Understanding and designing sensors in *E. coli*

by

Elizabeth Jane Clarke

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION

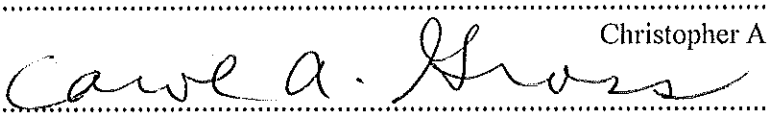
of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

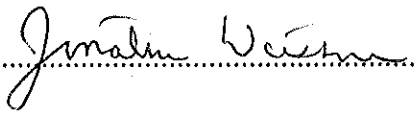
.....

Christopher A. Voigt



.....

Carol A. Gross



.....

Jonathan S. Weissman

Committee in Charge

Deposited in the Library, University of California, San Francisco

.....

Date University Librarian

Degree Conferred:.....

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

The text of the thesis 'Understanding and designing sensors in *E. coli*' is made up of both published and unpublished works. Chapter 3 is a revised reprint of part of the material as it appears in the Journal of Molecular Biology ("Kinetic Buffering of Cross Talk between Bacterial Two-Component Sensors", Volume 390, Issue 3, 17 July 2009) The coauthors, Eli Groban and Christopher Voigt, directed and supervised the research. Parts of Chapters 1, 2 and 4 are in submission at the Journal of Molecular Biology. The coauthor, Christopher Voigt, directed and supervised the research.

Personal Acknowledgements:

First and foremost, I thank my thesis advisor, Chris Voigt, without whom this work would not have been possible. I learned from him what it means to convey science in a way that is compelling and accessible. I am constantly amazed at his ability to take something that looks pretty bad to me (usually my data), and pull something positive out of it. It has been an awesome experience getting to see the lab evolve from (almost) day one, and I am so impressed by the caliber of people that Chris has been able to recruit. I look forward to seeing great things from the lab in the future!

The rest of my thesis committee, Jonathan Weissman and Carol Gross gave great support and showed insight into my project from day one. Carol, in particular, took me under her

wing long before my orals exam, encouraging me and giving great feedback when she welcomed me into her group meetings, which I attended for most of my time at UCSF.

My lab has been a fabulous place to work. Every single person who has come through the group has brought something positive to me. I have met people here that I will be friends with for a very long time.

Other people who have helped me scientifically or have meant a lot to me otherwise:

Chris Anderson

Vlad Denic

Erica Dishnica

Alisa Givental

Kris Kuchenbecker

My parents have made me who I am today. They have supported me in whatever I thought I wanted to do (artist, author, doctor, scientist) and told me that I was beautiful and brilliant. I am pretty lucky =)

Abstract

Understanding and designing sensors in *E. coli*

Elizabeth Jane Clarke

Two-component systems are a highly conserved signal transduction pathway that enable bacteria to sense changes in their environment and adjust gene expression to adapt to nutrients, stresses, and other signals. The body of this work seeks to determine the extent to which E. coli uses these sensors as a network to process their environment. This is broken into two parts: (1) Whether cross talk can occur at the phosphorelay level and (2) whether the sensors are able to function as a combinatorial sensor. A combinatorial sensor is made up of a set of sensors, each of which is activated to different degrees by many inputs such that the pattern of their activation defines the signal. Using promoter reporters and flow cytometry, we measured the response of three two-component osmosensors in *E. coli* (envZ/ompR, cpxA/cpxR, and rcsC/rcsD/rcsB) to 38 chemicals including known inducers of the systems, membrane perturbing agents, alcohols and chemicals of industrial relevance. We found that each system responded to a wide spectrum of conditions and that the three systems are uncorrelated, meaning that unique patterns of gene expression are generated by even closely related chemical compounds. Of the eight possible patterns generated by a three sensor system, we observe five. This data show that bacteria are able to use a limited set of sensory components to identify a diverse set of compounds and environmental conditions.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1. Preamble.

My work in the Voigt lab began with Chris Anderson, where we made tumor seeking bacteria ("Environmentally controlled invasion of cancer cells by engineered bacteria" JMB, 2006). Although that work is not presented in this thesis, it was an awesome project, and affected how I thought about bacterial sensing for the remainder of my time here.

The motivation behind that work was to come up with a way to get *E. coli* to invade cancer cells as a function of a certain condition (oxygen content, cell density, or small molecule). We were able to do this by transferring an invasion module from another bacterium into *E. coli*. At the end of the day, a big part of the problem will become how to specify the microenvironment of the cancer to the bacterium (so that it doesn't invade and kill healthy cells). Bacteria have evolved to be very good at this type of problem. They are able to survive in changing environments, rapidly and accurately. I started thinking about whether you could take this natural response, and use it to make new types of biosensors.

The next project I worked on was constructing a bacterial thermometer (appendix A). Though this work was never published, the idea was to look for *E. coli's* natural transcriptional response to temperature (from published microarray data) to find candidate temperature reporters. I was also able to do dome simple circuit engineering to make a better heat sensor.

Finally, the core of my thesis work revolved around drawing an analogy between the mammalian olfaction system and a set of sensors in *E. coli*.

1.2. Two component Systems.

E. coli can thrive in environments that differ in temperature, pH, osmolarity, oxidative stress, and antimicrobial agents [1, 2]. To survive, the cell must sense these stresses and respond by affecting necessary changes in gene expression for survival. One mechanism for linking environmental sensing to transcriptional regulation is two-component systems[3, 4]. Bacteria have many two-component systems (TCS; *E.coli* has ~32) that relay extracellular information to the cytoplasm where transcriptional responses are executed. Each system consists of a membrane bound sensor and a cognate cytoplasmic response regulator. Upon stimulation of the sensor, it autophosphorlylates and transfers the phosphate to its response regulator. The response regulator is often a transcription factor, binding to promoters where it activates or represses gene expression. In some cases, the sensor is very specific, responding to a single stimulus and regulating only a limited number of genes, appropriate to that stress. For example, *E. coli's* CusSR TCS is activated by copper, which at high levels is toxic to the cell, and CusR activates the genes encoding copper efflux pumps[5]. Other two component systems are stimulated

by multiple conditions, and regulate many genes of diverse function[6-8]. The most general sensors are often assigned a function of sensing osmolarity[9-11].

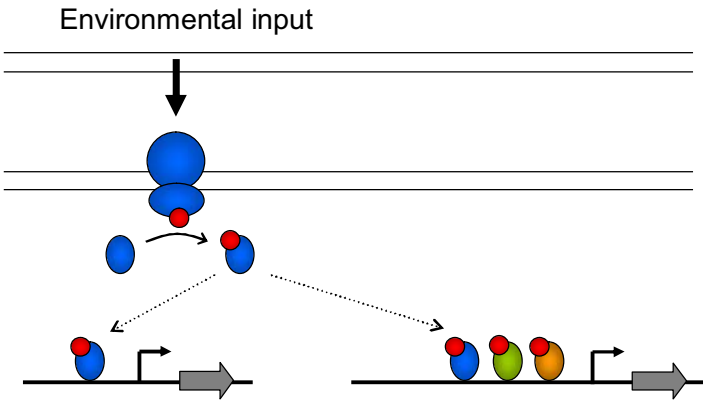


Figure 1.1: A two-component system

Two component systems function to communicate information from the outside to the inside of the cell in order to adapt to changes in conditions. The first component is in the in inner membrane. It auto-phosphorylates upon stimulation and transfers its phosphate to its partner response regulator protein. In this activated form, the response regulator binds to promoters, to up or down-regulate gene expression.

1.3. EnvZ/OmpR.

EnvZ/ompR is arguably the most well-characterized TCS, and is commonly labeled an “osmosensor”. EnvZ is activated by external stimuli, autophosphorylates, and transfers its phosphate to OmpR. Classically, EnvZ/ompR has been shown to alter the expression of outer membrane porins ompC and ompF in response to changes in osmotic pressure [12]. However, there are other signals, such cationic peptides, which also activate this system (Figure 1A). Stimulation of EnvZ can induce the membrane stress

response through over-expression of outer membrane porins which in turn leads to the de-repression of the alternate sigma factor, σ^E [13].

1.4. CpxA/CpxR.

CpxA/cpxR is the system closest in sequence to EnvZ/ompR[14]. In contrast to EnvZ, the primary role of CpxA is to de-phosphorylate CpxR[15]. It is subject to feedback amplification and repression, via up-regulation of its own expression, as well as up-regulation of its negative regulator, a periplamic putative chaperone protein, cpxP[16-18]. CpxAR responds to cell envelope stress, and activates genes to combat the stress, such as periplasmic protein folding and degrading factors. Multiple stresses feed into cpx activation through the periplasmic proteins NlpE and CpxP. Certain activators require transmission via the lipoprotein NlpE[19], whereas others require CpxP but not NlpE[16, 20, 21]. Interestingly, a few activators can stimulate CpxR in the absence of CpxA, through phosphorylation by the small molecule acetyl phosphate[15].

1.5. RcsC/RcsC/RcsB.

The RcsC/RcsC/RcsB signaling pathway consists of three proteins. The membrane-bound RcsC autophosphorylates and transfers the phosphate to the membrane-bound RcsD. RcsD then phosphorylates the response regulator RcsB. The specific signal activating is unknown, although most activating conditions, such as contact with surfaces and overexpression of periplasmic proteins[4] are likely to lead to envelope stress. Like CpxA, RcsC requires a lipoprotein, RcsF, to activate downstream signaling in response to certain signals[22, 23]. Some promoters are regulated by the RcsB homodimer, while

others require the expression of a second protein RcsA, which forms a dimer with a single RcsB[24, 25].

1.5. Combinatorial sensing.

We postulate that two-component sensors that respond to many inputs could participate in a combinatorial sensor, analogous to olfaction in higher organisms. A combinatorial sensor is made up of a set of receptors, each of which is activated by many inputs such that and the pattern of their activation defines the signal. The mammalian olfactory system is a canonical example of a combinatorial sensor. For example, the human nose is a remarkably flexible biosensor. It has ~400 different chemoreceptors, yet it can recognize ~10,000 different odorants[26, 27]. This is accomplished via a combinatorial recognition of odorants; the nose does not have a single receptor for every odorant, rather, there is a set of receptors that binds non-specifically to many chemicals with different affinities. A chemical produces a signature pattern of receptor activities, which is processed by the brain to identify the odor. This illustrates two notable advantages of such an organizing principle. First, a small number of receptors can respond to a large number of chemicals. Second, the system can be trained to respond to new inputs without the addition of new receptors through rewiring of the system's downstream components. Combinatorial sensing has been applied in industry, where electronic noses based on materials that bind to many chemicals non-specifically have been incorporated into analytical and control systems[28, 29].

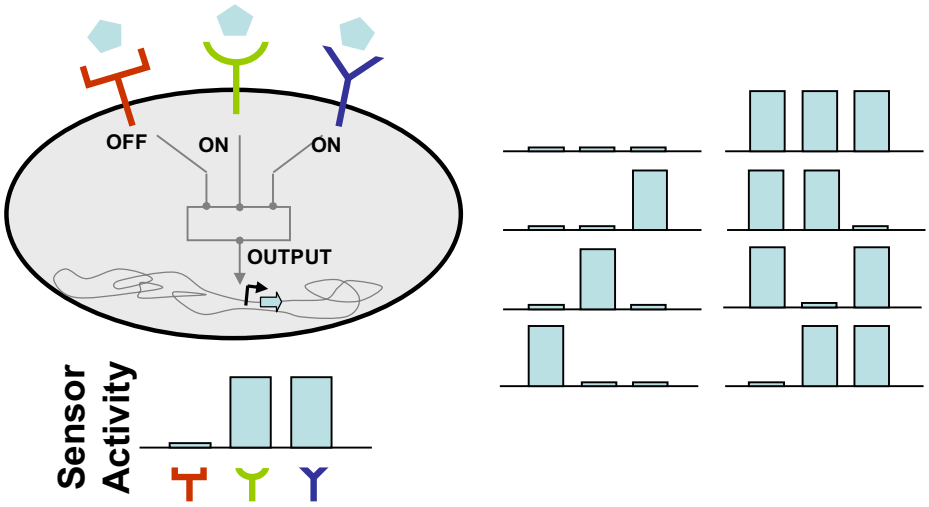


Figure 1.2: Combinatorial sensing

With three general binary sensors, a total of eight possible patterns can be generated. With intermediate states allowed, many more patterns are possible. The patterns are theoretically distinguished by signal processing in the cell, for example by complex promoters.

Combinatorial sensing with two-component systems could be an efficient strategy bacteria have evolved to sense multiple inputs with a limited set of sensory components, conceptually analogous to the mammalian olfactory system. Here, multiple two component systems would respond to different inputs, and the downstream regulatory network would interpret the incoming pattern. For combinatorial sensing to take place, the sensors should (i) respond non-specifically to multiple environmental signals; (ii) be un-coupled, such that different patterns are generated in response to the inputs; and (iii) there should also be a means for pattern recognition, such as downstream promoters with multiple response regulator binding sites.

Multiple two-component systems could interact with each other as part of a higher order network[30-36]. Interactions could be at the level of cross talk at the phosphorelay level, as integration of response regulator binding to promoters, or as gene product based feedback from one system into another[37, 38]. Although cross talk at the phosphorelay step is rare *in vivo*[39, 40], downstream interactions between two component systems are known to occur. For example, in *S. enterica*, a gene product controlled by the PhoPQ TCS protects PmrA in the PmrAB TCS from dephosphorylation[41, 42]. Parallel two component systems have also been shown to regulate common promoters, for example, CpxR and OmpR overlap binding sites at the *csgD* promoter, which is also regulated by the two component systems RcsAB, RstA, and its own gene product[10].

Previous work demonstrating that envZ/ompR, cpxA/cpxR, and rcsC/rcsD/rcsB are activated by many conditions primarily measured one system at a time, with different metrics, growth conditions, strains, and induction times. As a result, it is impossible to compare relative induction either across conditions within a single sensor, or between sensors. To address this, we optimized reporter system and assay conditions and used standardized measurements to compare induction by known inducers, alcohols and chemicals of industrial relevance both within and between sensors. This induction data allowed us to determine the degree to which each sensor is general, measure the correlation between each sensor and osmolarity, and importantly, between pairs of sensors. By visualizing the patterns, we could determine how the chemicals clustered in terms of their impact on the sensors, as well as the diversity of patterns generated

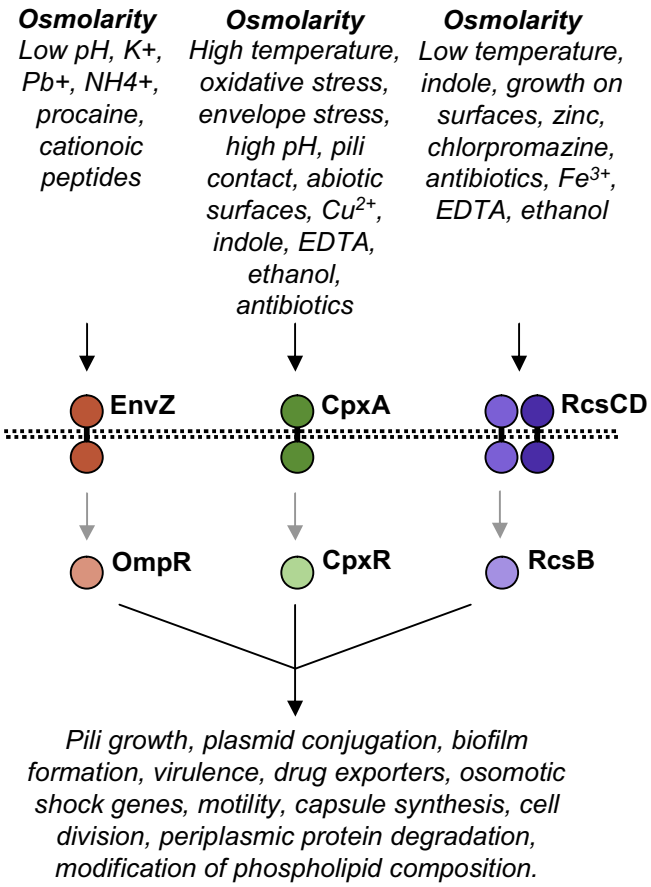


Figure 1.3: Overlapping inputs and overlapping regulons

Chapter 2

Standardizing two-component reporters and assay conditions

2.1. Preamble

This section of my thesis gets its own chapter. Though there isn't much data here, perfecting the assay conditions, and picking the reporters took up such a big part of my graduate career that I wanted it to stand on its own. Hopefully I will convey here the amount of toiling that went on to get it just right.

2.2. Selecting promoter reporters

To assay sensor activity, we measured the activity of promoters regulated by each of the three systems. Each reporter contains a regulated promoter driving green fluorescent protein (GFP) on a medium copy plasmid (Figure 1B). We tested a number of promoters activated by the response regulator for each system, and picked the reporter that had the largest dynamic range: envZ (P_{ompC}), cpx (P_{cpxP}), and rcs (P_{rprA}). For rcs, we used a reporter activated by the RcsB homodimer (rather than RcsB-RcsA heterodimer)

as this eliminates the possibility of a spurious signal from the histidine kinase independent RcsA. A full list of the promoters tested is provided below.

Table 2.1: Promoter reporters tested, and genomic locatios

EnvZ		Cpx		Rcs	
Gene	Location	Gene	Location	Gene	Location
ompC - p1	2310952...2310802	cpxP	4103753...4103858	rprA	1768254...1768396
ompF	986725...956205	cpxR	4103893...4103693	bdm	1554649...1554341
tppB	1710593...1710793	degP	180609...180884	osmC - p1	1554538...1554655
		acrD	2585417...2585617	rcsA	2021792...2022016
		ppiA-p1	3489862...3489747	spy	1823978...1823649
				wza	2135787...2135270

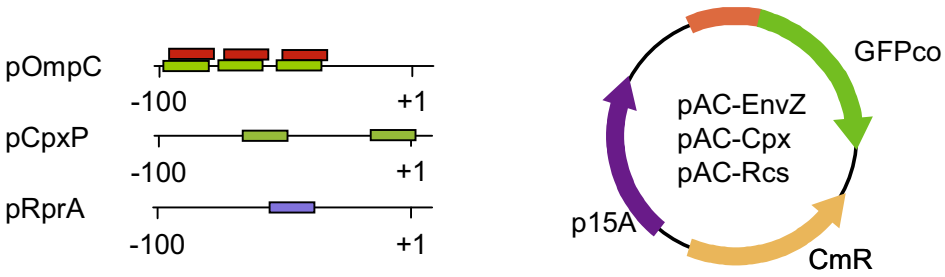


Figure 2.1: Promoter reporters and footprinting data

Known regulator binding sites are shown above. OmpR in red, CpxR in green, and RcsB in purrple. On the right is the plasmid map of the reporter plasmid.

2.3. Assay conditions

To measure promoter induction, cells containing a single reporter plasmid were grown up in minimal media, and split into either a control flask, or a flask containing inducing chemical. Each chemical was supplied at the concentration which led to a 50% growth rate defect as measured at the end of the 2 hours. Samples were collected for two hours after induction, and GFP content for un-induced and induced cells was measured using flow cytometry (Figure 1C). We noticed two things that directly affected how we

performed our induction assay. One, all of the reporters, especially those for envZ, became more active upon entry into stationary phase. And two, differences in temperature, growth rate, and media composition had a large impact on induction. This is consistent with the documented observations that certain chemicals induce some two component systems only under specific growth conditions such as low temperature or a specific carbon source[43]. Because of this, we measured induction only over logarithmic growth, kept temperature constant by growing in water bath shakers, and grew cells in minimal media, whose composition and pH we could carefully control. Under these conditions we found envZ/ompR to be less strongly induced than in a rich media, such as LB, an observation also made by Goulain and co-workers[44].

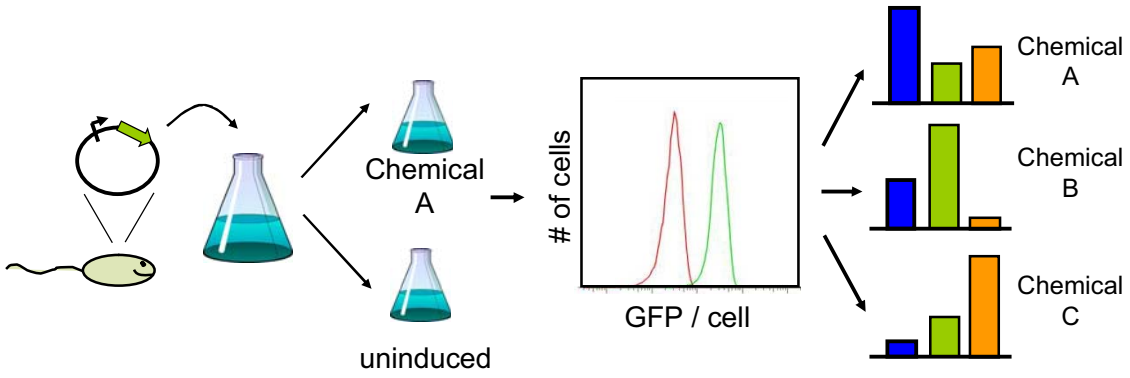


Figure 2.2: Assay

A culture was inoculated from a plate and grown for 14 hours, and then diluted 100x into fresh media. At OD_{600} = 0.4 cells were again diluted 100x into 150 mls. These cells were grown at 37°C in a shaking water bath at 160 rpm. When cells reached OD_{600} = 0.2, the culture was split into 6 flasks, one un-induced, and the rest containing various induction chemicals. Samples was taken at t = 0 min, 15min, 30min, 60min, 90min and 120min. Protein expression was immediately stopped in each sample with 2 mg/mL kanamycin, and the samples were stored on ice until flow cytometry was performed.

2.3 Reporter controls

We show an inducer dependent activation of each reporter only in cells that contain its response regulator (Figure 1D). Panel 1 of Figure 1D shows the induction of pAC-EnvZ in the presence (WT) and absence (ΔompR) of the response regulator ompR. Induction by indole occurs only in the WT strain, thus we conclude that the observed induction is a result of changes in levels of phosphorylated ompR. We similarly tested pAC-Cpx and pAC-Rcs in their response regulator knock out strains as shown in Panel 2 and Panel 3, and conclude that they are appropriate reporters. Known regulator interactions at each promoter are shown in Figure 1B. Interestingly, a combination of genetics and footprinting experiments supports cpxR activation of P_{ompC}[44]. Though we show that indole induces both envZ and cpx (Figure 3A), indole does not activate P_{ompC} in the presence of cpxR and absence of ompR. From this, we conclude that cpxR is not contributing significantly to P_{ompC} under our conditions, which is consistent with other observations[45].

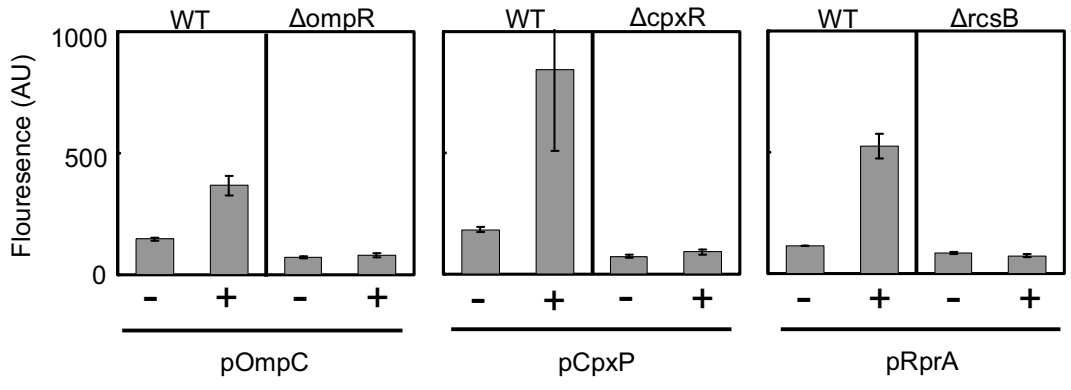


Figure 2.3: Reporter controls

For each system, induction occurs in the presence of the response regulator, but not in the absence. The controls were performed with a strong inducer for each: envZ/ompR: indole(3mM),

cpxA/cpxR: phenethyl alcohol(0.2%), and rcsC/rcsD/rcsB: NaCl (550mM). Error bars are the standard deviation from the mean of two runs performed on different days.

Chapter 3

Cross talk between two-component systems

3.1. Preamble

The data in this chapter is a small piece of a larger body of work spearheaded by Eli Gorban as a graduate student in the Voigt lab. In the paper (Kinetic Buffering of Cross Talk between Bacterial Two-Component Sensors, Journal of Molecular Biology Volume 390, Issue 3, 17 July 2009), we show a combination of *in vitro*, *in vivo*, and computational experiments to determine whether cross-talk between two two-component systems (EnvZ/ompR and CpxA/cpxR) exists, and if not, what buffering mechanism the cell uses to prevent it. This is particularly interesting in the context of my personal work, which demonstrates the promiscuity of the same set of sensors for their inputs. This leads to interesting speculation about the point at which signal integration may be occurring in the cell, if it is not happening at the level of phosphorelay.

I am also fascinated by the fact that with the artificial TAZ system used here (TAZ is the extracellular domain of tar fused to the cytoplamic domain of EnvZ) Eli sees much larger changes in phosphorylated response regulator than I ever come close to when making measurements with the wild type system. Imagine for a second that the

fraction of phosphorylated response regulator, going from 0 to 1, is drawn on a ruler going from 0 inches to 12 inches. Eli's system can go from 0 to 12 inches. MY data shows that the wild type cell, responding to normal stimuli, only ever experiences about an inch-worth of phophorylation states, if that. Remarkable!

3.2. Introduction

The *in vitro* and computation data predict that cross talk should only occur in a triple mutant: Δ*ompR* Δ*cpxA* Δ*actA-pta*. To test this prediction, all seven combinations of knockouts are made and only this particular strain demonstrates crosstalk, where the *cpxP* promoter is induced 280-fold upon the activation of EnvZ. Further, the behavior of the other knockouts agrees with the model predictions. This model points to a kinetic model of buffering where both the cognate phosphatase activity and competition between regulator proteins for phosphate prevents cross talk *in vivo*.

Because of the number of simultaneously expressed systems and the strong conservation in the sequence and structure, it seems plausible that cross reactions would occur frequently between systems[38]. This could either occur as a single kinase phosphorylating multiple response regulators, or conversely, a single regulator could be phosphorylated by multiple kinases. Cross reactions could even be exploited to create a neural network linking inputs and outputs such that a higher level of sensing power is achieved. In contrast, all of the two-component systems could behave linearly with one input being linked to one output. Given the similarity between systems, there would have to be some sort of buffering mechanism for linearity to be preserved.[38, 39, 46]

Previous *in vitro* studies demonstrate some promiscuity among histidine kinases and response regulator proteins. Ishihama and co-workers assayed 25 histidine kinases

from *E. coli* against the 34 *E. coli* response regulators.[47] They showed that after a 30 second incubation time, 11 of the 34 response regulators can be phosphorylated by more than one histidine kinase. There are only a few promiscuous kinases, however, so out of 692 possible cross talk pairs, only 3.0% of them showed *in vitro* cross talk. Skerker and Laub assayed the EnvZ, CpxA, and CheA kinases against a panel of all response regulators from *E. coli*.[48] Non-cognate transfer occurs between EnvZ and CxpR, but this is observed at a 60 minute timepoint. This is very slow compared to the cognate transfer, which occurs in 10 seconds. They suggest that the non-cognate transfer is too slow to be relevant *in vivo*.

Several theories have been proposed as to how two-component systems could buffer *in vivo* the slow crosstalk observed *in vitro*. Savageau proposed that buffering could emerge from the ability for the histidine kinase to both phosphorylate and dephosphorylate the response regulator (a bifunctional interaction).[46] Using a mathematical model, he demonstrated that the phosphotase function could decrease the background phosphorylation of the response regulator, thus reducing spurious phosphorylation.

Laub and co-workers argue that each kinase has a "kinetic preference" for its cognate substrate.[48] They postulate that subtle amino acid differences in the binding interface between the kinase and the response regulator affect the K_m. If the correct response regulator interacts for a longer time with its kinase, this both prevents access to the kinase by the incorrect substrate and drains the kinase of all available phosphate, resulting in phosphorylation of only the correct response regulator. To support this

hypothesis, they made small amino acid changes to the kinase at the binding interface, which resulted in a shift in kinetic preference that altered specificity.[49]

Here we characterize the interactions between the *E. coli* EnvZ/OmpR and CpxA/CpxR two-component systems. The phosphotransfer domains share the most similarity amongst *E. coli* two-component systems, sharing 31% amino acid identity for the kinase domain of the sensor histidine kinase and 50% identity between the receiver domain of the response regulators.[14] It has been shown previously that EnvZ can phosphorylate CpxR *in vitro*, albeit at a much slower rate than OmpR.[47, 48]

We perform a comprehensive kinetic study of the interactions between the EnvZ/OmpR and CpxA/CpxR two-component systems. First, we purified the kinases EnvZ and CpxA and their response regulators OmpR and CpxR. We phosphorylated either the kinase or the response regulator using radiolabeled phosphate and monitored phosphorylation kinetics *in vitro* (Figure 1). Moreover, we also synthesized radiolabeled ^{32}P acetyl phosphate and used this to phosphorylate the OmpR and CpxR proteins in the absence of kinase to determine autophosphoryation rates. Fully phosphorylated response regulator proteins were then exposed to unlabeled kinase to measure phosphatase activity. After obtaining rate constants for ten reactions, we parameterized a mathematical model and used this model to predict combinations of gene knock outs to make that could induce cross talk *in vivo*. To test the prediction, we constructed seven knock out strains and used promoter-GFP fusions to measure cross talk *in vivo*. Significant cross talk is only observed for a triple knock out suggested by the model.

3.3. Results

We monitor the *in vivo* phosphorylation state of OmpR and CpxR by following the activity of the *ompC* and *cpxP* promoters, respectively (Figures 1A and 1B). We selectively activate the EnvZ/OmpR pathway in order to determine whether *in vivo* cross talk occurs between it and CpxA/CpxR. As there are no known ligands for the EnvZ kinase, we used a second generation version of the hybrid kinase designed by Utsumi and Inoyue, which combines the transmembrane aspartate receptor (TAR) with the kinase domain of the EnvZ kinase (TAZ) (Figures 1C and 1D).[50] This hybrid kinase serves to selectively activate the EnvZ kinase domain, and phosphorylate its cognate response regulator OmpR in the presence of the small molecule aspartate. We activate the EnvZ kinase using the TAZ construct and monitor the output levels of both the EnvZ/OmpR and CpxA/CpxR system using the promoters *ompC* and *cpxP* to drive green fluorescent protein (GFP) expression, which is measured at the single-cell level using flow cytometry (Figures 1D and 1G).

TAZ phosphorylates the OmpR response regulator in the presence, but not the absence, of 5 mM aspartate *in vivo* in the wild type system (Figure 4E and Figure 2, first panel). Induction of TAZ via addition of 5 mM aspartate leads to a 34.7 (+/- 1.9) fold change in fluorescence at the *ompC* promoter showing that the presence of aspartate and TAZ provides an active EnvZ kinase. This activation is TAZ dependent, as a control strain lacking TAZ shows no activity at the *ompC* promoter (Figure 1F). Δ*ompR* does not respond to TAZ showing that the *ompC* promoter is specific to OmpR~P. Our use of GFP as a reporter for system activity, combined with flow cytometry, allowed us to

conduct single cell measurements. From this we see that cells containing the *ompC* promoter GFP fusions produce a single distribution in all cases (Figure 1E).

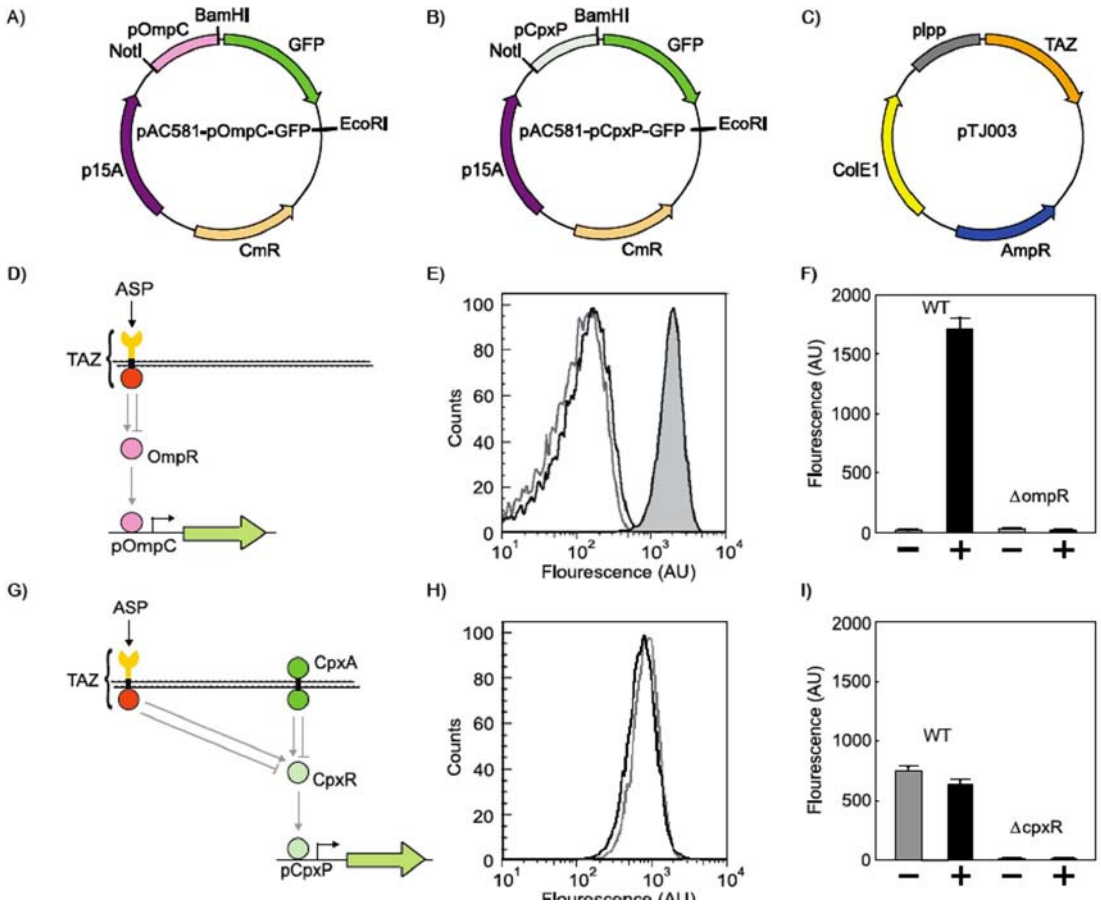


Figure 3.1: Experimental set-up

(A) pAC581-pOmpC-GFP and (B) pAC581-pCpxP-GFP, which contain the *ompC* and *cpxP* promoters driving GFP expression. (C) The reporters are co-transformed with a second plasmid pTJ003, which contains the TAZ protein under control of the constitutive lpp promoter. (D) TAZ phosphorylates OmpR in the presence of 5 mM asparate. (E) Gated cytometry data is shown for cells without TAZ (gray line) and with TAZ but without asparate (black line). Addition of 5 mM aspartate induces the system (black line, right). (F) The *ompC* reporter is not induced in the absence of TAZ (-), but is strongly induced when both TAZ and aspartate is present (+). When ompR is knocked out, TAZ has no affect on the system. (G) The ability of TAZ to induce CpxR was also measured in wild-type cells. (H) The cytometry distribution shows no difference between

cells with TAZ and aspartate (black line) and cells without TAZ (gray line). There is a basal level of activity from the *cpxP* promoter. (I) The activity of the *cpxP* promoter is the same in the absence (-) and presence (+) of TAZ and inducer. The basal activity is eliminated by knocking out CpxR. For both parts F and I, the mean of three experiments performed on different days is shown and the error bars are one standard deviation.

In the wild type system, active EnvZ does not affect the levels of CpxR~P response regulator *in vivo* (Figure 1G-I). By monitoring the abundance of phosphorylated CpxR via the *cpxP* promoter and activating TAZ, we are able to measure the presence or absence of cross talk between EnvZ and CpxR. We are unable to see a significant difference in *cpxP* activity in the presence of active EnvZ.

We then constructed seven knockout mutants containing all possible combinations of Δ*cpxA*, Δ*ompR*, and Δ*ackA-pta* (Methods). In addition, EnvZ is also knocked out and TAZ is introduced on a plasmid. For each of the knockouts, the activity of the promoters is compared for the Δ*envZ* strain (-) and the addition of the TAZ and 5 mM aspartate (+) (Figure 2).

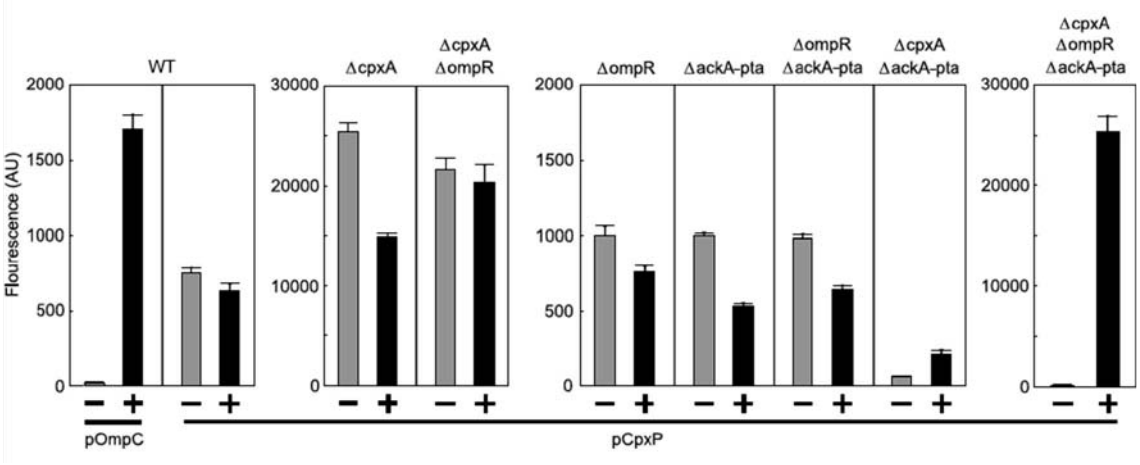


Figure 3.2: Response of seven knock-out strains

The knockouts are shown at the top of each panel and the promoter used as a reporter is shown at the bottom. For each knockout, the fluorescence in the absence of EnvZ (-) and presence of active EnvZ (TAZ and aspartate) (+) are shown. Panel 1 shows selective activation of the *ompC* promoter in the presence of TAZ and aspartate (black bars) but not in the absence of TAZ (gray bars). Also, TAZ is unable to activation the *cpxP* promoter. Different combinations of knock outs do not produce cross talk between EnvZ and CpxR, as shown from the inability of this kinase to activate the *cpxP* promoter. Only a triple knockout, Δ*ackA-pta* Δ*cpxA* Δ*ompR*, displays cross talk. Bars represent the mean of three different experiments and the error bars are one standard deviation from the mean.

The knockout Δ*cpxA* causes the level of CpxR~P to increase (Figure 2). In the absence of the CpxA kinase, CpxR autophosphorylates using acetyl phosphate as a substrate and reaches a higher steady state. Notably, the experiments recover the decrease in activity from the addition of active EnvZ as observed in the model (Figure 2). The single knockout Δ*ompR* behaves similarly to WT. The double knockout Δ*cpxA* Δ*ompR* strongly activates the *cpxP* promoter and is insensitive to active EnvZ.

The affect of disrupting the acetyl phosphate synthesis pathway was then explored. The Δ*ackA-pta* and Δ*ackA-pta* Δ*ompR* strains display similar behavior to WT. In contrast, the Δ*ackA-pta* Δ*cpxA* strain strongly decreases the amount of CpxR~P and the activity of the *cpxP* reporter. However, there is a small amount of crosstalk observed, where active EnvZ is able to phosphorylate CpxR and induce the *cpxP* promoter 3.6 (+/- 0.2) fold.

A Δ*ackA-pta* Δ*cpxA* Δ*ompR* triple knock out shows EnvZ dependent phosphorylation of CpxR *in vivo* (Figure 5, panel 9). Inducing TAZ in this system causes a 370.4 (+/- 79.8) fold increase in *cpxP* activity. This system lacks two crucial

mechanisms to prevent cross talk. First, it lacks the phosphatase activity of CpxA. Second, it does not have the OmpR protein, allowing CpxR access to the active EnvZ kinase. These two buffering mechanisms are quite effective. Therefore, cross talk is highly unlikely to be relevant *in vivo*. In the wild type system, CpxA serves as a phosphatase and OmpR quickly removes phosphate from active EnvZ. It is only by removing all possible mechanisms for insulation that we see a potent activation of CpxR by EnvZ *in vivo*.

3.4. Methods

Assay and Cytometry:

Cells were transformed with either pAC581-PompC-GFP, pAC581-PcpxP-GFP, or either plasmid and pTJ003 and grown for 12 – 16 hours at 37^{o} C on selective LB/Agar. Control cells containing no plasmid, used to normalize cell fluorescence, were grown for 12 – 16 hours at 37^{o} C on LB/Agar as well. Single colonies were used to inoculate overnight (12 – 16 hours) cultures of 2 mL of M9 media supplemented with appropriate antibiotics. Next, the cultures were diluted 100X into 50 mL of fresh M9 with appropriate antibiotics and grown at 37^{o} C in a shaking water bath to OD_{600} = 0.2. They were then split and 18 mL was put into each of two flasks and exposed to 2 mL of either pre-warmed M9 media or pre-warmed M9 media supplemented with 50 mM aspartate, for a final aspartate concentration of 5 mM. Cells were grown at 37^{o} C with shaking for two more hours post induction. Samples were taken, spun at 3,300 x g for 5 minutes and re-suspended in 1X PBS with Kanomycin (2 mg/mL) to stop translation. Cells were diluted into 1X PBS and single cell GFP measurements were made using a BD Biosciences LSRII, courtesy of the Gladstone Research Institute (Laser settings -- FSC: 577, SSC: 335, GFP: 607). Each

data set consisted of at least 30,000 bacteria. The FlowJo software package was used to gate the data by FSC-H and SSC-A before calculating a geometric mean GFP fluorescence value (Tree Star Inc.).

3.5. Discussion

Here, we started with two systems that that are share a high degree of sequence similarity and for which crosstalk had been previously demonstrated *in vitro*. Remarkably, it required three knockouts to induce crosstalk, which demonstrates the degree to which two-component systems are buffered from crosstalk. This is consistent with other recent observations, all of which point to phosphotransfer being linear in the native host. There are other mechanisms by which crosstalk can occur, including the inclusion of additional kinase domains in the sensor kinase which interact with other response regulators.[38, 51, 52] There are also examples of cyoplasmic adaptor proteins that integrate multiple sensors and then interact with a downstream response regulator. An example of this is the Spo0F protein in *B. subtilis*, which integrates multiple sensors, and activates the Spo0A response regulator through the histidine phosphotransferase Spo0B.[53, 54] However, unless there are additional domains or proteins involved in phosphotransfer, it would appear that the phosphotransfer that occurs in canonical two-component systems is remarkably linear.

Integrating two-component systems may play a critical role for bacteria to identify an environment. To this end, there are many promoters that contain multiple operators that bind different response regulators. There are also genetic circuits that act as logic to integrate multiple inputs. An emerging conclusion is that most of the signal integration between two-component systems occurs on the transcriptional level.

The remarkable degree of buffering occurs due to a combination of kinetic interactions. We find that the bifunctional phosphatase activities of the sensors are just fast enough to remove the phosphates that get transferred via non-cognate interactions. This supports the theory put forward by Savageau and co-workers. However, when this interaction is removed, crosstalk does not occur. Also, there is evidence that the response regulators act as phosphate sinks by preferentially interacting with their cognate sensor. Previous studies suggest that only 0.02% of EnvZ is phosphorylated *in vitro* and OmpR preferentially interacts with phosphorylated EnvZ, leaving only the unphosphorylated EnvZ to act as a phosphatase to interact with non-cognate response regulators.[47, 55, 56] We see evidence of this effect in the inverse induction when CpxA is knocked out. Also, the removal of OmpR is critical for significant crosstalk to occur.

Programmable sensors are a key component for engineering bacteria. The modularity of two-component systems makes them an intriguing target for engineering, where input signals can be rapidly connected to control different pathways through domain swapping.[57-62] In addition, two-component systems can generally be moved between species. This enables access to a diversity of sensing functions present in bacteria. However, it has been observed that after transfer, two-component systems can exhibit cross reactions to host systems, which can have deleterious effects.[48] Therefore, it will be critical to understand how natural systems buffer crosstalk such that these interactions can be engineered *de novo*.

Chapter 4

Combinatorial sensing in *E. coli*

4.1. Introduction

Two-component systems are a highly conserved signal transduction pathway that enable bacteria to sense changes in their environment and adjust gene expression to adapt to nutrients, stresses, and other signals. We have sought to determine whether bacteria use multiple two-component systems as a combinatorial sensor to discriminate environmental conditions. A combinatorial sensor is made up of a set of sensors, each of which is activated to different degrees by many inputs such that the pattern of their activation defines the signal. Using promoter reporters and flow cytometry, we measured the response of three two-component osmosensors in *E. coli* (envZ/ompR, cpxA/cpxR, and rcsC/rcsD/rcsB) to 38 chemicals including known inducers of the systems, membrane perturbing agents, alcohols and chemicals of industrial relevance. We found that each system responded to a wide spectrum of conditions and that the three systems are uncorrelated, meaning that unique patterns of gene expression are generated by even closely related chemical compounds. Of the eight possible patterns generated by a three sensor system, we observe five. This data show that bacteria are able to use a limited set

of sensory components to identify a diverse set of compounds and environmental conditions.

4.2. Chemical library

We compiled a panel of 38 chemicals consisting of those that are known to influence these two-component systems, induce membrane perturbations, or are of industrial interest. We included chemicals that have been directly demonstrated to induce a system, and also chemicals with indirect evidence such as a growth phenotype in the absence of the two-component system[36]. A full list of the chemicals tested can be found below.

Table 4.1: Chemicals that produced response above threshold

Chemical	OD at 2 hours
Acetone	0.49
Butanol	0.59
Carbenecillin [1]	0.51
EtOH	0.49
Indole	0.51
Isobutanol	0.49
KCl	0.51
Methanol	0.53
NaCl	0.50
Oxacillin [1]	0.64
pH 5	0.48
pH 9	0.6
Pentanol	0.46
Phenethyl alcohol	0.53
Procaine	0.50

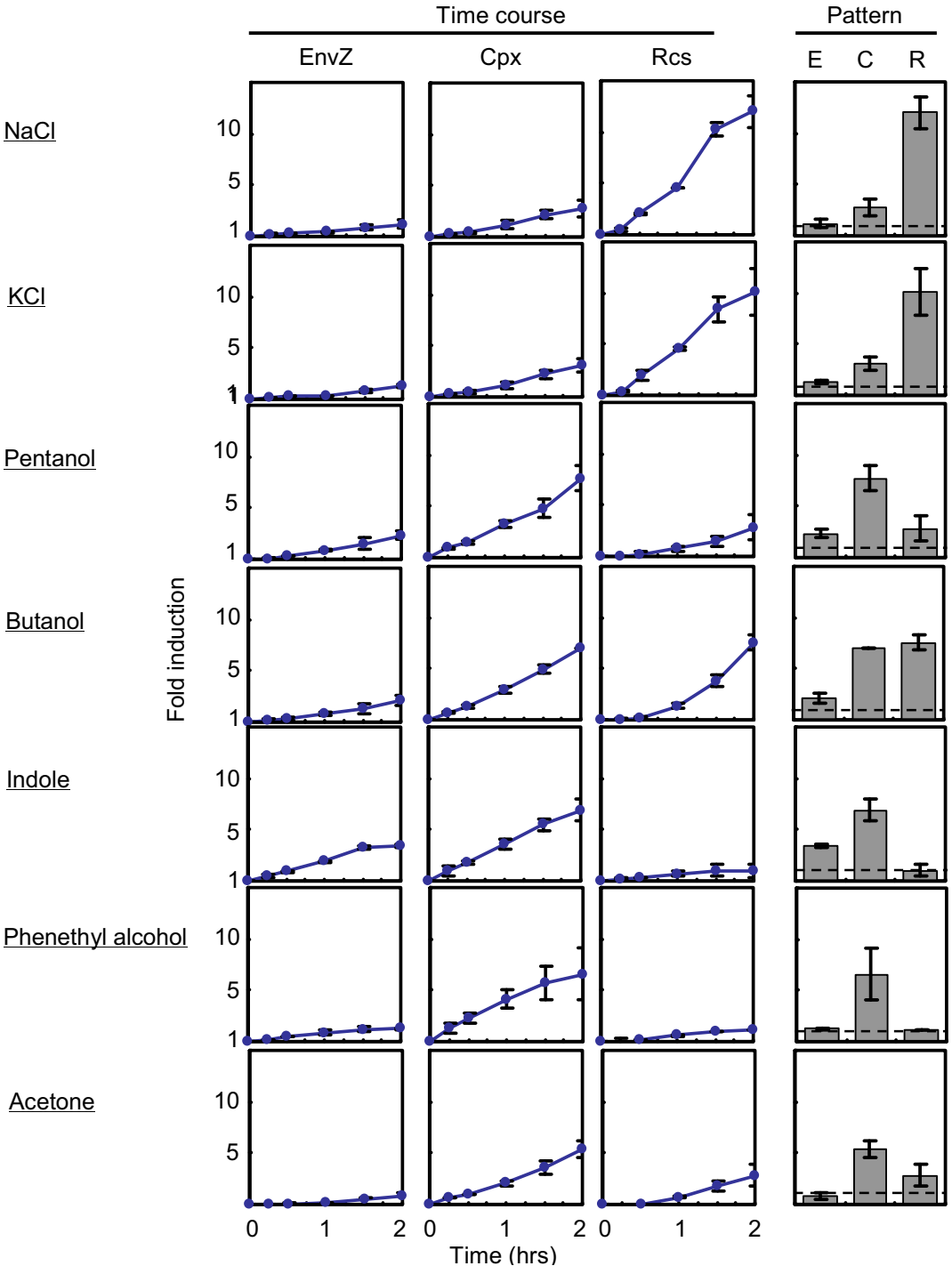
[1] Removed because constitutive control showed response

Table 4.2: Chemicals that produced response below threshold

Chemical	Source	Concentration
Aldrin	Supelco - 49000	100 µM
Anisole	Sigma - 296295	0.5 %
Butanetriol	Aldrich - B85803	6 %
Cellobiose	Fluka - 22150	400 mM
Chlorpromazine	MP Biomedicals - 190326	80 µM
Citral	Aldrich - C83007	0.002 %
CMC (cellulose)	Sigma - C5678	40 g/L
Cyclohexane	Riedel-de Haen - 15329	1%
EDTA	Sigma - E6758	1 mM
Ethyl butyrate	Aldrich - E15701	0.3 %
Ethyl propionate	Fluka - 96727	0.5 %
Heptanol	Fluka - 51790	0.03%
Hexanol	Fluka - 52830	0.05 %
Isoprene	Fluka - 59240	3.25 %
Limonene	Aldrich - 218367	0.7 %
m-Xylene	Flucka - 95670	0.1 %
Nitrofurazone	TCI - ALO1	1 µg/mL
Octanol	Sigma - 04500	0.02 %
Polymixin B	Fluka - 81271	0.4 µM
Sucrose	Fisher-BP220-212	545 mM
Thioridazine	Sigma - T9025	1 µg/mL
Toluene	Sigma - Aldrich - 650579	0.25 %
Trehalose	Acros organics - 309870250	200 mM
Trimethoprim	MP Biomedicals - 195527	2 µg/mL
ZnCl2	Sigma - Z0173	1 mM

4.3. Time-course inductions and flow cytometry

We next examined the effect of each compound on the expression of reporter genes. The three sensor pattern generated in response to a given chemical is a three bar histogram where each bar corresponds to the fold induction of one of each of the two-component systems at the two hour time point. Fold induction is calculated by dividing the mean GFP from cells grown in inducing conditions by the mean GFP value from uninduced cells. Raw time course data, patterns generated, and raw end point flow cytometry can be found below.


Time course
Pattern
EnvZ
Cpx
Rcs
E
C
R
NaCl
KCl
Pentanol
Butanol
Indole
Phenethyl alcohol
Acetone
Fold induction
10
5
1
0
1
2
Time (hrs)

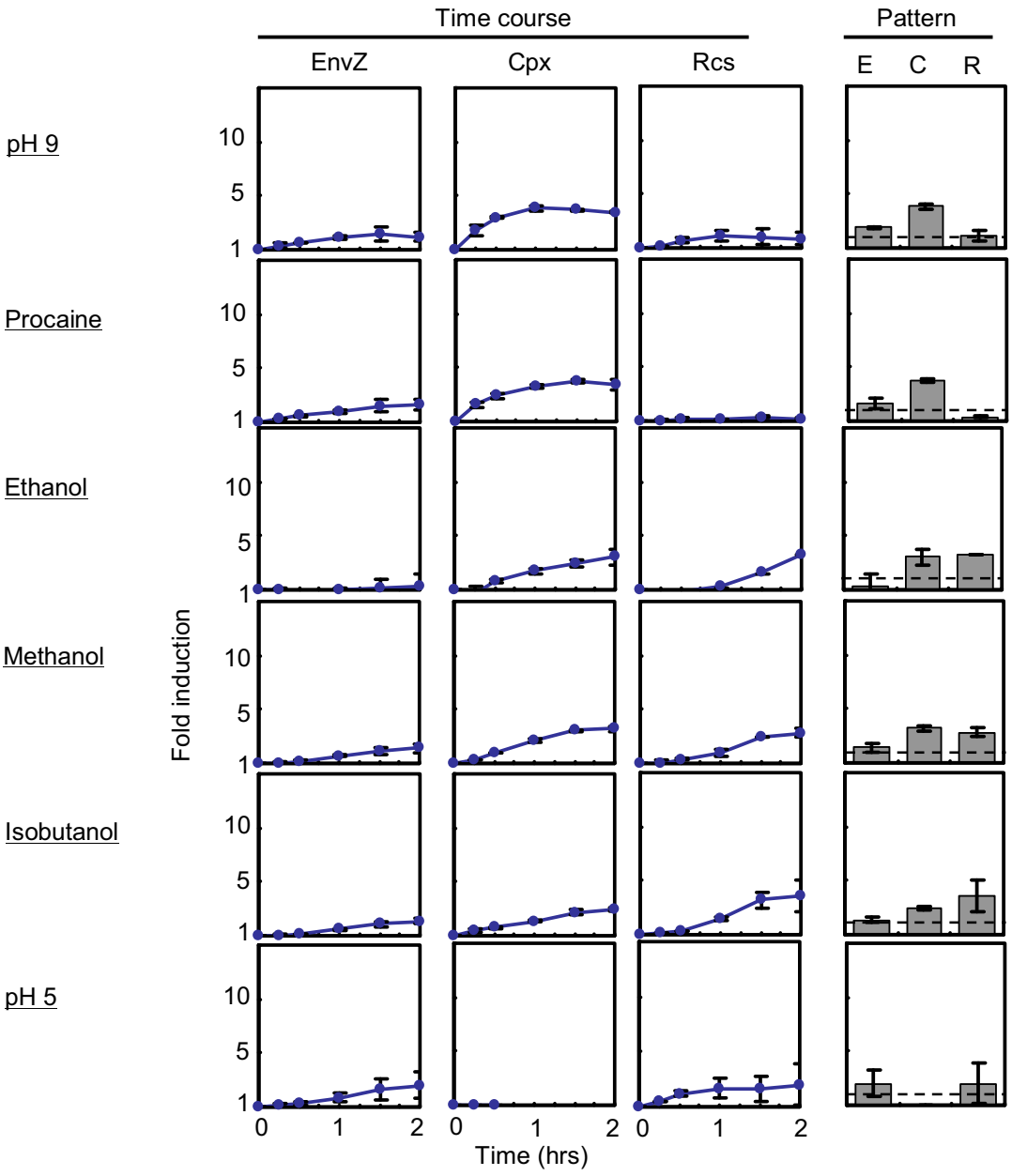
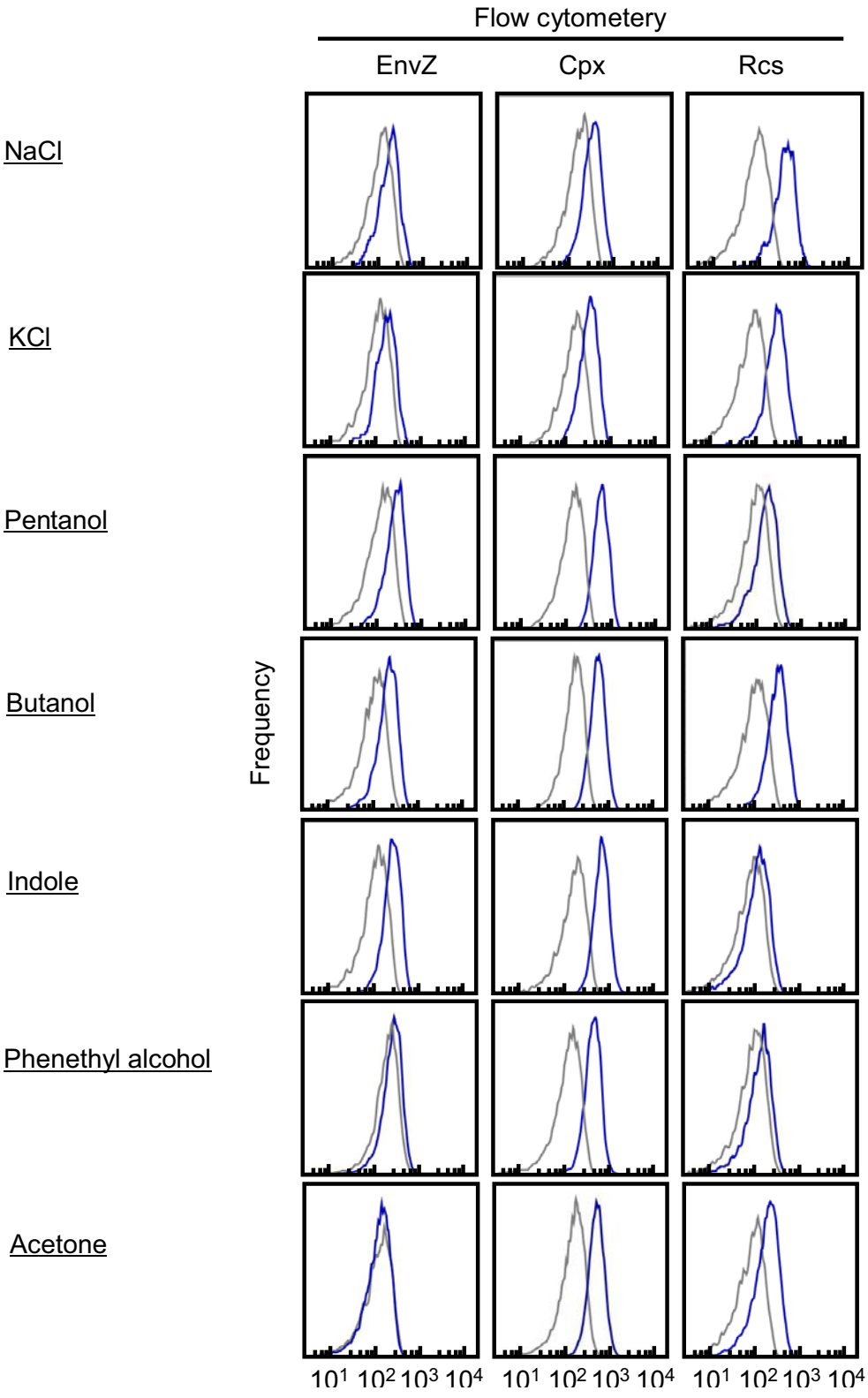


Figure 4.1: Time course data and corresponding patterns


Flow cytometery
EnvZ
Cpx
Rcs
NaCl
KCl
Pentanol
Butanol
Indole
Phenethyl alcohol
Acetone
Frequency
10^1 10^2 10^3 10^4
10^1 10^2 10^3 10^4
10^1 10^2 10^3 10^4

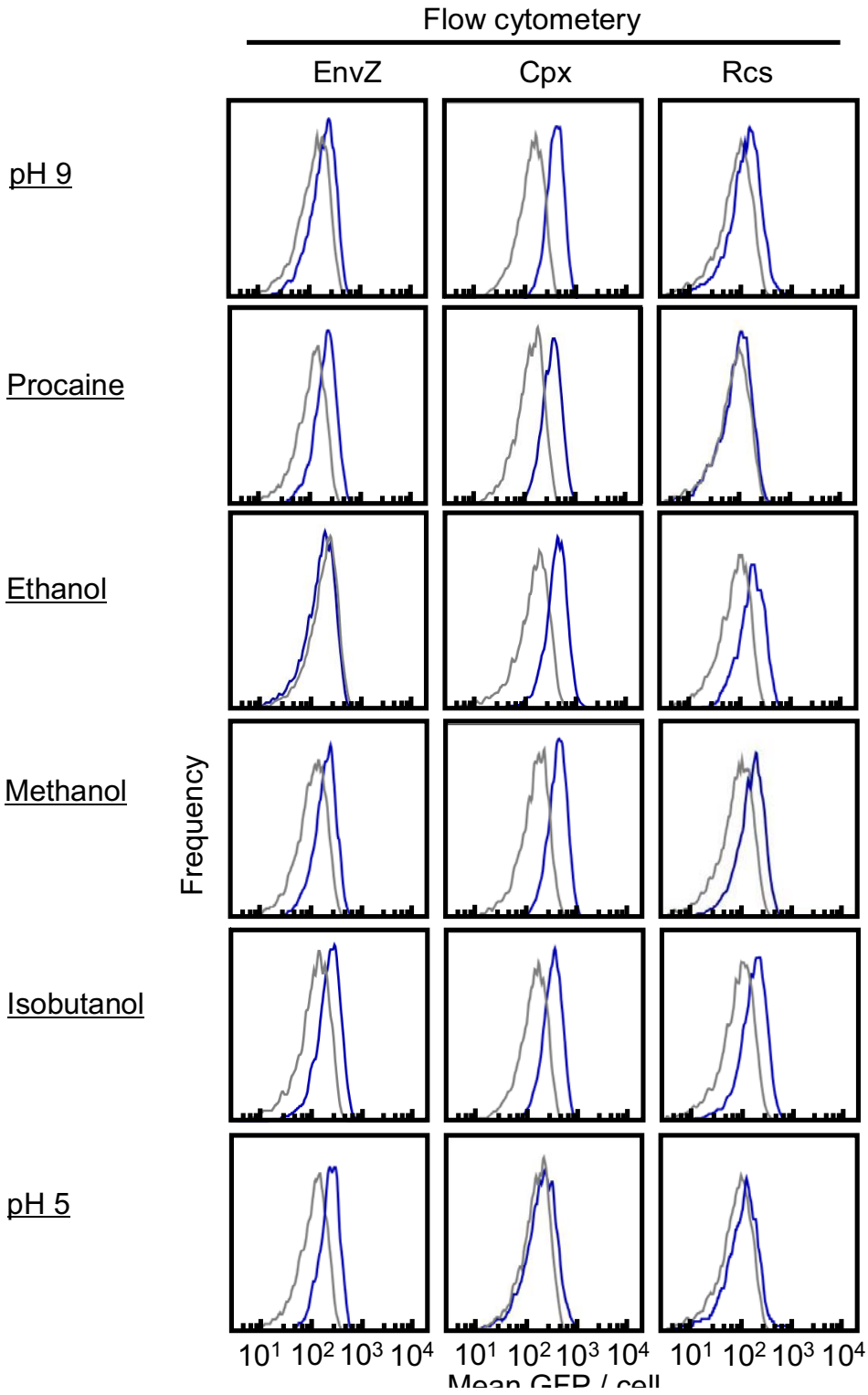


Figure 4.2: Flow cytometry data at 2-hour timepoint

We find that the kinetics of the responses vary both across chemicals, and across sensors. For each sensor, there are chemicals that lead to a fast induction and others that lead to slow induction. The differences in response time are not strictly a function of chemical; for example, the response of rcs to all of the alcohols is delayed for as much as an hour in some cases, whereas for cpx we see no delay for any of the alcohols. The kinetics of the response may be a result of how the chemical perturbs the membrane. Certain chemicals illicit responses in multiple sensors with similar kinetics (indole), whereas other chemicals induce with different kinetics. For example, rcs responds more slowly than cpx to butanol, ethanol and pentanol (time to half maximal induction occurs around 90 minutes, versus 60 minutes for cpx). One explanation for this is that the sensors have an intrinsically different response time. Though this might be the case, there are examples of chemicals that induce rcs faster than cpx, such as NaCl, KCl and sucrose. A second explanation is that the effect of alcohol on the membrane changes over time, and that the sensors respond differently to these time dependent perturbations.

4.4. Constitutive control

We excluded compounds from further analysis for one of two reasons: either the chemical induced all three systems by less than two-fold, or else it was non-specifically leading to GFP accumulation. This later criterion was determined by measuring the activity of a constitutive promoter under the same growth conditions; chemicals with control induction of greater than two fold were eliminated (Supplemental figure 2). There were two chemicals that fell into the latter category, both beta lactams. These drugs have been shown to affect the size and shape of cells; specifically they can lead to larger or

more rod-like cells[63]. Either deformation could contain more GFP, even with a constitutively expressed GFP. Of the 38 chemicals tested, 13 passed these two criteria and were chosen for further analysis.

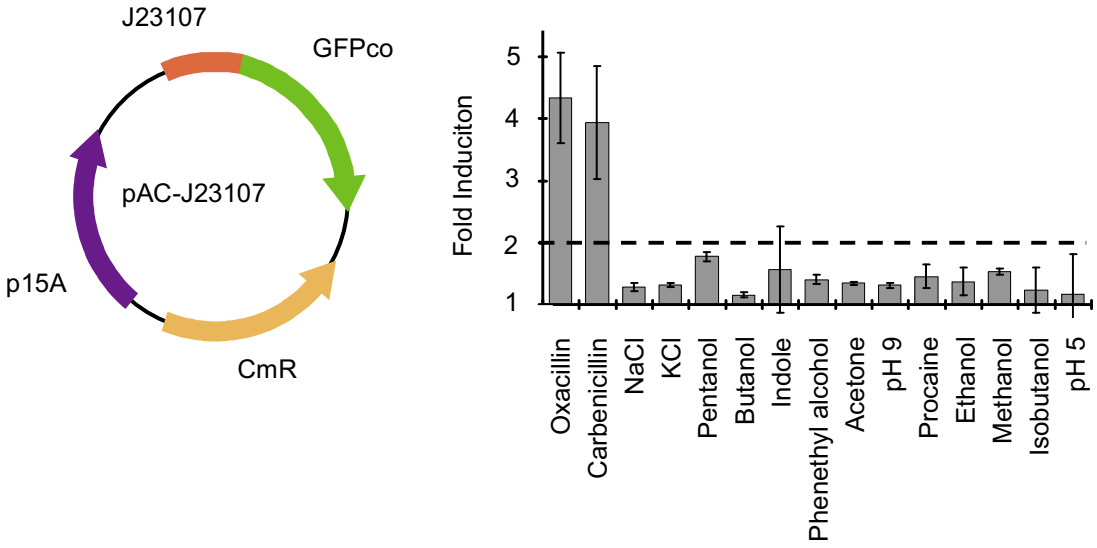


Figure 4.3: Constitutive expression control

4.5. Summary of fold inductions

Each two-component system is activated by the majority of the 13 chemicals. A summary of the fold inductions for all of the chemicals can be found in Figure 3A. Though we designed the chemical library to include compounds expected to influence these systems, and eliminated in the initial screen those chemicals that affected none of them, the degree of overlap in inducers of the three systems was surprising. In fact, as defined by a two-fold cut off, EnvZ/ompR, cpxA/cpxR, and rcsC/rcsD/rcsB responded to 10, 11, and 10 of the 13 chemicals, respectively. The top five inducers for each of envZ and cpx include alcohols, indole, and pH extremes. These seemingly do not have a

common mode of action; alcohol is known to fluidize the membrane, indole causes derangement of the outer membrane, and pH affects levels of both periplasmic and envelope proteins. [64, 65] Rcs is strongly induced by alcohols and also salt, two stresses that have opposite effects on membrane fluidity. Because each system is activated by many and diverse inputs, they meet the first requirement for a combinatorial sensor that the sensors be general.

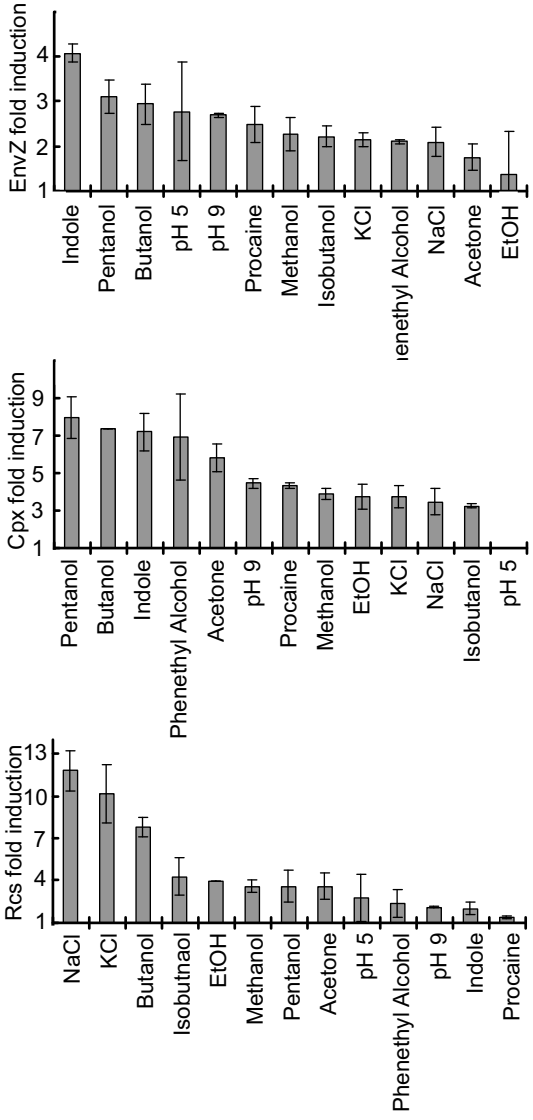


Figure 4.3: Summary of fold inducitons

4.6. Correlation of sensor inductions

Though these two-component systems respond to most of the 13 chemicals, a variety of three sensor patterns is generated. There is a slight correlation in induction between cpx-envZ, the two systems which share the highest sequence identity of any two-component systems in *E. coli[14]*, however little correlation between rcs-envZ and rcs-cpx (Figure 4). In all three correlation plots there are examples of chemicals that induce one system and not the other. For example, pH4 induces envZ but not cpx, and phenethyl alcohol induces cpx but not envZ. This suggests that the systems do not respond to the same inducing signal. Because these sensors respond for the most part independently of one another to each chemical, the second requirement for a combinatorial sensor is met.

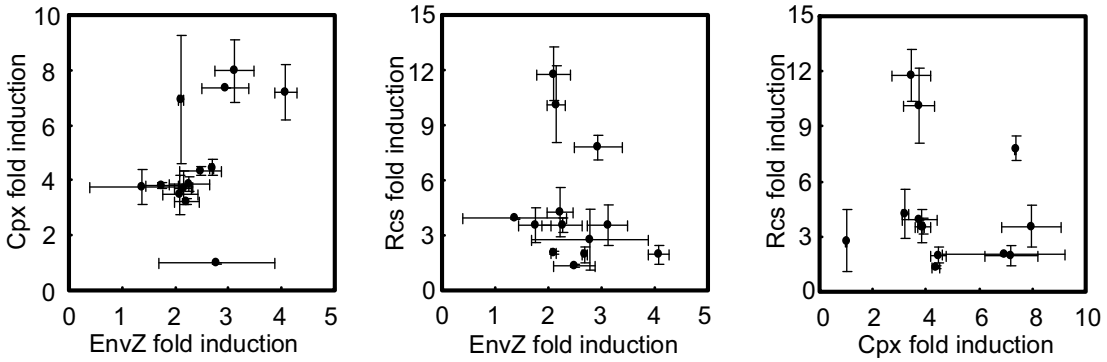


Figure 4.4: Correlation between sensors

Environmental conditions generate unique patterns of sensor responses. The sample correlation coefficients for each pair of sensors were calculated to be: cpx-envZ, R^2 =0.323; rcs-envZ, R^2 = 0.024; rcs-cpx, R^2 = 0.001. Points are the average of 2 or more experiments performed on different days, and error bars are the standard deviation from the mean.

4.7. Osmolarity and pH controls

Because env, cpx, and rcs are respond to high osmolarity, we determined whether the responses in gene expression were simply due to differences in osmolarity in the assay conditions(Figure 3B). If this were the case, we would expect the fold-induction to be correlated with osmotic strength. We added each chemical to media and measured the resulting osmolarity with an osmometer. We found little correlation between osmolarity and fold induction for envZ (R^2 = 0.066) and cpx (R^2 =0.095). For rcs, we find a significant correlation (R^2 =0.639). This is driven primarily by the strong response to the two highest osmolality conditions (NaCl and KCl), and when these two points are removed, there is no longer a correlation. To highlight this, there is a condition, namely sucrose, which is a high osmolality condition that does not induce (in fact it did not make the 2-fold cut off criterion for any sensor) and butanol, a low osmolality condition that induces strongly. Thus, though each system is induced by high salt, we can conclude that none of the systems respond strictly to osmolarity.

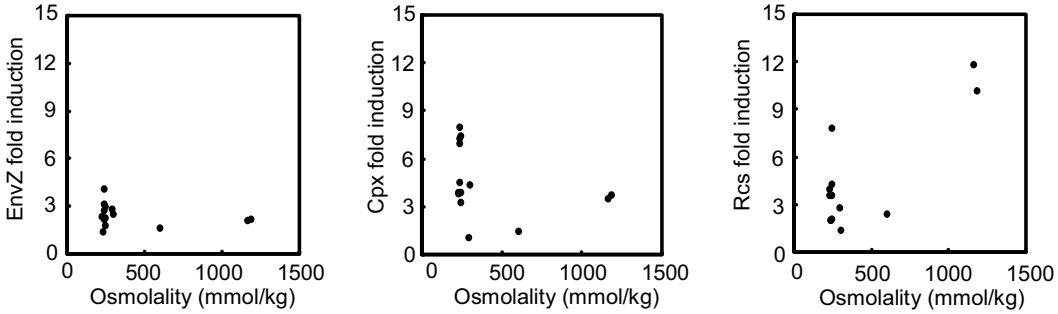


Figure 4.5: Correlation between fold induction and medium osmolarity

Correlation between fold-induction and osmolarity for each sensor. The sample correlation coefficients for each were calculated to be: envZ, R^2 = 0.066; cpx, R^2 =0.095. rcs, R^2 =0.639.

pH has a significant effect on induction, especially for envZ and cpx (Figures 2 and 3). To test whether any of the inductions to the chemicals were a result in a change in pH, we measured media pH after a two hour induction in the presence of each chemical. All pHs were within 0.11 pH units of an uninduced culture grown in parallel, thus pH is not a factor in the differences we observe.

4.8. Principle component analysis of pattern data

We applied principle component analysis (PCA) to the data set in order to visualize both the full diversity of patterns generated, as well as how chemicals cluster in terms of sensor pattern. PCA allows the user to reduce the dimensionality of data by finding new axes that capture the most information from the original data set[66]. In our data, each input pattern can be thought of as a vector with three elements, producing a three dimensional space where each chemical is represented by a point. In this case, PCA reduces the dimensions from three to two, but it could be applied to systems with many more sensors as well (e.g. as analysis of all 32 *E. coli* sensors). To capture differences in the patterns themselves, rather than differences in induction magnitudes, we normalized each pattern by its maximum before performing the PCA. As a result, a pattern made up of small inductions will cluster with a pattern of large inductions, as long as the pattern is the same. Though PCA allows us to visualize clusters, the components themselves do not intuitively correspond to a specific sensors activity; note that the point corresponding to no induction for all three sensors is not (0,0) in PCA space, rather is represented in Figure 5 by a black dot.

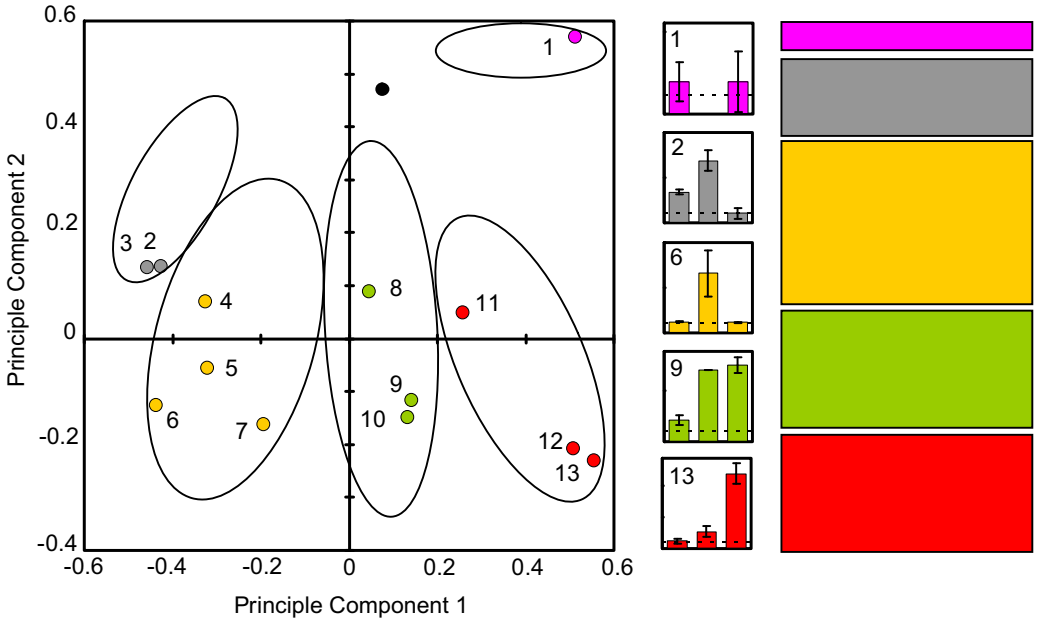


Figure 4.6: Principle component analysis of pattern data

Principle component analysis of sensor data. Each point in the left panel corresponds to the pattern generated by a single chemical projected onto a 2-dimenstional principle component space. Points are grouped by type of pattern, and representative patterns are shown on the right. The pattern corresponding to no induction for all three systems is represented on the graph as a black point.

Compounds in principle component space were grouped by the type of patterns they induce. Of the seven patterns possible with three binary sensors (excluding the all-OFF response) we observe four. Examples of each are shown to the right of the PCA plot. The most common pattern is one where envZ is off, and both cpx and rcs are on; 7 out of the 13 patterns fall into this category. Within this set, we divided the chemicals into two clusters. In one, rcs responds much more that cpx and in the other they respond more or less equally. We do not observe the pattern with all sensors on; though there are conditions such as butanol and NaCl which induce all three, they induce cpx and rcs so

much more than envZ that it is classified as being an envZ-off, cpx-on, rcs-on pattern. We also do not observe the two patterns corresponding to only envZ or only rcs on.

Certain clusters can be rationalized based on the effect that the cluster members have on the membrane. For example indole and procaine are in the same cluster, and both perturb the outer membrane. These two patterns are characterized by envZ and cpx being on, and rcs off. The pattern itself is surprising since rcs has been shown to respond to surface contact as well as over-expression of an outer membrane lipoprotein, both of which are likely to influence the outer membrane. We also found that some chemicals which increase membrane fluidity, such as alcohols and acetone[67], cluster.

We were surprised to see that certain chemicals did not cluster. For example, though the linear short chain alcohols cluster, pentanol and phenethyl alcohol induce very different patterns, and cluster with pH10. Also, instead of clustering with butanol, isobutanol clusters with the salts, a separation driven by a large difference in cpx induction. These alcohols could affect the membrane differently, and by extension the two-component systems, based on their specific properties such as chain length, bulkiness, and lipophilicity. pH4 is the only condition we tested that induces envZ and rcs but not cpx, and though envZ and rcs are also slightly induced at pH10, cpx is dramatically induced. This leads to two very different patterns, and thus different clusters.

4.9. Effect of concentration on response amplitude and pattern.

We examined sensor activity in the reporter strains over a range of short-chain alcohol (ethanol, butanol, pentanol) concentrations and a range of pH.

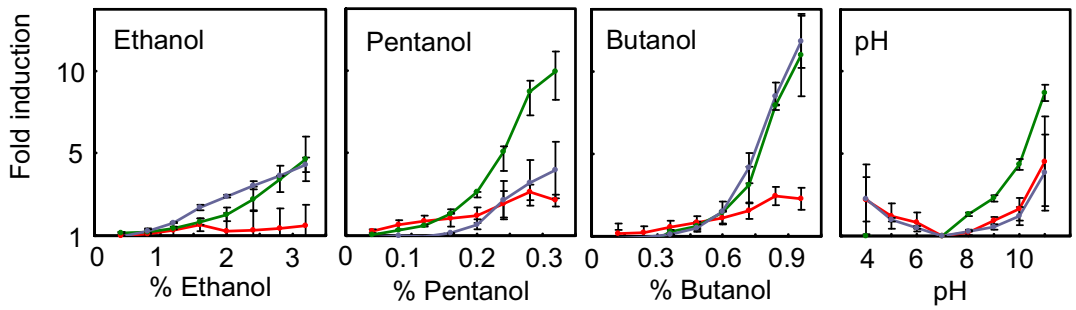


Figure 4.7: Response to chemical concentration gradient

For each of ethanol, pentanol, butanol and pH, a series of measurements were made, either varying concentration of the alcohol, or varying pH. EnvZ is red, cpx is green, rcs is blue. Data points are the average of two runs performed on different days, and error bars are the standard deviation from the mean

The response to alcohols and pH shows that the sensors are graded, rather than all-or-none. As cells are exposed to more inducer there is an incremental increase in activity for all sensors.

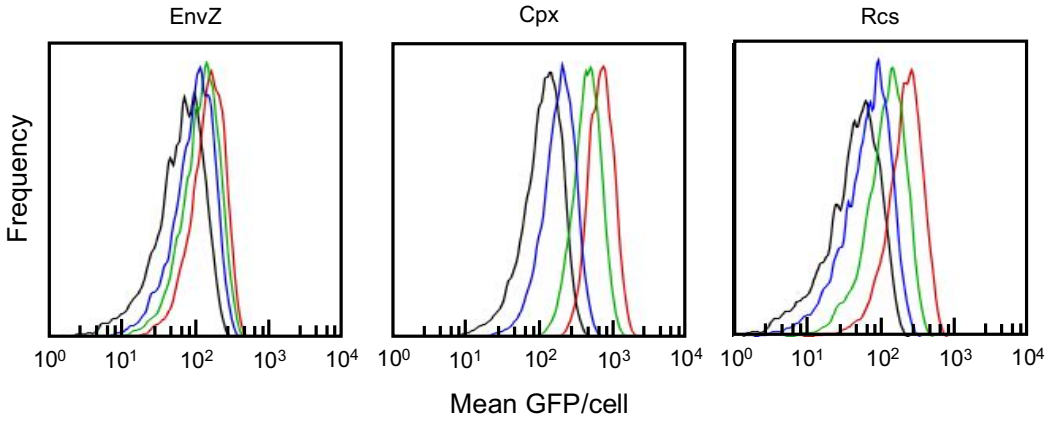


Figure 4.8: Graded sensors

On each panel is superimposed raw flow cytometry data from four difference butanol concentrations at the two hour time point. Histogram color in order of increasing butanol concentration: black, blue, green, red.

Each concentration has a corresponding induction pattern which can be visualized in PCA space, allowing us to observe how that pattern changes across a gradient of pH or inducer concentration. Bellow, each of the 4 trajectories corresponds to one inducer, with each point on the trajectory a different concentration. For the alcohols, the trajectory starts at low and moves towards high concentration, with conditions inducing all sensors less that 2-fold removed. A trajectory crossing the boundary of a cluster signifies that the pattern has changed.

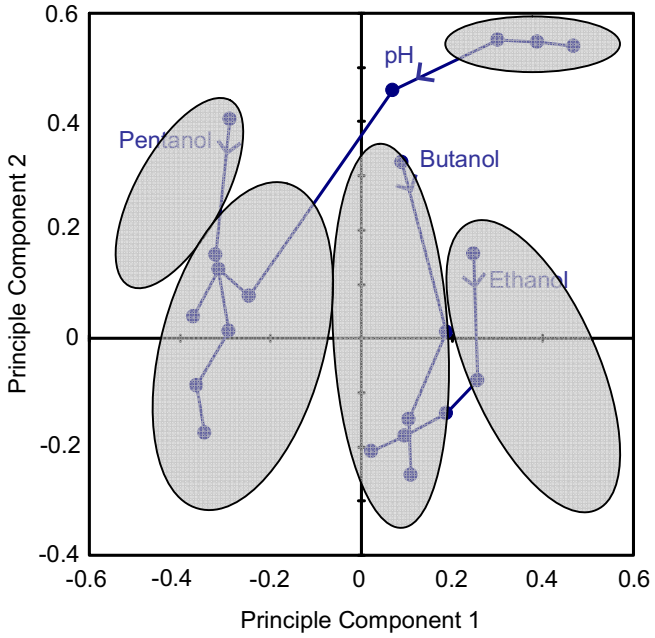


Figure 4.9: Projection of concentration data onto PC space

Projection of each sequence of patterns onto principle component space. A trajectory crossing the boundary of a cluster signifies that the pattern has changed. The patterns for the alcohols are insensitive to changes in concentration, whereas the pattern for pH changes dramatically. Ethanol and butanol are distinguishable from pentanol for all concentrations.

Interestingly, though the underlying sensor inductions vary dramatically, the patterns for the three alcohols remain remarkably constant. Though the trajectories of pentanol and ethanol do cross boundaries at their lower concentrations, they are only subtle changes in pattern. In contrast, varying pH does change the pattern. At the two extremes, the trajectory lies in clusters of dramatically different patterns, with the pattern changing from envZ and rcs on and cpx off at low pH, to cpx on at high pH.

The trajectories of butanol and ethanol are nearly overlapping across all concentrations which indicates that their patterns are the same independent of concentration. Interestingly, neither trajectory crosses that of pentanol. Thus the initial observation that pentanol is in a different cluster from the shorter chain alcohols is not an artifact of picking concentrations on a different part of the same trajectory. Based on pattern, envZ, cpx, and rcs can discriminate pentanol from ethanol and butanol, and are able to read out alcohol concentration from amplitudes of induction. The sensors can determine their position on the pH gradient using patterns alone.

4.10. Methods

Strains and plasmids

Strains BW28357 (a K12 derivative), BW29655 (BW28357ΔenvZompR), BW29849 (BW28357ΔcpxAcpxR) and BW27870 (BW28357ΔrcsB) were obtained from the Coli

Genetic Stock Center at Yale. All promoter reporters are variants of pAC581. This plasmid possesses a p15A origin, chloramphenocol resistance marker, and a codon optimized GFPmut3. GFP expression is under the control of various promoters. *PompC* (2310952...2310802), *PcpxP* (4103753...4103858) and *PrprA* (1768254...1768396) were amplified from MG1655 genomic DNA. The constitutive control plasmid with promoter J23107 was constructed by inserting annealed primers into pAC581. J23107 is a 35 base pair sequence from a library of constitutive promoters and assigned BBa_J23107 in the MIT registry of standard biological parts (JC Anderson, University of California, Berkeley).

Media

Assays were performed in M9 minimal media (48 mM Na2HPO4, 22 mM KH2PO4, 16 mM NH4Cl, 2 mM MgSO4, 0.1 mM CaCl2 and 8.6 mM NaCl, pH7.4) with 0.4% glucose as the carbon source plus 30 μg/mL chloramphenicol. Cells were plated on selective LB agar (25 g of Luria-Bertani broth (Difco) and 18 g Bacto agar (Difco) per L) plus 30 μg/mL chloramphenicol.

Induction Assay

Cells were streaked from glycerol stocks onto an LB agar plate and grown for 16 hours. A culture was inoculated from this plate and grown for 14 hours, and then diluted 100x into fresh media. At OD_{600} = 0.4 cells were again diluted 100x into 150 mls. These cells were grown at 37^{o}C in a shaking water bath at 160 rpm. When cells reached OD_{600} = 0.2, the culture was split into 6 flasks, one un-induced, and the rest containing various

induction chemicals. Samples was taken at t = 0 min, 15min, 30min, 60min, 90min and 120min. Protein expression was immediately stopped in each sample with 2 mg/mL kanamycin, and the samples were stored on ice until flow cytometry was performed. Control cells containing no plasmid were handled similarly. Data points are the average of at least 2 independent experiments with assays run on different days.

Chemicals

Acetone (4% ; Sigma-650501), Butanol (0.54% ; Sigma-B7906), Ethanol (3% ; Gold shield chemicals co), Indole (3mM ; Fisher-06104), Isobutanol (0.3% ; Fluka-82059), KCl (550mM ; Fisher-S77375), Methanol (4.8% ; Sigma-34898), NaCl (550mM ; Acros organics-327300010), Pentanol (0.3% ; Fluka-77597), Phenethyl Alcohol (0.2% ; Sigma-P13606), Procaine (17mM ; Sigma-P9879)

Measurig osmolaritys

Chemical was added at the appropriate concentration to M9, and the osmolarity was measured using a Wescor vapro pressure osmometer, model 5520. The machine was calibrated with three standards: 1000, 290, and 100mmol/kg. A clean test was performed prior to measurements to ensure linearity and accuracy.

Flow cytometry

Cells were diluted 100x into phosphate-buffered saline containing 2 mg/ml kanamycin and single-cell GFP measurements were made on a BD Biosciences LSRII flow cytometer (Courtesy of the Gladstone Resarch Institute) with laser settings: FSC,578;

SSC,335; GFP,607. Each data set consisted of at least 10,000 cells. The FlowJo software package (Tree Star Inc) was used to gate the data on the FSC-SSC scatter. From this, a histogram of GFP fluorescence was generated the mean GFP fluorescence calculated.

Data analysis

Principle component analysis was preformed using the multivariate analysis add-on in Microsoft Excel. Prior to analysis, each row of the matrix is normalized to its max, and then each column in mean centered. The contributions of each sensor to the principle components are: PC1 (0.025, -0.619, 0.785) and PC2 (0.988, -0.103, -0.112). PC1and PC2 together capture 85% of the variance of the original dataset. To generate Figure 6, we first removed conditions that induced all sensors by less than 2-fold. We then normalized each pattern to its maximum, as before. Each concentration pattern was then projected onto the PCA space of Figure 5.

4.11. Discussion

In this work, we demonstrate that the osmosensors envZ/ompR, cpxA/cpxR, and rcsC/rcsD/rcsB function as a combinatorial sensor of chemicals and environmental conditions including alcohols, salts, pH and membrane perturbing agents. Though previous work has demonstrated that these systems are activated by many conditions, our work shows that diverse patterns are generated, and that the sensor fold inductions are uncorrelated with each other, as well as with osmolarity. In PCA space, clustering of certain chemicals could be rationalized based on their membrane effect; however, we were surprised to find that pentanol was in a different cluster from the shorter chain alcohols. To explore this, we looked at how these patterns changed as a function of

inducer concentration and found both that the patterns generated by butanol, ethanol and pentanol were robust to inducer concentration, and that pentanol induced a distinct pattern from the other alcohols across all concentrations.

One way for combinatorial sensing to be physiologically relevant to the cell is for there to be a way for the cell to "read" the pattern. We hypothesized that pattern recognition might occur at the transcriptional level; there exist complex promoters that are regulated in different combinations of two response regulators [24, 68-70] as well as csgD which is regulated by all three. Multiple transcription factors regulating a single promoter can lead to predictable and physiologically sensible logic[71, 72]. To determine whether co-regulated promoters are capable of integrating signals in this system, a first step would be to map the cis-regulatory function of the complex promoters across combinations of three response regulator inputs. Another non-mutually exclusive possibility is that quantitative differences in timing and expression of genes could be an optimized and coordinated response to a given stress.

A combinatorial sensor composed of general sensors may be favored by evolution because it allows a large set of environmental conditions to be detected by a limited set of sensory components, in direct contrast to a system where each environmental condition necessitates a specialized sensor.[73] Also, rather than evolving a new sensor protein in order to sense a new environmental condition, the cell could rewire downstream signal processing in order to recognize a new pattern. This could be possible via modification of the promoter structure or transcription factor binding sites[74].

We do not expect that each input makes direct binding interactions with the sensor kinases. More likely, the sensors are responding to properties of the membrane

and envelope. Temperature, pH, various chemicals, oxygen, and osmolarity all have been shown to affect the membrane. For lipophilic compounds, these changes are often induced by a direct insertion of the compound into the lipid bilayer. Certain stresses can immediately affect membrane fluidity, whereas other stresses lead to a secondary response, such as changes in membrane permeability, potential, lipid structure and composition, and protein composition[64, 75-79]. Altered by a particular stimulus, it may be that these membrane properties, in some combination, are themselves the input to the two-component systems.

The impact of alcohol chain length on membrane perturbation is well studied and is particularly interesting in the context of our observations. Longer chains influence fluidity further into the membrane than do shorter chain length alcohols[64]. Also, the adaptive changes in lipid composition vary with chain length[80]. For alcohols with chain lengths 1-4, 18:1 fatty acid levels increase and 16:0 fatty acid levels decrease. For chain lengths 5-10 however, the opposite occurs with 18:1 fatty acid levels decreasing and 16:0 levels increasing. This correlates with our observation that pentanol clusters away from the shorter chain alcohols. The difference in patterns between pentanol and the lower chain alcohols is driven predominantly by a difference in the rcs response. One might speculate that rcs is sensing only shallow changes in membrane fluidity, or else that it is sensitive to the ratios of these specific lipids.

A combinatorial sensor composed of a small number of general sensors could be used for industrial applications to detect the presence of a specific chemical. For a given application, other candidate sensors from *E. coli* (such as rstBA, kdpDE, baeSR), or from other bacteria, could be screened to identify a subset that best identifies the chemical of

interest. Screening sensor promoters should be faster than other strategies, such as designing or evolving a protein. One or more general sensors could also be used in combination with existing engineered sensors to increase their specificity. In the simplest implementation, a color readout could indicate the abundance of a particular chemical, but one could envision engineering a more sophisticated system capable of pattern recognition, enabling the user to interface the output with other components in their design.

This systematic measurement of the response of two component systems to a variety of chemicals demonstrates that *E. coli* has general and uncoupled sensors. Combinatorial sensing, made possible by sensors with these properties, could be an efficient mechanism for organisms to perceive changing environments.

References

1. Groisman, E.A. and C. Mouslim, *Sensing by bacterial regulatory systems in host and non-host environments.* Nat Rev Microbiol, 2006. **4**(9): p. 705-9.
2. Ruiz, N. and T.J. Silhavy, *Sensing external stress: watchdogs of the Escherichia coli cell envelope.* Curr Opin Microbiol, 2005. **8**(2): p. 122-6.
3. Beier, D. and R. Gross, *Regulation of bacterial virulence by two-component systems.* Curr Opin Microbiol, 2006. **9**(2): p. 143-52.
4. Majdalani, N. and S. Gottesman, *The Rcs phosphorelay: a complex signal transduction system.* Annu Rev Microbiol, 2005. **59**: p. 379-405.
5. Munson, G.P., et al., *Identification of a Copper-Responsive Two-Component System on the Chromosome of Escherichia coli K-12.* J. Bacteriol. %R 10.1128/JB.182.20.5864-5871.2000, 2000. **182**(20): p. 5864-5871.
6. Prohinar, P., et al., *OmpR-dependent and OmpR-independent responses of Escherichia coli to sublethal attack by the neutrophil bactericidal/permeability increasing protein.* Mol Microbiol, 2002. **43**(6): p. 1493-504.

7. Rampersaud, A. and M. Inouye, *Procaine, a local anesthetic, signals through the EnvZ receptor to change the DNA binding affinity of the transcriptional activator protein OmpR.* J Bacteriol, 1991. **173**(21): p. 6882-8.

8. Thomas, A.D. and I.R. Booth, *The regulation of expression of the porin gene ompC by acid pH.* J Gen Microbiol, 1992. **138**(9): p. 1829-35.

9. Alphen, W.V. and B. Lugtenberg, *Influence of osmolarity of the growth medium on the outer membrane protein pattern of Escherichia coli.* J Bacteriol, 1977. **131**(2): p. 623-30.

10. Jubelin, G., et al., *CpxR/OmpR Interplay Regulates Curli Gene Expression in Response to Osmolarity in Escherichia coli.* J. Bacteriol. %R 10.1128/JB.187.6.2038-2049.2005, 2005. **187**(6): p. 2038-2049.

11. Sledjeski, D.D. and S. Gottesman, *Osmotic shock induction of capsule synthesis in Escherichia coli K-12.* J Bacteriol, 1996. **178**(4): p. 1204-6.

12. Aiba, H., T. Mizuno, and S. Mizushima, *Transfer of phosphoryl group between two regulatory proteins involved in osmoregulatory expression of the ompF and ompC genes in Escherichia coli.* J Biol Chem, 1989. **264**(15): p. 8563-7.

13. Mecsas, J., et al., *The activity of sigma E, an Escherichia coli heat-inducible sigma-factor, is modulated by expression of outer membrane proteins.* Genes Dev, 1993. **7**(12B): p. 2618-28.

14. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.

15. Wolfe, A.J., et al., *Signal integration by the two-component signal transduction response regulator CpxR.* J Bacteriol, 2008. **190**(7): p. 2314-22.

16. Danese, P.N. and T.J. Silhavy, *CpxP, a stress-combative member of the Cpx regulon.* J Bacteriol, 1998. **180**(4): p. 831-9.

17. Raivio, T.L., D.L. Popkin, and T.J. Silhavy, *The Cpx envelope stress response is controlled by amplification and feedback inhibition.* J Bacteriol, 1999. **181**(17): p. 5263-72.

18. De Wulf, P., O. Kwon, and E.C. Lin, *The CpxRA signal transduction system of Escherichia coli: growth-related autoactivation and control of unanticipated target operons.* J Bacteriol, 1999. **181**(21): p. 6772-8.

19. Otto, K. and T.J. Silhavy, *Surface sensing and adhesion of Escherichia coli controlled by the Cpx-signaling pathway.* Proc Natl Acad Sci U S A, 2002. **99**(4): p. 2287-92.

20. Nakayama, S. and H. Watanabe, *Involvement of cpxA, a sensor of a two-component regulatory system, in the pH-dependent regulation of expression of Shigella sonnei virF gene.* J Bacteriol, 1995. **177**(17): p. 5062-9.

21. DiGiuseppe, P.A. and T.J. Silhavy, *Signal detection and target gene induction by the CpxRA two-component system.* J Bacteriol, 2003. **185**(8): p. 2432-40.

22. Laubacher, M.E. and S.E. Ades, *The Rcs phosphorelay is a cell envelope stress response activated by peptidoglycan stress and contributes to intrinsic antibiotic resistance.* J Bacteriol, 2008. **190**(6): p. 2065-74.

23. Castanie-Cornet, M.P., K. Cam, and A. Jacq, *RcsF is an outer membrane lipoprotein involved in the RcsCDB phosphorelay signaling pathway in Escherichia coli.* J Bacteriol, 2006. **188**(12): p. 4264-70.

24. Francez-Charlot, A., et al., *RcsCDB His-Asp phosphorelay system negatively regulates the flhDC operon in Escherichia coli.* Mol Microbiol, 2003. **49**(3): p. 823-32.

25. Sturny, R., et al., *NhaR and RcsB independently regulate the osmCp1 promoter of Escherichia coli at overlapping regulatory sites.* J Bacteriol, 2003. **185**(15): p. 4298-304.

26. Ache, B.W. and J.M. Young, *Olfaction: diverse species, conserved principles.* Neuron, 2005. **48**(3): p. 417-30.

27. Bar-Cohen, Y., *Biomimetics--using nature to inspire human innovation.* Bioinspir Biomim, 2006. **1**(1): p. P1-P12.

28. Doleman, B.J., E.J. Severin, and N.S. Lewis, *Trends in odor intensity for human and electronic noses: relative roles of odorant vapor pressure vs. molecularly specific odorant binding.* Proc Natl Acad Sci U S A, 1998. **95**(10): p. 5442-7.

29. Freund, M.S. and N.S. Lewis, *A chemically diverse conducting polymer-based "electronic nose".* Proc Natl Acad Sci U S A, 1995. **92**(7): p. 2652-6.

30. Hellingwerf, K.J., *Bacterial observations: a rudimentary form of intelligence?* Trends Microbiol, 2005. **13**(4): p. 152-8.

31. Bijlsma, J.J. and E.A. Groisman, *Making informed decisions: regulatory interactions between two-component systems.* Trends Microbiol, 2003. **11**(8): p. 359-66.

32. Oshima, T., et al., *Transcriptome analysis of all two-component regulatory system mutants of Escherichia coli K-12.* Mol Microbiol, 2002. **46**(1): p. 281-91.

33. Hagiwara, D., et al., *Genome-wide analyses revealing a signaling network of the RcsC-YojN-RcsB phosphorelay system in Escherichia coli.* J Bacteriol, 2003. **185**(19): p. 5735-46.

34. Pruss, B.M., et al., *A complex transcription network controls the early stages of biofilm development by Escherichia coli.* J Bacteriol, 2006. **188**(11): p. 3731-9.

35. Hengge-Aronis, R., *Recent insights into the general stress response regulatory network in Escherichia coli.* J Mol Microbiol Biotechnol, 2002. **4**(3): p. 341-6.

36. Zhou, L., et al., *Phenotype microarray analysis of Escherichia coli K-12 mutants with deletions of all two-component systems.* J Bacteriol, 2003. **185**(16): p. 4956-72.

37. Mitrophanov, A.Y. and E.A. Groisman, *Signal integration in bacterial two-component regulatory systems.* Genes Dev, 2008. **22**(19): p. 2601-11.

38. Laub, M.T. and M. Goulian, *Specificity in two-component signal transduction pathways.* Annu Rev Genet, 2007. **41**: p. 121-45.

39. Siryaporn, A. and M. Goulian, *Cross-talk suppression between the CpxA-CpxR and EnvZ-OmpR two-component systems in E. coli.* Mol Microbiol, 2008. **70**(2): p. 494-506.

40. Groban, E.S., et al., *Kinetic buffering of cross talk between bacterial two-component sensors.* J Mol Biol, 2009. **390**(3): p. 380-93.

41. Kato, A., T. Latifi, and E.A. Groisman, *Closing the loop: the PmrA/PmrB two-component system negatively controls expression of its posttranscriptional activator PmrD.* Proc Natl Acad Sci U S A, 2003. **100**(8): p. 4706-11.

42. Kato, A. and E.A. Groisman, *Connecting two-component regulatory systems by a protein that protects a response regulator from dephosphorylation by its cognate sensor.* Genes Dev, 2004. **18**(18): p. 2302-13.

43. Hagiwara, D., et al., *Genome-Wide Analyses Revealing a Signaling Network of the RcsC-YojN-RcsB Phosphorelay System in Escherichia coli.* J. Bacteriol. %R 10.1128/JB.185.19.5735-5746.2003, 2003. **185**(19): p. 5735-5746.

44. Batchelor, E., et al., *The Escherichia coli CpxA-CpxR Envelope Stress Response System Regulates Expression of the Porins OmpF and OmpC.* J. Bacteriol. %R 10.1128/JB.187.16.5723-5731.2005, 2005. **187**(16): p. 5723-5731.

45. Price, N.L. and T.L. Raivio, *Characterization of the Cpx Regulon in Escherichia coli Strain MC4100.* J. Bacteriol. %R 10.1128/JB.00798-08, 2009. **191**(6): p. 1798-1815.

46. Alves, R. and M.A. Savageau, *Comparative analysis of prototype two-component systems with either bifunctional or monofunctional sensors: differences in molecular structure and physiological function.* Mol Microbiol, 2003. **48**(1): p. 25-51.

47. Yamamoto, K., et al., *Functional characterization in vitro of all two-component signal transduction systems from Escherichia coli.* J Biol Chem, 2005. **280**(2): p. 1448-56.

48. Skerker, J.M., et al., *Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis.* PLoS Biol, 2005. **3**(10): p. e334.

49. Skerker, J.M., et al., *Rewiring the specificity of two-component signal transduction systems.* Cell, 2008. **133**(6): p. 1043-54.

50. Jin, T. and M. Inouye, *Ligand binding to the receptor domain regulates the ratio of kinase to phosphatase activities of the signaling domain of the hybrid Escherichia coli transmembrane receptor, Taz1.* J Mol Biol, 1993. **232**(2): p. 484-92.

51. Parkinson, J.S. and E.C. Kofoid, *Communication modules in bacterial signaling proteins.* Annu Rev Genet, 1992. **26**: p. 71-112.

52. Kofoid, E.C. and J.S. Parkinson, *Transmitter and receiver modules in bacterial signaling proteins.* Proc Natl Acad Sci U S A, 1988. **85**(14): p. 4981-5.

53. Fabret, C., V.A. Feher, and J.A. Hoch, *Two-component signal transduction in Bacillus subtilis: how one organism sees its world.* J Bacteriol, 1999. **181**(7): p. 1975-83.

54. Jiang, M., et al., *Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis.* Mol Microbiol, 2000. **38**(3): p. 535-42.

55. Mattison, K. and L.J. Kenney, *Phosphorylation alters the interaction of the response regulator OmpR with its sensor kinase EnvZ.* J Biol Chem, 2002. **277**(13): p. 11143-8.

56. Yoshida, T., S. Cai, and M. Inouye, *Interaction of EnvZ, a sensory histidine kinase, with phosphorylated OmpR, the cognate response regulator.* Mol Microbiol, 2002. **46**(5): p. 1283-94.

57. Baumgartner, J.W., et al., *Transmembrane signalling by a hybrid protein: communication from the domain of chemoreceptor Trg that recognizes sugar-*

binding proteins to the kinase/phosphatase domain of osmosensor EnvZ. J Bacteriol, 1994. **176**(4): p. 1157-63.

58. Dwyer, M.A., L.L. Looger, and H.W. Hellinga, *Computational design of a Zn2+ receptor that controls bacterial gene expression.* Proc Natl Acad Sci U S A, 2003. **100**(20): p. 11255-60.

59. Levskaya, A., et al., *Synthetic biology: engineering Escherichia coli to see light.* Nature, 2005. **438**(7067): p. 441-2.

60. Looger, L.L., et al., *Computational design of receptor and sensor proteins with novel functions.* Nature, 2003. **423**(6936): p. 185-90.

61. Utsumi, R., et al., *Activation of bacterial porin gene expression by a chimeric signal transducer in response to aspartate.* Science, 1989. **245**(4923): p. 1246-9.

62. Ward, S.M., et al., *A NarX-Tar chimera mediates repellent chemotaxis to nitrate and nitrite.* Mol Microbiol, 2002. **44**(3): p. 709-19.

63. Tomasz, A., *The mechanism of the irreversible antimicrobial effects of penicillins: how the beta-lactam antibiotics kill and lyse bacteria.* Annu Rev Microbiol, 1979. **33**: p. 113-37.

64. Dombek, K.M. and L.O. Ingram, *Effects of ethanol on the Escherichia coli plasma membrane.* J. Bacteriol., 1984. **157**(1): p. 233-239.

65. Stancik, L.M., et al., *pH-Dependent Expression of Periplasmic Proteins and Amino Acid Catabolism in Escherichia coli.* J. Bacteriol. %R 10.1128/JB.184.15.4246-4258.2002, 2002. **184**(15): p. 4246-4258.

66. Jurs, P.C., G.A. Bakken, and H.E. McClelland, *Computational methods for the analysis of chemical sensor array data from volatile analytes.* Chem Rev, 2000. **100**(7): p. 2649-78.

67. Dombek, K.M. and L.O. Ingram, *Effects of ethanol on the Escherichia coli plasma membrane.* J Bacteriol, 1984. **157**(1): p. 233-9.

68. Ebel, W. and J.E. Trempy, *Escherichia coli RcsA, a positive activator of colanic acid capsular polysaccharide synthesis, functions To activate its own expression.* J Bacteriol, 1999. **181**(2): p. 577-84.

69. De Wulf, P., et al., *Genome-wide profiling of promoter recognition by the two-component response regulator CpxR-P in Escherichia coli.* J Biol Chem, 2002. **277**(29): p. 26652-61.

70. Shin, S. and C. Park, *Modulation of flagellar expression in Escherichia coli by acetyl phosphate and the osmoregulator OmpR.* J Bacteriol, 1995. **177**(16): p. 4696-702.

71. Setty, Y., et al., *Detailed map of a cis-regulatory input function.* Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7702-7.

72. Kalir, S. and U. Alon, *Using a quantitative blueprint to reprogram the dynamics of the flagella gene network.* Cell, 2004. **117**(6): p. 713-20.

73. Alm, E., K. Huang, and A. Arkin, *The evolution of two-component systems in bacteria reveals different strategies for niche adaptation.* PLoS Comput Biol, 2006. **2**(11): p. e143.

74. Perez, J.C. and E.A. Groisman, *Evolution of transcriptional regulatory circuits in bacteria.* Cell, 2009. **138**(2): p. 233-44.

75. Garbe, T.R., M. Kobayashi, and H. Yukawa, *Indole-inducible proteins in bacteria suggest membrane and oxidant toxicity.* Arch Microbiol, 2000. **173**(1): p. 78-82.

76. Kim, I.S., J.H. Shim, and Y.T. Suh, *Changes in membrane fluidity and fatty acid composition of Pseudomonas putida CN-T19 in response to toluene.* Biosci Biotechnol Biochem, 2002. **66**(9): p. 1945-50.

77. Sinensky, M., *Homeoviscous adaptation--a homeostatic process that regulates the viscosity of membrane lipids in Escherichia coli.* Proc Natl Acad Sci U S A, 1974. **71**(2): p. 522-5.

78. Zilberstein, D., et al., *Escherichia coli intracellular pH, membrane potential, and cell growth.* J Bacteriol, 1984. **158**(1): p. 246-52.

79. Farr, S.B., D. Touati, and T. Kogoma, *Effects of oxygen stress on membrane functions in Escherichia coli: role of HPI catalase.* J Bacteriol, 1988. **170**(4): p. 1837-42.

80. Ingram, L.O., *Adaptation of membrane lipids to alcohols.* J Bacteriol, 1976. **125**(2): p. 670-8.

Appendix A

Engineering temperature sensors in *E. coli*

A.1 Introduction

This work was performed over the course of 6 months leading up to November of 2005 as part of MIT's iGEM competition. Substantial work was contributed by Matt Eames in the Kortemme lab, as well as by two local high school students: Xiaoyan liu (Apple) and Nessa Ramos. We presented our findings at the iGEM meeting, and were awarded 'Best Part'.

Our goal was to engineer cells that would change color as a function of temperature. We envisioned spreading a lawn of engineered bacteria onto a plate, exposing the plate to a temperature gradient, and imaging the generated color.

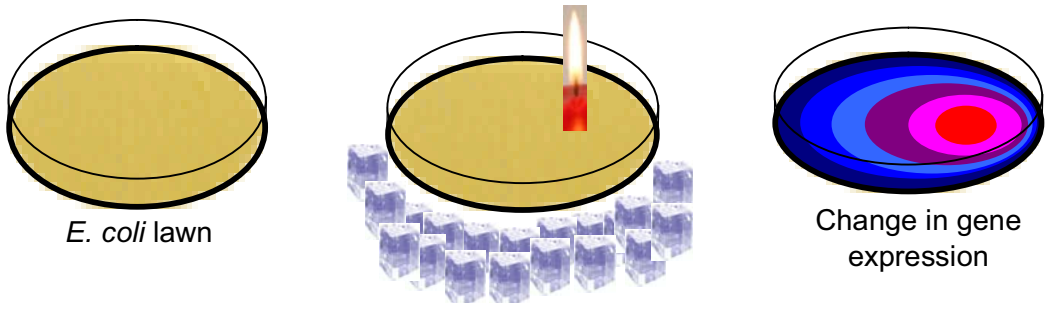


Figure A.1: Engineering goal

To do this, we thought that we could borrow from *E. coli's* natural response to hot and cold by controlling expression of fluorescent proteins with promoters unregulated during temperature stress.

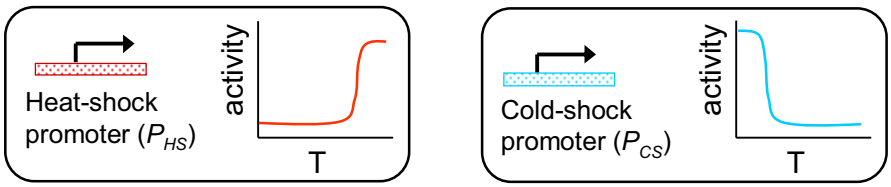


Figure A.2: Theoretical temperature sensors

A.2 Temperature sensitive promoter screening.

We examined a number of microarray papers that measured post-temperature-stress transcript levels, and picked a set of candidate promoters. They are as follows:

Table A.1: Candidate heat and cold shock promoters

Heat Shock	Cold Shock
clpB	*glpBC*
groEL	*ompT*
htpG	*nupG*
dnaK-p1	*katG*
dnaK-p2	*hypB*
dnaJ	*hybB*
grpE	*ansB*
phoBR	cspA
lon	cspA*
htpX	
ibpA	
*ibpA**	
ibpB	
yedU	
hslUV	
yccV	

* mRNA truncation

We constructed a series of high copy plasmids that each had one promoter driving GPF. We measured the transfer function for each (Output vs. Temperature) by performing the following series of experiments.

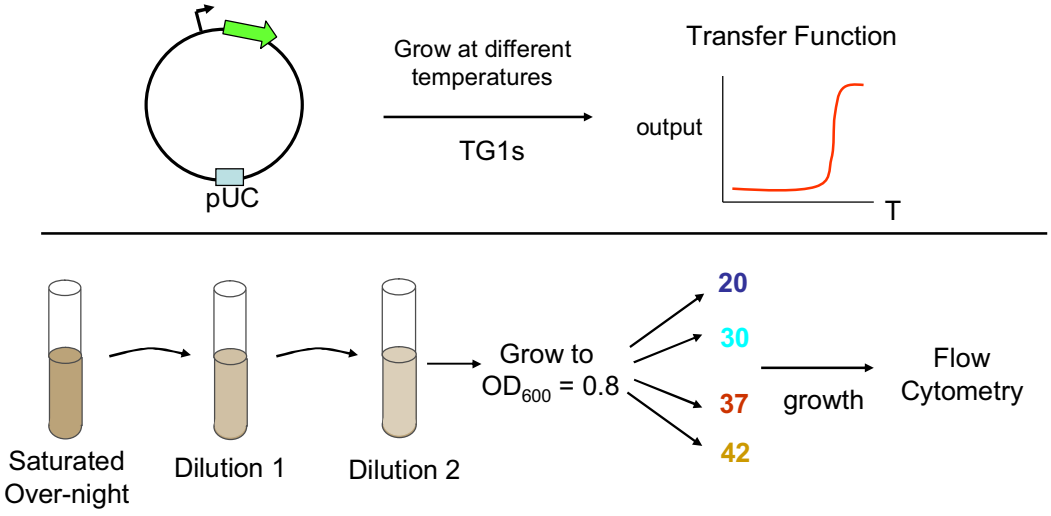


Figure A.3: Assay

By performing this for each promoter we were able to tabulate parameters such as on state, off state, switch point and gain. Examples of a few are given below.

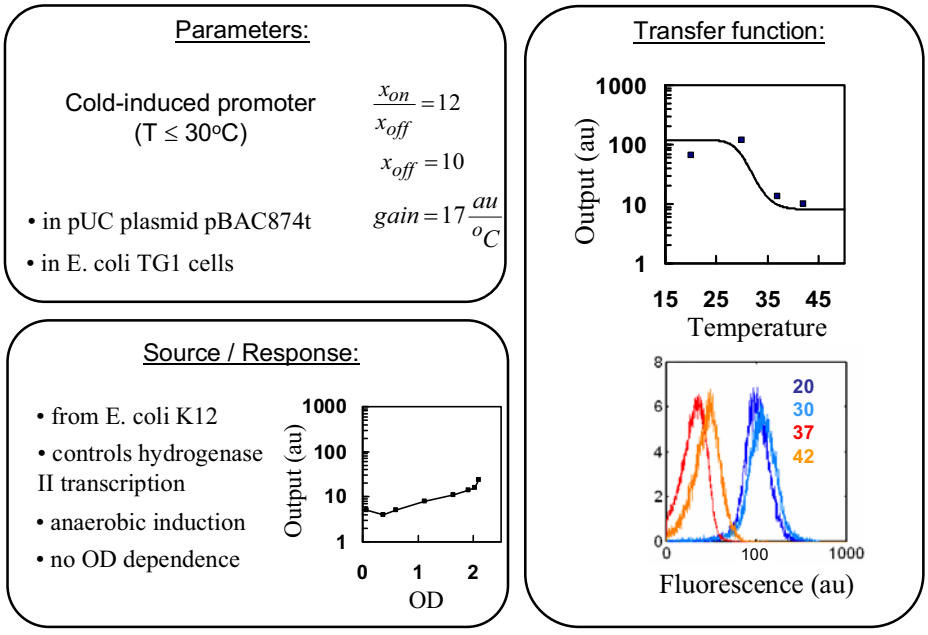


Figure A.4: HybB cold sensor

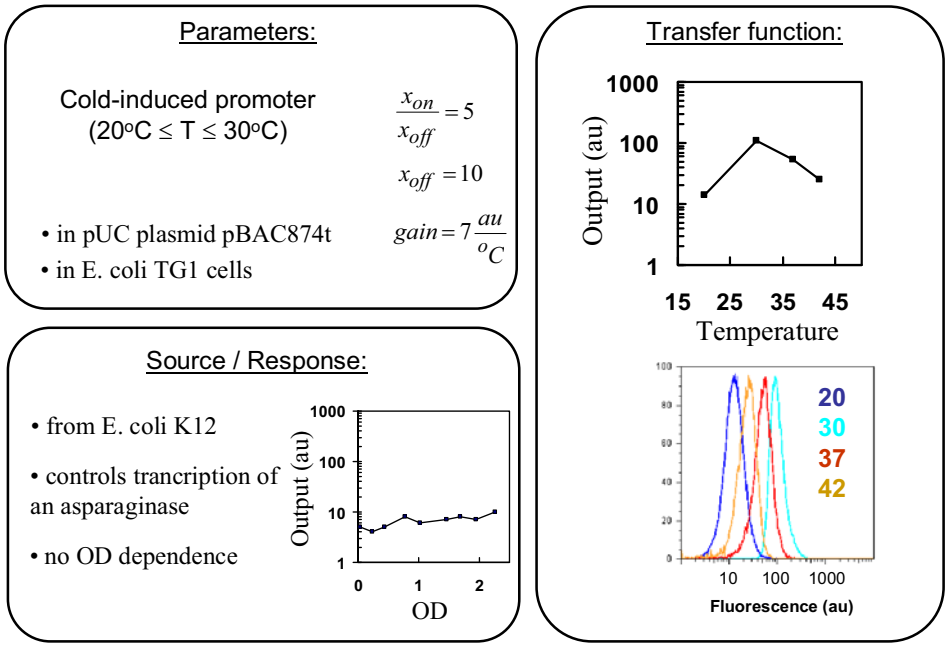


Figure A.5: AnsB cold sensor

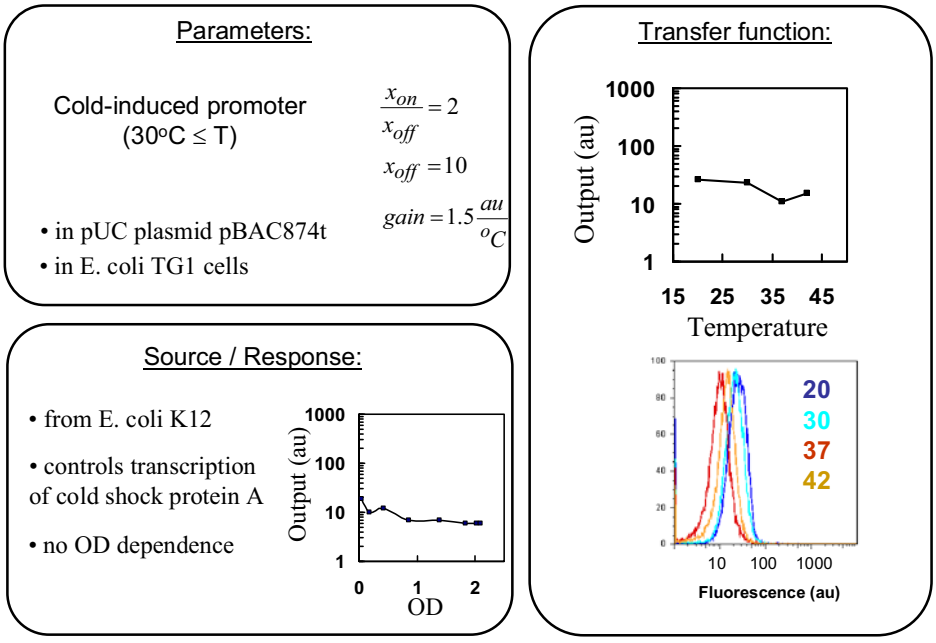


Figure A.6: CspA cold sensor

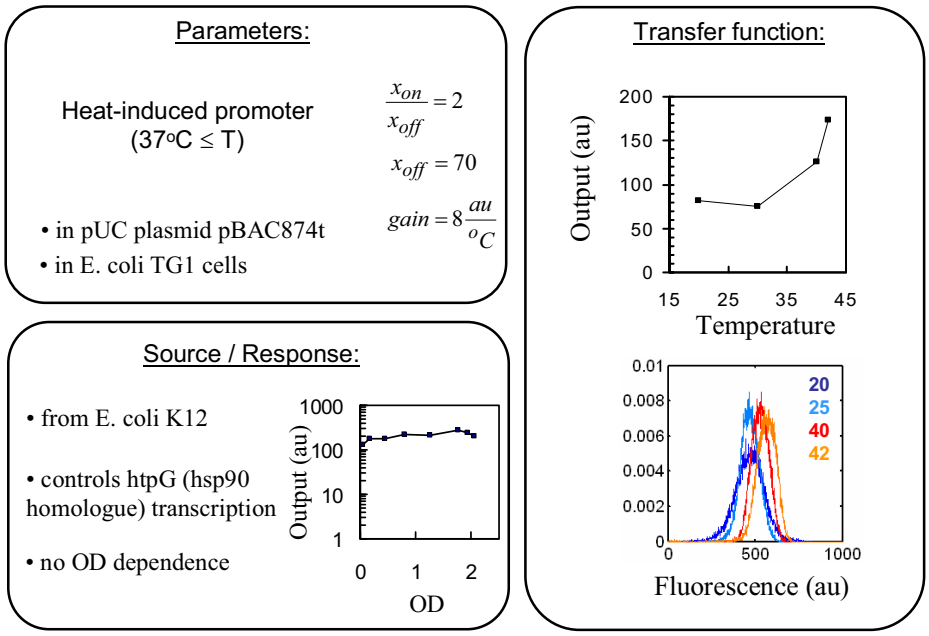


Figure A.7: HtpG hot sensor

Of the cold sensors, hybB performed the best, but the best heat sensor (htpG) was not good at all. So we decided to see if we could turn a cold sensor into a hot sensor.

A.3 Hot sensor engineering

To do this, we decided to invert the output of the best cold sensor (hybB). The design for this is shown below.

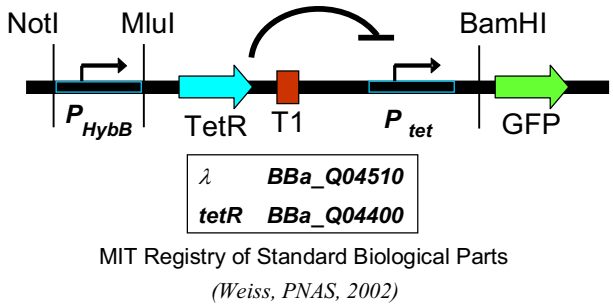


Figure A.8: Inverted htpG design

When we built both versions of this, we were disappointed to find that the circuit appeared to have no temperature dependence. Below is the flow cytometry data of the tetR based circuit.

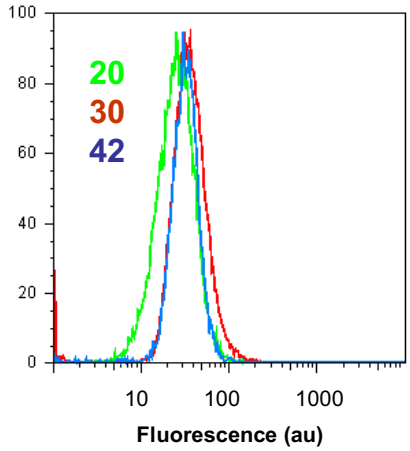


Figure A.9: Temperature response of version 1 of hot-sensing circuit

There were a few scenarios that could explain this failure. One is that the output from the hybB promoter (which is the input to the inverter) is too high (so that even when it is OFF it is ON enough to flip the inverter). The other option is that the hybB promoter never turns ON enough to flip the inverter. To test this, we added aTc, which binds to TetR and prevents it from binding well to DNA. Thus if the OFF state of hybB is too strong (and thus making too much TetR) this should fix the problem.

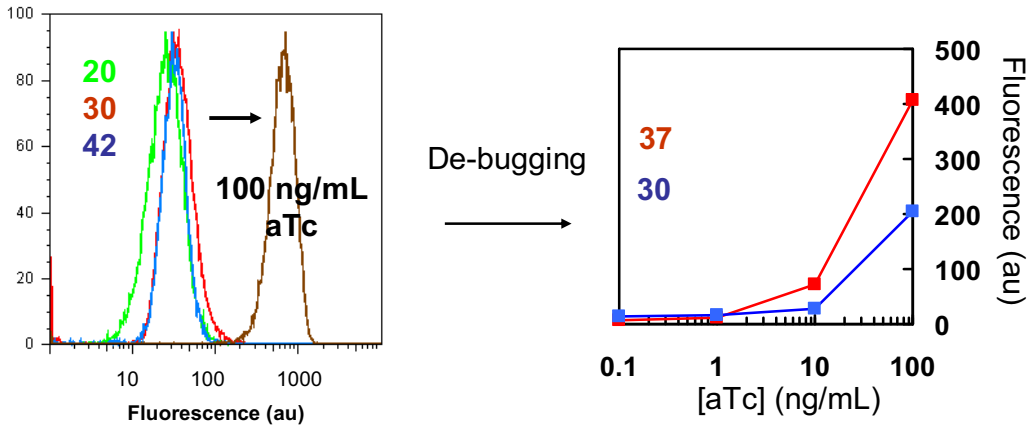


Figure A.10: Debugging the hot sensor

From this data, we can confirm that too much TetR is made, regardless of the temperature. In order to fix the circuit without having to add aTc, we made several variants of the circuit that had weaker ribosome binding sites in front of TetR.

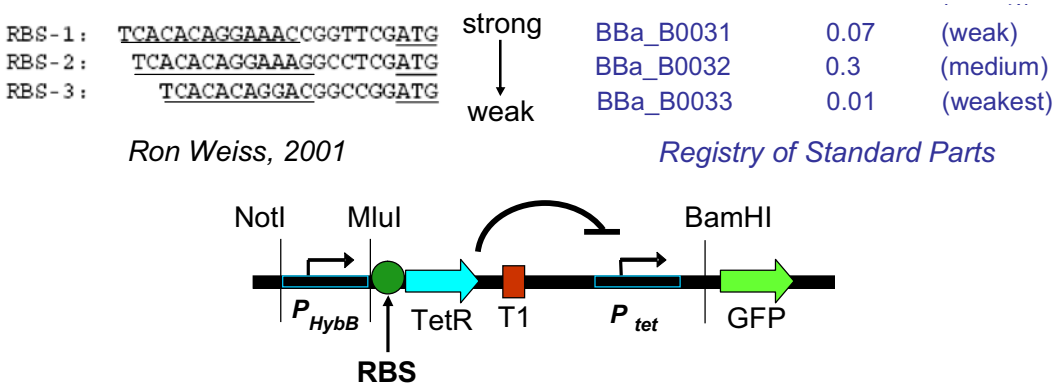


Figure A.11: Modifying the RBS of TetR

The circuit based on RBS-2 was successful in allowing for temperature dependent inversion. In order to test this version on a plate, as our original design had been envisioned, we poured tiny agar plates into the tubes for a 96 well PCR block. Onto each little agar slab we put cells either containing hybB, or containing the engineered temperature heat sensor. We then set a temperature gradient on the cycler, and let the

cells grow for 6 hours. On the top row of the figure below, is an image of the heat sensing cells. On the bottom is flow cytometry of cells scraped from the wells.

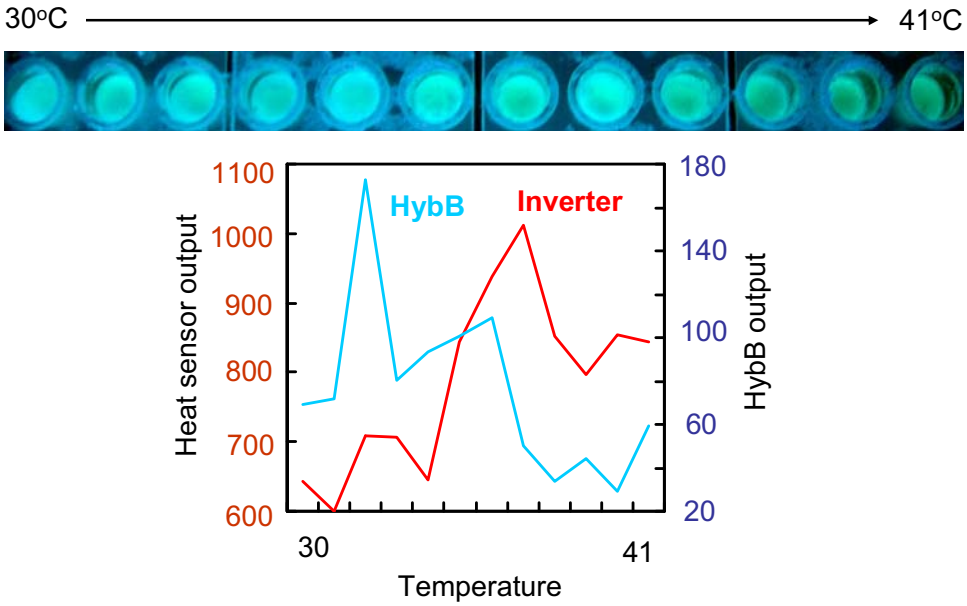


Figure A.12: Temperature sensing cells

A.4 Concluding thoughts

What is particularly interesting from this series of experiments, is that the best cold sensor was a promoter that had nothing to do with cold sensing. In fact, those promoters that we tested that regulated genes involved in the hot and cold response, were very poor temperature sensors. This is perhaps not surprising given our long assay time, and the fast feedback involved in E. coli's natural response to temperature changes.

Moreover, it speaks to the idea of being able to screen promoters, based on un-biased microarray data, for desired function, and should be broadly applicable to other engineering problems.

Appendix B
A timing circuit in *E. coli*

A1. Introduction

We set about to develop a tunable time delay circuit in *E. coli* derived from the *spv* system in *Salmonella*. Through exponential growth the output of the circuit is low. As the cells reach a specific density entering stationary phase, the circuit output increases rapidly. Linking the output of this circuit module to other modules in synthetic programs would give us dynamic control over cellular processes. Certain applications necessitate circuits that turn on earlier or later than defined by this circuit. To satisfy this need we wanted to generate a circuit library consisting of members that turn on over a range of different cell densities. Though this circuit did not ultimately meet our design goal, it could be applied in other settings.

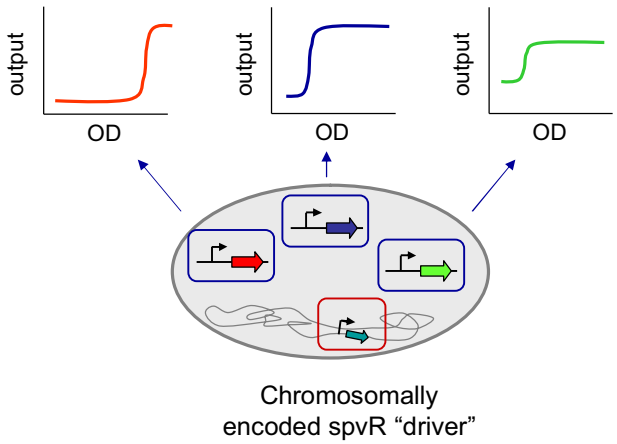


Figure B.1: A modular timer, capable of controlling multiple events

The *spv* circuit consists of a positive feedback loop embedded in a cascade. Specifically, the transcriptional activator SpvR activates both its own promoter, P_{spvR}, and another promoter, P_{spvA}. Upstream of the RNA polymerase binding site on P_{spvR} is a single spvR-dimer binding site necessary for circuit behavior. P_{spvA} has two SpvR-dimer operators, one upstream of and one overlapping with the RNA polymerase binding site. Binding of spvR to the upstream site recruits spvR to the downstream site. Binding of spvR to this second site is essential for RNA polymerase binding. Thus the induction of the system occurs as follows: a poorly understood signal increases the activity of P_{spvR}. This leads to an increase in cellular levels of spvR which is amplified by the positive feedback loop. This increase leads to occupancy of the upstream operator of P_{spvA} which leads in turn to binding at the downstream site. Finally, RNA polymerase binds the promoter and transcription is initiated. The timing of RNA polymerase binding can be altered by changing the binding constant of spvR to its operators on P_{spvA}.

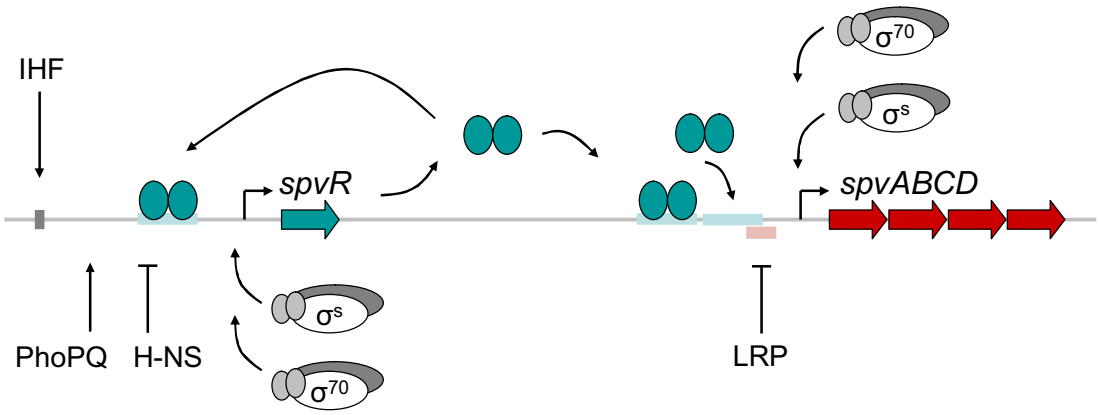


Figure B.2: Architecture of the wild type *spv* circuit

We designed our system to be both modular and capable of controlling the timing of multiple sequential events. To do this we have divided our circuit into two parts. The first is knocked into the genome at the *intC* locus and contains spvR, its endogenous promoter and ribosome binding site. The second is on a medium copy p15A plasmid that contains P_{spvA} driving GFP. This design has two advantages. First, the circuit library can be constructed entirely on the plasmid. Thus once the library has been constructed and characterized, incorporation into a new system requires only a knock in and the use of a desired library member promoter. Second, in any given system multiple different library members can be included to induce distinct processes at different times.

A2. Wild type circuit behavior

The behavior of the circuit with the driver knocked into the genome and the rest on a plasmid is shown below. The SpvR driver is knocked in the the intC locus of DH10B cells. The spvA promoter drives GFP, and the reporter is on a p15A plasmid. The circuit output is bimodal, the on state becoming more populated over cell growth

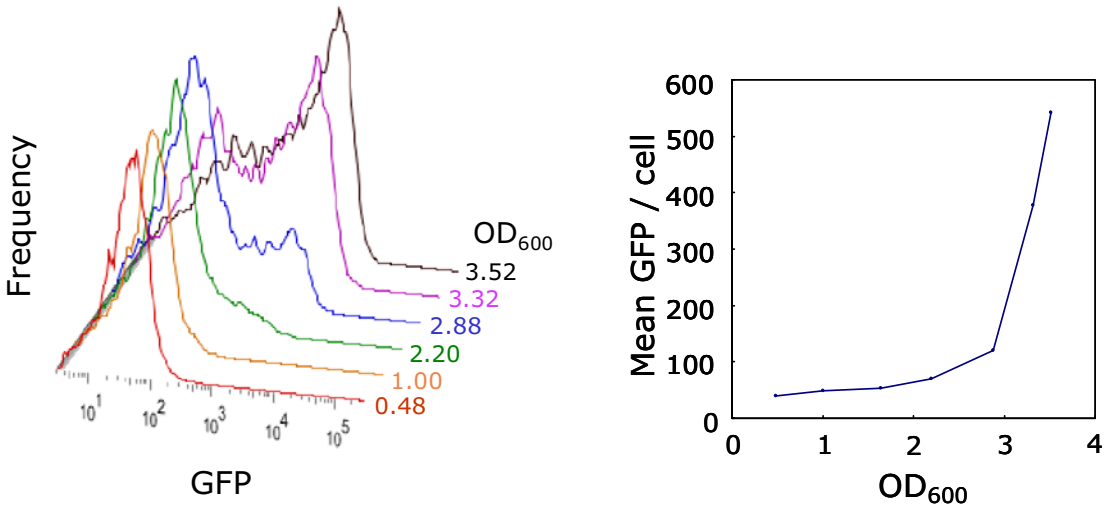


Figure B.3: Circuit output increases over bacterial growth

The switch point is a function of the growth medium. This is not surprising given that the OD at which cells enter stationary phase depends on growth media. Notice how sharply the circuit turns on and how tightly it is off before turning on. This is probably a function of both the positive feedback at the first promoter, and the cooperative binding of transcription factors at the second promoter.

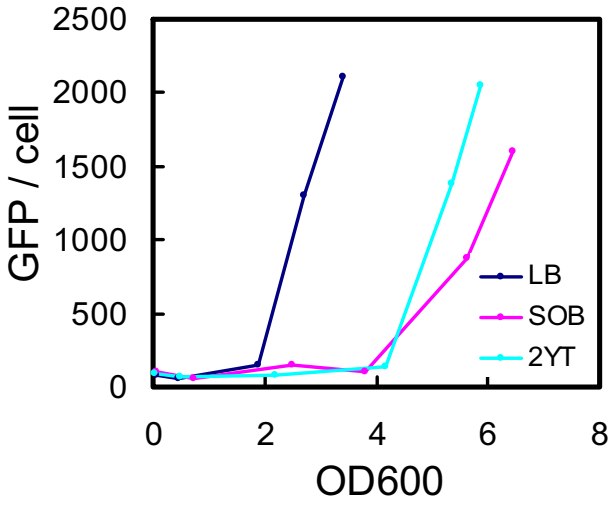


Figure B.4: Media dependence of switch point.

A3. Library construction

Two strategies were used generate the P_{spvA} promoter library. The first is targeted mutagenesis of the spvR operators. We expect that mutation in the upstream site will alter the concentration of cellular spvR necessary for occupancy without changing the ultimate promoter activity. (As this activity depends on the downstream site and RNA polymerase binding site neither of which are altered) A decrease in affinity of spvR for this operator should lead to a delay in circuit induction; an increase in affinity should lead to an earlier induction. Mutations in the second site should also alter circuit timing and may affect promoter activity as well. The second strategy was a random mutagenesis of the entire promoter. Here, though any given mutation may not hit an operator, other promoter properties might change. For example DNA supercoiling could indirectly affect the ability for spvR to bind; sigma factor binding strength, and propensity for open complex formation could affect the timing of RNA polymerase binding and its ability to begin transcription. This turned out not to work very well because the stretch of DNA was so short we got very low frequency of mutations.

To characterize the library we measured the activity of the promoter GFP reporter plasmids described above in knock in cells using 96-well format flow cytometry. In order to measure the dynamics of circuit response of this sub-library we took multiple time points over the course of cell growth, with increased frequency in late exponential growth and early stationary phase in order to finely distinguish the densities at which different mutants switch on.

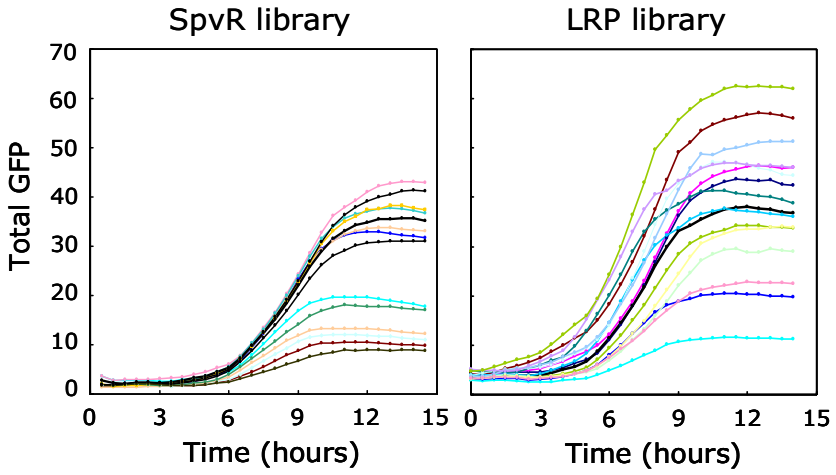


Figure B.5: Bulk fluorimetry screen for circuit mutants

We then moved a few interesting mutants to cytometry. We noticed and tabulated a few interesting characteristics of these mutants.

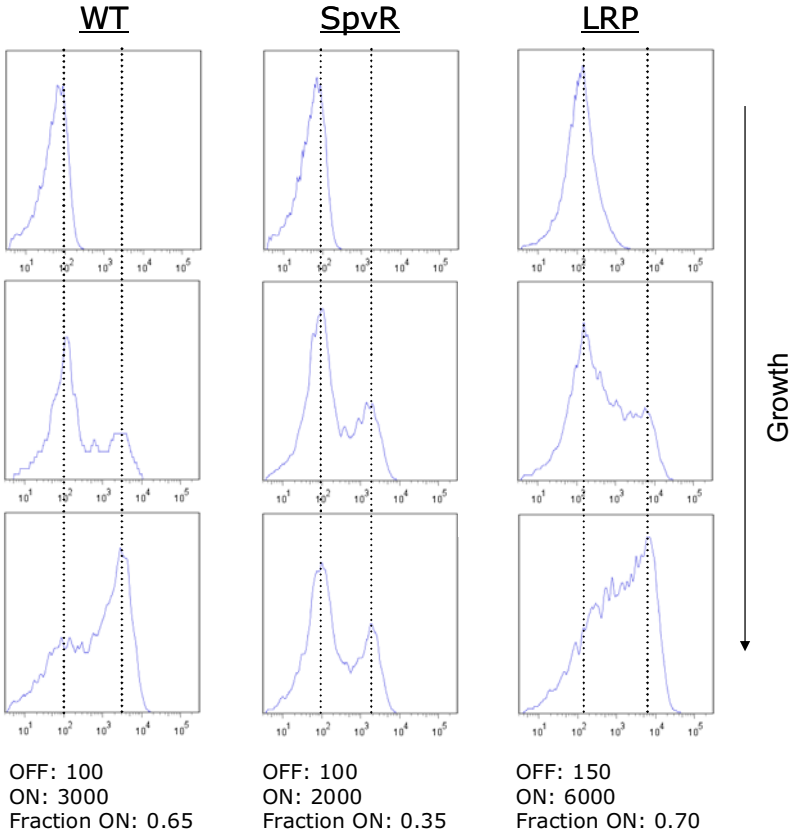


Figure B.6: Flow cytometry of circuit mutants

Mutants in the SpvR binding site show delayed induction and reach a lower fractional occupancy of the on state, while maintaining the same off state magnitude and slightly lower off state compared to the wild type circuit. Mutants in the LRP binding site have a higher off state, a significantly higher on state and are noisier.

A4. Concluding Thoughts

This circuit did not meet our design goal because we could not generate mutants that switched early. This is presumably because the inputs to both promoters are sigma S (the stationary phase alternate sigma factor). Two main thoughts come to mind:

(1) We could try to fix the problem.

Some possible ways of trying this: Stronger SpvR RBS; Mutations in the H-NS 'binding site'; Additional constitutive promoter upstream of spvR; Replacement of SpvR promoter with another promoter that is growth phase dependent but σs independent (there are microarray papers on this)

(2) Use it anyways.

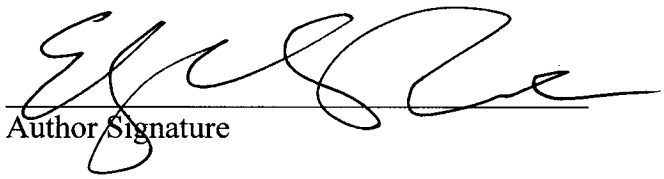
Some application may necessitate expression only when entering stationary phase. And this is a pretty nice performer, pretty tightly off, turns on sharply, high on state (which you could perhaps make even stronger) maybe some day someone will use it =)

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

4/28/10.

Date