## **SUPPLEMENTARY INFORMATION**

# **Cell-Free Characterization of Coherent Feed-Forward Loop-Based Synthetic Genetic Circuits**

Pascal A. Pieters<sup>†,‡</sup>, Bryan L. Nathalia<sup>‡</sup>, Ardjan J. van der Linden<sup>†,‡</sup>, Peng Yin<sup>§</sup>, Jongmin Kim<sup>|,\*</sup>, Wilhelm T.S. Huck<sup>¶,\*</sup> and Tom F. A. de Greef $\dagger$ , $\ddagger$ , $\P$ , $\ddagger$ , $\ast$ 

†Laboratory of Chemical Biology and Institute for Complex Molecular Systems, Department of Biomedical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

‡Computational Biology Group, Department of Biomedical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

§Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115,

USA; Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

<sup>ǁ</sup>Division of Integrative Biosciences and Biotechnology, Pohang University of Science and

Technology, 77 Cheongam-ro, Pohang, Gyeongbuk 37673, Republic of Korea

¶ Institute for Molecules and Materials, Radboud University, Heyendaalseweg 135, 6525 AJ,

Nijmegen, The Netherlands

#Center for Living Technologies, Eindhoven-Wageningen-Utrecht Alliance, Eindhoven, the **Netherlands** 

### *Corresponding Authors:*

 $E$ -mail: jongmin.kim $@$ postech.ac.kr, w.huck $@$ science.ru.nl and t.f.a.d.greef $@$ tue.nl

#### **Golden Gate Assembly**

```
plasmid...GCAT
                                (variant with<br>start codon)
GCAT...promoter...AAGC
                    AAGC...UTR...CATG
                                  CATG...CDS...TGAA
                    AAGC........trigger.......TGAA
                                 (variant 1 without
                                 start codon)
GCAT...promoter.....switch.....GTTG
                                  GTTG...CDS...TGAA
                                 (variant 2 without
                                start codon)
GCAT...promoter.....switch.....CGAA
                                  CGAA...CDS...TGAA
                                                 TGAA...terminator...GTCG
                                                                       GTCG...plasmid
```
**Figure S1:** Schematic of the type of DNA components and their combinations to create the relevant DNA constructs.<sup>1</sup> The 5- to 3-prime 4 nucleotide Golden Gate Assembly (GGA) overhang sequences are shown in distinct colors, whereas the construct types (plasmid, promoter, UTR, switch, trigger, CDS, terminator) are displayed in black in between the overhangs. The start codon in the concerned overhang sequence is underlined. Two variants of the blue overhang sequence were used in combination with the toehold switches to decrease background expression by omitting the obsolete start codon in the original sequence.



**Table S1:** Primer sequences used in this research. GGA = Golden Gate Assembly, GA = Gibson Assembly and OL*X* represents for overlap sequence *X* (see Figure S1 for the various overlap sequences).



**Figure S2:** Comparison of the expression levels and sequence of an initial toehold switch-bearing DNA construct design and an optimized version. The differences in sequence at the promoter-switch transition and switch-CDS transition are displayed using sequence alignment at the aforementioned positions. Deletions are shown as a dash (-) and mutations are displayed in bold. The most notable changes are the deletion of a cloning scar between the promoter and switch and the omission of the start codon at the start of the eGFP sequence, which only leaves the toehold switch-regulated start codon for initiation of translation. The expression levels with (pink) and without (gray) toehold trigger encoding DNA construct are shown in the graph. The bars represent the average expression after 14 hours and the points represent individual experiments. N=1 for the initial switch design, N=2 for the optimized switch with trigger and N=4 for the optimized switch without trigger. All experimental conditions are summarized in Table S3.



**Figure S3:** A) Endpoint eGFP output concentrations of a titration of the P70a-eGFP DNA construct, which functions as an expression positive control construct. Expression from this construct plateaus for DNA concentrations above  $\sim$  5 nM. B) Endpoint eGFP concentrations for a titration of trigger (left) and switch DNA constructs (right). Both experiments display an increase in eGFP output for increasing concentrations of the DNA constructs. N=3 for all experiments, except 10 nM trigger DNA + 2 nM Switch-eGFP DNA, for which N=10, and 10 nM trigger DNA + 5 nM Switch-eGFP DNA, for which N=8. All experimental conditions are summarized in Table S3.



**Figure S4:** Endpoint expression of a simple eGFP gene in the presence of various concentrations of  $σ<sup>28</sup>$ -producing DNA construct. The  $σ<sup>28</sup>$  protein produced by the DNA **R1** construct competes with  $σ<sup>70</sup>$ for the RNA polymerase. Nevertheless, the production of eGFP from the  $\sigma^{70}$ -controlled P70 promoter did not decrease for increasing concentrations of DNA *R1*, indicating that the concentration of  $\sigma^{28}$  protein that is produced is too low to significantly alter  $\sigma^{70}$ -regulated protein expression. N=1 for all conditions, except for the negative control without DNA *R1*, for which N=4. All experimental conditions are summarized in Table S3.



Figure S5: A) Endpoint expression of the expression cascade consisting of a toehold trigger that activates a toehold switch that regulates the translation of  $\sigma^{28}$ , which activates the transcription of an eGFP output gene. The cascade produced eGFP when an on-target toehold trigger was used, whilst expression was absent when an off-target trigger was used instead (N=2). B) The endpoint eGFP concentration and expression delay ( $t_{50}$ ) for a range of  $\sigma^{28}$  construct concentrations (N=3). Expression levels increased for increasing  $\sigma^{28}$  DNA concentrations, while the t<sub>50</sub> steadily dropped. All experimental conditions are summarized in Table S3.



Figure S6: Endpoint expression of the CFFL 1 and reference motif with appropriate controls. Background expression in absence of on-target trigger was lower for the CFFL than reference motif. Omission of a sigma-factor construct (Y1 or R1) completely reduces expression to background levels. All experiments were performed in triplicate, except for the off-target control for the reference motif, for which N=2.



**Figure S7:** a) Schematic representation of all DNA species and the DNA, RNA and protein level interactions that constitute the reference motif. The three main components are the toehold switch, marked by the 5'-adjacent RNA stem loop, and its corresponding RNA trigger (orange), the E. coli  $\sigma$ <sup>28</sup>-factor (blue) and fluorescent output protein (green). b) Time traces of eGFP expression of the reference motif with an on-target trigger (light traces are three distinct experiments and the dark solid lines are their average) and without trigger (dark gray). All experimental conditions are summarized in Table S3.



**Figure S8:** Endpoint eGFP (green) and eCFP (blue) concentrations of select subnetworks of the composite CFFL, where only the trigger and  $\sigma^{28}$  constructs from one side of the network and the trigger and output constructs from the other side were expressed. These experiments demonstrate that the two sides of the network can influence each other through the  $\sigma$ <sup>28</sup> factor. The full composite CFFL and the composite CFFL where all  $\sigma$ <sup>28</sup> constructs are left out are included as positive and negative control, respectively. Bars represent the average output protein endpoint concentration and the data points display the endpoint concentration of individual experiments. DNA concentrations are, unless omitted in the subnetwork, as shown in Figure 2B. All experimental conditions are summarized in Table S3.

#### **Reference Motif Constitutive Node**



Figure S9: Endpoint concentrations, t<sub>50</sub> and trade-offs for the reference motif with varying concentrations of  $\sigma^{28}$ -encoding DNA construct (DNA R1). In addition to the default construct (circles), a construct with a point mutation in the promoter to give a low efficiency  $\sigma^{28}$ -producing gene was evaluated (triangles). The trade-offs are shown for the default high-yield  $\sigma^{28}$  DNA construct. N=3 for all conditions. All experimental conditions are summarized in Table S3.

## **Ordinary differential equation (ODE) models**

We constructed ODE models of the synthetic CFFL and reference motif to simulate flow reactor experiments. Protein species and all RNA species and complexes were taken as state variables of the ODE models. DNA species were defined as parameters that can be adjusted to emulate the dynamic inputs given in the flow experiments. Complex formation of the RNA toehold switches and triggers, constitutive transcription and translation were modeled with a mass action term, while activation of the P28 promoter by the  $\sigma^{28}$  factor was modeled using Hill-type kinetics. Fluorescent protein maturation was not explicitly included, but is captured in the translation rate of the protein. Since the model was exclusively used to model flow reactor experiments, depletion of resources was assumed not to play a major role and was not represented in the ODE models.

## *ODE Model of the synthetic CFFL*

The CFFL model comprises a RNA trigger (*T*) species that is produced from DNA species *X1* (*DNA* $_T$ ) with rate  $b_T$ . SwitchA-S28 mRNA is produced from DNA species *Y1* (*DNA*<sub>Sw1</sub>) with rate *bSw1*. SwitchA-eGFP mRNA is produced by expression from DNA species *Z1* (*DNASw2*) with rate  $B_{\text{Sw2}}$  (background expression) and with rate  $b_{\text{Sw2}}$  upon activation by the  $\sigma^{28}$  factor (*S28*) through Hill kinetics ( $K_{S28}$  and  $N_{S28}$ ). The  $\sigma^{28}$  factor was previously shown to behave according to Michaelis-Menten kinetics,<sup>2</sup> so  $N_{528}$  was set to 1 for the parameter fitting procedure, but was varied during parameter sampling (Table S2). Binding of trigger RNA (*T*) to SwitchA-S28 mRNA (*Sw1*) and SwitchA-eGFP mRNA (*Sw2*) to form complexes *TSw1* and *TSw2*, respectively, was explicitly modeled using mass action kinetics (*kon,TSw* and *koff,TSw*). Translation from the trigger:switch complexes was represented by rates  $k_{t,l,0,28}$  and  $k_{t,l,GPP}$ . Since translation from unbound switch RNA species can also occur, but at a lower rate determined by the efficiency of the toehold switch, this was included in the model using *bgratiosw*. The resulting translation rate of unbound SwitchA-S28

mRNA is the product of *ktl,S28* and *bgratiosw* and the translation rate of unbound SwitchA-eGFP mRNA is the product of *ktl,GFP* and *bgratiosw*. The degradation rates of all RNA species was represented by parameter *aRNA* and of all protein species by parameter *aprotein*. Since the degradation rate of proteins in the TXTL reactions was negligible, as demonstrated by the constant plateaus after 10+ hours of batch TXTL reactions, the value of *aprotein* was set to zero. The flow rate of the experiment can be found in all equations as parameter  $k_{flow}$ , which was set as 1.6 h<sup>-1</sup> for all experiments, giving a residence time of 37.5 min. The resulting ODE model of the synthetic CFFL is given by Equations S1-8, where Equation S5 shows a helper function *f(..)* which is used in the Hill equation in Equation S6.

$$
\frac{dT(t)}{dt} = b_T \cdot DNA_T + k_{off, TSw} \cdot TSw1(t) - k_{on, TSw} \cdot T(t) \cdot Sw1(t) + k_{off, TSw} \cdot TSw2(t) \n- k_{on, TSw} \cdot T(t) \cdot Sw2(t) - (a_{RNA} + k_{flow}) \cdot T(t)
$$
\n(S1)

$$
\frac{dSw1(t)}{dt} = b_{\text{Sw1}} \cdot DNA_{\text{Sw1}} + k_{\text{off,TSw}} \cdot TSw1(t) - k_{\text{on,TSw}} \cdot T(t) \cdot Sw1(t) - (a_{\text{RNA}} + k_{\text{flow}}) \cdot Sw1(t)
$$
(S2)

$$
\frac{dTSw1(t)}{dt} = k_{on,TSw} \cdot T(t) \cdot Sw1(t) - k_{off,TSw} \cdot TSw1(t) - (a_{RNA} + k_{flow}) \cdot TSw1(t)
$$
\n(S3)

$$
\frac{dS28(t)}{dt} = k_{t1, S28} * TSw1(t) + k_{t1, S28} \cdot bgratio_{sw} \cdot Sw1(t) - (a_{protein} + k_{flow}) \cdot S28(t)
$$
\n(S4)

$$
f(S28(t)) = \left(\frac{S28(t)}{K_{S28}}\right)^{N_{S28}}
$$
\n(S5)

$$
\frac{dSw2(t)}{dt} = B_{\text{Sw2}} \cdot DNA_{\text{Sw2}} + b_{\text{Sw2}} \cdot DNA_{\text{Sw2}} \cdot \frac{f(S28(t))}{1 + f(S28(t))} + k_{\text{off,TSw}} \cdot TSw2(t) - k_{\text{on,TSw}} \cdot T(t) \cdot Sw2(t) - (a_{\text{RNA}} + k_{\text{flow}}) \cdot Sw2(t)
$$
\n(S6)

$$
\frac{dTSw2(t)}{dt} = k_{on, TSw} \cdot T(t) \cdot Sw2(t) - k_{off, TSw} \cdot TSw2(t) - (a_{RNA} + k_{flow}) \cdot TSw2(t)
$$
\n(S7)

$$
\frac{dGFP(t)}{dt} = k_{tl, GFP} \cdot TSw2(t) + k_{tl, GFP} \cdot bgratio_{sw} \cdot Sw2(t) - (a_{protein} + k_{flow}) \cdot GFP(t)
$$
\n(S8)

Where species *T* is the trigger RNA, *Sw1* the Switch-S28 RNA and *TSw1* the complex of bound Trigger and Switch-S28 RNA. *Sw2* is the Switch-eGFP RNA and *TSw2* the complex of Trigger and Switch-eGFP RNA. *S28* is the σ <sup>28</sup> protein species and *GFP* is the eGFP protein species. *DNAX* is the

concentrations of the DNA coding for RNA species **X** (nM). Further parameter descriptions and units can be found in Table S2.

#### *ODE Model of the reference motif*

The model of the reference motif is largely equal to the CFFL model, except for the absense of an interaction between RNA trigger *T* and mRNA species Sw1. In the reference motif model, RNA species Sw1 represents the S28 RNA (without toehold switch) that results from transcription of DNA species  $R1$  (DNA<sub>Sw1</sub>). Since parameter  $DNA<sub>Sw1</sub>$  now represents the concentration of a different physical DNA species (*R1* instead of *Y1*), the transcription rate of this species (*bSw1,ref*) is assumed to be different than for the CFFL model.

$$
\frac{dT(t)}{dt} = b_T \cdot DNA_T + k_{off, TSw} \cdot TSw2(t) - k_{on, TSw} \cdot T(t) \cdot Sw2(t)
$$
\n
$$
-(a_{RNA} + k_{flow}) \cdot T(t)
$$
\n(S9)

$$
\frac{dSw1(t)}{dt} = b_{\text{Sw1, ref}} \cdot DNA_{\text{Sw1}} - (a_{\text{RNA}} + k_{\text{flow}}) \cdot Sw1(t)
$$
\n(S10)

$$
\frac{dS28(t)}{dt} = k_{t1, S28} * Sw1(t) - (a_{protein} + k_{flow}) \cdot S28(t)
$$
\n(S11)

$$
\frac{dSw2(t)}{dt} = B_{\text{Sw2}} \cdot DNA_{\text{Sw2}} + b_{\text{Sw2}} \cdot DNA_{\text{Sw2}} \cdot \frac{f(S28(t))}{1 + f(S28(t))} + k_{\text{off,TSw}} \cdot TSw2(t) - k_{\text{on,TSw}} \cdot T(f) \cdot Sw2(t) - (a_{\text{RNA}} + k_{\text{flow}}) \cdot Sw2(t)
$$
\n(512)

$$
\frac{dTSw2(t)}{dt} = k_{on, TSw} \cdot T(t) \cdot Sw2(t) - k_{off, TSw} \cdot TSw2(t) - (a_{RNA} + k_{flow}) \cdot TSw2(t)
$$
\n(S13)

$$
\frac{dGFP(t)}{dt} = k_{t1, GFP} \cdot TSw2(t) + k_{t1, GFP} \cdot bgratio_{sw} \cdot Sw2(t) - (a_{protein} + k_{flow}) \cdot GFP(t)
$$
\n(S14)

Where species *T* is the trigger RNA and *Sw1* is the S28 RNA (without any toehold switch). *Sw2* is the Switch-eGFP RNA and *TSw2* the complex of Trigger and Switch-eGFP RNA. *S28* is the σ 28 protein species and *GFP* is the eGFP protein species. *DNAX* is the concentrations of the DNA coding for RNA species **X** (nM). Further parameter descriptions and units can be found in Table S2.

## **Temporal Ultrasensitivity Computation**

The temporal ultrasensitivity calculations were based on the  $\alpha$  and  $\beta$  measures, calculated from the maximum output amplitude per input pulse length (Figure 4C and D).<sup>3</sup>  $\alpha$  is the pulse duration at which the maximum response has increased 10% of the total difference between background expression (no input) and full activation (persistent step input), computed by linearly interpolating the nearest experimental or simulation data points. Similarly, β marks the pulse duration of 90% increase in response. The temporal ultrasensitivity is defined as follows:

temporal ultrasensitivity = 
$$
\frac{\alpha}{\beta}
$$
 (S15)

Thus, the temporal ultrasensitivity is low for a gradually increasing activation upon an increase in input pulse length and approaches 1 for a circuit that approaches immediate switch-like behavior at a given pulse length.



**Table S2:** Model parameter, their units, used values and sampled parameter range. Sample values marked with an asterisk (\* ) were obtained through parameter fitting. The DNA concentrations (DNAT, DNASw1 and DNASw2) are the concentrations used for parameter sampling. The concentrations used in other *in silico* experiments are given in Table S3.



**Figure S10:** Simulations of the influence of the presence of toehold switches at two positions in the CFFL circuit. A) Simulated concentration traces for the CFFL (pink) and reference motif (blue) of both the eGFP output (solid) and σ <sup>28</sup> protein concentration. B) The *ON*/*OFF* ratios of σ <sup>28</sup> and eGFP in the CFFL and reference circuits. The  $\sigma^{28}$  concentration of the reference motif remains constant after addition of an input, since it is constitutively produced. The eGFP output of the reference motif displays an *ON*/*OFF* ratio of 8.6, which is caused only by the toehold switch on the output construct, since the other switch is not present in the reference motif. The fold-change in the  $\sigma^{28}$ concentration of the CFFL mainly represents the activation of the toehold switch on the switch- $\sigma^{28}$ construct and was observed to be 5.3. Together, the activation of the switches largely accounts for the high fold-change in eGFP output of the CFFL, leaving a factor 1.7 to be accounted for by other factors, such as differences in parameter values between the reference motif and CFFL and nonlinearity of the  $\sigma^{28}$ -DNA interaction.



Probability  $0.04$ 



**Figure S11:** *In silico* selection of parameter samples of the CFFL model that display high temporal ultrasensitivity. A) Schematic that shows the selection procedure of the parameter samples. First, a latin hypercube sampling (LHS) of the 13-dimensional parameter space was taken in the logarithmic domain. For each parameter sample, the temporal ultrasensitivity and general output statistics (maximum output and relative activation upon when given a step input) of the CFFL were computed. An initial selection was performed based on the general output measures to ensure reasonable output protein concentrations (between 1 nM and 1 mM) and discernible activation when an input signal is given (at least 5% activation), which also narrows down the sampling to parameter values for which the temporal ultrasensitivity can reliably be determined. The resulting selection of parameter samples is denoted as 'Pre' in this figure. The subsequent step is to select for parameter samples that result in temporal ultrasensitivity higher than 0.2, yielding the 'Post' parameter collection. Comparison of the parameter values found in the Pre and Post collections gives information about the preference for certain parameter values for circuits with high temporal ultrasensititivity. B) The distribution of the values of each circuit parameter in the Pre and Post collections of parameter sets visualized as box plots with a dashed green line indicating the parameter value obtained from a fit to the experimental data. For most parameters, the Pre and Post distributions are comparable, meaning that those parameter values are equally likely to be found before and after selection for a high temporal ultrasensitivity. The parameters that display the largest shift in distribution between Pre and Post are bgratio<sub>sw</sub>, the fraction the translation rate of the bound switch:trigger complex that is observed as leakage in the unbound toehold switch, and  $N<sub>S28</sub>$ , which is the Hill coefficient of the binding of  $\sigma^{28}$  to its promoter. A high temporal ultrasensitivity is associated with a lower bgratio<sub>sw</sub> and a higher N<sub>S28</sub>. C) 2D distributions of parameter values of a select combination of parameters, displaying the Pre and Post collections. These plots show which combinations of parameter values are enriched when selecting for a high temporal ultrasensitivity.







**Table S3:** DNA construct concentrations of all batch, flow and *in silico* experiments.



**Table S4:** Composition of the DNA solutions used in flow experiments. A flow experiment of an initial fill of the reactor with TXTL reaction mixture and 35% reactor volume *No input* solution. Subsequently, every 15 minutes 40% of the reactor was refreshed with a mixture consisting of 65% TXTL mixture and the remaining 45% one of the DNA solutions. The sequence of DNA solutions was: 11 or 15 steps *No input*; 1 step *High input*; 0, 1, 3 or 7 steps *Low input* (to create 15 min, 30 min, 1 h and 2 h input pulses) and the remaining steps *No input* until 11 h of reaction time were reached. Additionally, a negative control without input signal was constructed by only supplying the *No input* solution. The persistent input experiments were conducted using the following sequence: 11 or 15 steps *No input*; 1 step *High input* and the remaining steps *Low input*.

## **References**

(1) Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V., and Murray, R. M. (2014) Linear DNA for Rapid Prototyping of Synthetic Biological Circuits in an *Escherichia coli* Based TX-TL Cell-Free System. *ACS Synthetic Biology 3*, 387–397.

(2) Yelleswarapu, M., van der Linden, A. J., van Sluijs, B., Pieters, P. A., Dubuc, E., de Greef, T. F. A., and Huck, W. T. S. (2018) Sigma Factor-Mediated Tuning of Bacterial Cell-Free Synthetic Genetic Oscillators. *ACS Synth. Biol. 7*, 2879–2887.

(3) Gerardin, J., Reddy, N. R., and Lim, W. A. (2019) The Design Principles of Biochemical Timers: Circuits that Discriminate between Transient and Sustained Stimulation. *Cell Systems 9*, 297- 308.e2.