Presented at Synthetic Biology: Engineering, Evolution and Design (SEED), 14-18 July 2014
http://www.cds.caltech.edu/~murray/papers/smmm14-seed.html

# Engineering Transcriptional Regulator Effector Specificity Through Rational Design and Rapid Prototyping

Emmanuel L.C. de los Santos,[*,†] Joseph T Meyerowitz,[†,‡] Stephen L Mayo,[†] and Richard M. Murray[†,¶]

*Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA, and Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA, USA*

E-mail: emzodls@caltech.edu

[*]To whom correspondence should be addressed
[†]Caltech Biology and Biological Engineering
[‡]Caltech Chemistry and Chemical Engineering
[¶]Caltech Control and Dynamical Systems

**Abstract**

The pursuit of circuits and metabolic pathways of increasing complexity and robustness in synthetic biology will require engineering new regulatory tools. Feedback control based on relevant molecules, including toxic intermediates and environmental signals, would enable genetic circuits to react appropriately to changing conditions. In this work, computational protein design was used to create functional variants of qacR, a tetR family repressor, responsive to a new targeted effector. The modified repressors target vanillin, a growth-inhibiting small molecule found in lignocellulosic hydrolysates and other industrial processes. A computationally designed library was screened using an *in vitro* transcription-translation (TX-TL) system. Leads from the *in vitro* screen were characterized *in vivo*. Preliminary results demonstrate dose-dependent regulation of a downstream fluorescent reporter by vanillin. These repressor designs provide a starting point for the evolution of improved variants. We believe this process can serve as a framework for designing new sensors for other target compounds.

# Introduction

Synthetic biology currently lacks wires. While we can control the expression of target genes with transcriptional regulators, triggers for these transcriptional regulators are limited to a small number of molecules and other inputs (e.g. light)[1]. As a consequence, most synthetic circuits right now are limited to proof-of-principle demonstrations without being extendable to real world applications. In order to design synthetic control circuits for real world applications, such as metabolic engineering or biofuel production, signals from these pathways, such as the level of a toxic metabolic intermediate need to be transmitted into existing transcriptional control machinery. This work intends to develop a framework to use a combination of computational protein design (CPD) and rapid prototyping using an *in vitro* transcription-translation (TX-TL) system to switch effector specificity of existing transcriptional regulators to respond to targeted small molecules of interest (Figure 1).

The tetR family of transcriptional regulators is a large family of transcriptional regulators found in bacteria. They are named after the tetR repressor, which controls the expression of tetA, an efflux pump for tetracycline[2]. They contain two domains, a helical-bundle ligand-binding domain and a

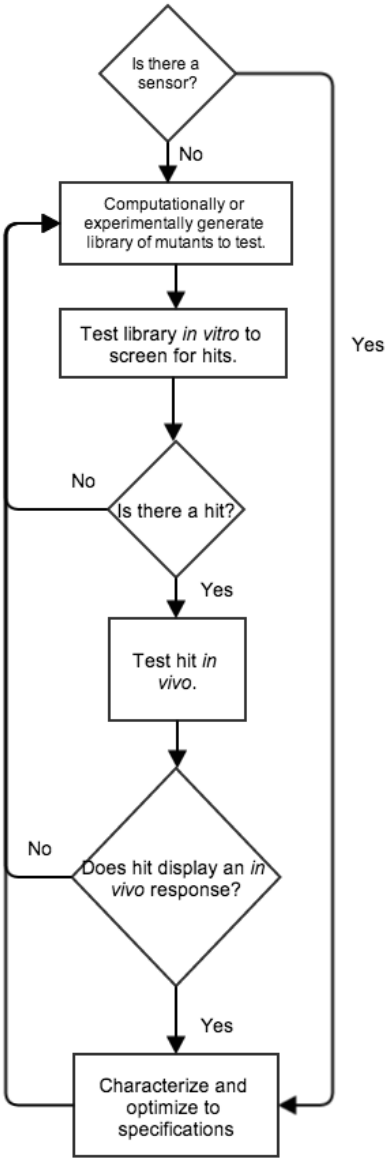



Figure 1: **Workflow for generating novel sensors.** The *in vitro* TX-TL platform allows for the rapid screening of an explicit sequence library that can be designed computationally. Hits from the in vitro screen can then be further characterized in vivo to see if the hit contains the desired specifications. Further refinement can then be done of the hits through directed evolution or further computational design until specifications necessary are achieved.

helix-turn-helix DNA-binding domain. In the absence of their inducing molecule, tetR repressors are bound to DNA, preventing the transcription of downstream genes. Inducer binding to the ligand-binding domain causes a conformational change in the DNA binding domain that causes dissociation from the DNA, allowing transcription of downstream genes. The tetR transcriptional

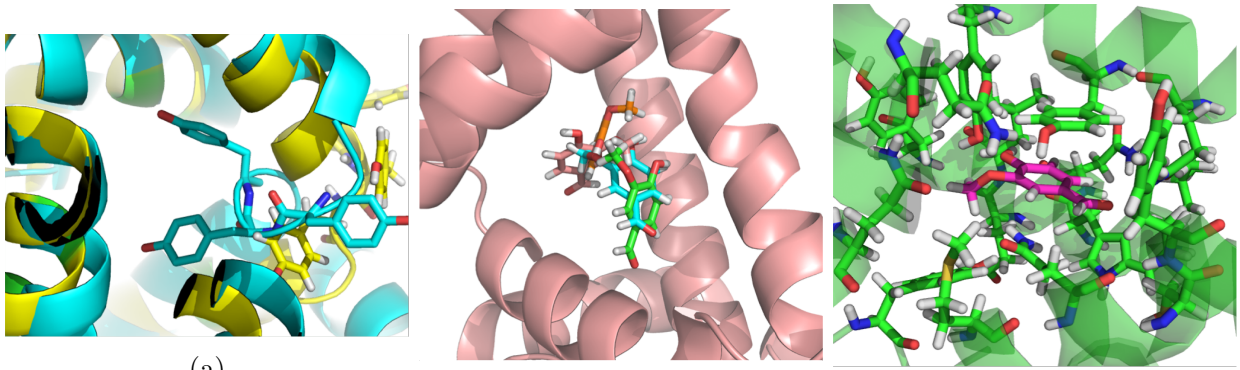
regulation machinery has been used in the design of synthetic circuits, including the repressilator[3] and the toggle switch[4].

qacR is a tetR-family repressor found in *S. aureus* that controls the transcription of qacA, an efflux pump that confers resistance to a large number of quaternary anionic compounds. qacR has been studied because it is induced by a broad range of structurally dissimilar compounds[5]. Structural examination of qacR in complex with different small molecules has shown that qacR has two different binding regions inside a large binding pocket. While qacR has multiple binding modes with its inducers, in all cases for which there are structures, binding of the inducer causes a tyrosine explusion moving one of the helices and altering the conformation of the DNA binding domain rendering qacR unable to bind DNA[6–8]. Crystal structures of inducer-bound forms of qacR and the qacR-DNA complex coupled with a definitive structural mechanism for qacR induction make it the ideal starting point for CPD of new transcriptional regulators. In this work, we describe our efforts to apply our framework to engineer qacR to sense vanillin, a phenolic growth inhibitor that is a byproduct of lignin breakdown to convert biomass into intermediate feedstock for biofuel production[9].

# Results and discussion

## Computationally Aided Protein Design of QacR Library

An *in silico* model of vanillin was constructed using the Schrödinger software suite. Partial charges for vanillin were computed using Optimization in Jaguar version 7.6[10] using HF/6-311G** as the basis set. Vanillin rotamers were chosen using chemical intuition by looking at the ideal angles for the carbon hybrid orbitals. A model of an idealized vanillin binding pocket was designed by looking at the protein data bank for proteins that bound small molecules similar to vanillin. Models of vanillin in the qacR binding pocket were generated using the Phoenix Match algorithm[11]. The algorithm was asked to find potential vanillin binding sites close to the location of the tyrosine expulsion in the binding pocket of qacR (Figure 2a) while being in proximity to amino acid positions to allow for favorable pi-stacking and hydrogen bonding interactions. The Phoenix Match algorithm yielded



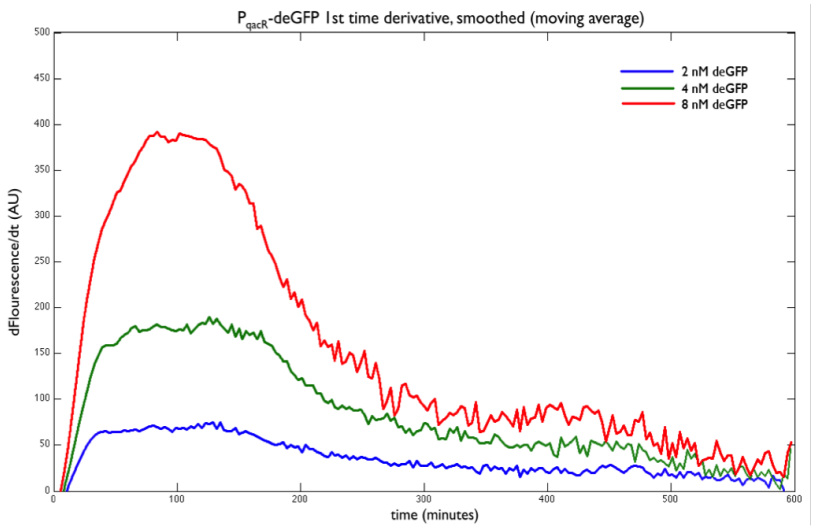
(a) (b) (c)

Figure 2: **Computational design of qacR variant library.** (A) Overlay of structures of the non-ligand bound (cyan, PDB ID: 1JTO) and ligand bound (yellow, PDB ID: 3BQZ) conformations of qacR. The binding of the ligand causes the movement of three tyrosine residues this causes a conformational shift in the DNA- binding domain of the protein making it unable to bind to DNA. (B) Potential vanillin binding sites in qacR found using Phoenix Match. (C) Computational model of a candidate qacR design with vanillin (magenta) in the binding pocket.

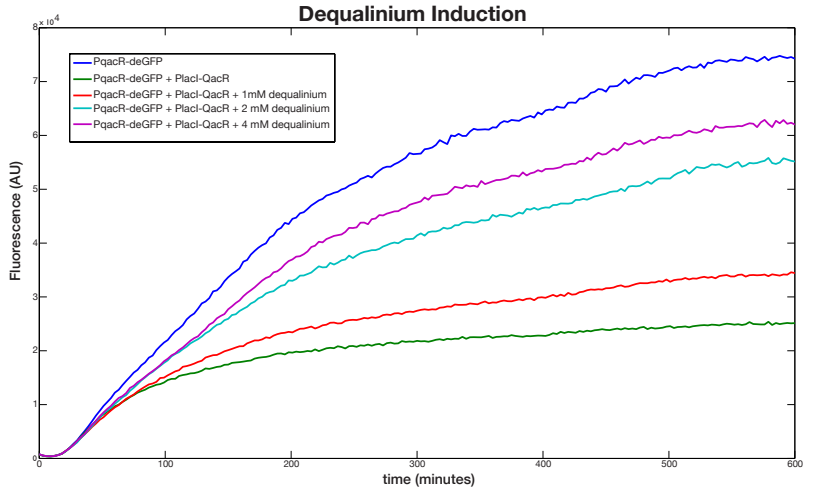
four potential binding positions for vanillin (Figure 2b). Computational protein sequence design was then used to select amino acid residues at the positions around the potential vanillin binding sites. Monte Carlo with simulated annealing[12] and FASTER[13] were used to sample conformational space. A backbone independent conformer library with a 1.0Å resolution was used for the designed residues[11]. Structures were scored using the PHOENIX energy function with the inclusion of an additional geometry bias term that favored pi-stacking and hydrogen bonding interactions[11]. In order to minimize the possibility of steric clahses in the protein, calculations that considered both the DNA-bound state and the ligand-bound state using a Multi-state design algorithm were also performed[14]. Finally, calculations that included an energy bias to favor the wild-type residue were also performed. The lowest energy sequences from these four calculations (single-state biased, single-state non-biased, multi-state biased, and multi-state non-biased) were analyzed to generate a set of ten mutants for *in vitro* testing (Figure 2c).

## *In vitro* screening of designs

We first decided to look at the response of the wild-type protein to one of its native inducers. This was done by placing Green Fluorescent Protein (GFP) downstream of the qacA promoter sequence




(a)




(b)

Figure 3: **Validation of TX-TL Screening.** (A) Time derivative of GFP fluoresence as a function of time. Plasmid encoding GFP downstream of the native qac promoter was added to the TX-TL platform. Higher concentrations of plasmid yielded more GFP signal. (B) Response of wild-type qacR to dequalinium. DNA encoding GFP and wild-type qacR was added to the TX-TL system. Increasing fluorescent signal is observed with increasing concentrations of dequalinium. The highest fluorescent signal is observed when there is no repressor in the system, demonstrating the ability of TX-TL to test for qacR repression and de-repression.

($P_{QacA}$). While we observed a hundred-fold decrease in fluoresence in cells containing plasmids encoding the wild-type qacR gene in addition to GFP, addition of berberine, a native qacR inducer yielded no observable difference in fluoresence (data not shown). We hypothesized that the inducer was not getting into the cells due to the differences in cell wall between gram-positive and gram-negative bacteria. Because of this, we decided to use an *in vitro* transcription-translation (TX-TL)

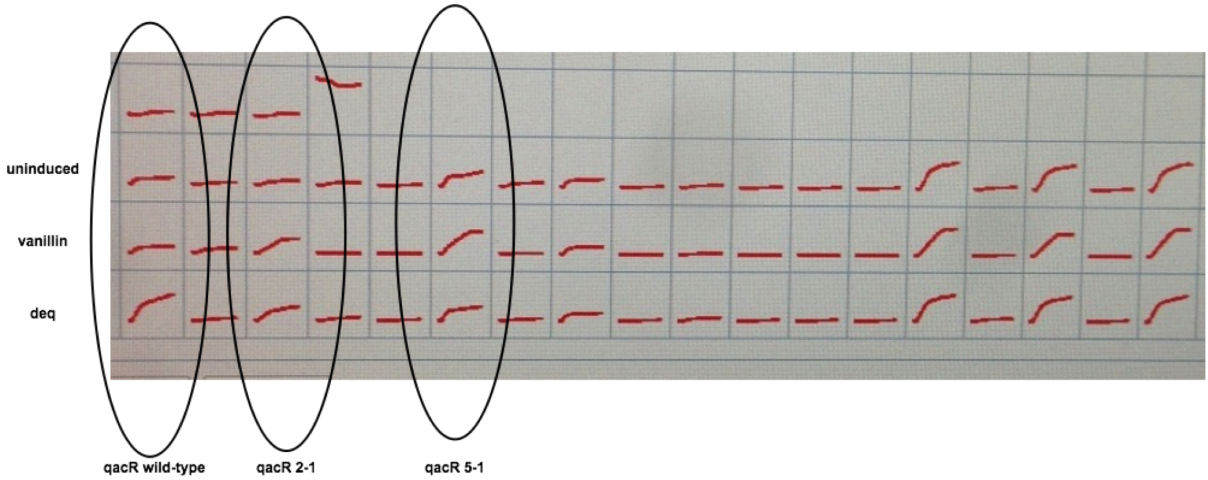



Figure 4: ***In vitro*** **TX-TL screen of qacR mutants found potential candidates for further testing.** Seventeen candidate qacR designs were screened using TX-TL. Plasmids containing DNA encoding each of the qacR variants were placed into the system along with water, dequilinium (native qacR inducer) and vanillin. To monitor qacR response, a plasmid encoding GFP downstream of the native qacA promoter was also added to the system. Two of the qacR mutants that seemed to display some level of vanillin induction were selected for further characterization.

system to test our designs[15]. Initial testing showed an increase in GFP fluoresence with increasing concentrations of plasmid encoding pQacA-GFP (Figure 3a). The addition of plasmid encoding qacR resulted in a decrease in fluoresence. The addition of dequalinium, a colorless native qacR inducer, resulted in an increase in fluoresence until about 85% of the fluoresence when no DNA encoding repressor was present (Figure 3b). Given the ability to show both repression and induction of the native qacR system, we used the TX-TL system to characterize our designs.

None of the initial designs showed any repression of GFP fluoresence. Analysis of the computational models of a design that contained only three mutations from a qacR mutant that was previously shown to be functional, showed the mutations causing the potential for steric clashes in the DNA bound state. We designed a second library reverting either the 50th and 54th positions (A50F/W54L) or the 119th position (Y119L) to their wild-type identity.

The second library was screened using the TX-TL platform to look for qacR variants that responded to vanillin. Plasmids containing DNA that encoded each of the qacR variants or the wild-type qacR sequence were placed into a TX-TL reaction containing either water, dequilinium or vanillin. QacR activity was monitored by a plasmid that encoding GFP downstream of $P_{QacA}$.

Two of the mutants, qacR 2-1 and qacR 5-1, displayed an increase in fluoresence in the presence of vanillin and dequilinium over water (Figure 4). We tested these mutants *in vivo* for further characterization.

## *In vivo* testing of qacR 2-1 and qacR 5-1

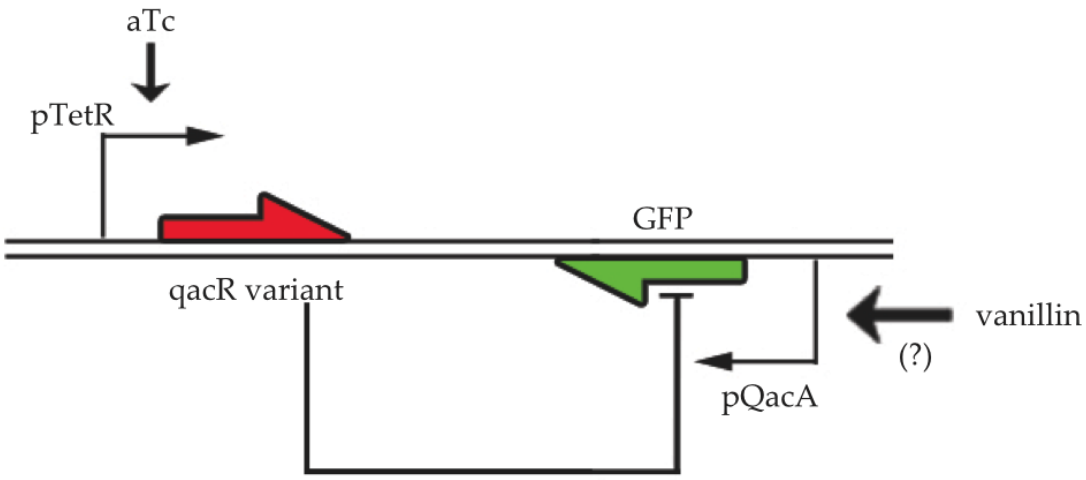



Figure 5: **Circuit layout for *in vivo* tests.** Genes encoding GFP under the control of the native qac promoter, and our QacR designs under the control of a tet-inducible promoter were placed in a single plasmid and transformed into DH5$\alpha$Z1 cells. qacR levels were controlled using aTc for varying vanillin concentrations. Candidate designs that are responsive to vanillin should show an increase in fluorescence with increasing vanillin concentrations

Plasmids that contained genes encoding the wild-type qacR sequence, qacR 2-1 and, qacR 5-1 were placed under the control of an inducible promoter, $P_{tet}$ and GFP downstream of $P_{QacA}$ were cloned into DH5$\alpha$Z1 cells (Figure 5). For each of the qacR variants, we compared differences in fluoresent signal across increasing vanillin concentrations. This was done for different inducer concentrations by varying the concentration of anhydrous tetracycline (aTc). Cells that were grown in higher aTc concentrations had a lower measured optical density (OD), indicating a slower doubling time. We hypothesize that this is due to the toxicity of the qacR repressor to the *E. coli* strain. Since qacR is not a native protein, it is possible that qacR is binding to locations in the *E. coli* genome. Interestingly, the differences in optical density measurements become less pronounced with

increasing vanillin concentration, suggesting that vanillin may provide a mitigating effect to this toxicity. In order to account for differences in OD, fluoresent measurements were normalized to OD. Figure 6, shows the fold change in normalized fluorescence for increasing vanillin concentrations. The lowest vanillin concentration,$250\mu M$, was used as a baseline for each aTc concentration. We observed an approximately 2.5 fold increase in fluorescence over the baseline for qacR 2-1 tth an aTc concetration of 6.25 ng/uL. About a two-fold increase in fluorescence was observed for qacR 5-1 at an aTc concentration of 10 ng/$\mu$L. A much lower increase in fluorescence was observed in cells that contained the wild-type protein.

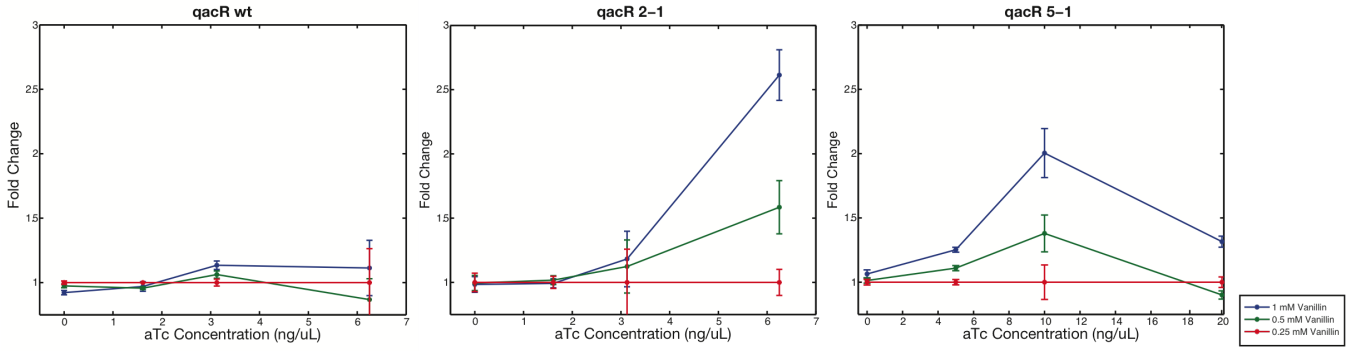



Figure 6: **qacR mutants show a vanillin-dependent increase in GFP fluorescence.** QacR variants were tested in vivo at different aTc and vanillin concentrations. The mutants show between a 2 and 2.5 fold change in normalized GFP expression when compared to a baseline vanillin concentration. The wild-type qacR protein does not display this behavior.

Figure 7 shows the normalized GFP fluorscence as a function of vanillin concentrations for each of the cultures at the aTc concentration where the highest fold-change was observed. Cells that contained the wild-type qacR were dimmer than cells that contained the qacR variants, suggesting that the wild-type qacR protein is more functional as a DNA repressor. We hypothesize that the mutations in the qacR variants also affect their ability to bind to DNA. This is supported by the fact that higher aTc concentrations are needed to see a decrease in fluoresence and repression of GFP expression. Since the DNA binding interaction is weakened, the qacR variants are more sensitive to vanillin resulting in the increase in dynamic range observed. We believe that these variants can serve as a starting point for directed evolution towards a vanillin sensor that meets the required specifications for use in a synthetic circuit.

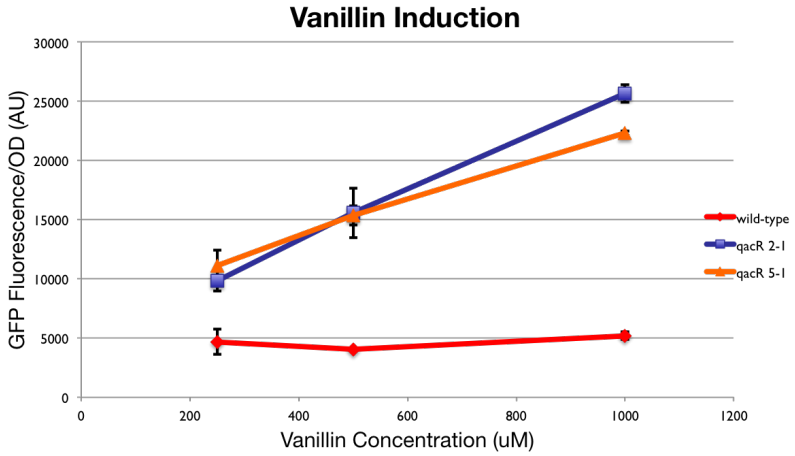



Figure 7: **Vanillin Dosage Response Curves.** GFP normalized to optical density plotted for the aTc concentration where the maximum fold change was observed. This is at 6.25 ng/$\mu$L for qacR 2-1 and 10 ng/$\mu$L for qacR 5-1.

### Future Work

More characterization is necessary to ensure the ability of the qacR variants to sense vanillin. Since the cells exhibit slower growth under certain vanillin and aTc concentrations, it is possible that observed differences in fluorescence are due to differences in protein expression because of the stage of growth of the cell (i.e. early-log vs. late-log) instead of due to a vanillin dependent response. Further confounding the results is the fact that the growth defect of the cells is mitigated at higher vanillin concentrations though vanillin has also been shown to be toxic to the cells. In order to separate the effects of protein toxicity and vanillin toxicity further *in vitro* characterization is required to demonstrate improved vanillin sensitivity of the qacR mutants.

## Materials and Methods

### Cell strain, media

The circuit was implemented in the *E.coli* cell strain DH5$\alpha$Z1, a variant of DH5a which contains a chromosomal integration of the Z1 cassette[16]. All cell culture was done in optically clear M9CA minimal media (Teknova M8010).

Anhydrotetracycline (aTc) was diluted in media at concentrations of 0, 1.6125, 3.125, 5, 6.25,

10, 20 ng/$\mu$L. Vanillin was diluted in the media at 1, 0.5, and 0.25 mM.

## Genes and Plasmids

DNA encoding the qacR genes was constructed using overlap extension PCR. Plasmids used contained chloramphenicol resistance with a p15a origin of replication.

## Plate reader experiments

Plate reader data were collected on a Biotek H1MF machine. Cell were grown in at least two consecutive overnight cultures in M9CA minimal media. On the day of the experiment, overnight cultures were diluted 1:100 and grown for 5 hours to ensure that the cells were in log phase. Cells were then diluted 1:100 into fresh media at the specified experimental condition. Cells were grown in these conditions at 37C for $10-12$ hours in M2P Lab Flower Plates while shaking at 1100 rpm. Measurements were taken after $10-12$ hours by transferring the cells to clear bottomed 96-well microplates (PerkinElmer, ViewPlate, 6005182) . GFP was read at 488/525 with gain 100.

Analysis of the data was done by taking fluorescence readings for each independent well. Experimental conditions were done in triplicate and repeats were averaged. Error bars shown are standard error of the mean.

## References


(1) Purnick, P. E. M.; Weiss, R. *Nature reviews. Molecular cell biology* **2009**, *10*, 410–422.

(2) Ramos, J. L.; Martinez-Bueno, M.; Molina-Henares, A. J.; Terán, W.; Watanabe, K.; Zhang, X. D.; Gallegos, M. T.; Brennan, R.; Tobes, R. *Microbiology and Molecular Biology Reviews* **2005**, *69*, 326–+.

(3) Elowitz, M. B.; Leibler, S. *Nature* **2000**, *403*, 335–338.

(4) Gardner, T. S.; Cantor, C. R.; Collins, J. J. *Nature* **2000**, *403*, 339–342.

(5) Grkovic, S.; Hardie, K. M.; Brown, M. H.; Skurray, R. A. *Biochemistry* **2003**, *42*, 15226–15236.

(6) Schumacher, M. A.; Miller, M. C.; Grkovic, S.; Brown, M. H.; Skurray, R. A.; Brennan, R. G. *Science* **2001**, *294*, 2158–2163.

(7) Schumacher, M. A.; Brennan, R. G. *Research in Microbiology* **2003**, *154*, 69–77.

(8) Peters, K. M.; Brooks, B. E.; Schumacher, M. A.; Skurray, R. A.; Brennan, R. G.; Brown, M. H. *PLoS ONE* **2011**, *6*, e15974.

(9) Klinke, H. B.; Thomsen, AB,; Ahring, B. K. *Applied Microbiology and Biotechnology* **2004**, *66*, 10–26.

(10) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. *International Journal of Quantum Chemistry* **2013**, *113*, 2110–2142.

(11) Lassila, J. K.; Privett, H. K.; Allen, B. D.; Mayo, S. L. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 16710–16715.

(12) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364–1368.

(13) Allen, B. D.; Mayo, S. L. *Journal of Computational Chemistry* **2006**, *27*, 1071–1075.

(14) Allen, B. D.; Mayo, S. L. *Journal of Computational Chemistry* **2010**, *31*, 904–916.

(15) Sun, Z. Z.; Hayes, C. A.; Shin, J.; Caschera, F.; Murray, R. M.; Noireaux, V. *Journal of visualized experiments : JoVE* **2013**, e50762.

(16) Lutz, R.; Bujard, H. *Nucleic Acids Research* **1997**, *25*, 1203–1210.