To appear, 2013 American Control Conference (ACC)
http://www.cds.caltech.edu/~murray/papers/dhm13-acc.html

DESIGN AND IMPLEMENTATION OF A BIOMOLECULAR CIRCUIT FOR TRACKING PROTEIN CONCENTRATION

Emmanuel L. C. de los Santos[1*], Victoria Hsiao[1*] and Richard M. Murray[1]

***Abstract*— This paper describes the design, modeling, synthesis and preliminary validation of a protein concentration regulator circuit. The circuit is designed to maintain the level of a target protein to a reference level, specified by the amount of another protein. This is implemented using a single negative feedback loop that inhibits the production of the target protein once its concentration is equal to the reference amount. A mathematical model consisting of a set of ODEs is derived from mass action laws and Hill function approximations of protein production. Steady-state analysis of the model is used to predict parameter sensitivity and experimental behavior. We implemented this circuit in *E. coli* using scaffold-based sequestration and transcriptional activation. Preliminary experimental results show the system matching predictions from our model and performing the expected task.**

I. INTRODUCTION

As biomolecular circuits in synthetic biology increase in complexity, the incorporation of closed loop feedback controllers into circuit designs will become necessary for robust performance. In order to do this in a biochemical circuit, a mechanism that can track the difference in the amount of an output to a given reference value is required. With this in mind, we designed a sequestration-based negative feedback circuit, in which one protein tracks the level of another protein, which we consider to be our reference.

This paper discusses the design, modeling and biochemical implementation of this circuit. In this circuit, we set the level of our reference, protein A. The presence of protein A conditionally activates the synthesis of protein B, our desired output. Protein B is designed to have regions that interact with protein A, creating a negative feedback loop that shuts down the synthesis of protein B when the levels of A and B are equal. The design of this circuit was partially inspired by a circuit designed by Franco *et al.* [1] where an *in vitro* RNA transcription circuit coupled the levels of two double stranded DNA species through sequestration and transcriptional regulation.

Synthetic feedback regulation is a critical hurdle for the design and engineering of complex synthetic circuits. Recent theoretical work by Ang *et al.* [2] proposed and modeled a two-promoter transcription-level system for implementing integral control. Stapleton *et al.* [3]demonstrated a transcription-level negative feedback circuit in mammalian cells, in which an RNA binding protein repressed translation of its own mRNA. Here, we model and demonstrate a protein-level negative feedback circuit.

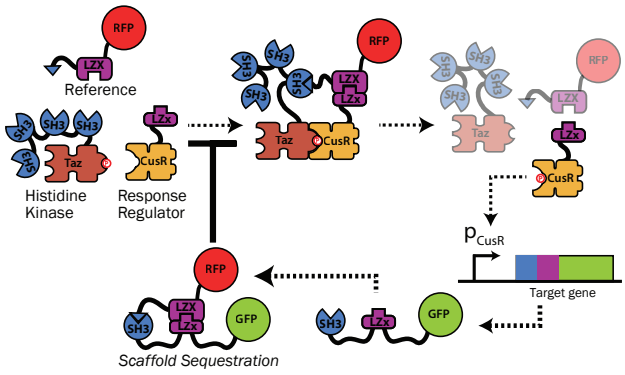

Fig. 1. A visual schematic of our circuit design. The circuit regulates the production of the amount of target protein with respect to the amount of reference protein. This is done using programmable scaffold domains. Expression of the target is dependent on the amount of free scaffold. The target contains domains which sequester free scaffold creating a negative feedback loop.

Previously, Dueber *et al.* showed that synthetic scaffold proteins could be used to co-localize intermediates in the production of a metabolic product [4], [5], [6]. They designed synthetic scaffolds out of single protein binding domains linked together by short repeating peptide sequences. Further work by Whitaker *et al.* showed that prokaryotic two-component systems could be selectively phosphorylated using these eukaryotic scaffolds [7].

Two-component systems are naturally found in bacteria as environmental sensor and response systems [7], [8]. A transmembrane histidine kinase senses environmental conditions, such as osmolarity or pH, and phosphorylates a response regulator protein. The phosphorylated response regulator becomes an active transcription factor, allowing the transcription and translation of target genes.

Building off of this previous work [7], our circuit uses a two-domain synthetic protein scaffold as our reference protein (scaffold) that activates transcription of our target (Fig. 1). The target is designed to bind to the scaffold, creating a negative feedback loop via sequestration.

The main contribution of this paper is the design of a biochemical reaction network that implements closed-loop tracking of the level a reference protein. The circuit was

Research supported in part by the Benjamin M. Rosen Bioengineering Center, the California Institute of Technology, and the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

* These authors contributed equally to the work
[1] E de los Santos, V Hsiao and R Murray are in the Department of Bioengineering, California Institute of Technology, Pasadena, CA 91125.

designed from biochemical reactions that could be realized experimentally. Mathematical modeling was done using the biochemical reactions in the circuit design. Preliminary experimental results show good agreement with our model suggesting that the presence of the negative feedback loop regulates the production of our target protein in the correct way. Further characterization is needed in order to confirm this.

II. Circuit description

The circuit consists of four major components: a synthetic scaffold protein (Sc), a histidine kinase (HK), a response regulator (RR), and a target protein. The synthetic scaffold serves as our reference level. It consists of two pairs of small protein-protein binding domains: the SH3-peptide/SH3-domain and the leucine zipper LZx/LZX, which have been previously used and characterized [6]. The HK and RR are commonly found in bacterial signaling systems. In our system, they serve as the mechanism for conditional activation of the target protein based on the reference levels.

In natural systems, there are hundreds of HK/RR pairs which perform signaling functions with minimal crosstalk. Since we desire activation of the RR to be conditional on our reference protein, we have selected a non-cognate pair of histidine kinase and response regulator, known to be orthogonal *in vivo*. Based on previous designs by Whitaker *et al.*, the HK is bound to a SH3 ligand and the RR is bound to a LZx domain, which both bind to Sc [7]. When both HK and RR are bound to the scaffold, the RR becomes an active transcription factor, resulting in the expression of our target protein [7]. The target protein contains domains for both the SH3 ligand and LZx, which are complementary to the scaffold binding domains. This allows the target to out-compete both the HK and RR for binding to the scaffold, effectively sequestering the scaffold, repressing its own production. We expect that with this circuit design, we can regulate the relative steady-state concentrations of scaffold to anti-scaffold). Figure 1 shows a schematic description of our circuit design.

III. Mathematical model

We constructed an ODE-based dynamical model of our circuit from the chemical reactions occurring in our system [9]. The model consists of 19 species corresponding to the active and non-active states of the HK and RR, and the different two-species and three-species complexes that can be formed. The model contains 56 different reactions that describe production and degradation of our basic species (Sc, HK, RR and target), activation of HK and RR, binding of HK and RR to Sc, and sequestration of Sc by our target. With the exception of the formation of our target protein (AS), mass-action kinetics was used to describe the rate laws of the reactions. Parameters were selected from experimental values found in the literature where available [6], [8], [10], [11].

The HK is activated via auto-phosphorylation to form $\mathrm{HK_p}$, transfer of this phosphate group activates the RR to $\mathrm{RR_p}$. $\mathrm{RR_p}$ is inactivated either through phosphatase activity of the HK or auto-dephosphorylation. Key reactions that describe this process in our model are:

$$\mathrm{HK} \underset{k^r_{\mathrm{HK_p}}}{\overset{k^f_{\mathrm{HK_p}}}{\rightleftharpoons}} \mathrm{HK_p} \tag{1}$$

$$\mathrm{HK_p} + \mathrm{RR} \underset{k^r_{\mathrm{NCP}}}{\overset{k^f_{\mathrm{NCP}}}{\rightleftharpoons}} \mathrm{HK} + \mathrm{RR_p} \tag{2}$$

$$\mathrm{RR_p} \xrightarrow{k_{\mathrm{dephos}}} \mathrm{RR} \tag{3}$$

Since our design is dependent on activatation of the RR with respect to the amount of reference protein (scaffold) that is available for binding, we expect the non-scaffold mediated rate of RR activation (described by reaction 2) to be low. Accordingly, rate constants were chosen to be similar to those between a HK and its non-cognate RR [8]. We assume that our reference protein has two independent binding sites to which a RR and HK can bind, and that this binding is independent of the phosphorylation state of HK and RR. When both $\mathrm{HK_p}$ and RR are bound to a scaffold, RR activation occurs at a faster rate, chosen to be similar to the rate of phosphorylation between a HK and its cognate RR. Some of the reactions that describe this process in our model are:

$$\mathrm{Sc} + \mathrm{HK_p} \underset{k^r_{\mathrm{SH_3}}}{\overset{k^f_{\mathrm{SH_3}}}{\rightleftharpoons}} \mathrm{ScHK_p} \tag{4}$$

$$\mathrm{ScHK_p} + \mathrm{RR} \underset{k^r_{\mathrm{LZX}}}{\overset{k^f_{\mathrm{LZX}}}{\rightleftharpoons}} \mathrm{ScHK_pRR} \tag{5}$$

$$\mathrm{ScHK_pRR} \underset{k^r_{\mathrm{CP}}}{\overset{k^f_{\mathrm{CP}}}{\rightleftharpoons}} \mathrm{ScHKRR_p} \tag{6}$$

Activated RR, $\mathrm{RR_p}$ allows for the production of our target protein, AS with the reaction:

$$\mathrm{RR_p} \xrightarrow{\beta_{AS}} \mathrm{RR_p} + \mathrm{AS} \tag{7}$$

The creation of the anti-scaffold uses a Hill equation to describe its kinetic rate. We assume that the rate of production of AS from $\mathrm{RR_p}$ is unaffected by any other molecules $\mathrm{RR_p}$ is bound to. Therefore, the rate of production of AS is:

$$\begin{aligned} \frac{d[\mathrm{AS}]}{dt} =& \beta_{AS}\left[\beta_0 + K_d\left(\frac{f}{g}\right)\right] \\ f =& \mathrm{RR_p} + \mathrm{ScRR_p} + \mathrm{ScHKRR_p} \\ &+ \mathrm{ScHK_pRR_p} + \mathrm{ScRR_pAS} \\ g =& K_d + \mathrm{RR_p} + \mathrm{ScRR_p} + \mathrm{ScHKRR_p} \\ &+ \mathrm{ScHK_pRR_p} + \mathrm{ScRR_pAS} \end{aligned} \tag{8}$$

AS contains the same protein domains as RR and HK, enabling it to competitively bind to and sequester Sc. As with HK and RR binding to Sc, we assume that As can bind to either of these sites independently. However, once AS is bound to Sc, if an occupied binding site becomes available,

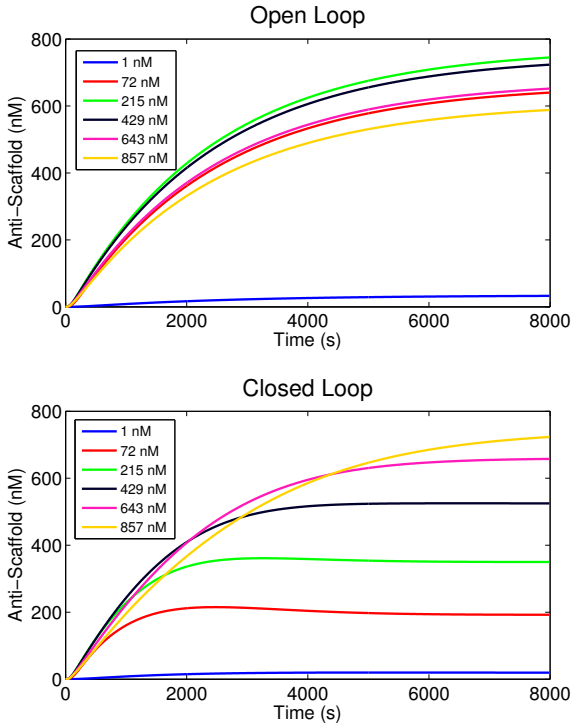

Fig. 2. Time trace of total anti-scaffold (AS) concentration for varying scaffold(Sc) concentrations at a fixed response regulator concentration (350 nM).

we assume that AS is able to bind to and sequester this site very rapidly. Furthermore, we assume that once AS is bound to both of the sites in Sc, the sequestration of Sc by AS is irreversible. Some of the reactions that describe AS sequestration of Sc are:

$$\mathrm{Sc} + \mathrm{AS} \xrightarrow{k^f_{\mathrm{seq}}} \mathrm{ScAS} \tag{9}$$

$$\mathrm{ScHK} + \mathrm{AS} \underset{k^r_{\mathrm{LZX}}}{\overset{k^f_{\mathrm{LZX}}}{\rightleftharpoons}} \mathrm{ScHKAS} \xrightarrow{k^r_{\mathrm{SH_3}}} \mathrm{ScAS} + \mathrm{HK} \tag{10}$$

$$\mathrm{ScRR} + \mathrm{AS} \underset{k^r_{\mathrm{SH_3}}}{\overset{k^f_{\mathrm{SH_3}}}{\rightleftharpoons}} \mathrm{ScRRAS} \xrightarrow{k^r_{\mathrm{LZX}}} \mathrm{ScAS} + \mathrm{RR} \tag{11}$$

We have implemented our model using the SimBiology toolkit in MATLAB, which converts the reactions into a set of ODEs. Using the numerical solver ode23t, the simulations were run until each of the species reached steady-state. To evaluate the behavior of our circuit, simulations were performed over a range of steady-state Sc and RR concentrations. This was done by changing the rate of synthesis of the corresponding species. We compared the results of this simulation to the open loop circuit, which used the same model without the sequestration reactions. Figures 2 and 3 summarize the results of these simulations.

In the simulations, the addition of the negative feedback regulates the level of our target protein dependent on the amount of Sc in the system. The maximum reference level

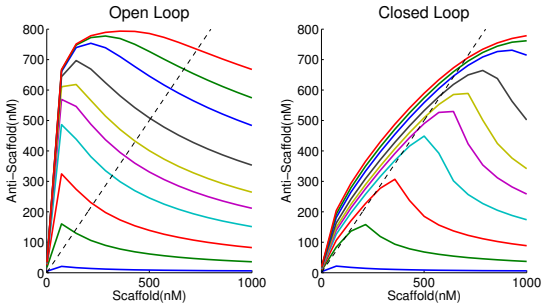

Fig. 3. Steady-state simulations for different Sc and RR concentrations. The presence of the negative feedback loop is able to regulate the level of our target molecule, the anti-scaffold (AS), allowing it to track the level of our reference, the scaffold (Sc).

which our target is capable of tracking is dependent on the amount of RR in the system, which mediates the scaffold-dependent expression of our target protein. As we increase the RR levels, we achieve the maximum amount of target protein that can be produced. The range in which the target protein is able to track the reference protein also increases. For a fixed RR concentration, the amount of target protein decreases from a maximum value as Sc increases. This is because co-localization of the HK and RR decreases if there is too much scaffold in the system. The regulation of the production of the target protein by the feedback loop is evident when compared to the open loop case where the target protein level quickly reaches its maximum concentration, and then decreases due to the loss of co-localization with increasing Sc.

In order to further characterize our circuit, and to determine which parameters our system was most sensitive to we used the SimBiology toolkit to numerically calculate the sensitivity of our output, the total target protein concentration (x_{TOT}) with respect to the different parameters in our model. Specifically we calculated:

$$\frac{dx_{\mathrm{TOT}}}{x_{\mathrm{TOT}}} \Big/ \frac{d\theta_e}{\theta_e}$$

Figure 4 shows the normalized sensitivities of the total amount of target protein with respect to different parameters for a fixed scaffold concentration (572 nM) and increasing RR concentrations. We reasoned that increasing RR concentrations would be analogous to increasing the gain of our feedback circuit. Our circuit is most sensitive to parameters affecting the phosphorylation rates of the HK and RR, $k^f_{\mathrm{HK^P}}$,$k^r_{\mathrm{HK^P}}$, k_{dephos}, parameters dealing with sequestration, $k^f_{\mathrm{SH_3}}$, $k^r_{\mathrm{SH_3}}$, k^r_{LZX}, k^r_{LZX}, k^r_{seq}, and parameters affecting the production of HK, Sc and RR, β_{Sc}, β_{RR}, and β_{HK}. The sensitivity analysis informs us on which parameters have the most impact on circuit performance. By choosing to focus on parameters which are more easily tuned experimentally, we can utilize the model to optimize the circuit design process.

IV. Synthesis of the Circuit

Using plasmids obtained from WR Whitaker and JE Dueber, standard cloning practices were used to assemble the cir-

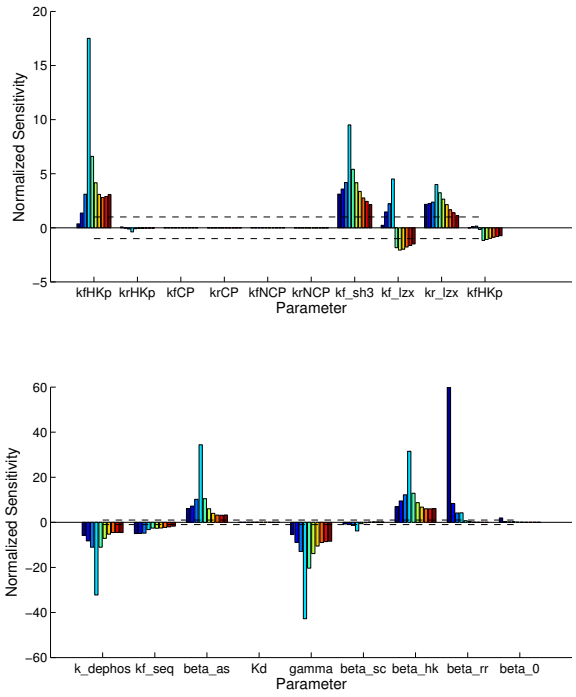

Fig. 4. Sensitivity analysis showing the normalized sensitivity of total anti-scaffold to the different parameters in our model. The total amount of anti-scaffold is most dependent on the production and degradation rates of the various species, the affinity of the programmable scaffolds, and the deactivation rate of the response regulator. For the level of histidine kinase used in these simulations (2 nM), the total anti-scaffold concentration is invariant with the auto-activation rates of the histidine kinase.

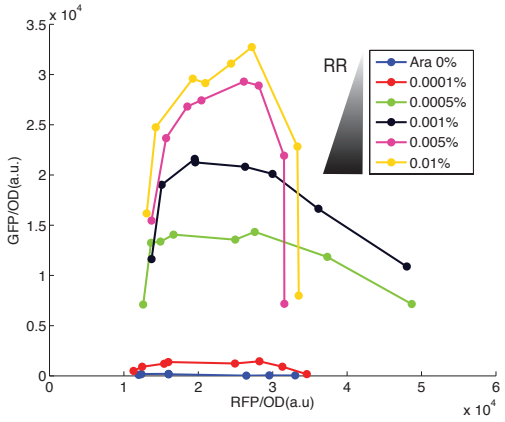

Fig. 5. Steady state fluorescence values for the open loop circuit with scaffold-mCherry fluorescence on the x-axis (0 - 150nM aTc induction, L-R) and GFP values on the y-axis. Increasing arabinose concentrations indicate increasing response regulator concentrations.

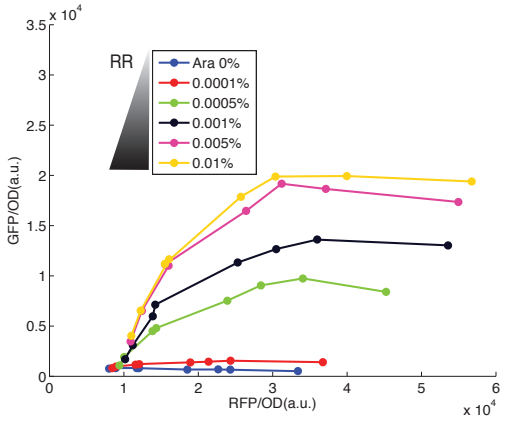

Fig. 6. Steady state fluorescence values for the closed loop circuit with scaffold-mCherry fluorescence on the x-axis (0 - 150nM aTc induction, L-R) and anti-scaffold-GFP values on the y-axis. Increasing arabinose concentrations indicate increasing response regulator concentrations.

cuit onto three plasmids. The reference (scaffold) component was cloned into a high-copy plasmid under a tetracycline-inducible promoter. The histidine kinase was cloned into a low-copy plasmid under a constitutive promoter. The response regulator and target (anti-scaffold) were cloned into a medium-copy plasmid, with the response regulator under an arabinose-inducible promoter and the anti-scaffold under a response-regulator-activated promoter.

We implemented our circuit *in vivo* using a strain of *E. coli* with the two endogenous HK/RR pairs used in our system knocked out [7]. This was done to reduce any non-scaffold mediated activation of the RR.

Characterization of circuit behavior was done by independently varying the levels of Sc and RR. The inducible promoters allowed for the tunable expression of Sc and RR by anhydrous tetracyline (aTc) and arabinose, respectively. The Sc and AS were fused to fluorescent reporters, providing real-time circuit behavior data in bulk and single cell measurements. Cells were induced with 0 - 150 nM of anhydrous tetracycline (aTc) and 0 - 0.01% arabinose. A modified version of MOPS EZ Rich media (Teknova, M2105) is used to reduce the effect of media coloration on fluorescence measurements. Induced cultures are grown overnight (20 hrs, 37C) and bulk fluorescence measurements (GFP:488/520nm, RFP:580/610nm) are taken with a BioTek H1F microplate reader. An open loop version of the circuit, which contained a fluorescent reporter but none of the domains required to sequester the Sc was tested in parallel with our closed loop circuit.

V. Experimental Results

Overnight cultures for the open and closed loop circuits were grown overnight and steady state fluorescence values for the scaffold-mCherry, GFP (open loop) and anti-scaffold-GFP (closed loop) fusion proteins were measured. Figures 5-6 show the steady state values of scaffold versus anti-scaffold in terms of total normalized RFP versus GFP fluorescence.

In the open loop circuit (Fig. 5), we see that with no response regulator (Ara 0%), GFP induction is close to zero. As response regulator concentration increases, the steady state level of GFP peaks at some optimal concentration of scaffold ($2 - 3 \times 10^4$ a.u) and then decreases with increasing scaffold induction. We hypothesize that as scaffold

concentrations exceed those of the response regulator and histidine kinase, each free scaffold will be bound to only one or the other, effectively reducing the amount of activated response regulators. This single-occupancy effect is one that has previously been shown [6]. We also see that with higher concentrations of response regulator, the scaffold seemed to saturate at a lower maximum concentration, resulting in the curves shown for 0.005% and 0.01% arabinose. Finally, the effect of leaky promoter activation should also be considered. In the case of no scaffold induction (0 nM aTc, first data point), we would expect to still see some background level of scaffold production, which, when combined with an excess of response regulator and histidine kinases to occupy those scaffolds, results in increasing background levels of anti-scaffold. Looking at the x-axis of Fig. 5, we see that scaffold concentrations start at 1×10^4 a.u. with 0 nM aTc induction. As the arabinose induction of response regulator increases, so too does the steady state concentration of anti-scaffold.

In the closed loop circuit (Fig. 6), we expect the steady state anti-scaffold concentrations to increase more gradually with increasing scaffold concentrations, which is apparent in the experimental data.

Rather than peaking at 2×10^4 a.u. scaffold fluorescence, the closed loop circuit asymptotes closer to 4×10^4 a.u. and does not yet show the effects of single scaffold occupancy. We believe this is consistent with our model, since anti-scaffold sequestration effectively reduces the amount of free scaffold such that it actually lessens the probability of single-scaffold occupancy. Based on the model, we predict that if the scaffold induction range were increased beyond 150 nM, the steady state anti-scaffold concentration would show a similar decrease as the open loop, just at higher scaffold concentrations. Finally we see that although it has nearly the same level of leaky scaffold production, the anti-scaffold levels are much lower than their open-loop counterparts.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have mathematically modeled and shown preliminary experimental results for a protein-level concentration regulator circuit. We have shown that several experimental parameters, most notably those related to phosphorylation and sequestration rates, are the most influential on circuit performance. Experimentally, we have constructed and tested both the concentration regulator circuit and an open loop version for comparison, and shown that steady state behavior of both circuits matches the predictions of our model. In future studies, we would design and characterize additional circuits to test each of these parameters, and compare the results with model predictions. Additionally, it may be interesting to further model and experimentally explore time-dependent behavior of the circuit.

ACKNOWLEDGMENTS

The authors would like to thank John E. Dueber and Weston R. Whitaker from the University of California, Berkeley, for providing the initial plasmid constructs, the *E. coli* strain WW62, and continuing experimental advice.

REFERENCES

[1] Franco, E., Forsberg, P-O., Murray RM. Design, modeling and synthesis of an *in vitro* transcription rate regulatory circuit. ACC 2008

[2] Ang, J., Bagh, S., Ingalls, B.P., McMillen, D.R. Considerations for using integral feedback control to construct a perfectly adapting synthetic gene network. J Theor Biol. 2010 Oct 21;266(4):723-38.

[3] Stapleton, J.A., Endo, K., Fujita, Y., Hayashi, K., Takinoue, M., Saito, H., and Inoue, T. Feedback Control of Protein Expression in Mammalian Cells by Tunable Synthetic Translational Inhibition. ACS Synth. Biol., 2012, 1 (3), pp 8388

[4] Dueber, J.E., Wu, G.C, Malmirchegini, G.R., Moon, T.S., Petzold C.J., Ullal, A.V., Prather, K.L., Keasling, J.D. Synthetic protein scaffolds provide modular control over flux through an engineered metabolic pathway. Nature Biotechnology. 2009 Aug; 27(8):753-9.

[5] Moon, T.S., Dueber, J.E., Shiue, E. Prather, K.L. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. Metabolic engineering. May 2010. 12(3):298-305.

[6] Whitaker, W.R., and Dueber, J.E. Metabolic Pathway Flux Enhancement by Synthetic Protein Scaffolding. Methods in Enzymology, 2011, 497, 448-466.

[7] Whitaker, W.R., Davis, S.A., Arkin, A.P., and Dueber, J.E. Engineering Robust Control of Two-Component System Phosphotransfer Using Modular Scaffolds. Proc Natl Acad Sci USA. 2012 Oct 30;109(44):18090-5.

[8] Groban, E.S., Clarke, E.J., Salis, H.M., Miller, S.M., Voigt, C.A. Kinetic Buffering of Cross Talk between Bacterial Two-Component Sensors. J. Mol. Biol., 2009, 390, 380-393

[9] Alon, U. An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman & Hall/CRC, 2006

[10] Pazy, Y., Wollish, A.C., Thomas, S.A., Miller, P.J., Collins, E.J., Bourret, R.B., Silversmith, R.E. Matching biochemical Reaction Kinetics to the Timescales of Life: Structural Determinants That Influence the Autodephosphorylation Rate of Response Regulator Proteins. J. Mol. Biol.,2009, 392, 1205-1220

[11] Solomaha, E., Szeto, F.L., Yousef, M.A., Palfrey, H.C. Kinetics of Src homology 3 domain association with the proline-rich domain of dynamins: specificity, occlusion, and the effects of phosphorylation. J Biol Chem., 2005, 280(24),23147-56