2 \cdot \left(K_2 \cdot \left(b_1 + \frac{a_1 \cdot (K_1 \cdot M)^{n_1}}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}{1 + \left(K_2 \cdot \left(b_1 + \frac{a_1 \cdot (K_1 \cdot M)^{n_1}}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}. \tag{6}$$

It can be shown that the composite functions in (3)–(6) have the following properties: they have one minimum and one maximum, they are monotonically decreasing (AR and RA) or increasing (AA and RR), and they have a single inflection point. As a result, the composite functions are qualitatively similar to sigmoid functions, and hence we can precisely define the dose-response parameters, as shown next.

2 Formulas for the dose-response parameters

Basal output (b). The basal expression, defined as the minimum value of the dose-response curve, can be computed by evaluating the expressions in (3)–(6) for $M = 0$ or $M \to \infty$:

$$b = \begin{cases} f_2(f_1(0)) \,, & \text{for AA and RR architectures,} \\ \\ f_2(f_1(\infty)) \,, & \text{for RA and AR architectures,} \end{cases} \tag{7}$$

which leads to

$$b = \begin{cases} b_2 + \dfrac{a_2}{1 + K_2^{n_2}(b_1 + a_1)^{n_2}} \,, & \text{for AR and RR architectures,} \\ \\ b_2 + \dfrac{a_2 \cdot (K_2 b_1)^{n_2}}{1 + (K_2 b_1)^{n_2}} \,, & \text{for RA and AA architectures.} \end{cases} \tag{8}$$

Maximum change in output (a). The maximum change in output is the difference between the maximum and minimum value of the dose-response curve. Similarly as for the basal expression (b), we derive the formula for a from the definition:

$$a = \begin{cases} f_2(f_1(\infty)) - f_2(f_1(0)) \,, & \text{for AA and RR architectures,} \\ \\ f_2(f_1(0)) - f_2(f_1(\infty)) \,, & \text{for RA and AR architectures,} \end{cases} \tag{9}$$

which leads to a formula for a that is identical for all four biosensor architectures

$$a = a_2 \cdot K_2^{n_2} \cdot \left(\frac{(b_1 + a_1)^{n_2} - b_1^{n_2}}{(1 + K_2^{n_2}(b_1 + a_1)^{n_2}) \cdot (1 + (K_2 b_1)^{n_2})} \right). \tag{10}$$

Biosensor dynamic range (μ). From the previous definitions, the dynamic range can be simply computed as $\mu = a/b$, with a and b as given in equations (8) and (10). The resulting expressions are

$$\mu = \begin{cases} \mu_2 \cdot \frac{((b_1+a_1)^{n_2} - b_1^{n_2})K_2^{n_2}}{(1+\mu_2+(K_2(b_1+a_1))^{n_2}) \cdot (1+(b_1 K_2)^{n_2})} \,, & \text{for AR and RR architectures,} \\ \\ \mu_2 \cdot \frac{((b_1+a_1)^{n_2} - b_1^{n_2})K_2^{n_2}}{(1+(K_2(b_1+a_1))^{n_2}) \cdot (1+(1+\mu_2)(b_1 K_2)^{n_2})} \,, & \text{for RA and AA architectures.} \end{cases} \tag{11}$$

The formula for μ in the repressed-repressor architecture in equation (11) is the same as equation (5) in the main text.

Biosensor threshold (θ). We define the threshold as the concentration of metabolite for which the level of biosensor output expression is 50% of the maximum, relative to the basal. This definition corresponds to solving the following equation for θ,

$$f_2(f_1(\theta)) = b + \frac{a}{2}, \tag{12}$$

where b and a are given in equations (8) and (10), respectively. Solving equation (12) for each architecture in (3)–(6), we get

$$\theta = \begin{cases} \dfrac{1}{K_1} \cdot \sqrt[n_1]{\dfrac{a_1 K_2}{\sqrt[n_2]{A-1} - b_1 K_2} - 1}\,, & \text{for RA and RR architectures,} \\ \dfrac{1}{K_1} \cdot \sqrt[n_1]{\dfrac{1}{\dfrac{a_1 K_2}{\sqrt[n_2]{A-1} - b_1 K_2} - 1}}\,, & \text{for AR and AA architectures,} \end{cases} \tag{13}$$

where A is a function of the model parameters

$$A = 2 \cdot \frac{(1 + (K_2(b_1 + a_1))^{n_2}) \cdot (1 + (b_1 K_2)^{n_2})}{2 + ((b_1 + a_1)^{n_2} + b_1^{n_2}) K_2^{n_2}}\,. \tag{14}$$

Biosensor sensitivity (n). We define the sensitivity as the slope of the normalized dose-response curve

$$n = 4\theta \cdot \left| \frac{\mathrm{d}}{\mathrm{d}M} \left(\frac{f_2(f_1(M)) - b}{a} \right)_{M=\theta} \right|, \tag{15}$$

where $|\cdot|$ is the absolute value, and (a, b, θ) are the dose-response parameters computed in equations (8), (10) and (13). The definition in (15) applies to the four biosensor architectures and is inspired by the observation that in a Hill equation of the form $f(x) = (x/\theta)^n / (1 + (x/\theta)^n)$, it can be shown that $n = 4\theta\, \mathrm{d}f/\mathrm{d}x|_{x=\theta}$. The definition in (15) leads to explicit formulas for the sensitivity as a function of tunable parameters, but the resulting expressions are involved and we omit them for brevity.

3 Bounds for dose-response parameters

3.1 Constraints between dynamic range and threshold

In this work we consider the TF-operator affinity (K_2) as the main tuneable parameter to adjust biosensor function. Thus in this section we consider the dose-response parameters as functions of K_2, that is, from the expressions in equations (8)–(13) we write $b = b(K_2)$, $\mu = \mu(K_2)$ and $\theta = \theta(K_2)$. The formulas for the dose-response parameter reveal that dynamic range and threshold are inherently coupled to each other through the K_2 parameter. In Figure SF1 we plot $\mu(K_2)$ vs $\theta(K_2)$ and observe that the shape of the curve depends strongly on the b_1 parameter, which represents the basal level of TF activity. Numerical exploration of the parameter space suggests that for $b_1 = 0$, the (μ, θ)-curve is monotonic, while a non-zero b_1 value leads to a non-monotonic relationship between threshold and dynamic range. This phenomenon appears in all four biosensor architectures.

From the expression for the basal output in (8), it can be shown that $b(K_2)$ is a decreasing function of K_2 for the AR and RR architectures, and an increasing function of K_2 for the RA and AA architectures. In contrast, we found that the function $a(K_2)$ defined in (10) is non-monotonic, and reaches a maximum

$$a_{\max} = \frac{a_2 \cdot ((1 + \mu_1)^{n_2} - 1)}{\left((1 + \mu_1)^{\frac{n_2}{2}} + 1\right)^2}, \tag{16}$$

for an optimal operator affinity

$$K_2^{\max,\, a} = \frac{1}{b_1} \cdot \frac{1}{\sqrt{1 + \mu_1}}, \tag{17}$$

where $\mu_1 = a_1/b_1$. Note that when $b_1 = 0$, the optimal operator affinity $K_2^{\max,a} \to \infty$, and thus a becomes a monotonic function of K_2.

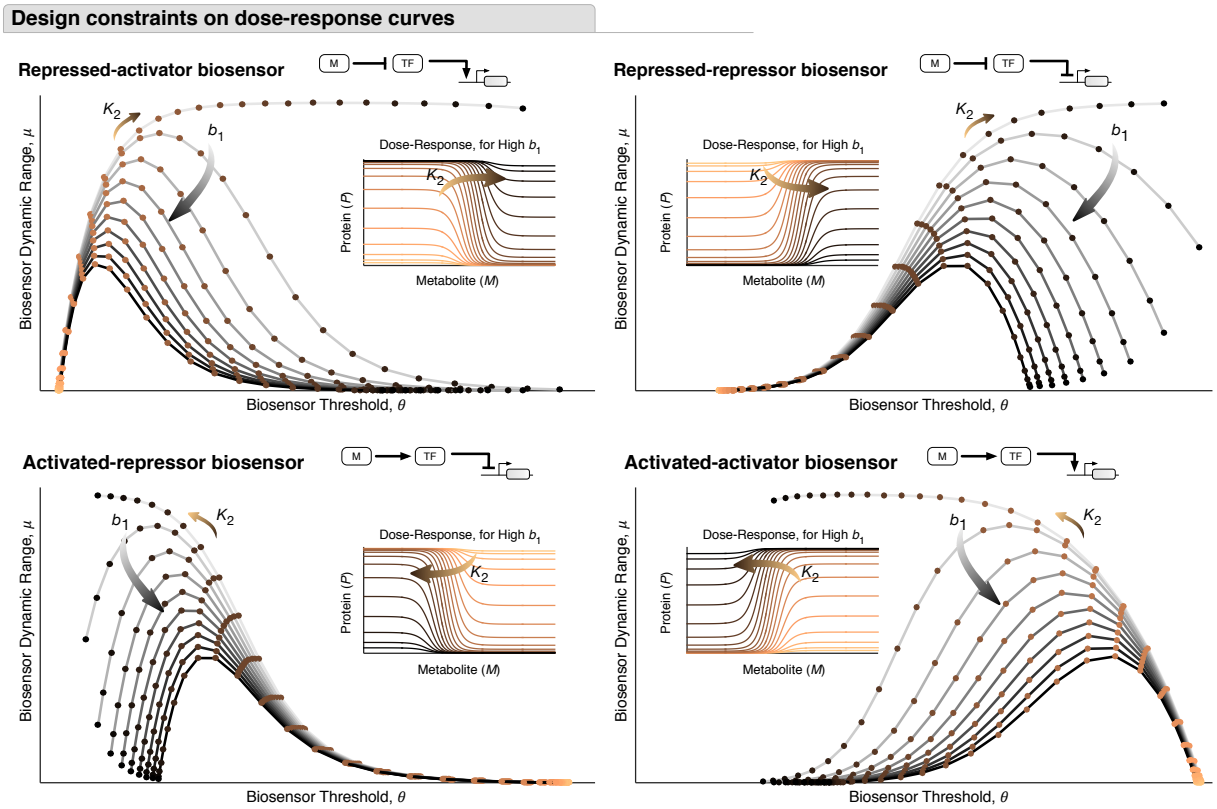

Figure SF 1: Plots of the $(\mu(K_2), \theta(K_2))$-curves for each biosensor architecture and increasing values of the b_1 parameter. The insets show the dose-response curves for different K_2 values for a non-zero value of b_1.

3.2 Maximal dynamic range

To compute the operator affinity that maximizes the dynamic range, we solved the equation $\mathrm{d}\mu(K_2)/\mathrm{d}K_2 = 0$ with $\mu(K_2) = a(K_2)/b(K_2)$ and $a(K_2)$ and $b(K_2)$ given in (8) and (10), respectively. We found that the dynamic range reaches a maximum at

$$K_2^{\max,\mu} = \begin{cases} \dfrac{1}{b_1} \cdot \dfrac{1}{\sqrt{1+\mu_1}} \cdot (1+\mu_2)^{\frac{1}{2n_2}} & \text{for AR and RR architectures,} \\ \dfrac{1}{b_1} \cdot \dfrac{1}{\sqrt{1+\mu_1}} \cdot \dfrac{1}{(1+\mu_2)^{\frac{1}{2n_2}}} & \text{for RA and AA architectures.} \end{cases} \tag{18}$$

The maximal dynamic range is then

$$\mu_{\max} = \mu_2 \cdot \frac{(1+\mu_1)^{n_2} - 1}{\left((1+\mu_1)^{\frac{n_2}{2}} + (1+\mu_2)^{\frac{1}{2}}\right)^2}, \tag{19}$$

where $\mu_2 = a_2/b_2$. The expression in (19) is valid for all four biosensor architectures. Importantly, we also note that in the limit $b_1 \to 0$, the optimal affinity $K_2^{\max,\mu} \to \infty$, which means that dynamic range becomes a monotonically increasing function of K_2, in agreement with the numerical observations in Figure SF1.

3.3 Bounds for the dose-response threshold

From the expression for θ in equation (13), we note that θ has minimal and maximal values depending on the extremal values of the function A in (14). It can be shown that $\mathrm{d}A/\mathrm{d}K_2 < 0$ for all $K_2 > 0$ and

hence A is a monotonically decreasing function of K_2, with a maximum at $K_2 = 0$ and a minimum at $K_2 \to \infty$. Assuming all parameters other than K_2 are fixed, positive real values, we evaluated $A(K_2)$ for $K_2 = 0$ and $K_2 \to \infty$, to get the expressions for $A_{\max}$ and $A_{\min}$, respectively, as

$$A_{\max} = A(0) = 2^{\frac{-1}{n_2}} \cdot b_1 \cdot (1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}}, \tag{20}$$

$$A_{\min} = A(\infty) = 2^{\frac{1}{n_2}} \cdot (b_1 + a_1) \cdot \frac{1}{(1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}}}. \tag{21}$$

We then computed the maximum and minimum bounds of the response threshold, $\theta_{\min}$ and $\theta_{\max}$, by substituting $A_{\max}$ and $A_{\min}$ into expression for the threshold θ in (13). The bounds for the threshold are

- For RA and RR architectures:

$$\theta_{\min} = \frac{1}{K_1} \cdot \sqrt[n1]{\frac{\mu_1}{2^{-\frac{1}{n_2}} \cdot (1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}} - 1} - 1}, \tag{22}$$

$$\theta_{\max} = \frac{1}{K_1} \cdot \sqrt[n1]{\frac{\mu_1}{2^{\frac{1}{n_2}} \cdot \left(\frac{(1+\mu_1)}{(1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}}} \right) - 1} - 1}. \tag{23}$$

- For AR and AA architectures:

$$\theta_{\min} = \frac{1}{K_1} \cdot \sqrt[n1]{\frac{1}{\frac{\mu_1}{2^{\frac{1}{n_2}} \cdot \left(\frac{(1+\mu_1)}{(1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}}} \right)^{-1}} - 1}}, \tag{24}$$

$$\theta_{\max} = \frac{1}{K_1} \cdot \sqrt[n1]{\frac{1}{\frac{\mu_1}{2^{-\frac{1}{n_2}} \cdot (1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}} - 1} - 1}}. \tag{25}$$

4 Parameter fitting

The dose-response parameters in Figures 1D and 3C of the main text were obtained from the promoter characterization data. For each induction curve, we computed the dynamic range (μ) and basal expression level (b) from the data at zero and full induction with IPTG or TMG. The response threshold was determined from Hill functions fitted to the data using the `Solver` routine in MS Excel for Mac 2016.

The model fits in Figures 1D (green curve) and 3C (blue and yellow curves) were produced as explained in the Methods section. We used the optimizer `fmincon` from the Global Optimization Toolbox in Matlab 2016a. We ran the optimization 500 times with varying initial guesses determined by the `multistart` routine in Matlab 2016a. The convergence of the objective function, shown in Fig. SF2, suggests the optimal solutions are global minimizers in all three parameter fittings. In Table SF1, we report the parametric constraints employed and the ensemble average across all the runs of the optimization routine.

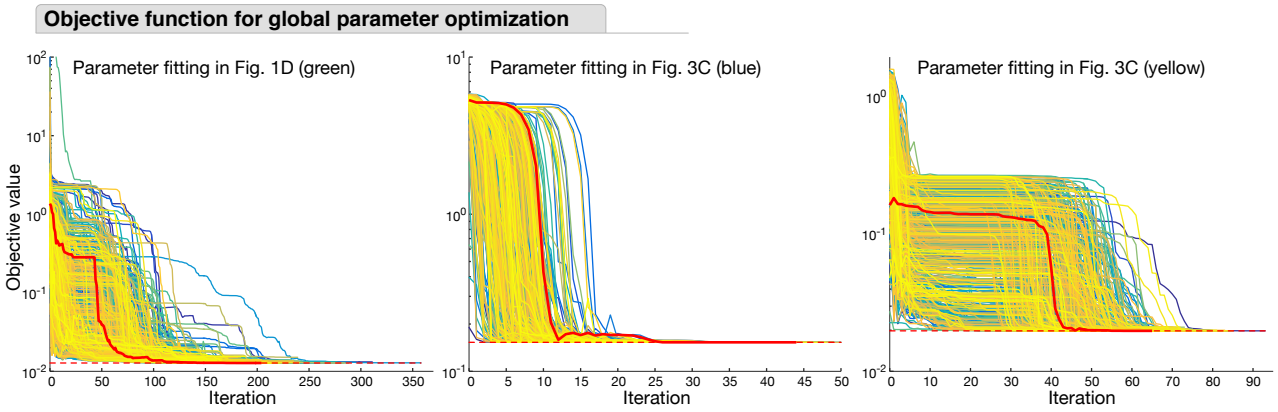

Figure SF 2: Evolution of objective function as a function of the iteration number of the optimization routine. The results show 500 runs of the optimization solver for each model in Fig. 1D and 3C in the main text, with varying initial conditions. The trace highlighted in red is of the convergence to the optimal solution with the lowest objective value, the corresponding parameters of which are reported in the last column of Table SF1. These parameters were used to generate the solid curves in Figs. 1D (green curve) and 3C (blue and yellow curves).

Table SF 1: Results of parameter fitting. Lower and upper bounds used for parameter optimization, plus their mean and standard error across 500 runs of the optimization algorithm. The last column are the parameter values corresponding to the optimal solution with the lowest objective value, and were used to generate the solid curves in Figs. 1D (green curve) and 3C (blue and yellow curves). All values are reported to 4 significant figures.

Parameter	Lower bound	Upper bound	Mean ± SE	Reported
Model in Figure 1D (green)				
b_1	0	0.1	$3.339{\cdot}10^{-2} \pm 1.961{\cdot}10^{-3}$	$2.588 \cdot 10^{-8}$
a_1	0	500	348.4 ± 6.448	62.63
K_1	10^{-4}	1	$1.804{\cdot}10^{-1} \pm 1.228{\cdot}10^{-3}$	$2.112 \cdot 10^{-1}$
n_1	1	5	$2.436 \pm 2.132{\cdot}10^{-2}$	3.039
b_2	0	20	$1.231 \pm 5.216{\cdot}10^{-2}$	$4.132 \cdot 10^{-6}$
a_2	0	$3 \cdot 10^3$	$2149 \pm 3.933{\cdot}10^{-1}$	2147
n_2	1	3	$1.526 \pm 2.409{\cdot}10^{-2}$	1.008
Model in Figure 3C (blue)				
b_1	0	1	$9.482{\cdot}10^{-2} \pm 8.352{\cdot}10^{-4}$	$7.931{\cdot}10^{-2}$
K_1	10^{-6}	1	$3.454{\cdot}10^{-3} \pm 4.001{\cdot}10^{-8}$	$3.454{\cdot}10^{-3}$
Model in Figure 3C (yellow)				
b_2	0	5	$4.065{\cdot}10^{-6} \pm 7.199{\cdot}10^{-7}$	$3.110{\cdot}10^{-9}$
a_2	10^3	10^4	$3002 \pm 4.071{\cdot}10^{-1}$	3005
n_2	1	3	$1.189 \pm 2.535{\cdot}10^{-4}$	1.189

5 Biosensor architectures

Table SF 2: Some examples of engineered biosensors based on metabolite-responsive transcription factors. Each biosensor can be abstracted into one of the four architectures shown in Fig. 1A. [1]TyrR is a dual-regulator that has been used as an activator and repressor of gene expression in response to tyrosine. [2]BetI has been re-engineered to respond positively or negatively to choline.

Architecture	Metabolite (M)	Transcription Factor (TF)
Repressed-Activator	Isopentenyl diphosphate	AraC-Idi [1]
	NADPH	SoxR [8]
Activated-Repressor	Tyrosine	TyrR[1] [1]
	Choline	BetI[2] [7]
Repressed-Repressor	Acyl-CoA	FadR [11]
	Choline	BetI[2] [7]
	IPTG	LacI [6]
	Malonyl-CoA	FapR [3]
Activated-Activator	Tyrosine	TyrR[1] [11]
	Mevalonate	AraC [9]
	Alkanes	AlkS [5]
	Benzoate	BenR [10]
	L-methionine	Lrp [4]
	o-acetyl-L-(homo)serine	CysR [2]

References

[1] Chou, H. H., and Keasling, J. D. Programming adaptive control to evolve increased metabolite production. *Nature Communications 4* (2013), 1–8.

[2] Hoffmann, K., Grünberger, A., Lausberg, F., Bott, M., and Eggeling, L. Visualization of imbalances in sulfur assimilation and synthesis of sulfur-containing amino acids at the single-cell level. *Applied and Environmental Microbiology 79*, 21 (2013), 6730–6736.

[3] Liu, D., Xiao, Y., Evans, B. S., and Zhang, F. Negative Feedback Regulation of Fatty Acid Production Based on a Malonyl-CoA SensorActuator. *ACS Synthetic Biology 4*, 2 (2013), 132–140.

[4] Mustafi, N., Grünberger, A., Kohlheyer, D., Bott, M., and Frunzke, J. The development and application of a single-cell biosensor for the detection of l-methionine and branched-chain amino acids. *Metabolic Engineering 14*, 4 (2012), 449–457.

[5] Reed, B., Blazeck, J., and Alper, H. Evolution of an alkane-inducible biosensor for increased responsiveness to short-chain alkanes. *Journal of Biotechnology 158*, 3 (2012), 75–79.

[6] Richards, D. H., Meyer, S., and Wilson, C. J. Fourteen Ways to Reroute Cooperative Communication in the Lactose Repressor: Engineering Regulatory Proteins with Alternate Repressive Functions. *ACS Synthetic Biology 6*, 1 (2017), 6–12.

[7] Saeki, K., Tominaga, M., Kawai-noma, S., Saito, K., and Umeno, D. The rapid diversification of BetI-based transcriptional switches for the 2 control of biosynthetic pathways and genetic circuits. 3. *ACS Synthetic Biology 5*, 11 (2015), 1201–1210.

[8] Siedler, S., Schendzielorz, G., Binder, S., Eggeling, L., Bringer, S., and Bott, M. SoxR as a single-cell biosensor for NADPH-consuming enzymes in Escherichia coli. *ACS Synthetic Biology 3*, 1 (2014), 41–47.

[9] Tang, S.-Y., and Cirino, P. C. Design and application of a mevalonate-responsive regulatory protein. *Angewandte Chemie Int. Ed Engl. 50*, 5 (2011), 1084–1086.

[10] Uchiyama, T., and Miyazaki, K. Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Applied and Environmental Microbiology 76*, 21 (2010), 7029–7035.

[11] Xiao, Y., Bowen, C. H., Liu, D., and Zhang, F. Exploiting non-genetic, cell-to-cell variation for enhanced biosynthesis. *Nature Chemical Biology 12*, 5 (2016), 339–344.