NORTHWESTERN UNIVERSITY

Development of a Generalizable Strategy for Converting Metabolite Binding Proteins into Metabolite Responsive Transcription Factors

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Interdisciplinary Biological Sciences Program

By

Andrew K. D. Younger

EVANSTON, IL

June 2017

Abstract

Development of a Generalizable Strategy for Converting Metabolite Binding Proteins into Metabolite Responsive Transcription Factors

Andrew K. D. Younger

Metabolite biosensors are powerful tools for basic biological research, medical diagnostics, and biotechnological applications. However, a generalizable strategy for developing new metabolite biosensors when an existing sensor cannot be found in nature, is a persistent challenge. Furthermore, while transcription factor biosensors have the broadest range of applications, the pool of naturally occurring transcription factor biosensors is small. There is however, a wealth of metabolite binding proteins that can be found in nature, that bind many metabolites, but are unable to regulate transcription.

Therefore, the primary objective of my thesis was to develop a methodology by which a metabolite binding protein could be converted into a metabolite responsive transcription factor. Toward this goal, two hypothesis and literature driven approaches were investigated in order to determine the feasibly of converting a model metabolite binding protein into a metabolite responsive transcription factor. The split protein (SP) strategy ultimately resulted in a functional metabolite responsive transcription factor by fusing the BCR-ABL1 zinc finger DNA-binding domain (ZFP) internally to maltose binding protein (MBP) at amino acid 316R. This initial demonstration validated the feasibility of the idea that a metabolite binding protein can be converted into a metabolite responsive transcription factor.

Next, in order to investigate the generalizability of the SP conversion strategy, the biosensor engineering by random domain insertion (BERDI) method was developed to construct and test all possible insertions of the ZFP into MBP. Because the original biosensor developed by the SP strategy used a previously published split of MBP (316R) that resulted in an enzymatic biosensor, the BERDI method was developed so that reliance on this type of information would not be required for future biosensor development. In addition, the BERDI method is not specific to MBP as the metabolite binder, therefore the method can be used on any metabolite binding protein enabling this method to be as generalizable as possible. Using the BERDI method, three new splits of MBP were found to generate maltose responsive biosensor illustrating not only that the method can generate novel biosensors with no reliance of previously published information, but also that one metabolite binding protein can result in several functional biosensors.

Finally, to apply the BERDI method to metabolite binding enzyme for a biotechnologically relevant metabolite, farnesyl pyrophosphate (FPP), a screening strain for the inducible overproduction of FPP was developed. This strain enables the same cell to inducible over produce FPP such that cells containing potential FPP responsive biosensors experience a change in internal FPP levels to enable screening for functional FPP responsive biosensors. This strain enables the BERDI method to be applied to FPP binding enzymes to continue the investigation into the generalizability of the SP biosensor conversion strategy.

Overall, this work demonstrates the feasibility of a potentially transformative technology and lays the ground work for future investigations and applications of the BERDI method for biosensor engineering and development.

Acknowledgements

This work would not have been possible without the help and support of a large group of people. First I would like to thank my thesis advisor Josh Leonard. From day one you have fostered an incredibly positive and intellectually stimulating environment. I looked forward to lab every day and will sorely miss our frequent interactions. What I am most proud of is that as I look forward to what is next, I feel completely scientifically and academically prepared for whatever comes next. The PhD is a training program and I feel I have gotten trained exceptionally well, and that is primarily due to your dedication to being an excellent thesis advisor.

I would also like to thank my thesis committee: John Marko, Neda Bagheri, Keith Tyo, and Yun Wang. One of the reasons I joined the Interdisciplinary Biological Sciences (IBiS) Program at Northwestern was the ability to have a diverse learning and advising experience that spans many departments and academic disciplines. I believe that is a major strength of the IBiS program, and my thesis committee is a perfect example of the diversity of expertise that cannot be found in other graduate programs. I have greatly appreciated all of your questions, ideas, and advice. Your continued support has always reassured me that I was on the right path both scientifically and professionally.

The Leonard lab has been a second home to me over the course of my PhD and my overwhelmingly positive experience had a lot to do with some amazing scientists and people. Nichole Daringer and Yishan Chuang, you both showed me the level of effort and dedication required to ultimately be successful in a PhD. Andy Scarpelli and Rachel Dudek, you both kept the lab fun and lively and thoroughly contributed to the fantastic lab culture that made coming into lab everyday enjoyable. Michelle Hung, you were my IBiS partner is crime and trailblazer for anything and everything IBiS related. You made my life so much easier on many levels. Our pranks

on Josh (and others) were always fun, and I knew I could always count on you for sound advice about anything. Kelly Schwarz, over the course of my PhD you have been my best friend. From our cooking challenges, to love of sports, fantasy football, March madness, and fantasy bachelor I could not have asked for a better friend and lab office mate. You too, made my life so much easier by providing an excellent example of what to do, especially as I was writing my thesis. Joe Muldoon and Patrick Donahue, you both always asked pertinent questions in lab meeting that I could never anticipate, and for that I am forever grateful as it made me think about my project from new angles. Taylor Dolberg and Devin Stranford, thank you for being good friends and desk mates. Peter Su, I could not have asked for a harder working or more motivated collaborator to take on my project. You have picked up everything quickly, added valuable insight, and took lead on parts of the project that I would have had no chance of accomplishing. I will miss our trips to get awesome Asian food, pokemon go adventures (double diglett and scyther porygon taking on the jaded grad student in the A wing), and discussion of all sports.

In addition to my lab mates, I have had the privilege of working with five fantastic undergraduates, that have deeply enriched my experience here at Northwestern. Neil Dalvie, Austin Rottinghaus, Shreya Udani, Andrea Shepard, and Praneet Polineni you have all been fantastic students and it has been my absolute pleasure to working beside you during my Phd.

I would also like to thank my parents and sister for their unwavering support of me during my Phd. I know it wasn't always easy to understand what it was that I was doing, but your unconditional love and support was critical for me to stay level and ultimately be successful in this endeavor. I love you all very much.

This thesis is dedicated to my best friend and love of my life Jeci. I can't say enough about how much you mean to me and how much you daily love and support has made this whole process

easier. Your dedication and support to both of our educations is something I deeply value, and having someone who understand the aches, pains, and successes of a PhD has helped me tremendously. You're the best, I love you.

Abbreviations:

aTc – Anhydrotetracycline
BERDI – biosensor engineering by random domain insertion
BFP – Blue fluorescent protein
Bla – TEM1 β-lactamase
CFP – Cyan fluorescent protein
DMSO – Dimethylsulfoxide
EMSA – electrophoretic mobility shift assay
FACS – Fluorescence activated cell sorting
FAEE – Fatty acid ethyl ester
FPP – Farnesyl pyrophosphate
FRET – Förster resonance energy transfer
GC-MS – Gas chromatography mass spectrometry
GFP – Green fluorescent protein
IPTG – Isopropyl β-D-1 thiogalatopyranoside
LB – Lysogeny broth
MBP – Maltose binding protein
MSE – Mean squared error
ORF – Open reading frame
PBS – phosphate buffered saline
PDB – Protein data bank
PLSR – Partial least squares regression
RCF – Relative centrifugal force
RBS – Ribosome binding site
SP – Split protein
SZF – Split zinc finger
TALE – Transcription activator-like effector
YFP – Yellow fluorescent protein
ZFP – Zinc finger protein

Table of Contents

List of Tables and Figures

Chapter 1. Introduction and Background

1.1 The need for and applications of metabolite biosensors

Cellular metabolism is complex; even the simplest of organisms have thousands of nodes connected by tens of thousands of edges. While the connections and the chemical transformations they describe are well known, the dynamics of metabolism are hard to observe in real time, on a single celled level. Bulk measurements of metabolites from whole populations of cells can be evaluated using analytic equipment such as Gas Chromatography Mass Spectrometry (GC-MS). However, this is relatively low throughput (~10^3 samples per day) and because the data represent a population average, information on cellular heterogeneity is lost[1]. By contrast, metabolite biosensors can provide information on the levels of metabolites at the single cell level in real time. Here, a metabolite biosensor will be defined as a protein[2-15] or RNA[16, 17] species that can interact with a metabolite and can transduce this interaction into an output signal. Here I will survey both the applications of metabolite biosensors in addition to the various types of biosensors that exist and how they were developed.

Metabolite biosensors can provide a window into the complex state of cellular metabolism through a variety of output modalities. The simplest use of this information is to build a better understanding of how a metabolite flows through its pathway by having the output modality be a measureable signal, or linking its output to such a signal[3, 8, 15]. Another use for metabolite biosensors is the screening of very large (>10^8) libraries of cells by fluorescence-activated cell sorting (FACS)[1, 4]. Finally, transcription factor biosensors can enable the processing of the metabolite signal into more advanced outputs such as controlling the expression of enzymes to modulate the flux of metabolites through a pathway of interest. Below, I will survey three distinct applications of metabolite biosensors, to illustrate their use as tools for basic biological research

and discovery, and then I will describe the potential to use biosensors for the development of highly productive cell systems.

1.1.1 Biosensors for fundamental research and discovery

Here I will describe two examples of how metabolite biosensors have been used to answer basic biological questions. Metabolite biosensors can provide new information on the dynamics of a variety of intermediate metabolites to deepen the understanding of these natural systems.

Nicotinamide adenine dinucleotide (NADH) is the most important molecule involved in many cellular redox reactions. In order to probe the spatiotemporal activity of NADH and its oxidized form NAD+, a fluorescent biosensor was made by the fusion of the NADH responsive repressor Rex from *Staphylococcus aureus*[18], and a circularly permutated fluorescent protein GFP T-Sapphire[19]. This fusion, named Peredox by the authors, increases green fluorescence when bound to NADH, but remains dim when bound to NAD+. Using a second, unaffected red fluorescent protein (mCherry), the ratio of green to red fluorescence could be quantified to gain insight into the levels of NADH/NAD+. The relative expression of the protein and the GFP output signal can be normalized by dividing this signal by the mCherry signal, this way only the change in GFP signal that is attributable to the NADH/NAD+ ratio is analyzed. By adding different ratios of lactate and pyruvate to the cell media, which get interconverted using a NADH/NAD+ intermediate, the intracellular ratio of NADH/NAD+ could be modulated to allow for the calibration of Peredox. As a demonstration of the utility of the Peredox sensor, the PI3K pathway was inhibited, and the sensor's fluorescence was monitored. Over the next hour, fluorescence of the sensor dropped, indicating a decrease in NADH levels associated with glycolytic inhibition, which is hypothesized to occur following PI3K blockade. Furthermore, this sensor was then used

to probe the native NADH/NAD+ ratios within several mammalian cell types with the ability to measure dynamic changes in the cell with a temporal resolution on the order of seconds. In addition to NADH, other important metabolic intermediates such as ATP, glucose, lactate, pyruvate, and amino acid sensors have been similarly developed and employed for single celled metabolic monitoring[8, 20-30]. These sensors have proved especially effective at distinguishing intracellular differences in concentrations by imaging highly compartmentalized eukaryotic cells. Furthermore, the data these biosensors provide can be analyzed on a living single cell level, in real time.

In a dramatically different example, the CdaR transcription factor biosensor, native to *Escherichia coli*[3], was utilized as a tool to evaluate the performance of several non-native enzymes in the glucarate metabolic pathway. CdaR is a transcriptional activator that regulates transcription of its cognate promoter in response to glucarate, galactarate, and glycerate [31]. Glucarate is important as it can be used as a renewable replacement for plastics such as nylon. Glucarate can be produced from myo-inositol, however the conversion of myo-inositol to glucuronate by the native myo-inositol oxygenase enzyme (MIOX) is extremely slow. The subsequent transformation of glucoronate into glucarate by the glucoronate dehydrogenase enzyme (Udh) is fast. Therefore, the ability to monitor glucarate production is vital to be able to find new enzymes that can take the place of the slow native MIOX enzyme. A glucarate reporter was assembled by placing a CdaR responsive promoter in front of Green Fluorescent Protein (GFP) and introduced into *E. coli.* Next, four MIOX orthologs were tested and compared. The authors found that the ortholog from *Mus musculus* performed better than the native enzyme, and the other three orthologs. Importantly, the fluorescent reporter for glucarate production trended with the mass spectrometry, when glucarate was measured directly[3].

In both examples, metabolite biosensors quickly and easily provided new insights into native metabolism, helped tackle new scientific questions, and do so in real time on a single cell level.

1.1.2 For biotechnology

1.1.2.1 Medical and diagnostic applications

Another area where metabolite biosensors have shown promise is in the field of biotechnology. From low cost, noninvasive medical diagnostic sensors to sensors that enable the high throughput screening of large genetic libraries, and transcription factor biosensors that can transduce their signal into complex genetic programing, metabolite biosensors have been very successful in driving these biotechnological applications forward.

One major advantage of biologically inspired biosensors for medical diagnostic purposes is the inherent high degree of specificity that the biological sensors can achieve. Zinc deficiency is a common problem in the developing world and leads to a wide range of medical problems, especially in children under five[32]. Blood serum testing is the state of the art; however, this process requires cold storage and transport along with electricity, neither of which can be guaranteed in the developing world or following natural disasters. A cell based sensor of Zinc was constructed utilizing *E. coli*'s natural zinc responsive system of promoters and repressors[10]. In this proof of concept, an initial purple pigment, violacein, was produced under low zinc levels, then above a threshold, the cells would switch from violacein to β-carotene, an orange pigment. The cells could then be lyophilized for long term storage, and transport, then be rehydrated in blood samples to provide a colorimetric readout. While the system suffered from leaky expression and hard to

control pigment thresholds, it is an example of a low-cost implementation of a metabolite biosensor that could address issues for developing world medical diagnostic tools.

Viral outbreaks such as the Zika virus transmitted quickly by mosquitoes can rapidly become enormous global health challenges in detection and containment, especially in heavily populated areas. Traditionally, detection and diagnostic tools lag the spread of the disease because the development of these tools is often slow for novel pathogens. To attempt to address this need, an RNA based reporter system was developed. Utilizing RNA toehold switches, a lacZ reporter mRNA was translationally-inactivated by an upstream hairpin formation in the basal state. In the presence of the trigger RNA from the species intended to be detected, the secondary structure can unfold and translation of the reporter enzyme can take place[33]. The trigger and hairpin sequence can be designed to respond to nearly any viral DNA or RNA species. In addition, this sensor mRNA can be lyophilized on paper with cell free translation machinery and stored at room temperature. The user then adds fluid to be tested, rehydrating the paper based system, and waits for a color change indicating whether the trigger DNA/RNA is found in the sample. As a proof of concept, the development of new sensors for Ebola, and Zika viruses was accomplished under a week, and such paper diagnostics are stable at room temperature for over a year[34, 35]. These sensors could differentiate between several stains of each virus, and importantly, not give false positives for similar viruses such as the dengue virus. For less than $50 and one week to develop, yielding diagnostic paper that costs less than $1 to produce, these RNA-based sensors could rapidly have an impact on global health problems and diagnostics.

Metabolite biosensors have also been applied to challenges related to disease states in the gastrointestinal track. Changes in the gastrointestinal microbiota can lead to or be a symptom of a diverse host of disease states[36]. However, techniques such as colonoscopies are invasive,

expensive, and disruptive to the patient. Therefore, a whole cell biosensor was developed with a genetic memory circuit to detect the presence of the metabolite cue anhydrotetracycline (aTc)[37]. A lacZ reporter is tripped on in the presence of aTc, in addition, the aTc also triggers a stable memory circuit that produces lacZ after the aTc has passed. This idea was validated in the gut of mice, where the mice were treated -/+ aTc and then *E. coli* containing the memory circuit were fed to the mice. Fecal samples were collected and analyzed for lacZ positive colonies. Cells containing the circuit that had been exposed to aTc in the gut were still producing the lacZ reporter a week after the stimulus was provided. The modularity of the system also enables the rapid switching of the input trigger to sense other biologically relevant cues (other than aTc). This technology could enable the noninvasive detection of many disease biomarkers by utilizing cell-based metabolite biosensors linked to long lasting memory circuits. Additionally, it is possible to envision a cell based therapy that not only provides a diagnostic readout in the gut but then activates a therapeutic program in response to the disease state signal.

1.1.2.2 Screening applications

Metabolite biosensors can sense metabolites that are hard to measure via traditional analytical chemistry techniques. Furthermore, metabolite biosensors can monitor metabolites on a single celled level, not bulk population averages. Therefore, these biosensors can provide a unique opportunity to screen large libraries of genetic variants for the presence of hard to detect metabolites in single cells in a timescale that's on the order of magnitude of days, not weeks or months. Screening is typically accomplished by linking the output of the metabolite biosensor to either an antibiotic resistance marker to give a growth advantage or a fluorescent signal that can be analyzed and sorted using fluorescence-activated cell sorting (FACS).

Branched chain amino acids are valuable products and precursors in the drug and herbicide industries, but due to the lack appropriate sensors, high producing strains were difficult to screen for. However, a transcriptional regulator Lrp from *Corynebacterium glutamicum* was found to activate transcription of its cognate promoter in response to branched chain amino acids or methionine. This system was constructed to express Yellow Fluorescent Protein (YFP) to allow for the screening of high producing cells[38]. To find mutants that would allow for the over production of branched chain amino acids, wild type *C. glutamicum* cells were chemically mutagenized and subjected to FACS screening. After several rounds of sorting and enriching, mutants were isolated that were producing up to 8 mM valine, 2 mM leucine and 1 mM isoleucine. Given that wild type levels of these amino acids are in the nM to μM range, this strategy greatly enriched for high producing mutants. Similar strategies that couple a fluorescent reporter followed by directed evolution have been implemented to (1) increase the responsiveness of an existing alkane biosensor by 5-fold[39]; (2) find new genomic targets for improved malonyl-CoA production[7]; (3) find mutants with increased Lysine synthesis[40]; and (4) find mutants with increased triacetic acid lactone production[41].

In addition to utilizing fluorescence as a screening method, resistance to an antibiotic, and therefore growth selection, can provide an additional level of selection by which large libraries can be sorted. 1-butanol is a potential fuel alternative to both gasoline and ethanol, however, high throughput screening of alcohols like 1-butanol is a challenge due to its high volatility. BmoR, from *Thauera butanivorans* was found to activate transcription of its cognate promoter P_{BMO} in response to 1-butanol, among other alcohols[4]. In this system, BmoR drove the expression of a tetA such that high producing strains would be more resistant to the antibiotic tetracycline. The heterologous pathway to 1-butanol was then subjected to directed evolution under the selective

pressure of tetracycline and new strains producing as much as 120-fold more than the initial strains were isolated. The strategy of coupling cell fitness to production of an important molecule is an effective way of screening for higher producing strains and a quick way to test how individual perturbations of the pathway effect product titers.

1.1.2.3 Dynamic transcriptional control

Metabolic engineering has benefitted from the use of metabolite biosensors to do more than just monitor metabolism or screen for high producing strains. Transcription factor biosensors can not only sense and report on levels of a metabolite, but also use this signal as an input to control the behavior of a system dynamically. The ability to dynamically control certain elements of a metabolic pathway has led to increased yields and titers over traditional static configurations where the expression of the pathway components is constitutively expressed[9, 11]. The advantage of dynamic control is that the expression of enzymes in the pathway can be increased and decreased in response to the changing level of the metabolite being sensed. This is crucial for increased titers because static overexpression often results in the imbalance of co-factors, toxic accumulation of pathway intermediates, decreased growth, and or unproductive byproduct formation[5, 9, 11, 12].

An elegant example of dynamic control was constructed using the FadR transcription factor that naturally binds fatty acids to balance the production of fatty acid ethyl ester (FAEE)[5]. FAEE is biologically produced fuel that is a drop-in replacement for petroleum-derived diesel. FAEE is formed by the condensation reaction of ethanol and fatty acyl-CoA by the atfA enzyme. When there are high amounts of ethanol and fatty acyl-CoA, there needs to be high levels of the atfA enzyme, however when those reactants drop, the amount of atfA can be repressed. In addition, an overabundance of ethanol is toxic, so the ethanol production must match the fatty acyl-CoA

production. Further, if there is an overabundance of fatty acyl-CoA, this product will be converted into unproductive side products. Therefore, FadR was a perfect candidate to control this system as it is a repressor that alleviates repression upon binding to fatty acid. When the levels of fatty acid are low, the conversion of fatty acid to fatty acyl-CoA is repressed, ethanol is produced, and atfA is expressed. However, when fatty acids are being produced, FadR releases repression and thus the fatty acids are converted into fatty acyl-CoA, ethanol is produced, and these products are condensed by atfA to form FAEE. If the system becomes unbalanced, it will be reflected in the fatty acid pool and FadR will respond accordingly. By using FadR to control the metabolic pathway and balance the formation of intermediates, FAEE titers were increased three-fold to within 28% of the theoretical maximum.

A different style of control involves separating the growth stage of the organism from the production stage. The advantage of this strategy is that in many cases, the production of some desired product often draws on resources that the cells would normally use to grow and divide, slowing both processes down. By segregating the growth from production phases, a cell can first use all its resources to grow to a high density, and then metabolism can be switched to a production stage after most cell growth has ceased. This process allows for a potentially better time-dependent allocation of resources to maximize yields in both growth and production phases. In a recent example, the conversion of glucose through glycolysis was inducibly interrupted to start production phase to shunt glucose-6-phosphate to a two-step reaction ending in the desired product, myo-inositol[42]. To increase titers, several side reactions from glycolysis were knocked out and Pfk-1, the enzyme that converts fructose-6-phosphate into fructose-1,6-bisphosphate in glycolysis, was modified to allow for inducible degradation of the enzyme. Therefore, when the cells grew, glycolysis functioned normally, and then upon the addition of a small molecule inducer,

the Pfk-1 enzyme was degraded, allowing much of the glucose-6-phosphate to be converted into myo-inositol. By varying the growth phase at which the degradation of Pfk-1 was induced, thereby shifting the point at which the culture shifted from growth phase to production phase, myo-inositol titers were doubled.

There are numerous other examples (reviewed here[9, 11, 43]) of transcription factor biosensors dynamically balancing pathway intermediates or controlling growth/production phase transitions resulting in the increased production of lycopene[44], malonyl-CoA[45], gluconate[46], amorphadiene[12], and isopropanol[47], compared to the traditional static overexpression/knockout systems. Through a variety of mechanism of action, metabolite biosensors have a diverse set of applications from basic research and discovery, to medical devices, and cells with modified metabolic processes. Furthermore, metabolite biosensors have provided live single celled analysis of intracellular dynamics of a particular metabolite. In many cases, the utilization of a metabolite biosensor allowed researchers to answer questions or address needs that would have otherwise been infeasible.

1.2 Current types of biosensors and their method of development

Given the wide range of biosensor applications that exist, there also exists a diversity of methods by which biosensors are developed and act. Different applications require different operating parameters and output modalities which gives rise to the many different classes of biosensors that have been described. In addition, the generalizability varies greatly with the type of development method, i.e. how simple is it to develop a new biosensor in a class of sensors. Therefore, the following sections are an overview of these different classes of biosensors and a commentary on their generalizability.

1.2.1 Fluorescent, antibiotic resistant, and FRET-based sensors.

1.2.1.1 Fluorescent biosensors

Fluorescent biosensors are a commonly used sensor architecture that is useful for cell staining and imaging in addition to FACS-based screens. In this class of sensor, a florescent protein is fused directly to the sensing domain. This class does not include transcription factors that regulate the expression of fluorescent reporter genes, which will be covered in a later section.

In all cases where a fluorescent protein is being fused to a sensing domain, the way the fusion occurs is the differentiating feature. The most commonly used method is to circularly permute the florescent protein and the sensing domain. This process involves inserting the permutated fluorescent protein, with its original N and C termini fused by a short (2-8 amino acid) flexible linker into a loop or unstructured region of the sensing domain. Circularly permuted GFP or YFP has been inserted into (1) a NADH binding protein Rex to generate an NADH/NAD+ sensor[19], (2) a hydrogen peroxide binding protein OxyR to generate a hydrogen peroxide sensor for mammalian live cell imaging[48], (3) a ATP binding protein GlnK1 to provide a sensors for ATP:ADP levels[49], (4) calmodulin/M13 to generate a Ca^{2+} biosensor [50], (5) three separate cGMP binding fragments of PKG1 to detect changes in cGMP:cAMP ratios[51]. However, in all cases except the cGMP sensor, the chimeric protein underwent directed evolution by error prone PCR or site directed mutagenesis after the initial fusion to improve or optimize performance. Typically, several mutations were required before the chimera functioned in an acceptable range for the desired use.

A distinct strategy has been described and validated by generating a maltose biosensor. Instead of permuting and inserting the fluorescent protein into the sensing domain, in this example, GFP was split into two, non-fluorescent fragments, and fused to the N and C termini of maltose

binding protein (MBP)[52]. Upon binding maltose, MBP experiences a major change in conformation, which results in the reorientation of the two fragments of GFP in a closer proximity to allow for the reconstitution of the fluorescent protein. While different in construction from the circularly permutated fluorescent biosensors, the resulting biosensor is functionally very similar.

Fusing metabolite responsive domains with fluorescent proteins has proven to be a successful strategy for generating biosensors. However, the successes of these biosensors are due to how flexible the coding regions of the fluorescent proteins are, and how conformationally sensitive their chromophores are, not necessarily that the metabolite sensing domains fused to the fluorescent proteins are well suited for this application. Additionally, the common reliance on directed evolution following fusion, highlights the challenges associated with developing this type of a sensor. Furthermore, in the case of MBP, the requirement of the large conformational change, makes this strategy not necessarily straightforward or broadly generalizable.

1.2.1.2 Antibiotic resistant biosensors

Instead of a fluorescence output, another commonly used output modality is the ability to grow in the presence of an antibiotic. The TEM-1 β-lactamase enzyme (bla) confers resistance to ampicillin and penicillin, two commonly used gram-negative antibiotics, and this resistance is the output modality for this class of biosensors.

Similar to the fluorescent biosensors, a metabolite sensing domain was fused to bla by circularly permuting either bla, or both bla and the sensing domain. Since it is simple to screen for constructs that still retain bla activity after fusion (growth on ampicillin), large libraries of fusion proteins may be constructed and evaluated. The bla coding region is isolated, circularized, then both this species and the coding region of the sensing domain are digested with DNase1 or S1 nuclease to randomly cut each piece once. These two linear fragments are then ligated to each

other, transformed into *E. coli,* and screened for their ability to grow in the presence of ampicillin and the metabolite. The constructs that grow are then tested for metabolite dependent growth, which are then isolated as metabolite responsive biosensors. This method was successfully used to transform (1) maltose binding protein (MBP)[53, 54], (2) ribose binding protein[55], (3) glucose binding protein[56] and (3) xylose binding protein[56], into sugar dependent biosensors. Interestingly, while all these proteins are from a closely related class of sugar binding proteins, no insights, trends, or general rules about where the insertion would be successful were gleaned; all four proteins needed to be made via library generation, screening, and linker optimization[56].

In addition to the circular permutation method, the MuA transposase has been used to generate similarly functioning bla based biosensors. In this method, bla is transposed with a transposon sequence. Following transposition, the transposon can be replaced with the ligand-binding domain. In this case, the transposon was replaced with cytochrome b_{562} (cyt b)[57]. Cyt b is small peptide that undergoes a large conformation change upon binding to haem. This strategy was successful at generating several insertional positions where cyt b resulted in haem-dependent growth on ampicillin.

Finally, utilizing the many examples of successful splits of the bla enzyme, eight published splits were analyzed for their ability to accept the calmodulin, a sensor domain for Ca^{2+}, and the M13 peptide[58], a calmodulin binding peptide originally from skeletal muscle myosin light chain kinase . Of the eight, two could constitutively hydrolyze ampicillin, five were never able to do so, and one could do so in a Ca^{2+}-dependent fashion. While a successful biosensor was generated, it remains unclear how translatable insertions found by one library can be co-opted to incorporate new ligand-binding domains in a general sense. Furthermore, while these sensors are very sensitive

to their ligand, the switch characteristics are effectively digital; these sensors do not discriminate between different levels of their ligand.

1.2.1.2 FRET based biosensors

Förster resonance energy transfer (FRET) is a phenomenon by which two light sensitive moieties transfer energy in a distance-dependent manner. For fluorescent proteins, FRET occurs when the first fluorescent protein's emission wavelength matches the excitation wavelength of the second fluorescent protein. The resulting interaction is highly sensitive to the distance between the two proteins. Therefore, when the two proteins are close, excitation of the first protein results in the transfer of a virtual photon to the second protein, exciting it, and causing emission of the second proteins wavelength. However, if the proteins are far apart, the excitation of the first protein leads to its standard emission, and no energy is transferred. Therefore, by monitoring the emission of each fluorescent protein in the pair, it can be determined if the two proteins are close enough to experience FRET, or not, by analyzing the ratio of the two emissions. Similar to the split and reconstituted GFP-MBP fusion[52] mentioned earlier, FRET pairs can be fused to metabolite responsive domains in order to form a switch, if and only if, the ligand-binding changes the orientation of the FRET pairs in or out of a geometry permissive for energy transfer.

Cyan fluorescent protein (CFP) paired with yellow fluorescent protein (YFP), or blue fluorescent protein (BFP) paired with green fluorescent protein (GFP), are often used as FRET pairs. Calmodulin/M13 was once again utilized as a protein that undergoes a large conformational change in response to Ca^{2+}. Either the CFP/YFP pair or the BFP/GFP pair was fused on either side of calmodulin/M13, and to enable visualization of eukaryotic compartment differences, different versions were constructed (in addition to the cytosolic form) that included either an endoplasmic

reticulum localization signal or a nuclear localization signal[59]. Additionally, known mutations in calmodulin were made to change the sensitivity of the fusion protein, which was necessary since different subcellular compartments can have very different Ca^{2+} levels, and thus different sensor sensitivities are required to monitor changes in Ca^{2+} levels. These sensors successful functioned to quickly provide a readout for intracellular Ca^{2+} changes due to exogenously added signals such as histamine, ATP, ionomycin, and $CaCl_2$. While FRET sensors can be extremely sensitive to changes in protein conformation, the protein must undergo a conformational change upon ligand-binding in order for such a sensor modality to work. Furthermore, the attachments points of the pairs must be in a location that is close enough such that when the protein changes conformation, the pairs are either moved in, or out, of a geometry that enables FRET.

One solution to this challenge is to use families of proteins that are all structurally similar. Indeed, the sugar-binding family of periplasmic proteins, of which maltose binding protein is a member, all undergo significant conformational changes upon ligand-binding and have appropriated spaced N and C termini to accommodate a FRET pair. Using this strategy FRET sensors for glucose[60, 61], maltose[62, 63], and ribose[22, 64] have been developed and used in both prokaryotic and eukaryotic model systems. A caveat to these sensors, like the calmodulin-based sensors, is that the ligand-binding proteins must undergo mutagenesis to their binding pockets to alter the affinity of the protein for its ligand such that the range is suitable for the desired application. While there are successful examples of FRET-based biosensors, the stringent conditions on the sensing domain that must be met to qualify the protein to be a successful FRET sensor limits pool of potential candidates and may not be a broadly applicable strategy for generating new metabolite biosensors.

1.2.2 *De novo* protein design

The ability to computationally design a protein that can bind a metabolite *de novo* has, until recently, been an unsolved challenge. A computational method was developed based on forming favorable hydrogen bonds and van der Waals interactions to the ligand in addition to having a binding pocket that structurally fit the ligand. This method was evaluated to design a protein that binds to the steroid digoxigenin (DIG)[65]. Of the 17 computationally selected designs, two could bind DIG, and the best had an affinity for DIG in the low μM range. After further tweaking of the hydrogen bond network, the protein could selectively distinguish between DIG and two related steroid progesterone and β-oestradiol. In addition, the crystal structures of the protein showed very close agreement to the computationally predicted structure.

To convert this ligand-binding domain into a metabolite biosensor, the protein was fused to GFP. In the absence of DIG the ligand-binding domain was so unstable that it caused the degradation of the entire biosensors. Adding DIG-mediated stabilization of the biosensor, which led to a longer half-life and steady fluorescence resulting in a 5-fold induction of fluorescence over the unbound state[66]. This mechanism was also utilized to construct a transcription factor by fusing the unstable ligand-binding domain to a DNA-binding domain and transcriptional activator domain. In this scenario, only in the presence of DIG would the transcription factor be stabilized enough to activate transcription of a reporter gene. In the evaluation of this construct, high levels of DIG resulted in nearly a 60-fold induction of the reporter gene[66]. While the success of this computational method is a tour de force, and several ligand-binding proteins have been designed, whether this approach is readily generalizable remains to be determined.

1.2.3 RNA-based biosensors

RNA, much like a protein polypeptide, can fold into a huge diversity of structures, and have cis and trans catalytic activity. Therefore, RNA has been a medium by which natural metabolite responsive system have been discovered and characterized and a platform for which new sensors have been developed.

Thiamine pyrophosphate (TPP) and flavin mononucleotide (FMN) were discovered to have a novel translational attenuation mechanism for regulating their own biosynthesis pathways[67, 68]. The mRNAs for thiM and thiC, two enzymes responsible for thiamine biosynthesis, were both found to have sequences in the 5' untranslated region of their mRNA that formed a hairpin structure when bound to TPP. This hairpin, or riboswitch, prevents the translation by the ribosome and substantially downregulated the enzymes expression in the presence of thiamine, without the need for any protein co-factors. It was also demonstrated that there are riboflavin riboswitches in the biosynthesis pathway enzymes that undergo a similar form of translational control upon binding to FMN[67].

In addition to riboswitches, there are also ribozymes, RNA molecules that not only bind a small molecule, but then perform a catalytic activity such as cleavage. The *glmS* gene, which encodes the GlmS enzyme, is responsible for the formation of glucosamine-6-phosphate from frusctose-6-phoshate and glutamine. Like the riboswitches, there is a hairpin forming region in the 5' untranslated region of the mRNA that can bind the product of the enzyme, glucosamine-6-phosphate. Under high concentrations, glucosamine-6-phosphate binds the glmS mRNA, undergoes self-cleavage and is degraded, demonstrating a simple negative feedback loop[69]. Some researchers have even suggested that the presence of the natural RNA-based switches and enzymes

are potential evidence for an RNA world hypothesis about the development of life, as examples of situations where proteins are not required[68, 69].

To develop a strategy to create new ligand-binding RNA oligomers, a method called systematic evolution of ligand by exponential enrichment (SELEX) was created[70]. A large (10^{14}) pool of DNA molecules, each 100 nucleotides long and of random sequence, was generated, transcribed into RNA, then washed over columns containing a small molecule. After washing the column, and eluting remaining bound RNA molecules, the selected RNA molecules were reverse transcribed back into DNA, where the process would begin again. This way, a very large library of RNA molecules could be screened, and enriched for their ability to bind small molecules. Interestingly, 10^2-10^5 molecules were isolated for their ability to bind to one of the tested ligands. Considering the size of the initial library, this number is an incredibly small fraction, indicating how rare this phenomenon is. Yet, simultaneously, the number of molecules was much larger than zero, indicating there are many sequences that can accomplish the goal of ligand-binding.

In fact, it was using SELEX that an incredibly popular RNA aptamer for theophylline was developed[71]. The theophylline riboswitch has been utilized in a variety of contexts[16, 17]. A self-splicing intron was engineered to incorporate the theophylline aptamer, allowing the splicing decision to be regulated by the presence or absence of theophylline[72]. As a small molecule inducer for gene expression, the theophylline aptamer and a tetracycline aptamer were designed to interact with a GFP mRNA[73]. In this example, both aptamers were developed to be antiswitches, meaning that without the small molecule ligand, the RNA switch is unbound to the GFP mRNA, and fluorescence can be observed. However, in the presence of the small molecule, the RNA antiswitch's hairpins are rearranged, enabling the aptamer to bind to the GFP mRNA and inhibit translation. Further work has demonstrated the flexibility of the theophylline aptamer sequences

by mutating the RNA sequence to make it more sensitive to other related ligands such as caffeine and 3-methylxanthine[74]. Additionally, the theophylline aptamer has been integrated into ribozymes[75], and ncRNAs[76], to endow these RNAs with theophylline-responsiveness.

Other RNA aptamers to lysine and tryptophan have been implemented in screens[77] and as regulators for balancing metabolic flux[78]. In one example, the gene for citrate synthase, an essential TCA cycle enzyme, and a competing reaction for the biosynthesis of lysine, was put under the control of the lysine riboswitch. In this scenario, when there are low levels of lysine, citrate synthase is produced. However, when lysine levels are elevated, translation of citrate synthase is stopped by the bound lysine aptamer. This dramatically slows cell growth, but also improved lysine production by 63%.

RNA based biosensors have been shown to have very high specificity to their substrate, have a detection threshold than can be adjusted by varying the RNA sequence, and be flexible to mutation-based reengineering for different ligands[79]. However, because the sequence of the aptamer and its output domain must interact with each other so that their activities are linked, this method frequently requires a redesign by screening for new aptamer and output domain fusions, limiting the modularity of RNA base biosensors[16]. Another downside is the limited number of ligands that RNA aptamers have been developed for, suggesting that there may be classes of molecules that are unable to be bound by RNA species[80]. Therefore, these factors limit the generalizability of RNA aptamers as the backbone for the development of new metabolite biosensors.

1.2.4 Naturally occurring transcription factor biosensors

Genetic feedback control is a basic, commonly used mechanism for controlling metabolite levels in natural systems. Naturally occurring transcription factor biosensors often balance the production of pathway enzymes by monitoring the amount of a key intermediate or final product of a pathway. In native systems, genetic feedback control is commonly a cost-savings mechanism, by which if the cell has an abundance of a product, it can tune down the expression of the enzymes that produce that product, to conserve resources. The opposite is also true: if there is not enough of a given product, the enzymes to produce the product can be upregulated. This process is primarily accomplished by transcription factors, some of which are directly metabolite responsive. LacI and TetR are classic examples of this mechanism, responding to lactose/IPTG and the antibiotic tetracycline respectively. These two biosensors have been used countless times (reviewed here[81-83]) to control the expression of exogenously added open reading frames with their cognate small molecule inducers.

While LacI and TetR have been extensively used to control the expression of heterologous genes, there exists many situations in which controlling the gene expression of a reporter or a pathway would be better served by a native transcription factor that can dynamically control expression in response to a key intermediate or product of interest rather than exogenously added small molecule inducers such as IPTG or tetracycline. One method to find such a biosensor is to do a global transcriptional analysis of promoter expression that varies with the presence or absence of the metabolite of interest. In an example of this approach, cells expressing an engineered pathway to overproduce farnesyl pyrophosphate (FPP) were compared to a wild type strain, and promoters that were differentially regulated in those two conditions were identified[12]. Using transcriptome analysis, the GadE and RstA transcriptional activators, were found to downregulate,

or upregulated (respectively) their native targets in the presence of FPP. These FPP overexpression-responsive transcriptional activators and their cognate promoters were then implemented to control the FPP overexpression pathway. By dynamically controlling the pathway with the GadE transcriptional activator, production of the final product doubled compared to the initial, constitutive overexpression strain. This is likely due to the GadE-mediated downregulation of the enzymes that produce FPP, upon high levels of intracellular FPP. FPP accumulation is toxic, and therefore balancing the level of FPP production with the rate of FPP conversion into amorphadiene can help prevent the buildup of FPP. While GadE was successful at dynamically balancing FPP production and conversion, it is not a direct biosensor for FPP, and more directly involved in acid stress response, and pH regulation[84].

There are many examples of previously characterized transcription factor biosensors that can be applied to new systems to enable high throughput screening or dynamic control. BmoR, an alcohol-responsive transcriptional activator from *T. butanivorans,* was implemented to screen for cells that contained 1-buanol[4]. By using BmoR's congnate promoter $P_{BMO,}$ the expression of tetA, and therefore growth under tetracycline section, was induced in the presence of 1-butanol. Thus, cells producing more 1-butantol would be more fit under tetracycline selection and could be serially enriched and isolated. Using a similar tetA-based screening system, the NahR transcriptional activator from *P. putida* was implemented to screen for the presence of benzoate and 2-hydroxybenzoate[85]. Using a known high producing strain of benzoate, it was demonstrated that the NahR biosensor could discriminate between this strain and a low producing strain even after a $1{:}10^6$ dilution of the high producing strain to low producing strain. Other examples of known transcription factor biosensors being co-opted for strain discovery and development goals include: (1) LysG, the lysine responsive transcription factor from *C. glutamicum*[40]; (2) FadR, the

fatty acid responsive transcription factor from *E. coli*[5]; (3) FapR, the malonyl-CoA responsive transcription factor from *Bacillus subtilis*[7, 13, 86]; and (4) NR1, the acetyl phosphatase responsive transcription factor from *E. coli*[44].

A commonly required alteration to these natural systems is the reengineering of the native cognate promoter that accompanies the transcription factor[5, 7, 13, 40, 44, 86]. Native promoters rarely have the dynamic range or high expression that is commonly needed for the expression of the desired open reading frames. In all these cases, hybrid promoters were engineered by the incorporation of the native binding sites for the transcription factor, often from several different native promoters, all into a single synthetic promoter.

Despite these alterations prior to usage, naturally occurring transcription factor biosensors are a powerful tool for strain discovery and development. The generalizability of these biosensors is the inherently limited to known or previously characterized transcription factors. However, as was demonstrated with the FPP-responsive biosensors[12], new biosensors can be screened for using global transcriptional analysis if it is possible to overproduce the metabolite of interest.

1.2.5 Re-engineering natural sensors for different ligands or new activities

Transcription factor biosensors can be a powerful tool for metabolite biosensing, however when a suitable transcription factor biosensor is not known, a variety of techniques have been employed to re-engineer natural transcription factor biosensors for new ligands.

The simplest example of this technique is the ability to leverage a family of similarly structured transcription factors in the LacI/GalR family. These transcription factors have a ligand-binding domain and a DNA-binding domain and rely on an allosteric interaction between these two domains for regulating transcription in response to the binding of the metabolite[87]. Therefore,

two orthogonal DNA-binding domains and their cognate promoters (P_{LAC} and P_{TAN}) were mixed with ligand-binding domains that recognize different metabolites (lactose, fructose, fucose, ribose, trehalose, guanine, galatactose, cellobiose, and cytidine)[87, 88]. These chimeras all showed ligand-specific transcriptional regulation of the matching cognate promoter, corresponding to the attached DNA-binding domain. Interestingly, the native allosteric interaction was still active in the chimeras despite a smaller linker region in each fusion. This finding argues that this regulation does not require precise control or positioning of the ligand-binding domain relative to the DNA-binding domain[87] Furthermore, it was shown that these new chimeric transcription factors could be arranged into logic gates with the different ligands being the various inputs[88]. Using a similar concept, the DNA-binding domain from AraC, a native regulator of arabinose metabolism, and the DNA-binding domain Gal4 were each fused to the enzyme *idi*, which dimerizes in response to binding its ligand isopentyl pyrophosphate (IPP)[89]. Following fusion, mutagenesis and screening was performed to isolate clones that could regulated an AraC-responsive promoter in an IPP-dependent fashion.

Another strategy for generating new transcription factor biosensors is to use error prone PCR mutagenesis and high throughput screening of a known transcription factor to evolve it to sense a new ligand. The LuxR quorum sensing transcription factor natively recognizes oxohexanoyl-homoserine lactone (oxohexanoyl-HSL), and it very weakly recognizes octanoyl-HSL. After several rounds of mutagenesis and screening, eight LuxR variants were isolated that had a 100-fold increase in sensitivity to octanoyl-HSL[90]. Seven out of eight of the variants had mutants in the ligand-binding domain region, indicating that there is inherent plasticity in this pocket. However, these variants still recognized the original oxohexanoyl-HSL. Therefore, in a follow up study, the screen following mutagenesis was altered to included negative selection

against the original ligand in addition to positive selection for the new ligand. Under these conditions, a new variant was isolated that only responded to octanoyl-HSL and not oxohexanoyl-HSL[91]. Using a similar strategy of saturating mutagenesis in the ligand-binding pocket, AraC was evolved to recognize $_{D}$-arabinose instead of $_{L}$-arabinose[92], triacetic acid lactone[41], and mevalonate[93]. This strategy is not specific for AraC, XlyR, a native xylose responsive transcription factor as similarly mutated to bind a new ligand 1,2,4-trichlorobenzene[94].

By taking advantage of a structurally similar family of transcription factors or using saturating mutagenesis and high throughput screening, it is possible to re-engineer natural transcription factors to bind new ligands. In all these examples, the new ligand was either already partially recognized by the transcription factor, or the new ligand was structurally similar. It remains an open question whether any given transcription factor can be reengineered to sense a highly structurally divergent ligand.

1.2 Conclusions and thesis overview

From a broad perspective, there are many challenges associated with developing a truly generalizable method for creating, or finding, new metabolite biosensors. The primary reason for this is that there exists a large range of applications for metabolite biosensors, which potentially necessitates a number of different sensor properties. It may certainly be the case that no one method could generate any biosensor, for any application. For example, a transcription factor biosensor is highly desirable when the goal is to dynamically control a metabolic pathway at the transcriptional level. However, this same sensor would do a poor job of sensing metabolite levels in different cellular compartments, nor would it have an output that could monitor fast (seconds or less) changes in metabolite levels, such as a fluorescent or FRET-based sensor could. Additionally, for

diagnostic applications, it may be useful for the output of the biosensor to be highly visible such that the level of output can be determined qualitatively without any specialized lab equipment, whereas in a research setting, even dim outputs can be readily quantified using flow cytometry or microscopy paired with image processing. Therefore, a major challenge associated with developing a fully generalizable method is the diversity of the applications and biosensor properties required for each application.

With this in mind, I chose to focus my efforts to developing a generalizable strategy for generating transcription factor biosensors to narrow the search window and parameter space. Through the transcriptional regulation of fluorescent or enzymatic reporters, transcription factor biosensors can provide information on the levels of a particular metabolite, in single cells, in real time. In addition, transcription factor biosensor can also dynamically regulate the transcription of any open reading frame (ORF) with its cognate promoter. These properties make transcription factor biosensors amenable to a range of applications, ranging from basic discovery-based research to application-driven medical and biotechnology projects.

The upsides associated with a generalizable method for developing transcription factors are the potential for modularity and the ability for high-throughput construction and screening methodologies. From a modularity standpoint transcription activator-like effectors (TALEs) and zinc finger proteins (ZFP) are both classes of DNA-binding domains that have designable binding sites, allowing the user to specify or pick from a range of binding site sequence options. Utilizing these proteins as part of a method to generate new biosensors streamlines the DNA-binding domain aspect of the transcription factor by enabling the binding site to be modularly changed without substantially effecting the three-dimensional structure of the biosensor. From the DNA-binding domain up, developing a method that could combine the chosen DNA-binding domain with the

metabolite sensor domain can also be modular if the construction method does not depend on some feature of the sensing domain. By utilizing a known DNA-binding domain, and a construction method that is not dependent on the properties of the sensing domain, a further upside is the ability to assemble and screen libraries of potential biosensors in a high-throughput fashion. Since the parts are modular and the method in generalizable, this allows for the high-throughput sampling of many biosensor libraries without substantially changing the biosensor development process. Using DNA-binding domains with known binding sites also enables one to screen any biosensor made with a particular DNA-binding domain with a single promoter/reporter system. Even if downstream applications would require a different binding site sequence, the modular nature of both TALEs and ZFPs enable binding-site specificity to be readily changed after the biosensor has been developed. Therefore, the ability to both assemble and screen biosensors in a uniform way, regardless of the metabolite being sensed, is an upside that would make such a method very powerful.

In summary, harnessing transcription factor biosensors is currently limited to the utilization of naturally occurring transcription factors, or to the evolution of these natural sensors to sense new ligands. Furthermore, although the number of known transcription factors is limited, there is a wealth of known metabolite-binding proteins such as transporters and enzymes. This thesis is therefore motivated by the following open questions: (1) Can a metabolite binding protein be fused with a modular DNA-binding protein to generate a metabolite-dependent regulator of transcription? (2) If a metabolite-binding protein can be converted into a transcription factor, can this be done in a generalizable manner that enables the conversion of other metabolite-binding proteins into biosensors? (3) Are there classes of metabolite-binding proteins that are more amenable to conversion than are others? To address these questions, this thesis aims to investigate

several strategies for converting metabolite binding proteins into metabolite-responsive transcription factors, to provide a demonstration of a generalizable, high-throughput method for this conversion, and to generate a platform for evaluating the conversion of novel metabolite-binding proteins into metabolite-responsive transcription factors.

Chapter 2. Engineering modular biosensors to confer metabolite-responsive regulation of transcription

2.1 Context

Toward the goal of developing a generalizable strategy for converting the large pool of metabolite binding proteins into metabolite responsive transcription factors, this chapter addresses the feasibility of this idea by using maltose binding protein (MBP) and the BCR-ABL1 zinc finger DNA-binding domain (ZFP) as a model system. Prior to biosensor development, the design rules for zinc finger responsive promoters was first investigated. Understanding how zinc finger binding site placement would affect the ability of the ZFP or ZFP based biosensor to regulate transcription was an important open question to address, as all future biosensors could be based on ZFP DNA-binding domains. The design features for one ZFP should be true for all ZFPs in the Cys_2-His_2 class, regardless of binding site, and thus lead to generalizable design features for all ZFP based biosensors.

Another open question was how could MBP and the ZFP be combined in such a way to enable, maltose responsive regulation of transcription. In this chapter, we evaluated two distinct fusion strategies for converting MBP into a maltose responsive transcription factor. Utilizing the prior literature of biosensors made from MBP, a split zinc finger approach (SZF) was first investigated based off of successful FRET biosensor made by fusing fluorescent FRET pairs to the N and C terminal of MBP[95]. It was also known that the BCR-ABL1 ZFP had been shown to be split and reconstituted using self splicing inteins[96]. Taken together, I hypothesized that splitting the ZFP and fusing one finger to the N terminal and two fingers to the C terminal would enable maltose dependent change in conformation of the ZFP fragments in and out of a geometry

favorable to DNA-binding. The second conversion strategy was inspired by a fusion of MBP to TEM1 β-lactamase (bla), where bla was circularly permutated and fused internally to MBP at 316R[53]. The insertion of bla into MBP at 316R created an allosterically regulated bla enzymatic biosensor that would only cleave its substrate ampicillin in the presence of maltose. I hypothesized that this insertional position in MBP would also produce an allosterically regulated ZFP protein. This split protein (SP) conversion strategy was ultimately successful at generating a maltose responsive biosensor.

I lead the work described in this chapter including the design and execution of all experiments. This chapter is published as a paper in *ACS Synthetic Biology*[97]. This manuscript and chapter would not have been possible without the help of two talented undergraduate researchers Neil Dalvie and Austin Rottinghaus who helped with experimental execution, and writing of the manuscript.

2.2 Abstract

Efforts to engineer microbial factories have benefitted from mining biological diversity and high throughput synthesis of novel enzymatic pathways, yet screening and optimizing metabolic pathways remain rate-limiting steps. Metabolite-responsive biosensors may help to address these persistent challenges by enabling the monitoring of metabolite levels in individual cells and metabolite-responsive feedback control. We are currently limited to naturally-evolved biosensors, which are insufficient for monitoring many metabolites of interest. Thus, a method for engineering novel biosensors would be powerful, yet we lack a generalizable approach that enables the construction of a wide range of biosensors. As a step towards this goal, we here explore several strategies for converting a metabolite-binding protein into a metabolite-responsive transcriptional

regulator. By pairing a modular protein design approach with a library of synthetic promoters and applying robust statistical analyses, we identified strategies for engineering biosensor-regulated bacterial promoters and for achieving design-driven improvements of biosensor performance. We demonstrated the feasibility of this strategy by fusing a programmable DNA-binding motif (zinc finger module) with a model ligand binding protein (maltose binding protein), to generate a novel biosensor conferring maltose-regulated gene expression. This systematic investigation provides insights that may guide the development of additional novel biosensors for diverse synthetic biology applications.

2.3 Introduction

Cells evaluate and respond to their internal states through a range of mechanisms, including the wide use of molecular biosensors. In a general sense, a biosensor may be understood to comprise a species that senses one or more analytes, typically through a molecular recognition event involving binding to the analyte, such that recognition of the analyte is transduced into a change in the biosensor that enables it to effect a change in cell state. Biosensors may be composed of a range of biomolecules, most commonly including RNA [16, 17, 78] or protein [2-13]. Early applications included the generation of whole-cell biosensors, in which an environmental analyte enters a cell through active or passive transport. Upon recognition of the analyte by an intracellular biosensor, an output signal such as fluorescence, luminescence, or color-change is generated, most commonly by biosensor-induced expression of a reporter gene or by analyte binding-induced changes in the activity of a fluorescent or enzymatic biosensor protein. A particularly exciting frontier is the use of biosensors to sense not external factors, but rather a cell's internal metabolic state.

Metabolite-responsive biosensors may help to address several pervasive and persistent challenges in the fields of synthetic biology and metabolic engineering [2-5, 7, 9, 11-13, 44, 78]. First, biosensors may help overcome the costliest and rate-limiting step in the development of new biosynthetic pathways – screening and evaluating pathway or strain variants to both identify well-performing constructs and glean insights into pathway function that may be utilized in subsequent iterative rounds of the design-build-test engineering cycle. By coupling metabolite-binding to outputs such as fluorescence or antibiotic resistance, biosensors can enable the screening of large libraries (e.g., $>10^8$ members), which remain beyond the capacity of even contemporary automated platforms for performing clonal evaluations. For example, the naturally occurring transcription factor BmoR was harnessed to confer growth in the presence of butanol, which enabled the screening of a plasmid library to identify strains exhibiting robust production of 1-butanol [4]. Similar approaches have been harnessed to screen plasmid libraries to achieve enhanced production of mevalonate [93], triacetic acid lactone [41], and L-lysine [77]. Such an approach may be extended to screen for high-performing variants generated through genomic mutation, including both random mutagenesis, which has been utilized to optimize L-lysine production via mutation of endogenous enzymes [40], and targeted genome-wide mutagenesis, which has been used to optimize naringenin and glucaric acid production via combinatorial perturbation of endogenous gene regulation [2]. While most investigations to date have applied these methods to bacterial chasses, such approaches may also be extended to yeast and other organisms [7]. In general, "digital" biosensor outputs, such as expression of antibiotic resistance, are most useful for screening, while "analog" biosensor outputs, such as fluorescence, enable both screening and characterization of internal metabolite concentrations, potentially at the single-cell level, to guide construct analysis and iterative refinement.

A powerful, yet less explored extension of this approach is the use of metabolite-responsive transcriptional regulators to implement feedback control in order to optimize system performance. An early demonstration of this opportunity was the use of an acetyl phosphate biosensor to sense excess glycolytic flux, and in response, regulate the expression of limiting genes in the lycopene biosynthesis pathway. This activity resulted in both enhanced lycopene production and diminished growth defects [44]. More recently, feedback control was used to achieve balanced flux through several pathways that led to enhanced yields and improved cell survival during the production of a biofuel (fatty acid ethyl ester) [5]. This investigation made use of the natural FadR biosensor, which is antagonized by Acyl-CoA, paired with synthetic promoters engineered to achieve robust regulation by FadR. Similarly, lysine-responsive riboswitches were utilized to control the expression of citrate synthase and thereby increase lysine production by controlling flux in the TCA cycle [78]. The potential utility of biosensor-mediated feedback control is now widely recognized [1, 9, 98], and further implementation is currently limited largely by the pool of suitable biosensors.

A general challenge in the use of biosensors is that the pool of metabolites one would like to measure and potentially utilize for feedback control is much larger than the pool of metabolite-responsive transcriptional regulators that have been identified. Bioinformatic approaches and surveys of published literature may identify a number of useful biosensors that have simply not yet been utilized as such. For example, a recent study elegantly applied a systematic characterization of known metabolite-responsive transcriptional regulators to generate quantitative fingerprints enabling these biological "parts" to be harnessed for engineering applications [3]. However, since the entire pool of naturally-evolved biosensors is likely much smaller than the pool of metabolite targets, it would be attractive to develop approaches for engineering novel

biosensors. Ideally, such a biosensor could be constructed to recognize an analyte of interest (with some practical degree of specificity), would exhibit a dynamic range suitable (or tunable) to the application of interest, and could be directed to regulate a gene (or genes) of interest in a ligand-dependent fashion. Although this comprises a daunting protein engineering challenge, a number of smaller-scale successes suggest strategies that may help to achieve this goal.

The most widely used approach for engineering novel biosensors is to genetically fuse a ligand-binding protein with a distinct functional domain, such that the fusion causes the activity of the functional domain to be conditional upon the presence or absence of the ligand of interest [8, 14, 52-54, 57, 58, 76]. Most commonly, the functional domain comprises a fluorescent protein or enzyme conferring antibiotic resistance, each of which comprises an output amenable to screening the large libraries required to identify functional fusion proteins. For example, maltose binding protein (MBP) and β-lactamase (BLA) were circularly permutated to generate a library of fusion proteins, such that successful fusions exhibited high BLA activity only in the presence of maltose [53]. Calmodulin, which experiences a conformational change upon binding to Ca^{2+}, is similarly amenable to such a fusion strategy to create fusion proteins based upon BLA [58] or GFP and its derivatives [59]. Indeed, many similar approaches have harnessed proteins in which ligand binding induces a conformational change in order to generate biosensors in which fluorescence, often via FRET, provides a metric of intracellular metabolite concentration (reviewed in [8, 15, 43, 99]). Furthermore, zinc finger proteins (ZFP), transcription activator-like effectors (TALE), and CRISPR-based DNA-binding domains have been fused to putative repressor and activator domains to create novel transcription factors to regulate both prokaryotic and eukaryotic transcription, although such functions are not generally regulated by ligand binding to the transcription factor ([100-105]). However, recently an allosterically regulated version of Cas9 has been developed by

fusing the estrogen receptor-α to create a protein that represses transcription in the presence of the ligand, 4-hydroxytamoxifen[106]. Given the broad homology within the LacI/GalR family of ligand-responsive transcription factors, novel biosensors have also been constructed by fusing the ligand-binding domains from LacI paralogs to the LacI DNA-binding domain, conferring regulation of the *lac* promoter by fructose, ribose, or other species [41, 87, 88, 93]. Ultimately, computational protein design could guide the development of novel biosensors. To date, such methods have been used primarily to shift ligand specificity of existing biosensor proteins [90, 93, 107-110], although *de novo* design of novel ligand-binding proteins and biosensors is another promising frontier [65, 66]. Overall, these approaches bespeak the promise of engineering novel biosensor proteins, but to date no generalizable approach for engineering novel metabolite-responsive transcriptional regulators has been described.

In this study, we investigated, validated, and developed a strategy for engineering novel metabolite-responsive transcriptional regulators. Our central goals were to quantitatively evaluate several strategies for converting a ligand-binding protein into a functioning biosensor that regulates transcription, and to elucidate design principles governing the performance of biosensors constructed in such a fashion. To this end, we leveraged the facts that MBP is a well-characterized ligand-binding protein, and that zinc finger proteins (ZFP) are well-characterized and programmable DNA-binding domains. Furthermore, we applied quantitative analyses to identify rules for designing biosensor-regulated promoters and quantitatively characterize these new biological parts. This systematic investigation establishes a foundation for applying a potentially generalizable strategy towards the ultimate goal of engineering customized metabolite-responsive biosensors.

2.4 Materials and methods

2.4.1 Bacterial strains and culturing

All experiments were conducted in TOP10 *Escherichia coli* cells (F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ^-) (Life Technologies). Cells were maintained in Lysogeny Broth (LB) Lennox formulation (10 g/L of tryptone, 5 g/L of yeast extract, 5 g/L of NaCl) supplemented with appropriate antibiotics (Ampicillin 100 μg/mL or Kanamycin 50 μg/mL). All experimental analysis was conducted in M9 minimal media (1X M9 salts, 0.2% Casamino Acids, 2 mM $MgSO^4$, 0.1 mM $CaCl^2$, 1 mM Thiamine HCl) containing glycerol (0.4%) as the primary carbon source. 1% arabinose and variable amounts of maltose monohydrate and isopropyl β-D-1-thiogalactopyranoside (IPTG) were added as indicated. M9 medium containing both Ampicillin and Kanamycin was used to maintain the strains that contained both a reporter plasmid and a biosensor plasmid.

2.4.2 Plasmid construction

All plasmids were assembled using standard molecular biology techniques. Plasmid backbones containing "plug-and-play" multiple cloning sites and compatible plasmids containing synthetic parts (mCherry, GFPmut3b, pBAD, AraC, pTrc2) were generously provided by Jim Collins (MIT) [111]. Custom RBS sequences were designed using the RBS Calculator [112]. The pA15 low copy number origin was obtained from the Registry of Standard Biological Parts, plasmid pSB3K3. Template sequences derived from published descriptions were used for terminators [113], BCR-ABL1 [96], and the Zif268 portion of human EGR1 [114] (AddGene #52724). MBP was PCR amplified directly from TOP10 genomic DNA. The library of constitutive reporters was cloned in

a low copy number pA15 backbone (~10 copies per cell) with the ampicillin resistance cassette. All of the pBAD-based inducible ZFP and pTrc-based inducible biosensor expression constructs included a ColE1 backbone (~300 copies per cell) and kanamycin resistance cassette. The mCherry gene was cloned behind each ZFP or biosensor gene to act as a co-cistronic reporter to confirm arabinose and IPTG mediated induction of gene expression. Representative plasmid maps are included in Appendix A2.1.

2.4.3 Microplate-based fluorescence assays and analysis

Cultures were inoculated from single colonies into 2 mL of M9 media and grown overnight to stationary phase. Overnight cultures were diluted 1:20 and grown into exponential phase (OD_{600} ~0.5). Cultures were again diluted to an OD_{600} ~0.05, plated in black-walled clear bottom 96-well plates in biological triplicate, and induced with 1% arabinose (to drive expression of the ZFP) or IPTG as indicated (to drive expression of the biosensor), +/- maltose as specified. In each experiment, IPTG-induced expression of biosensor constructs was confirmed via the co-cistronic expression of mCherry (data not shown). Plates with lids were incubated and shaken in a continuous double orbital pattern at 548 cpm (2 mm) inside a BioTek Synergy H1 plate reader for 10 h with GFP, mCherry, and OD_{600} measurements taken every 15 min. Monochrometer settings were 481/511 nm for GFP and 585/620 nm for mCherry.

To quantify reporter output, GFP fluorescence per OD_{600} was quantified and averaged over 7 time points that span ~1.5 h of exponential growth (unless otherwise indicated). The specific fluorescence of each sample was defined as the mean (GFP/ OD_{600}) averaged across these 7 time points, and this specific fluorescence was averaged across 3 biological replicate samples. To quantify fluorescence attributable to GFP, each sample was background-subtracted using a control

sample comprising cells expressing no fluorescent proteins. To enable comparisons between promoters, each specific fluorescence value from the arabinose or IPTG-induced condition was normalized to the specific fluorescence of the uninduced condition, yielding "relative specific fluorescence". To normalize this metric of promoter performance to the base case, the relative specific fluorescence calculated for each promoter-ZFP (or promoter-biosensor) combination was then normalized to the same value calculated for that repressor using the "No sites" promoter, yielding a quantity we termed, "relative expression". This normalization strategy was utilized in order to implicitly correct for any minor effects that arabinose, IPTG, or maltose many confer on GFP/ OD_{600} in a manner that is unrelated to expression of the ZFP or biosensor. Thus, when quantifying biosensor performance, relevant control samples for the "+maltose" case (e.g., the uninduced case and No sites control case) were also quantified in the presence of maltose. For each metric, error was propagated according to the division rule to generate reported standard deviations.

2.4.4 Flow cytometry

Flow cytometry was used to quantify fluorescent reporter output on a single cell basis. Cells were grown and induced as described for the microplate-based fluorescence assays. Samples were collected after 5 hours of growth. Cells were then placed on ice, diluted 1:2 in chilled phosphate buffer saline (PBS) supplemented with 5 mM EDTA, and analyzed on an LSR II flow cytometer (BD). A minimum of 100,000 events were collected per sample. Mean fluorescent intensity was calculated using a minimum of 20,000 cells per sample using FlowJo software (Treestar), and relative expression calculations and error propagation were conducted as described for the microplate assays.

2.4.5 Statistical analysis of promoter design features

In order to use computational analysis to compare promoter designs, it was necessary to define quantitative descriptors, or features, that capture distinguishing architectural aspects of each promoter. Because our goal was to elucidate general design principles, we chose to limit our features to those describing the quantity and location of the various 9 bp ZFP binding sites. Following this approach, we defined 17 features that describe the locations of ZFP binding sites relative to both the -10 box (TATA box) and -35 region and relative positioning amongst the ZFP binding sites. In order to determine which promoter features were important for explaining variation in performance between promoters, several feature selection methods were applied to each set of input and output data (both of which were mean-centered and variance-scaled) to generate rankings of feature importance, noting that feature independence was not assumed. For these analyses, features always served as the regression inputs. The output was the "repressibility", which we defined as the negative of relative expression, such that a promoter-ZFP combination with a high repressibility exhibits low relative expression.

Three feature selection techniques were utilized: partial least squares regression (PLSR), Random Forest, and Lasso. PLSR was executed using the built-in MATLAB function, *plsregress*. To determine feature importance, a permutation test was used [115]. Briefly, the output vector was randomly permuted, and PLSR was executed for this meaningless output vector, such that when this process was repeated multiple times, we calculated the standard deviation associated with each coefficient (one coefficient per feature); thereby, the ratio of true coefficient magnitude to the standard deviation associated with this coefficient provided a metric by which features can be ranked in order of importance. To implement the Random Forest method, we modified a MATLAB script developed by Jaiantilal (https://code.google.com/p/randomforest-matlab/), which

was based upon a method originally described by Breiman and Cutler (http://www.stat.berkeley.edu/~breiman/RandomForests/). The last feature selection method used was Lasso regression, also known as sparse or regularized regression [116]. Lasso feature selection is generally considered more robust than a permutation test or Random Forest, because the selection is built into model generation and does not require removing features from a predictive model. Each of these methods is described in full detail in Supplementary Methods.

2.5 Results

2.5.1 Developing novel zinc finger protein-regulated constitutive promoters

In this investigation, we sought to develop a readily generalizable strategy for engineering novel biosensor proteins from the ground up. We hypothesized that such a goal might be achieved by first using an orthogonal DNA-binding protein to regulate transcription of an engineered promoter, and then fusing this DNA-binding domain to a distinct protein capable of binding the target ligand, such that when the fusion protein binds ligand, DNA-binding (and thus transcriptional regulation) is either disrupted or enhanced.

To begin investigating this overall strategy, we first sought to engineer a novel transcriptional regulator by leveraging the modular, programmable DNA-binding properties conferred by the zinc finger protein (ZFP) architecture [117-119]. ZFPs are small (compared to alternative architectures such as TALEs [104, 105]), easy to manipulate, and can be designed to bind to nearly any sequence. The ZFP architecture has previously been utilized to create novel transcription factors in *E. coli* [102], as well as in eukaryotes [80, 100, 101]. The Cys_2-His_2 class of ZFPs is an attractive DNA-binding domain, since each "finger" of the ZFP binds to a distinct 3 bp DNA sequence (Figure 2.1A) [120], and thus sequence specific binding is achieved by engineering ZFPs

comprising multiple "fingers" fused in tandem. Therefore, to initially investigate our strategy for biosensor engineering, we utilized the BCR-ABL1 ZFP as our DNA-binding domain. BCR-ABL1 is well-characterized, exhibits a low equilibrium dissociation constant when binding its cognate 9 bp DNA-binding site with three tandem fingers (K_d ~78 pM), and has been shown to exhibit conditional DNA-binding when genetically split and reconstituted [96, 117]. Notably, although the BCR-ABL1 consensus binding sequence is known, no *E. coli* promoters have been previously repressed by BCR-ABL1.

In order to begin elucidating the rules for building novel ZFP-regulatable promoters, a library of 68 different constitutive promoters was designed (Appendix A2.2). The library was built by inserting BCR-ABL1 binding site(s) at various locations around the consensus -10 box (TATAAT) and -35 region (TTGACA). The -10 box and -35 region are critical for recruitment of the σ^{70} factor of RNA polymerase, and these sequences are necessary and sufficient to create a constitutive promoter[121, 122]. Design features that varied across the library included BCR-ABL1 binding site location(s), relative to the consensus elements, and spacing between BCR-ABL1 binding sites. Although this design did not presuppose that ZFP-binding would repress transcription from these constitutive promoters, this was the anticipated mechanism because BCR-ABL1 was not fused to a transactivation domain [102]. Each promoter was encoded on a low copy number plasmid and drove expression of *E. coli*-optimized GFP as a reporter.

To quantify the extent to which each promoter was repressed or activated by BCR-ABL1 during exponential growth, we defined a metric of "relative expression" that describes how induction of BCR-ABL1 expression impacts expression from the promoter as compared to a "No Sites" control promoter lacking BCR-ABL1 binding sites (see Material and Methods). This relative expression normalization strategy was utilized in order to implicitly correct for any effects

that arabinose many confer on GFP/ OD_{600} in a manner unrelated to expression of the ZFP. Thus, low relative expression indicates that a promoter is highly repressed by BCR-ABL1. The library of promoters exhibited wide ranges of basal expression (Appendix A2.3) and repressibility by BCR-ABL1 (Figure 2.1B). Nearly all of the promoters exhibited a decrease in GFP expression upon the induction of the BCR-ABL1 ZFP, and no promoters showed any level of activation upon BCR-ABL1 induction. When relative expression was quantified at 10 h post-induction, at which point cultures had reached stationary phase, the observed repressibility was much more pronounced (Figure 2.1C). To investigate how relative expression patterns differed between cells within each population, we also examined several representative promoter cases at the single cell level using flow cytometry (Figure 2.1D). Overall, responses were unimodal, such that population-averaged fluorescence measured by flow cytometry corresponded well to comparable metrics obtained by microplate-based assays (Appendix A2.4), and thus the latter method was used for subsequent analyses. Given this wide range of phenotypes, we next investigated the relationship between promoter design and repressibility.

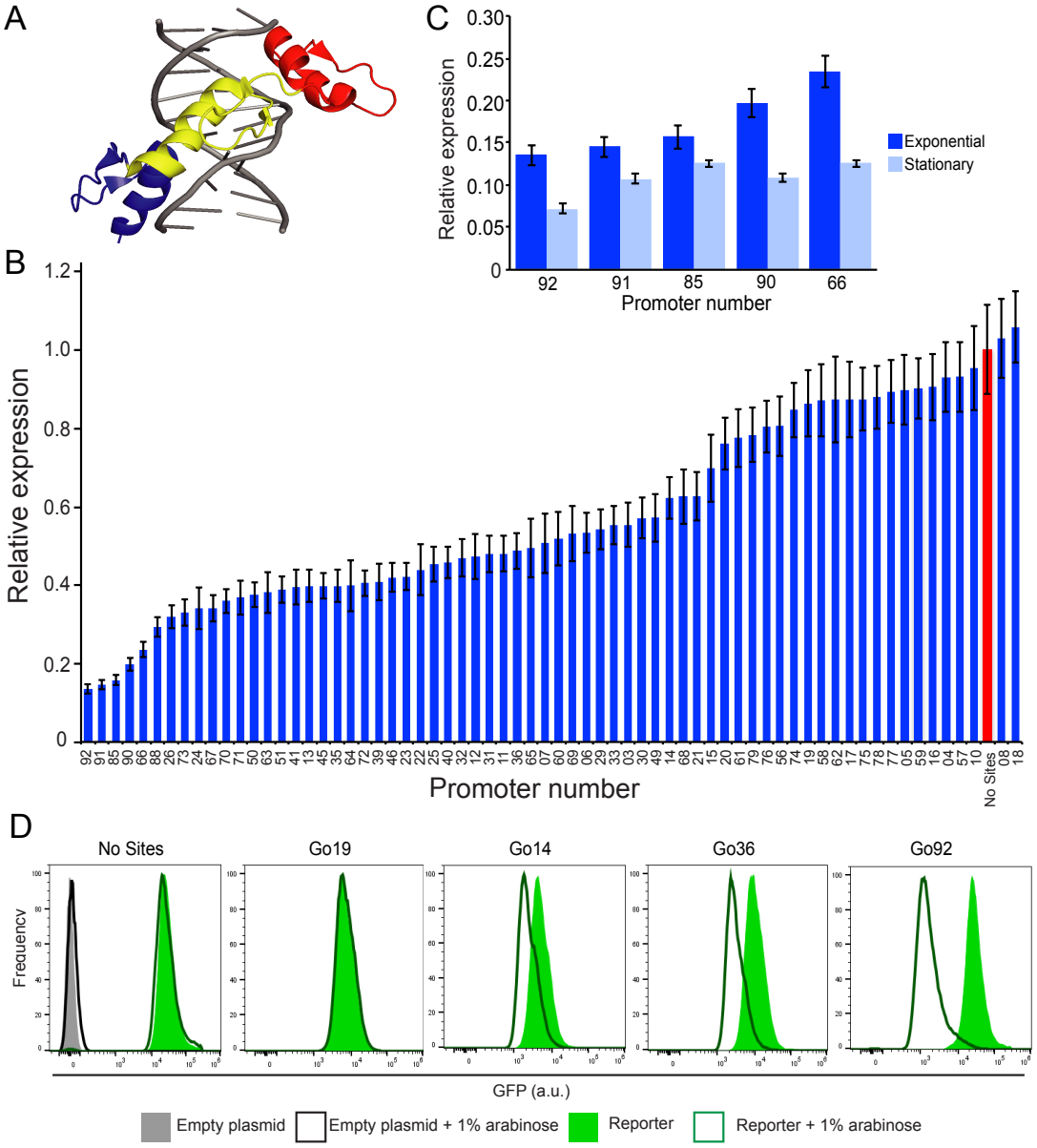


Figure 2.1 Developing novel zinc finger protein-regulated constitutive promoters
(A) Crystal structure of a Cys_2-His_2 class of zinc finger binding to its cognate 9 bp DNA sequence (PDB #4R2A) (B) Repression of the constitutive promoter library by BCR-ABL1 (normalized to No Sites control, shown in red, which lacks BCR-ABL1 binding sites) (C) Relative expression of the top 5 most repressible promoters was evaluated during both exponential growth (as in panel B) and after reaching stationary phase. (D) Select promoter constructs were evaluated by flow cytometry to assess variation in expression and repression across the population; one plot representative of three biological replicates is shown for each condition. A concentration of 1% arabinose was used to induce the expression of the BCR-ABL1 zinc finger. Relative expression is defined as the ratio of GFP/OD600 (for any given promoter) of the induced case relative to that of the uninduced case, divided by this same ratio for the No Sites promoter (a full description and rationale can be found in the Materials and Methods section). Relative expression values calculated from these data are explicitly compared to comparable microplate assay-based metrics in Appendix A2.4. Microplate data were collected over 7 sequential time points, spanning ~1.5 h of mid-exponential phase growth, and averaged. All data represent mean values calculated from three independent experiments, and error bars represent one standard deviation.

We first examined the impact of promoter design on repressibility by inspection. As depicted in Figure 2.2, promoters representing variations on a particular design feature were first grouped together in order to identify simple trends, with the caveat that such trends are potentially restricted to the scope of promoters evaluated in our library. Generally, placing two ZFP binding sites in between the -10 box and the -35 region (the promoter "core") led to greater repressibility than did insertion of a single ZFP binding site, regardless of its position (Figure 2.2A). Go15 is an exception to this trend, as it was not dramatically less repressible than was Go14. Furthermore, promoters with two ZFP binding sites in the core were more repressible than were comparable promoters that lacked these core sites but shared other ZFP binding sites (Figure 2.2B). Compared to promoters including all three ZFP binding sites required for recognition by BCR-ABL1, promoters in which only two sites were present exhibited reduced repressibility by BCR-ABL1 (Figure 2.2C). Adding additional ZFP binding sites downstream of the -10 box did not increase repressibility (Figure 2.2D). However, adding additional binding sites upstream of the -35 box did effectively increase repressibility (Figure 2.2E). Decreasing the spacer distance between the -10 box and the first downstream binding site increased repressibility (Figure 2.2F). Similarly, removing any spacer between the ZFP binding sites and the -10 box and -35 region resulted in greater repressibility than did removing spacers adjacent to either the -10 box or -35 region alone (Figure 2.2G). To explain the increase in repressibility observed for Go85 compared to Go66, we hypothesize that BCR-ABL1 cannot simultaneously occupy adjacent binding sites, and furthermore, that binding to the sites directly upstream of the -10 box and downstream of the -35 region results in greater repressibility than is conferred by BCR-ABL1 binding to the more distal binding sites present in Go66. Altogether, promoter Go92 exhibited the greatest repressibility of

any library member, and this promoter appears to follow many of the design rules suggested by the above inspection-based analysis. However, since only a subset of our promoter library was amenable to this direct inspection-based analysis, we next pursued a systematic analysis of the entire library to further refine our understanding of the applicable design rules.

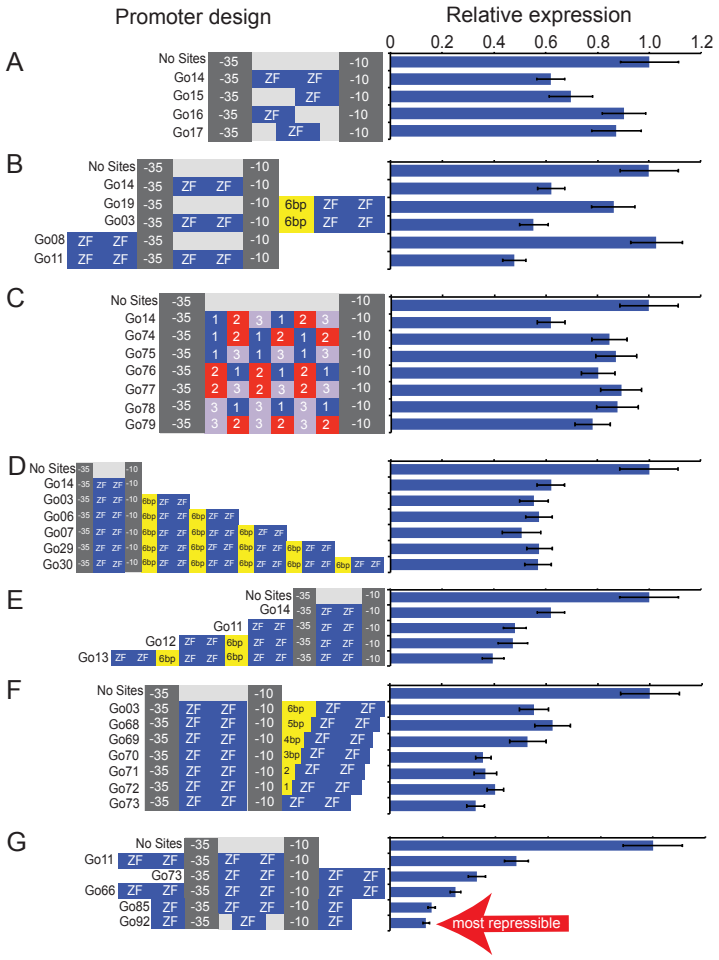


Figure 2.2 Inspection-based evaluation of promoter design rules
Promoters were manually grouped to represent exploration of design features including (A) presence and location of ZFP binding sites between the -10 box and -35 region, (B) combinatorial effects of having a pair of binding sites between the -10 box and -35 region, (C) variations in the locations of individual ZFP binding sites within the core, (D, E) contributions of additional BCR-ABL1 binding sites either downstream or upstream of the -10 box and -35 region, (F) spacing between the -10 box and the downstream ZFP binding sites, (G) combinatorial effects of directly flanking the -10 box and -35 region with ZFP binding sites. All data are re-plotted from Figure 2.1B. Abbreviations and conventions: ZF is the 9 bp BCR-ABL1 binding site; 1, 2, and 3 represent the first, second, or third, 3 bp finger binding sites within the BCR-ABL1 binding site; yellow boxes represent spacer sequences of the indicated length. Microplate data were collected over 7 sequential time points, spanning ~1.5 h of mid-exponential phase growth, and averaged. All data represent mean values calculated from three independent experiments, and error bars represent one standard deviation.

2.5.2 Computational identification of promoter design features conferring ZFP-mediated repression

Statistical methods can provide insights into large or diverse data sets that are difficult to compare qualitatively or by inspection alone. Therefore, we performed a series of statistical analyses termed computational "feature selection" in order to determine which promoter features are important for predicting the relative repressibility of a given promoter in the presence of the ZFP. Given a set of feature "inputs," feature selection seeks to eliminate those features that are redundant or irrelevant to the prediction of a particular output. In our analysis, the output was defined as the repressibility (the negative of relative expression) exhibited by each promoter in the library. To generate the input list, we defined a set of 17 quantitative features that described each promoter in the library. Because we sought to elucidate general design rules and avoid over-fitting our particular promoter library, we defined the 17 features strictly on the basis of describing the locations of each ZFP binding site relative to the -10 box, the -35 region, and to other ZFP binding sites (Figure 2.3A).

Three different feature selection methods were applied to analyze BCR-ABL1-mediated repression of our promoter library. We first used partial least squares regression (PLSR), in which the regression coefficient associated with each feature indicates the degree to which that feature explains variations in repressibility within our dataset. Each coefficient was scaled using a permutation test to correct for the coefficient one would calculate for a randomized (meaningless) output vector [115], and these corrected coefficients were used to generate a ranked lists of features (for PLSR coefficients, see Appendix A2.5). By repeating the PLSR using only fixed numbers of most important features, we found that the first few features explained much of the variance in the overall dataset (Figure 2.3B). A similar trend was observed when repeating the PLSR with a

limited number of principal components, indicating that principal components do not provide deeper understanding of this system than do single features. The second method used was Random Forest, which creates a predictive model of the system using decision trees. By iteratively generating decision trees using a subset of features, this method calculates the mean loss of accuracy when a given feature is removed, and this quantity is used to generate a ranked list of features by importance (Appendix A2.5). The final method used was Lasso regression, which performs least squares regression with an additional penalty placed on the magnitudes of regression coefficients [116]. This penalty is weighted by a parameter, λ, such that as λ is increased, the coefficients of unimportant features shrink to zero; features are thus ranked by the number of iterations that they retain a non-zero coefficient as λ is increased. For each value of λ, the regression fit was tested with 10-fold cross validation, to obtain a mean squared error (MSE) and number of retained coefficients for each value of λ (Figure 2.3C). A reasonable fit of the system was obtained using 3-5 features, with lower MSE for larger feature numbers likely representing over-fitting of noise in the dataset.

Overall, the three feature selection methods (PLSR, Random Forest, and Lasso) generated similar but not identical ranked lists of features (Figure 2.3D). The most important features included *spacer inside -10* and *spacer -10*, which were undesirable for repressibility, and *BS in core, core middle space,* and *pairs in core,* which were desirable for repressibility. Together, these findings indicate that placing binding sites as close as possible to the -10 box and -35 region are predicted to confer strong repressibility. Interpreting other features is less straightforward, such as *core middle space*, which describes whether there exist base pairs in the core other than ZFP binding sites; the three methods appear to disagree on the importance of *core middle space*. One possible explanation for this discrepancy is the influence of Go22 and Go92 on this analysis, since

both of these promoters contain space in the core but no space on either side of the -10 box (Go22 and Go92) and no space on either side of the -35 region (Go92). As such, for these promoters, containing space in the core may be merely associated with the presence of other promoter features conferring repressibility. These cases may well emphasize the importance of *spacer inside -10* and *spacer -10*. Moreover, Lasso, which is considered the most robust method for feature selection, ranked *core middle space* as unimportant, potentially by resolving this contradiction better than did the other methods. Altogether, this analysis enabled us to leverage our diverse promoter library to glean general, quantitative design principles for engineering ZFP-repressible promoters. However, it was not yet clear whether such rules would extend to the design of promoters regulated by ZFP-based biosensors.

A

Shorthand	Description of feature
BS in core	Number of binding sites in the core
pairs in core	Number of pairs in the core
BS upstream	Number of binding sites upstream of -35 region
pairs upstream	Number of pairs upstream of -35 region
BS downstream	Number of binding sites downstream of -10 box
pairs downstream	Number of pairs downstream of -10 box
length	Length of the promoter in base pairs
spacer -10	Base pairs downstream of the -10 box, before the first binding site
spacer -35	Base pairs upstream of the -35 region, before the first binding site
spacer inside -10	Base pairs upstream of the -10 box, before the first binding site
spacer inside -35	Base pairs downstream of the -35 region, before the first binding site
core middle space	Non-binding site base pairs in the core
ave down space	Average number of base pairs between binding sites downstream
ave up space	Average number of base pairs between binding sites up stream
down length	Base pairs downstream of the -10 box
up length	Base pairs upstream of the -35 region
lopsidedness	Absolute value of (binding sites upstream) - (binding sites downstream)

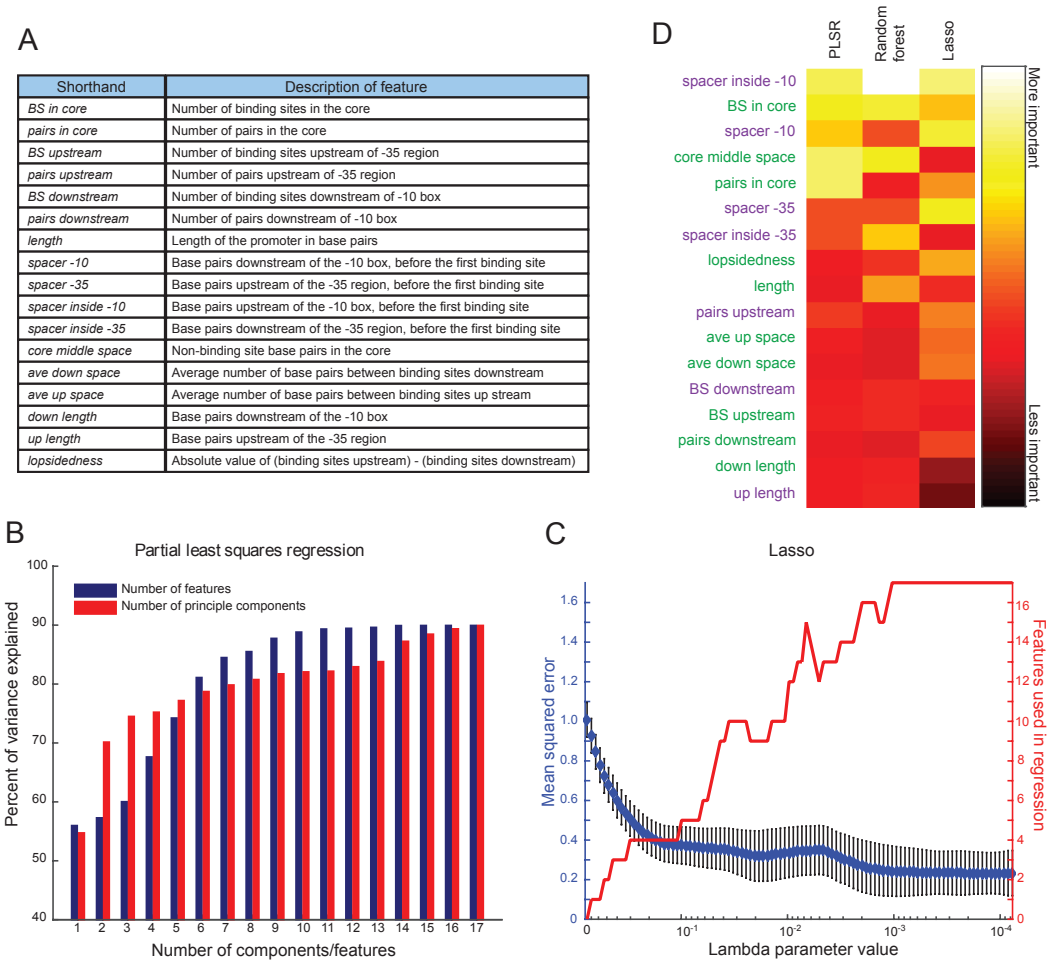


Figure 2.3 Computational identification of promoter design features conferring ZFP-mediated repressibility
(A) Shorthand names and descriptions of the 17 features chosen to describe the promoter library. One binding site is defined as the nine base pair ZFP binding site. One pair is defined as two adjacent binding site sequences. (B) PLSR analysis of the degree to which promoter features explain variance in the relative expression data (BCR-ABL-mediated repressibility) reported in Figure 2.1. Each series evaluates the explanatory power achieved using an increasing number of features or principle components, each of which is added to the set in ranked order from most to least important. (C) Lasso regression analysis of the degree to which promoter features explain variance in the relative expression data reported in Figure 2.1. This plot displays the number of features with non-zero coefficients (and the resulting mean squared error) as λ is increased, causing less important features to be eliminated from the regression analysis. The mean squared error was obtained through 10-fold cross validation, which also produced a standard error for the mean squared error, which is shown as error bars. (D) Relative importance of each feature, vis-à-vis explaining BCR-ABL1-mediated repressibility as determined by PLSR, Random Forest, and Lasso regression, with the overall order listed here determined by average rank across the three feature selection methods. The color each feature name indicates the sign of its regression coefficient: green indicates positive coefficients (large feature values confer more repressibility) and pink indicates negative coefficients (small feature values confer more repressibility). Detailed regression coefficients and importance values are provided in Appendix A2.5.

2.5.3 Conversion of transcriptional repressors into ligand-responsive biosensors

Having established that BCR-ABL1 functions as a transcriptional repressor, we next investigated two strategies for converting this repressor into a biosensor. Here, the primary goal was to investigate general strategies for converting a ligand-binding protein into a ligand-responsive transcription factor. As described above, we hypothesized that such conditional regulation of gene expression may be achieved by fusing BCR-ABL1 to a ligand-binding domain. To investigate the feasibility of this approach, we chose the uniquely well-studied maltose binding protein (MBP), in part because this protein experiences a substantial and well-characterized conformational change (~9 Å decrease in separation between N and C termini) upon ligand binding [123-128]. The first strategy we explored was termed the Split Zinc Finger (SZF) approach, in which BCR-ABL1 was split genetically such that the N and C termini of MBP were fused to BCR-ABL1-derived ZFPs. This strategy leverages prior observations in which the N and C termini of MBP were fused to FRET-paired fluorophores [22, 62, 63, 129], split GFP fragments [52], or context-dependent fluorophores [109], each enabling the monitoring of ligand binding-induced conformational changes in MBP. Further rationale for this strategy is that ZFPs exhibited conditional DNA-binding when this domain was genetically split and then reconstituted using either self-splicing inteins or protein-protein interactions [96, 130]. The second strategy we explored was termed the Split Protein (SP) approach, in which MBP was genetically split, with the halves fused to the N and C termini of intact BCR-ABL1. We hypothesized that such a construct may permit ZFP-DNA interactions in a manner that depends upon whether MBP is bound to maltose. The three most repressible reporters (Go66, Go85, and Go92) were used to evaluate the feasibility of each of these proposed biosensor mechanisms.

To investigate the SZF strategy, BCR-ABL1 was split between the first and second zinc fingers as previously described [96] (see Figure 2.4A for the proposed mechanism). None of the reporters evaluated exhibited repression upon the induction of SZF biosensor expression (Figure 2.4B). One potential explanation is that the three zinc fingers could not localize with the spacing or geometric orientation required to simultaneously bind a single BCR-ABL1 DNA-binding site. While it may be possible to improve SZF biosensor-mediated repression by identifying promoters that place BCR-ABL1 binding sites in geometries that enable a split ZFP to bind, there is no guarantee that such a configuration would exist. Moreover, even if such a promoter were identified, such a design is likely to be a "one-off" solution specific to the MBP SZF biosensor. Therefore, we next evaluated the alternative SP biosensor design strategy, which has the potential to be more generalizable.

To initially investigate the SP strategy, MBP was genetically split at the point previously reported to generate a functional chimera with beta-lactamase (BLA), termed "RG13" [53], and BCR-ABL1 was inserted between these N and C terminal fragments (see Figure 2.4C for the proposed mechanism). We hypothesized that such a split site might support the SP mechanism because (a) in RG13, MBP retains the capacity to bind maltose and (b) in RG13, binding of maltose to MBP likely induces a conformational rearrangement of the overall fusion protein (or at least the stabilization of a conformation that alleviates disruption of the BLA active site [131]). This SP biosensor was again evaluated against a select set of reporters. Notably, induction of biosensor expression suppressed reporter output in a manner that was significantly alleviated in the presence of maltose (Figure 2.4D). These data thus support the fundamental feasibility of the SP strategy. This observation is also consistent with our hypothesis that DNA-binding was impaired in the SZF architecture, at least compared to the SP architecture. Notably, repressibility conferred by the SP

biosensor was only slightly less than that observed with BCR-ABL1 ZFP alone (Figure 2.4E). This correspondence suggests that the rules for predicting robust ZFP-mediated repression of reporter output (Figures 2.2, 2.3) seem to hold true for the SP biosensor, which also supports the generalizability of the SP strategy with respect to reporter construct design. As was observed for BCR-ABL1-mediated repression (Figure 2.1D), SP biosensor-mediated repression and alleviation was also unimodal when analyzed by flow cytometry (Appendix A2.6).

Given these promising initial results, we next investigated how the method of SP biosensor implementation impacts the performance of the system. We first investigated how biosensor expression levels impact performance (Figure 2.4F). With increasing biosensor expression levels, repression of the promoter increased (red series), although maltose-mediated alleviation of reporter output also decreased (blue series). Such a trend is the expected behavior of this system if, at higher levels of biosensor expression, intracellular concentrations of maltose are insufficient to drive all of the biosensors into the ligand-bound (alleviated) state. We would propose two hypotheses that could explain this phenomenon. First, at higher concentrations of maltose, transport limitations may confer an upper bound on the intracellular concentration of maltose, irrespective of the extracellular maltose concentration. Second, it remains possible that even at saturating intracellular concentrations of maltose, the maltose-bound form of our biosensor may still bind DNA (and repress transcription) to a lesser but finite extent, compared to the apo form of the biosensor. Thus, we determined that for the case of a relatively high extracellular concentration of maltose (100 mM in medium, which is expected to be substantially higher than the concentration in the cytoplasm), induction of biosensor expression with 30 μM IPTG conferred a robust balance between repression and maltose-mediated alleviation of reporter output. To exclude the possibility that some other aspect of cell biology or ZFP function could explain the maltose-mediated

alleviation of reporter output, a similar analysis was performed using cells expressing the ZFP alone (i.e., in place of the SP biosensor). As expected, the addition of maltose had no significant impact on ZFP-mediated repression of the reporter (Appendix A2.7). Moreover, the addition of maltose did not substantially impact cell growth (Appendix A2.8). We next investigated how extracellular maltose concentration impacts biosensor performance (Figure 2.4G). With the caveat that intracellular concentrations of maltose are expected to be substantially lower than are extracellular concentrations, we observed that biosensor responsiveness varied with extracellular maltose concentration, and for the most sensitive reporters (Go92 and Go85), modest but significant alleviation was observed at extracellular maltose concentrations as low as 0.1 mM. To further investigate how maltose binding to MBP domains impacts SP biosensor performance, we repeated this dose-response analysis after making a mutation in the MBP domain (W340A), which has been reported to substantially weaken, but not ablate, MBP binding to maltose [132] (Figure 2.4H). As expected, a substantially higher (~400x) extracellular concentration of maltose was required to alleviate reporter output in the W340A SP mutant case. Altogether, these observations demonstrate the feasibility of engineering novel biosensors using the SP strategy. Although the MBP split site used in this initial construct was not selected to optimally implement the SP strategy, our overall goal was to evaluate the SP strategy in general. Therfore, we carried this functional SP biosensor forward for further development.

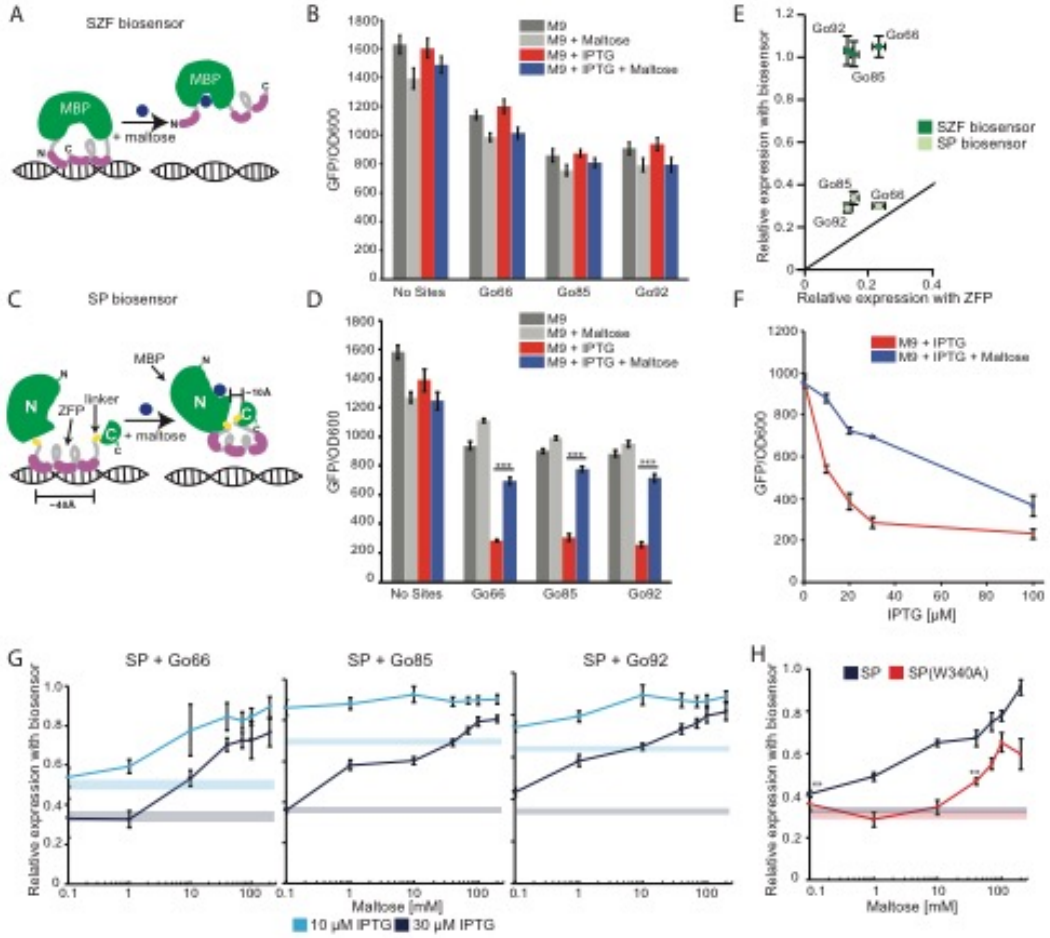


Figure 2.4 Engineering novel biosensors using the split zinc finger (SZF) and Split Protein (SP) strategies
(A) This cartoon illustrates the proposed mechanism of action of an SZF biosensor. Gray loops indicate the ZFP secondary structure, and purple regions indicate α-helices that mediate DNA recognition. (B) SZF biosensor performance, when paired with the reporter plasmids indicated, was evaluated by inducing biosensor expression (30 μM IPTG) and evaluating alleviation of repression upon addition of maltose (100 mM). (C) This cartoon illustrates the proposed mechanism of action of an SP biosensor. The small yellow regions indicate the positions of linker amino acids. Gray loops indicate the ZFP secondary structure, and purple regions indicate α-helices that mediate DNA recognition. (D) SP biosensor performance was evaluated as in panel B. (E) Comparison of the repression (i.e., reduction of relative expression) of reporter output upon expression of the SZF and SP biosensors compared to that mediated by inducing expression of the BCR-ABL1 ZFP alone. (F) Tradeoff between level of expression of the SP biosensor (IPTG dose) with both repression (-maltose) and alleviation (+100 mM maltose) of expression from the Go66 reporter. (G) Response of reporter output to various *extracellular* concentrations of maltose, under two levels of biosensor expression (IPTG doses). Light and dark blue horizontal bars indicate the 0 mM maltose case (for 10 μM or 30 μM IPTG, respectively), with the width of each bar indicating one standard deviation. (H) The impact of the W340A mutation, which is reported to diminish maltose binding [132], on biosensor performance was evaluated using the Go92 reporter and analyses paralleling those used in panel G. Colored bars correspond to the indicated biosensor with 0 mM maltose, with the width of each bar indicating one standard deviation. Microplate data were collected over 7 sequential time points, spanning ~1.5 h of mid-exponential phase growth, and averaged. Relative expression was utilized in order to implicitly correct for any minor effects that IPTG or maltose many confer on GFP/ OD_{600} in a manner that is unrelated to expression of the ZFP or biosensor (see Materials and Methods for details). All data points represent mean values calculated from two independent experiments, each run in biological triplicate, and error bars represent one standard deviation (**$p \leq 0.01$, ***$p \leq 0.001$).

2.5.4 Contributions of biosensor biophysical properties to biosensor performance.

In order to investigate how general biophysical properties of a biosensor impact its performance, we next performed a series of rational modifications of the SP biosensor protein. First, we hypothesized that if DNA-binding affinity limits the degree to which our biosensors repress transcription, then replacing the BCR-ABL1 domain with a ZFP that binds to DNA with higher affinity would improve transcriptional repressibility in the absence of maltose. However, since BCR-ABL1 interacts with its binding site with a K_d ~78 pM [117], a simple model of binding equilibrium would suggest that promoter occupancy should not vary much with changes in this high affinity binding constant. As a point of reference, we note that the dimeric tetracycline repressor (TetR) binds to its operator sequence (tetO) with a similar K_d ~20 pM [133], although tetR is understood to achieve exquisite transcriptional repression through contorting the target DNA rather than through high affinity binding alone[134]. In order to directly investigate the relationship between affinity and repression in our system, and to investigate the modularity of our biosensor vis-à-vis ZFP domain choice, we replaced BCR-ABL1 with the Zif268 ZFP domain from the human EGR1 protein. Zif268 binds its 9 bp binding site with a K_d ~8 pM (~10 times tighter than that of BCR-ABL1) [117]. Go92 was converted to Zif268-responsive promoter by replacing the BCR-ABL1 binding sites with Zif268 binding sites (GCAGAAGCC versus GCGTGGGCG, respectively). The SP biosensor was also modified to replace the BCR-ABL1 ZFP with Zif268 (SP-Zif268). SP-Zif268 did not exhibit an enhanced capacity to suppress reporter output, although it instead exhibited reduced fold-alleviation in the presence of maltose compared to the original SP biosensor (Figure 2.5A). Altogether, these observations are consistent with a simple model wherein increasing the affinity of the biosensor for its target DNA did not increase repression,

presumably because such a change would not impact promoter occupancy. Moreover, the fact that the SP-Zif268 biosensor exhibited significant (if somewhat diminished) functionality indicates that the ZFP domains within SP biosensors may be exchanged in a modular fashion.

We next investigated how biosensors size may impact reporter repression, for example by sterically occluding RNA polymerase binding to the -10 box and -35 region. To this end, the fluorescent protein mCherry was fused to either the N- or C- terminus of the SP biosensor to generate mC-SP or SP-mC, respectively. In this experiment, mCherry was selected as a functionally "neutral" fusion partner in order to investigate the impact of increasing the bulk of the biosensor alone. Although the SP-mC modification did not improve biosensor performance, the mC-SP construct notably exhibited both improved repression and increased fold-induction of reporter output upon the addition of maltose (Figure 2.5). Altogether, these data suggest a useful strategy for building novel biosensors in which the tradeoff between desired performance characteristics may be optimized for a particular application. In general, adding steric "bulk" may outperform enhancing DNA-binding for increasing repression in the "off" state without sacrificing the degree to which the regulated gene is expressed when in the "on" state.

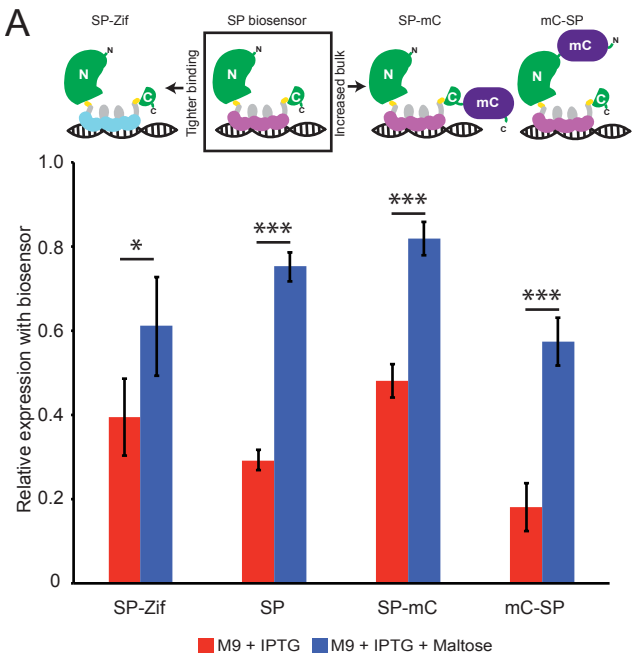


Figure 2.5 Contributions of biosensor biophysical properties to biosensor performance
(A) At top, the illustration summarizes the biosensor design space explorations described in this figure, using the SP biosensor as a reference case. SP-Zif268 incorporates the tighter binding Zif268 ZFP in place of BCR-ABL1. In SP-mC and mC-SP, mCherry was fused to the C-terminus or N-terminus of the SP biosensor, respectively. Below, biosensor performance, when paired with the Go92 reporter, was evaluated by inducing biosensor expression (30 μM IPTG) and evaluating alleviation of repression upon addition of maltose (100 mM). Microplate data were collected over 7 sequential time points, spanning ~1.5 h of mid-exponential phase growth, and averaged. All data represent mean values calculated from two independent experiments, each run in biological triplicate, and error bars represent one standard deviation (*$p \leq 0.05$, *** $p \leq 0.001$).

2.6 Discussion

In this study, we investigated a potentially generalizable strategy for converting metabolite-binding proteins into metabolite-responsive transcription factors. By systematically and quantitatively evaluating the design principles governing the performance of such biosensors, which was the focus of this investigation, this work establishes a foundation for pursuing the long-term goal of engineering repertoires of customized metabolite-responsive biosensors. By leveraging modular design of both promoter libraries and biosensor proteins, these investigations elucidated a number of design principles that are useful for both explaining the variations observed in our libraries and for guiding the design of novel biosensors in subsequent work.

Using a library of engineered promoters, we identified several important rules by which binding of a ZFP to DNA confers a repression of transcription. Interestingly, nearly all promoters evaluated were repressed, at least to some degree, and none exhibited increased expression in the presence of the ZFP. The BCR-AB1 ZFP alone was sufficient to achieve significant transcriptional repression, even though this protein is smaller than canonical natural transcription factors, such as TetR and LacI (106 aa compared to 221 aa (TetR) and 374 aa (LacI)). This minimal ZFP also regulated gene expression in manner somewhat different from that conferred by a previously described fusion between a ZFP and a transactivation domain from CRP. Lee *at al.* observed that this ZFP-CRP fusion conferred transcriptional activation when bound upstream of the +1 site and repression when bound downstream of the +1 site [102], while our minimal ZFP (which lacks a transactivation domain) conferred repression even when bound upstream of the +1 site (Figure 2.2). We determined that placing ZFP binding sites as close as possible to the consensus -10 box and -35 region of the promoter yielded the highest level of transcriptional repression. The -10 box and -35 region are the sites at which the transcription initiation factor σ70 binds in order to mediate recruitment and assembly of the RNA polymerase (RNAP) complex. Therefore, we hypothesize that placing the ZFP binding sites very close to the -10 box and -35 region effectively prevents the σ70 from binding to this region of DNA and/or σ70-mediated recruitment of RNAP. The greatest repression was observed when both the -10 box and -35 region were abutted with ZFP binding sites, and blocking the -10 box may confer greater repression than does blocking the -35 region (Figures 2.2G, 2.3D). It is possible that these observations may be leveraged to achieve greater repression by overlapping the ZFP binding sites with the conserved -10 box and -35 region, since binding of biosensors to these sites may more efficiently block σ70-mediated recruitment of RNAP. One potential challenge associated with this strategy is the potential to repress endogenous

genes that share the consensus -10 box and -35 regions, although minimizing overlap with these consensus sequences could mitigate this problem.

Our comparison of two potential biosensor engineering strategies – the SP (split protein) and SZF (split zinc finger) architectures – revealed several insights into the feasibility and generalizability of each approach. The SP biosensors repressed the most-repressible reporters to nearly the same extent as did the ZFPs alone, suggesting that the rules governing promoter design may be generalizable across SP biosensors (Figure 2.4D). In contract, the SZF biosensor evaluated conferred no repression of reporter output (Figure 2.4B). We hypothesize that insertion of MBP between the fingers of BCR-ABL1 precluded simultaneous binding of DNA by all three zinc fingers. Even if this geometric constraint were alleviated by modulating the protein or DNA sequences, it is likely that such a solution would be unique to each biosensor. Therefore, the SZF approach may be generalizable, but not readily so. In contrast, the SP approach was both more effective and may also be more readily generalizable.

Our investigation also provided several insights into the mechanism by which this initial SP biosensor functions and the prospects for extending this approach to generate novel biosensors. In many ways, these insights leverage the wealth of information available to describe our model ligand-binding domain, MBP. The SP biosensors utilized the MBP split sites that were identified by using a random domain insertion approach to generate the "RG13" MBP/BLA fusion protein [53]; the N terminal half of SP comprises the first 316 aa of MBP, and the C terminal half comprises residues 319-370 of MBP. In the crystal structures of both MBP and RG13, residues 316R and 319A are ~10 Å apart [125, 131]. However, it should be noted that the RG13 crystal structure was obtained in the presence of saturating Zn^{2+}, a condition which ablated the activity of the BLA subdomain of the protein, and that no maltose-bound (or zinc-free) structure of RG13 has been

obtained. Thus, these distances should be treated as estimates as to how RG13 residues 316R and 319A are positioned when the protein is expressed under physiological conditions. When MBP binds maltose, the separation of these residues increases by no more than ~3 Å [123]. In contrast, when a Cys_2-His_2 class ZFP binds to its cognate 9 bp of DNA, the distance separating the N- and C-termini of the ZFP is ~40 Å [135]. Therefore, we hypothesize that in order for the ZFP domain of the SP biosensor to adopt a conformation capable of binding its 9 bp DNA target, residues 316R and 319A may be separated by as much as 40 Å. Furthermore, since the addition of maltose alleviates biosensor-mediated repression of transcription (and therefore impairs or ablates DNA-binding), we hypothesize that maltose binding to the SP biosensor stabilizes interactions between the split MBP fragments, such that residues 316R and 319A are retained in a close (~13 Å) spacing, which prevents the ZFP domain from adopting a conformation capable of DNA-binding (Figure 2.4C). Importantly, if the SP biosensor operates via this ligand binding-induced stabilization mechanism, then biosensor function need not rely upon a ligand-binding induced conformational change in MBP. Thus, this mechanism could be extended to ligand-binding proteins that do not experience a ligand binding-induced conformation change as dramatic as that exhibited by MBP. Moreover, the proposed ligand binding-induced stabilization mechanism is consistent with the "induced fit" model of substrate binding, in which ligand binding causes a shift in protein structure that results in an increase in the stability of the ligand-bound complex. Indeed, ligand binding-induced stabilization may confer allosteric regulation of many proteins, and this property may even be engineerable [136]. Thus, we speculate that the mechanism of the MBP-based SP biosensor may be extended to biosensors based upon distinct ligand-binding domains. Moreover, there exist many methods by which proteins can be split, fused, and screened, including *in vitro* methods such as circular permutation and domain insertions, as well as computational methods for predicting

effective split sites, such that evaluating whether a given ligand-binding protein is amenable to conversion into a biosensor using the SP approach is relatively straightforward [54, 137-140]. In fact, periplasmic binding proteins such as MBP may be generally amenable to conversion into molecular switches via domain insertion, as was demonstrated by the insertion of TEM-1 beta-lactamase into ribose binding protein, glucose binding protein, and xylose binding protein[56]. Additionally, an allosterically regulated version of Cas9 has been developed by using domain insertion to fuse Cas9 to estrogen receptor-α to create a repressor that is inducible by the addition of 4-hydroxytamoxifen[106]. In this study, a site in Cas9 that is permissive to protein insertion was first identified using random domain insertion of a PDZ domain. Thereafter, the estrogen receptor-α was inserted into this site following the rationale that this receptor undergoes a substantial conformation change upon ligand binding, bringing its termini within 21 Å of one another in the presence of ligand, such that only the ligand bound conformation of the receptor may exhibit a structure that avoids disruption of the Cas9 structure (and, presumably, function). Thus, while this technology is of substantial utility for regulating Cas9, it is not yet clear whether or how this approach may be extended to generate Cas9-based regulators that are responsive to a range of metabolites or ligands. Altogether, the SP strategy appears to be a promising and potentially generalizable method for generating novel biosensors, although further investigation is required to determine which types of ligand-binding proteins may be most readily converted into biosensors via this approach.

Our investigation also provided several insights into how biophysical properties of the biosensor itself could impact its overall performance. First, comparing SP biosensors based upon BCR-ABL1 to those based upon Zif268, the latter of which binds its cognate DNA with approximately 10-fold greater affinity, we observed that the SP-Zif268 biosensor repressed

transcription to a similar extent but exhibited a reduced response to the addition of maltose. As discussed above, the observed comparable degree of repression is consistent with a simple model of high affinity binding, in which both SP and SP-Zif268 biosensors achieve a similar level of promoter occupancy. To interpret the reduced response to maltose, we hypothesize that due to the tighter binding of Zif268 to DNA, even the maltose-bound state may interact with DNA to some extent that represses reporter output (indeed, the same may be true to a lesser extent for the original SP biosensor). For example, if each maltose-bound biosensor exists in an equilibrium between states that are competent (disfavored) versus incompetent (favored) for DNA-binding, then the higher affinity with which Zif268 binds DNA may cause biosensors based upon this protein to become "trapped" in a DNA-bound state, even when bound to maltose. Finally, the fact that the SP-Zif268 biosensor nonetheless exhibited significant (if somewhat diminished) functionality indicates that, within the SP framework, the ZFP domains may be exchanged to tune biosensor performance or to regulate novel reporter constructs.

We also investigated the role of biosensor size on performance, which provided some insights into how biosensor performance may be tuned. We observed that fusing mCherry to the N terminus of the SP biosensor (mC-SP) improved both reporter repression and fold induction upon the addition of maltose, although no such effect was observed when mCherry was fused to the C terminus of the SP biosensor (SP-mC). While it is not possible to provide a specific structural explanation for these effects, a reasonable speculation is that the mC-SP biosensor sterically occludes recruitment of the RNAP to a greater extent than does the original SP biosensor. If this were true, it could be possible to achieve even greater repression of reporter expression by exploring the addition of "bulky" domains of various sizes, shapes, and linker geometries to a candidate SP biosensor.

We also attempted to compare the performance of our initial SP biosensors to that of some naturally-evolved biosensors, using the systematic characterization of the latter that was recently reported by Rogers *et al.* [3]. Rogers et al. evaluated fold-induction after cells had reached stationary phase, while we evaluated both repression and alleviation during exponential growth, so we re-analyzed our data from the experiments reported in Figure 2.5 using a later time point at which cell growth had slowed, to facilitate this comparison (Appendix A2.9). While this analysis did indeed lead to higher calculated values for degree of repression and fold-alleviation upon the addition of ligand (6 ± 0.4 fold-repression and 3.8 ± 0.3 fold-alleviation for the mC-SP biosensor), the natural biosensors generally achieved a greater fold-induction, due in large part to the more efficient suppression of output gene expression when in the ligand-free "off" state. While such nuances reflect the manner in which biosensor performance is evaluated to some extent, this investigation more importantly identifies specific performance attributes of SP biosensors that might be targeted to better approach the performance of naturally-evolved biosensors.

Although we evaluated and identified several promising strategies for improving biosensor performance, it is possible that when extending the SP approach to target applications, biosensor performance may be further improved by either design-driven or screening-based methods. Some strategies could entail refining reporter design. Depending on the application requirements, fold-induction may be improved by locating the reporters on single-copy plasmids or on chromosomal DNA (instead of on low copy number plasmids as described here) to increase promoter occupancy for a given quantity of biosensors. Alternatively, the promoter sequence could be altered to partially diminish interactions with σ70, potentially using either targeted mutations or random promoter mutagenesis followed by selection to "tune" a promoter to match the properties of a given biosensor. Other strategies could improve the biosensor proteins. In particular, although

utilizing the RG13 split site for MBP proved to be feasible for generating our initial SP biosensors, it is likely that evaluating all possible ZFP insertion sites into a ligand binding protein may identify fusion proteins that are specifically suited to the SP mechanism. Based upon our observations of the factors limiting SP biosensor performance, candidate biosensors may also be improved by random mutation and directed evolution (e.g., optimizing allosteric regulation to enhance ligand binding-induced alleviation of DNA-binding). A final strategy could be to process the output of our existing biosensor/reporter(s) system to achieve preferable overall performance characteristics. For example, reporter output could be coupled to additional genetic circuitry, such as RNA-based toe hold switches or positive feedback circuits to amplify reporter output, and with some tuning, increase fold-induction[33, 141]. By leveraging the modularity conferred by programmable ZFP binding [117, 135], it may also be possible to implement multiple SP biosensors in a single cell. Moreover, high throughput genome engineering approaches such as MAGE [142] could make it possible to place even endogenous genes under partial or total control of such engineered biosensors. In sum, a modular approach to biosensor engineering is likely to accelerate the generation of novel biosensors, iterative improvement of biosensor performance, and adaptation of biosensors for novel applications in metabolic engineering and synthetic biology.

2.7 Acknowledgements

Flow cytometry experiments were conducted at the Robert H. Lurie Flow Cytometry Core Facility. Traditional sequencing services were performed at the Northwestern University Genomics Core Facility. This work was supported by the National Science Foundation (MCB-1341414 to JNL); the Environmental Protection Agency (STAR Fellowship F13A30124 to AKDY); and Northwestern University (Undergraduate Research Grant to NCD; Summer Research

Opportunity Program Fellowship to AGR). AKDY was supported in part by National Institutes of Health T32 Training Grant GM 008449 through Northwestern University's Biotechnology Training Program.

Chapter 3. Development of novel metabolite responsive transcription factors via, transposon mediated, high throughput protein fusion

3.1 Context

In the previous chapter, we successfully demonstrated the conversion of a metabolite binding protein into a metabolite responsive transcription factor using maltose binding protein and a zinc finger DNA-binding domain. In principle, the split protein (SP) strategy demonstrated in chapter 2 is generalizable to any protein. However, the amino acid position where MBP was split had been previously published and demonstrated to make a maltose responsive enzymatic biosensor out of MBP and the enzyme TEM1 β-lactamase (bla). Therefore, our next goal was to develop a method by which all possible insertions of the ZFP into any protein could be created and evaluated for biosensing capabilities such that the reliance on previously publish information could be minimized. To accomplish this task, the MuA transposase was utilized for its ability to randomly insert a transposon into a target DNA sequence. The insertions that occurred inside the gene of interest could then be isolated, and the transposon exchanged for the ZFP. This strategy, in theory, could allow us to sample all possible insertions of a ZFP into any gene of interest. Therefore, if the SP strategy for the conversion of metabolite binding proteins into metabolite responsive transcription factors can be applied to proteins other than MBP, this transposase based method would be able to evaluate that. Since we had already developed a biosensor for MBP, any new method for finding biosensors, if successful, should be able to pull out the original split of MBP. Therefore, to test our new transposase based method, we once again turned to MBP as a model system for which we know there is a minimum of one biosensor that can be made via the fusion of MBP with the ZFP. This work was a collaboration between myself and Peter Su. Together

we designed and executed all experiments and wrote the manuscript. This paper would not have been possible with the help of two talented undergraduate research assistants Andrea Shepard and Shreya Udani who helped conduct some of the experiments, and Ted Cybulski who helped Peter Su with the python based NGS analysis pipeline. This paper will be submitted to Protein Engineering, Design and Selection.

3.2 Abstract

Metabolic engineering has benefitted from using naturally occurring metabolite biosensors to dynamically regulate and balance heterologous pathways. However, the pool of biosensors that can accomplish this is small, and there exist many metabolites for which a biosensor does not exist. To address this challenge and take advantage of the wealth of metabolite binding proteins that exist, we developed a high-throughput method for Biosensor Engineering by Random Domain Insertion (BERDI). Our approach takes advantage of an unbiased *in vitro* transposon insertion reaction to examine all possible insertions of a DNA-binding domain into a metabolite-binding protein and uses FACS to sort for functional biosensors. The advantage of this approach is that it efficiently evaluates all possible metabolite-responsive transcription factors stemming from a parent metabolite binder with a DNA-binding domain. We developed and evaluated this method by creating a library of insertions of a zinc finger DNA binding domain into maltose binding protein, characterized the insertional landscape of the library, and ultimately discover several functional biosensors. Our results validate a generalizable method that may be applied towards converting a wide range of metabolite binding proteins into novel biosensors for applications in metabolic engineering and synthetic biology.

3.3 Introduction

Metabolite biosensors have a wide variety of uses, from basic research and discovery, to diagnostics, and application-driven biotechnology [79]. There are many classes of metabolite biosensors that can accomplish these tasks, ranging from fluorescent and FRET based biosensors [6, 15], to RNA based biosensors [16, 17], and transcription factor biosensors [1, 9, 11]. Transcription factor biosensors have proved to be especially powerful as they enable basic metabolic profiling such as monitoring the levels of glucarate [3] and malonyl-CoA [7], in additional to high-throughput screening of large genetic libraries for 1-butantol, succinate, and adipate [4], benzoic acids [85], and L-Lysine [40]. Finally, transcription factor biosensors have demonstrated substantial utility to balance metabolic flux to increase production titers and yields by implementing dynamic feedback control of pathway intermediates for the production of lycopene [44], fatty acid ethyl ester [5], amorphadiene [12], 1-butanol [4], and malonyl-CoA [13].

However, these examples rely on the existence of naturally occurring transcription factor biosensors. New transcription factor biosensors have been generated through the chimeric fusion of the ligand binding domain from one transcription factor to the DNA binding domain from a different transcription factor, however, these biosensors are generally limited to families of structurally related transcription factors such as the LacI/GalR family to preserve ligand responsive allosteric regulation [87, 88]. Additionally, metabolite binding proteins have been fused to known transcription factors such as AraC and Gal4 to rewire the transcription factor's natural regulation [66, 89]. Finally, the binding pocket of transcription factors such as LuxR [90, 91], AraC [41, 92, 93], and XylR [94], have mutagenized and evolved to bind new ligands; however, this was limited to structurally similar ligands. In all cases mentioned above, a naturally occurring metabolite

responsive transcription factor was a required starting point to generate a new biosensor. However, there exist many metabolites for which transcription factors are not known, or cannot be evolved.

Fortunately, nature has evolved a wealth of transporters and enzymes that sense and subsequently bind an enormous range of metabolites; they don't however, regulate transcription. Recently, a strategy for fusing the native maltose transporter in *Escherichia coli*, maltose binding protein (MBP), with a zinc finger DNA binding domain (ZFP), was implemented to generate a maltose responsive transcription factor [97]. The zinc finger was fused internally at a split of MBP (316R) that was previously optimized for the fusion of MBP and TEM1 β-lactamase (bla) [53]. However, for this to be a generalizable technique, we need to be able to generate novel biosensors without the reliance on previously published information. Therefore, we developed a general method by which all insertions of a ZFP into a ligand binding protein could be generated in a high-throughput fashion, and be screened for functional biosensors.

This method, Biosensor Engineering by Random Domain Insertion (BERDI), utilizes the MuA transposase to insert a transposon nonspecifically into a given plasmid encoding the metabolite binding protein. Transposon mutagenesis provides a simple and efficient method for generating a library of insertions due to its nonspecific and single insertion into each target DNA molecule, minimal scar sequence, and ability to generate > 10^5 variants in a single pot *in vitro* reaction. The transposon is later simply exchanged for a ZFP coding sequence to generate a library of potential biosensors. This method has been used to successfully circular permute proteins as well as profile proteins for permissible insertion points [57, 106, 138, 143, 144]. Here, we demonstrate that this high-throughput, protein fusion technique can be used to generate a library of fusions between a ZFP and MBP to test if any other sites in the MBP can accept the ZFP insertion in a manner that produces a maltose responsive transcription factor biosensor. By combining this library with a

previously described zinc finger responsive fluorescent reporter [97] and FACS, we successfully enriched for three novel maltose-responsive biosensors. The generalizability of this method enables the rapid and high-throughput creation of new biosensors from metabolite binding proteins.

3.4 Materials and methods

3.4.1 Bacterial strains and culturing

All experiments were conducted in DS941 Z1 Escherichia coli cells (AB1157, recF143, $lacI^q$ lacZ ΔM15, P_{laciq}-LacI, P_{N25}-TetR). Cells were maintained in Lysogeny Broth (LB) Lennox formulation (10 g/L of tryptone, 5 g/L of yeast extract, 5 g/L of NaCl) supplemented with appropriate antibiotics (Ampicillin 100 μg/mL, Kanamycin 50 μg/mL, and or Chloramphenicol 34 μg/mL). All experimental analysis was conducted in M9 minimal media (1X M9 salts, 0.2% Casamino Acids, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1 mM Thiamine HCl) containing glycerol (0.4%) as the primary carbon source. Variable amounts of isopropyl β-D-1-thiogalactopyranoside (IPTG) were added as indicated to induce biosensor expression. Maltose monohydrate was added to the media at a final concentration of 100 mM.

The biosensor expression vector was built using standard molecular biology techniques using synthetic parts generously provided by Jim Collins (MIT)[111]. The green fluorescent protein (GFP) reporter plasmid driven by the pGo92 zinc finger responsive promoter was used a previously described [97]. Custom RBS sequences for the biosensor and reporter plasmids were designed using the RBS calculator [112]. The camR and sacB ORF was acquired from pKM154 as a gift from Kenan Murphy [145] (Addgene plasmid #13036) and cloned into a storage vector containing MuA transposon recognition sequences, flanked by BglII restriction sites (pAY438). The BCR-

ABL1 zinc finger protein was subcloned into a storage vector flanked by NotI restriction sites (pAY437). Description of all plasmids used in this study can be found in Supplemental Table B1.5, and electronic plasmid sequence files are included in the supplementary materials.

3.4.2 Library construction and transposition reactions

The MuA transposase inserts the transposon randomly, and in either a forward or reverse direction. Furthermore, the transposon can be inserted in any of the 3 possible codon frames in MBP. Therefore, three frames multiplied by a forward or reverse insertion yields six possible insertions for a given codon of target DNA, but only one combination (forward and in-frame) will produce a *productive* insertion. The transposase also leaves a partially controllable scar, therefore to keep the ZFP in frame with the rest of MBP, and to avoid undesirable scar products, a three alanine linker on either side was selected. A detailed description of the transposon sequence and potential scar options can be found in Supplemental Figure B1.1.

A transposon conferring chloramphenicol resistance as well as containing the *sacB* gene for negative selection with sucrose was digested out of a storage plasmid (pAY438) using BglII, gel extracted, and subjected to an ethanol precipitation to maximize purity. *In vitro* transposition reactions were carried out according to the protocol within Thermo Scientific's Mutation Generation System kit (catalog # F701). 100 ng of the purified transposon was combined with 200 ng of the target plasmid containing MBP (pAY447), and incubated with the MuA transposase for 4 hours at 30^{o} C. Care was taken to use clean nuclease-free water, as the reaction is sensitive to contamination. After the reaction was heat inactivated (10 min at 75 ^{o}C), a PCR cleanup was conducted on the reaction contents. The entire contents were then electroporated into two tubes (~250 uL each) of electrically competent *E. coli* cells. The cells were selected on plates containing

chloramphenicol (transposon) as well as ampicillin (plasmid backbone) to reach ~8 x 10^5 colonies, over 100x the possible library size of 6,288 (the number of directions the transposon can insert in, 2 by the total length of the target plasmid, 3144 bp). A summary of these values can be found in Supplemental Table B1.6. Serial dilutions were made at each cloning step and extrapolated to estimate library size. The MBP gene was digested out with restriction enzymes KpnI and SphI and gel purified to size-select for genes that had a successful insertion, and moved to an expression plasmid under the control of a lac-inducible promoter pTrc (pAY431). Restriction digestion using the NotI site present in the transposon scar was used to replace the transposon with the ZFP gene, and the resulting ligation was transformed into competent *E. coli* cells that already contain the zinc finger responsive GFP reporter plasmid (pAY430). Cells were selected with ampicillin and kanamycin for both plasmids as well as 10% sucrose to maximize loss of the transposon. See Figure 3.1 for a visual description of this process.

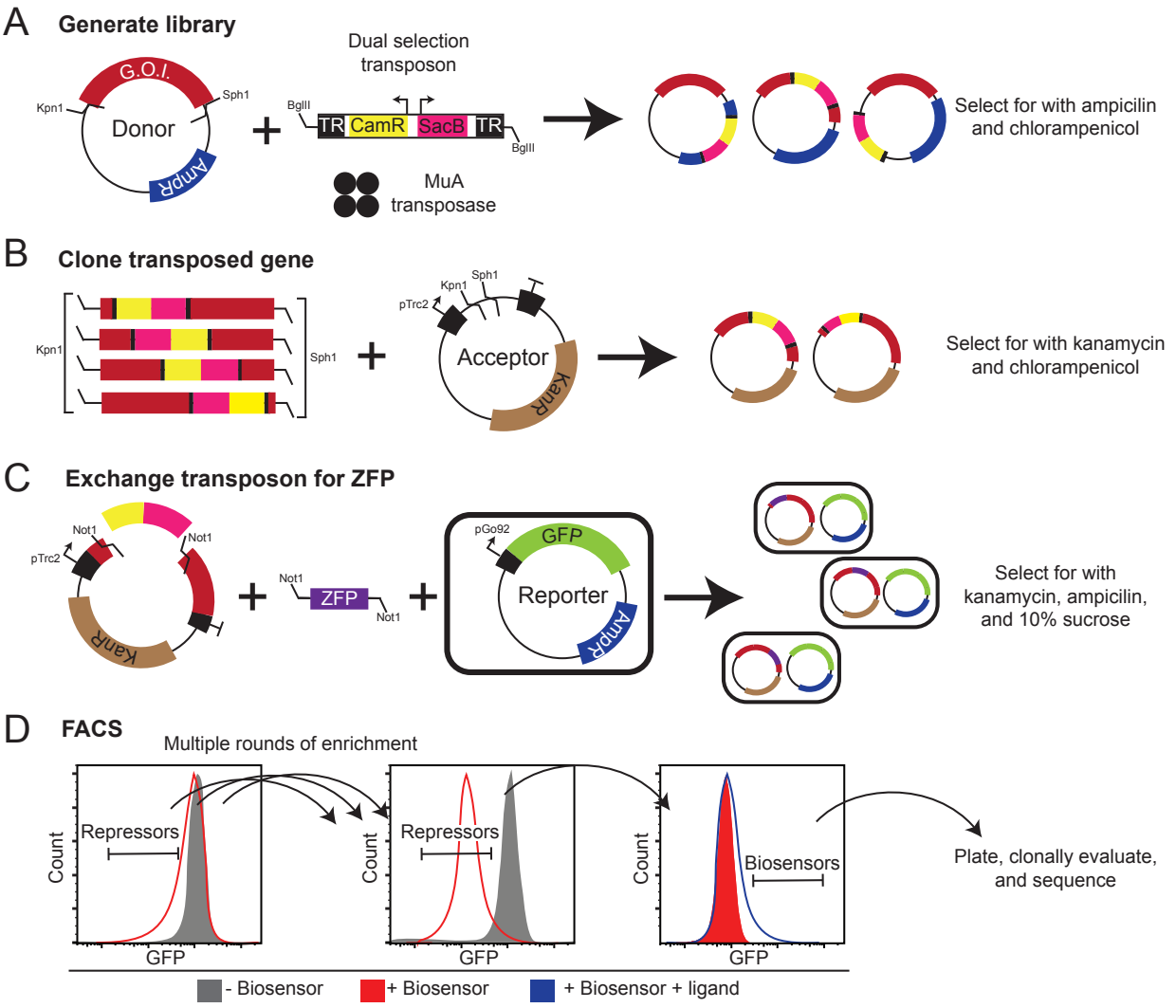


Figure 3.1 Transposon based method high throughput generation of fusion proteins.
(A) Library generation. The donor plasmid containing the gene of interest, the transposon, and the transposase enzymes are incubated for 4 hours at 30° to create a library of random insertions. The plasmid pool is transformed and selected for using chloramphenicol and ampicillin markers. (B) Cloning of the transposed gene. Gene containing the inserted transposon is isolated using the restriction enzymes KpnI and SphI, gel extracted, and cloned into a similarly digested expression plasmid. This pool of plasmids is transformed and selected for using kanamycin and chloramphenicol markers. (C) Exchanging the transposon for the ZFP. The transposon is replaced with ZFP using the NotI sites, and is the resulting plasmids are transformed into DS941 cells containing the reporter GFP plasmid. Cells containing the biosensor plasmid, without the transposon, and the reporter plasmids are selected for using kanamycin, ampicillin, and 10% sucrose. (D) Cartoon of potential enrichment strategy for metabolite responsive biosensors using FACS. The gray distribution represents the biosensor and reporter plasmid's GFP expression prior to induction of the biosensor protein. The red distribution is the GFP expression with the biosensor after induction. The blue distribution is the GFP expression with the induced biosensor in the presence of ligand.

3.4.3 Gel electrophoresis and ImageJ analysis

All gel electrophoresis experiments were conducted with a 1% agarose gel and run in 1x TAE (tris acetate EDTA) at 120 volts. DNA was stained using SYBR Safe (Thermo Scientific) and imaged under blue light. Approximate band sizes were estimated using a 1 kb ladder (New England BioLabs). Plot profiles of resulting gel images were analyzed using ImageJ's plot profile function. Gray values for each lane were background subtracted from an empty gel lane.

3.4.4 Microplate-based fluorescence assays and analysis

Cultures were inoculated from single colonies into 2 mL of M9 media and grown overnight to stationary phase. Overnight cultures were diluted 1:10 and grown for 1-2 h (OD600 ~ 0.5). Cultures were again diluted 1:10 (OD600 ~0.05), plated in black-walled clear bottom 96-well plates in biological triplicate, and induced with 30 μM IPTG and or 100 mM maltose. Plates with lids were incubated and shaken in a continuous double orbital pattern at 548 cpm (2 mm) inside a BioTek Synergy H1 plate reader for 10 h with GFP fluorescence and OD600 absorption measurements taken every 15 min. Monochrometer settings were 485/515 nm for GFP.

3.4.5 Flow cytometry and fluorescence activated cell sorting (FACS)

Overnight cultures were diluted 1:10 and grown for 1-2 h (OD600 ~ 0.5). Cultures were again diluted 1:10 (OD600 ~0.05) in either M9 media, or M9 media containing 100 μM IPTG. Cultures were grown for 4 h post-induction prior to FACS sorting. Cells were then diluted down to a concentration of 10^7 cells/mL in 4° PBS.

Sorting was performed on a BD FACS Aria II instrument (BD Biosciences, San Jose, USA) using an 85 μm tip with a 488 nm excitation laser and a FITC emission filter (530/30 nm). This FITC channel was used for analysis of GFP expression. Cells were first gated based on forward and side scatter, then the population of single cells were plotted on a GFP histogram. Uninduced cells with the constitutive level of GFP expression were analyzed prior to sorting (100,000 events), and gating was set such that no more than 1% of this population would be selected for. 100 μM IPTG was used to achieve maximal expression of the library to assay for members that can repress the GFP reporter. The gating threshold was set such that at most 1% of the uninduced population was used to sort the induced population for repressors. Cells were sorted into 3 mL of M9 minimal media containing ampicillin and kanamycin. For each round of sorting, 100,000 cells were collected in this media and subsequently inoculated into 75 mL of M9 and grown overnight at 37°C. Subsequent sorts were done the next day using this sorted population, repeating the steps above.

Traditional flow cytometry was performed on a LSRII flow cytometer (BD Biosciences, San Jose, USA). For all flow cytometry analyses, mean fluorescent intensity was calculated based on the GFP histograms of single cells (gated by forward and side scatter) using FlowJo Software (Tree Star)

3.4.6 Next generation sequencing and analysis

The naïve library of fusions was first digested out of the expression vector using KpnI and SphI. This DNA fragment was gel extracted and subjected to probe sonication to shear the library into fragments less than 500 bp. Library preparation was done by PCR amplifying four equally sized regions of MBP, with common sequences attached to the primers. The PCR primers were

designed to bind either the MBP or ZFP sequence, and an MBP primer was always paired with a ZFP primer for library preparation such that the insertion site could be determined from each amplicon. In total, 8 unique PCRs were run on the naive library: (four evenly spaced MBP primers) x (forward or reverse ZFP primer). See Supplemental Figure B1.2 for a visual layout of the primers and Supplemental Table B1.7 for a list of PCR reactions. Two biological replicates of the library were prepared for a total of 16 total reactions. Samples were then submitted to the University of Illinois at Chicago (UIC) Sequencing Core, where adapters and barcodes were further appended to the amplicons via another PCR reaction. All samples were run together on an Illumina MiSeq lane using paired end DNA sequencing. In total, ~8M reads were generated from the 16 PCRs.

All data analysis was done in with a custom pipeline in Python. Briefly, reads were filtered for presence of the transposon scar sequence as well as for the primers that generated that reads' amplicons. Then, forward reads were filtered for those with a paired-end match of at least 12 contiguous bases, and then filtered out reads that had base calls that fell below a minimum Phred quality score of 20. Next, the reads were assigned to specific bins depending on their read length, in roughly 50 bp ranges, to allow for length-appropriate score filtering Reads were then aligned to the MBP template sequence using the Needleman-Wunsch algorithm (using a gap opening penalty of 10.0, gap extension penalty of 0.5, and the EDNAFULL scoring matrix), and filtered based on length-adjusted alignment scores, as scores generally increase with length of alignment. These alignments were then analyzed for insertion sites based on a contiguous region of alignment to the MBP template via a regular expression search. A graphical representation of this pipeline can be found in Supplemental Figure B1.3. In all cases, the insertion site refers to the last base of MBP upstream (5') of the transposon insertion. All of the NGS analysis code and data is available at https://github.com/PeterSu92/BERDI-NGS-insertion-analysis.

3.5 Results

3.5.1 Generation of random domain insertion libraries via transposon mutagenesis

The target library generation numbers were calculated with the prior knowledge that MuA inserts randomly into target DNA molecules as well as in either forward or reverse, as the transposon recognition sites on either end of the transposon are palindromic. Therefore, for a plasmid of length *n* bases, the total number of possible insertions is 2*n*. Finally, given that the insertions are random and independent of one another, we aimed to achieve at least 10x of this number of colonies in each step to ensure library diversity. All calculations of library sizes can be found in Supplemental Table S2. The first step in the process is the incubation of a plasmid containing MBP with the transposon and transposase enzyme (Figure 3.1A). After the transposition reaction, the library of transposed plasmids is transformed and selected for using ampicillin and chloramphenicol. Our initial transposition library had well over 8 x 10^5 colonies, or ~125x the possible library size (Supplemental Table B1.6). Next, MBP containing the transposon sequence is digested out of the initial plasmid using the restriction enzymes KpnI and SphI and cloned into a expression plasmid containing the Lac promoter pTrc2 (Figure 3.1B). The library was again transformed and selected for successful ligation products using kanamycin and chloramphenicol. The transposon was then excised using NotI and replaced with the ZFP. This ligation was transformed into cells containing the zinc finger responsive GFP reporter and selection for plasmids that have lost the transposon, but have both plasmids, was done by using kanamycin, ampicillin, and 10% sucrose (Figure 3.1C). Both cloning steps in Figures 3.1B and 3.1C also achieved the target 10x oversampling, as shown in Supplemental Table B1.6. The library could then be sorted using FACS for functional repressors, followed by sorting in the presence of ligand for functional biosensors as described by the Figure 3.1D cartoon.

3.5.2 Analyzing diversity landscape of the naïve library

In order to confirm library diversity prior to sorting for functional biosensors, we analyzed the naïve library using three distinct methods. First, the library was digested out of the expression vector and subsequently digested with a restriction site unique to the ZFP (Figure 3.2A). The plot profiles, made in ImageJ, from each lane is plotted in Figure 3.2B. The gel image and plot profile show a distribution of DNA sizes consistent with diverse insertions. To evaluate insertions that were too rare to show up on a gel, we next performed both Sanger sequencing on 46 colonies and next-generation sequencing (NGS) on the library by amplifying regions containing both the MBP and ZFP. The insertions identified by both Sanger and NGS are plotted according to their position in MBP (Figure 3.2C). The insertions found by NGS are plotted on top of MBP in black arrows. The insertions that are unique to Sanger sequencing are plotted below MBP in red, and the insertions that were found common to both methods are plotted in blue. A full list of quantified insertions by Sanger sequencing and NGS can be found in Supplemental Table B1.8.

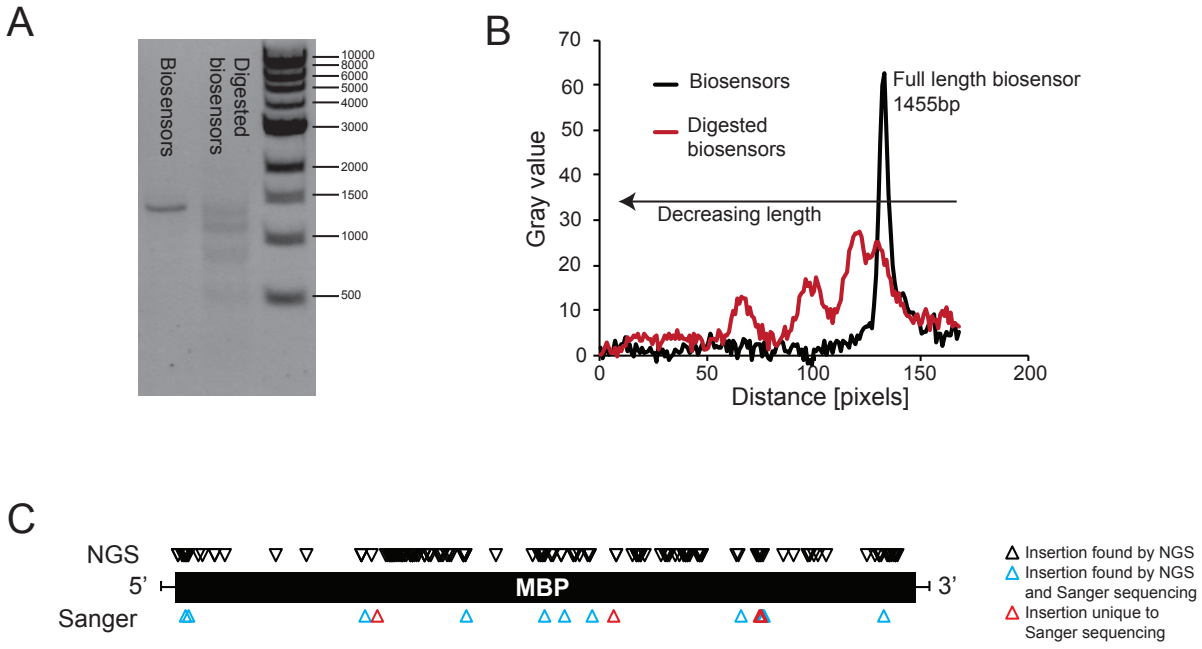


Figure 3.2 Comprehensive analysis of naïve library topology.
(A) Gel electrophoresis image comparing the naive library of biosensors, to the same sample now treated with a restriction enzyme unique to the ZFP. Lane 1, full length biosensors (1455bp). Lane 2, digested biosensors. Lane 3, DNA ladder with the corresponding bp values listed. (B) Plot profile of both lanes from panel A created in ImageJ. (C) ZFP insertion positions into MBP. Insertions found via NGS are displayed in black. Insertions found via traditional Sanger sequencing that were common to NGS are displayed in blue, and unique to Sanger sequencing are displayed in red.

Based on our amplification method and conservative NGS analysis parameters, we can confidently identify at least 153 insertions across MBP, or 13.7% of all possible insertions. Out of these, 48 (31%) were in frame, and 102 (49%) contain a forward-facing ZFP, matching well with the expected distribution from MuA's random insertions. (Table 3.1) We confidently identified 37 *productive* insertions (forward-facing and in-frame) from these analyses. Additionally, 12 insertions (71%) found by Sanger sequencing individual colonies were also present in the NGS analysis. We observe a dynamic range in insertion counts of 10^6 in our NGS data, but the fact that the transposon insertion ratios (forward/reverse, in-frame vs. out of frame) demonstrate proper biochemistry carried out by the transposase (Table 3.1) lead us to believe our library is unbiased.

Additionally, since only 1 of the colonies that were Sanger sequenced showed any of the top 5 most frequent (63%) of the insertions calculated from our NGS, and 5 colonies were not found by NGS, we hypothesize that PCR bias significantly affected our NGS results. This is further backed by the fact that 66% of our insertions were present at 5 counts or less.

Therefore, to determine whether our NGS method produced the same distribution as Sanger sequencing of individual colonies, we performed the Chi-squared statistical test on both the distribution and counts of insertions determined from the both methods (Supplemental Table B1.9). The P-value obtained from this study was $<< 10^{-10}$, affirming that these two distributions are statistically different. The fact that we found 5 unique insertions from 17 total insertions in Sanger sequencing not only contributes to this statistical difference, but also gives us confidence that our library is in fact diverse.

Table 3.1 Naïve library insertion statistics

Condition	Expected if unbiased	Observed by NGS and Sanger
In frame	33%	31%
Out of frame	66%	69%
Forward ZFP	50%	49%
Reverse ZFP	50%	51%

3.5.3 Biosensors found via screening of the naïve library

The naïve library was first sorted for the ability to repress the GFP reporter. The uninduced library was compared to the library that had the biosensor induced with 100 μM IPTG. The gating selection for repressors was determined selecting only the dimmest 1% of the uninduced library, then sorting the induced library with that gate. This process was iterated on three consecutive days to enrich for repressors from the initial library. Next, the population of reversible repressors was sorted for to minimize false positive repressors. Following four rounds of sorting, the cells were plated and individual clones were grown, and assayed for maltose responsiveness, using a microplate reader with 30 μM IPTG and 100 mM maltose. Library members that were both at least a 2-fold repressor as well as maltose responsive were sent for Sanger sequencing (Figure 3.3A). Following sequencing, the insertional position of the ZFP was visualized by highlighting the amino acid that was disrupted by the insertion of the ZFP on the crystal structure of MBP (Figure 3.3B)(PDB# 1ANF)[123]. The previously described reference biosensor, ZFP insertion at 316R, is highlighted in yellow for comparison purposes [97]. Three distinct biosensors were enriched: an insertion at 277A, an insertion of two ZFPs at 270A, and an insertion at 335P. Interestingly, the four insertion points are distributed in three distinct regions of MBP. All four are on the outside of the protein and are either in a loop (270A) or at the end of an α-helix, near a loop (316R, 277A, and 335P). Given the sample size, and the lack of crystal structures of the new biosensors, it is difficult to predict if they share other features that lend themselves to maltose-responsive transcriptional regulation. Performance of the new biosensors was evaluated in comparison to the reference biosensor by flow cytometry (Figure 3.3C). Biosensors were evaluated with 30 μM IPTG to induce the expression of the biosensor and 100 mM maltose to determine the extent of the maltose responsiveness. This IPTG dose lead to the largest maltose sensitivity of the reference

biosensor, therefore it was chosen as a starting point [97]. Both the 277A and 270A biosensors have similar repression compared to 316R (~3 to 4-fold), however, neither are as sensitive to maltose. The 335P biosensor on the other hand is a better repressor (~10-fold) and is substantially responsive to maltose. The enrichment of the 270A double ZFP insertion but not the 270A single ZFP insertion prompted us to generate the 270A single to better understand why the double ZFP, presumably a much rarer ligation product, compared to the single, would have been enriched over single. This double insertion is a potential product (albeit rare) given the cloning to replace the transposon with the ZFP uses the same NotI restriction site on both the 5' and 3' end of the fragment, allowing for double ZFP insertions to be cloned into a single transposon insertional position. The 270A single construct is a milder repressor compared to the 270A double at high levels of IPTG (60 and 100 μM) which explains why it was not as enriched as the 270A double during FACS (Supplemental Figure B1.4). Furthermore, at 30 μM IPTG where the maltose responsiveness was measure on the microplate reader, the 270A double outperforms the single as well. Taken together, this explains why the enrichment strategy would not have enriched for 270A single, but does enrich for the 270A double.

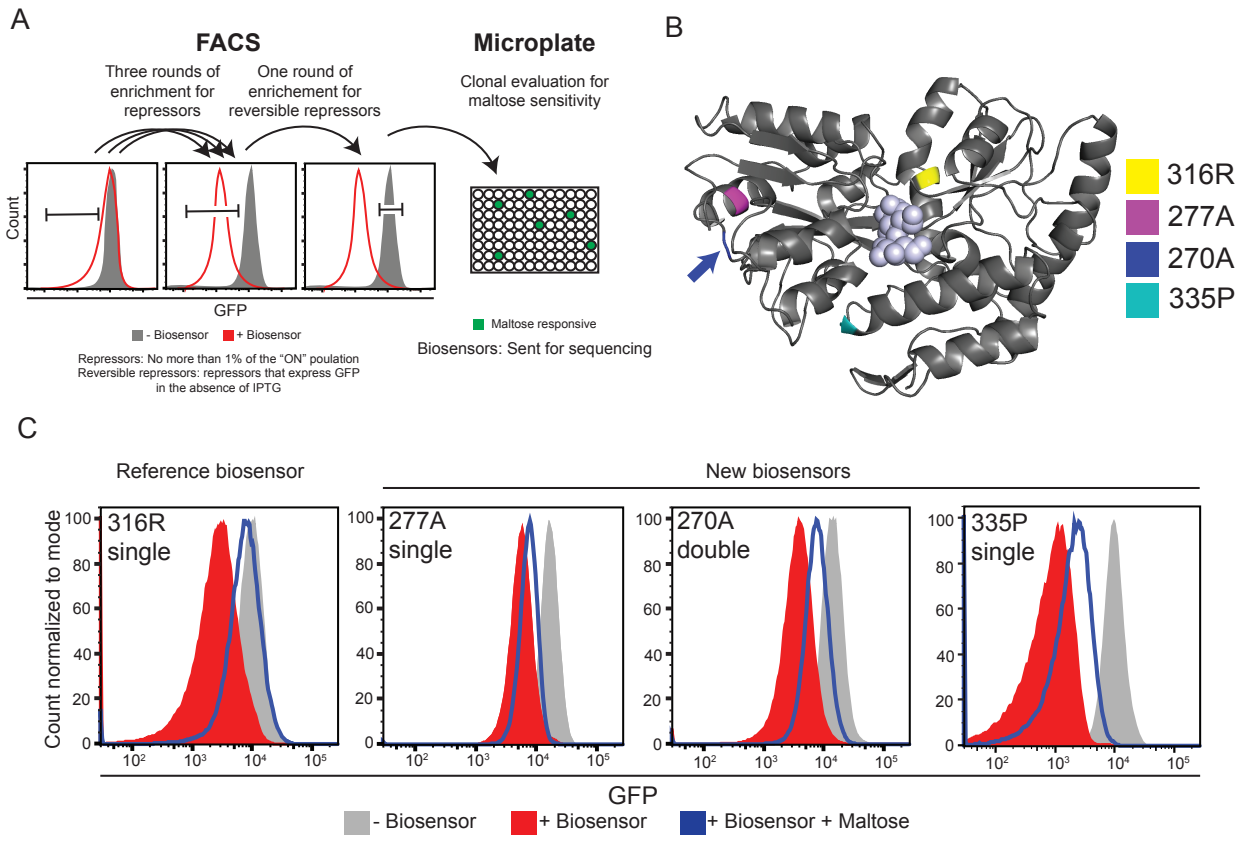


Figure 3.3 Functional biosensors found via screening of the naïve library.
(A) Cartoon of the sorting strategy. Briefly, three rounds of FACS were done, each time isolating no more than 1% of the "ON" population. One subsequent round of sorting was done to isolate only repressors that were reversible. Cells were then plated, and clonally evaluated for maltose responsiveness. Successful biosensors were sent for Sanger sequencing to determine the insertions position. (B) Crystal structure of MBP is shown in gray, with the insertional positions (in amino acid number) of each biosensor labeled. The reference biosensor insertional position is highlighted in yellow, compared to the three new biosensors highlighted in purple, blue, and red. Space filling spheres represent the position of the ligand maltose. (C) Flow cytometry of the reference biosensor compared to the three new biosensors. The gray curve represents the biosensor and reporter plasmid's GFP expression prior to induction of the biosensor protein. The red curve is the GFP expression with the biosensor induced with 30μM IPTG. The blue curve is the GFP expression with the biosensor induced with 30μM IPTG and 100mM maltose. The insertional position (in amino acid number), and whether the ZFP is a single, or double insertion, is listed in the top left corner of each plot. Plots represent a minimum of 10,000 cells in each condition.

3.5.4 Biosensor dose and linker analysis

To investigate the impact of biosensor dose on the overall repression and maltose sensitivity, the strongest repressor, 335P, was induced with a range of IPTG concentrations (Figure

3.4). At a high concentration (60 μM), the repression is not significantly higher compared to at 25 or 30 μM IPTG, indicating that at these lower concentrations, there are already saturating levels of the biosensor to achieve that level of repression. However, at 60 μM IPTG, the system is no longer sensitive to maltose, indicating there is an overwhelming level of biosensor not bound to maltose. As the level of IPTG (and thus biosensor expression) decreases, the sensitivity to maltose increases, indicating there is an optimal window of biosensor expression for maximal maltose sensitivity while maintaining sufficient repressibility around 25 μM IPTG. However, once the IPTG level drops below 20 μM IPTG, the overall repression of the reporter decreases, as expected given the decrease in biosensor protein. The vertical line at 10^3 is a visual aid to ease the comparison of the different conditions.

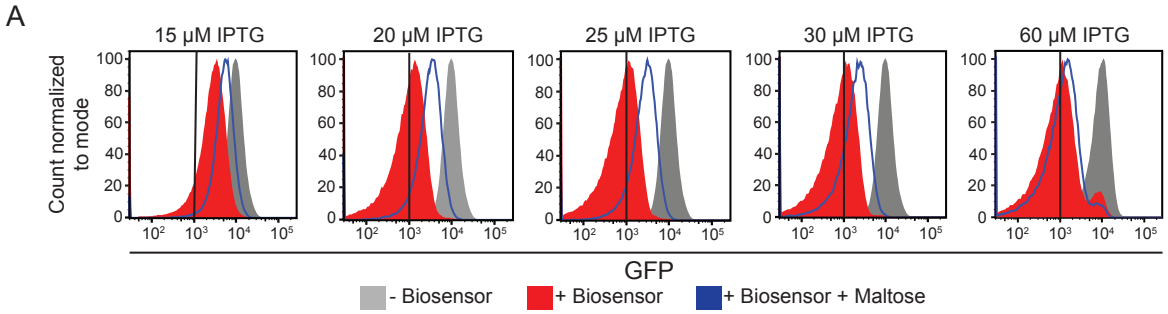


Figure 3.4 Impact of 335P biosensor dose on overall performance.
Response of reporter output to the addition of IPTG and IPTG along with maltose measured by flow cytometry. The gray curve represents the biosensor and reporter plasmid's GFP expression prior to induction of the biosensor protein. The red curve is the GFP expression with the biosensor induced with the indicated amount of IPTG. The blue curve is the GFP expression with the biosensor induced with the indicated amount of IPTG and 100mM maltose.

Intriguingly, the original reference biosensor, 316R was not enriched in the screen. One possible explanation for this is that the linkers introduced by the transposase were altering the performance of the biosensors. The reference biosensor has two amino linkers on either side, lysine and leucine on the 5' end of the ZFP insertion and an asparagine and valine on the 3' end, whereas the transposon would leave three alanines on either side. To investigate the impact of this

difference, the 316R biosensor with three alanines on each side of the ZFP was constructed and compared to the original 316R reference biosensors (Figure 3.5A). Surprisingly, this new reference biosensor with the three alanines had completely lost its ability to repress the GFP reporter. This demonstrates the importance of both the length and composition of the linker sequences to overall biosensor performance. To further investigate how linker length effects performance, three variants of the 335P biosensor were made, each containing a different linker length. The original 335P contains three alanines on each side of the ZFP (3AA), so variants with two alanines (2AA), one alanine (1AA), or no linkers (0AA) were constructed and analyzed via flow cytometry (Figure 3.5B and 3.5C). As the linker length shortened, the repression of the biosensor decreased indicating that the ZFP was less able to bind to the promoter region, presumably due to more constrained nature of the folding of the ZFP portion of the biosensor. The maltose responsiveness varied non-linearly as a function of linker length. The 1AA 335P biosensor offered the best combination of repressive abilities and maltose responsiveness. This is further evidence that the linker length is important to overall biosensor performance, and that changing the linkers results in the alteration of performance.

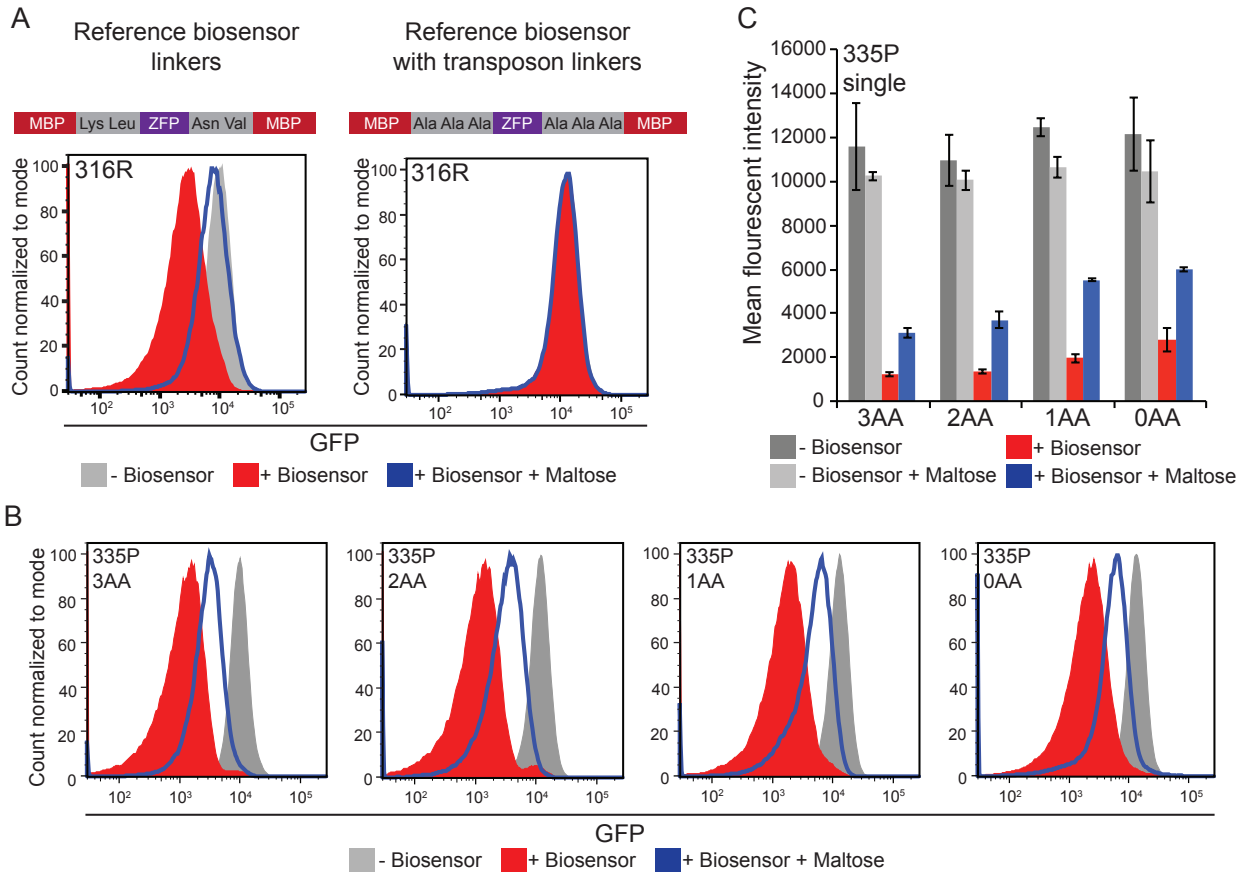


Figure 3.5 Impact of biosensor linkers on overall performance.
(A) Comparison of the effect of amino acid linkers between reference biosensor and its transposon-created counterpart (B) Flow cytometry data of reporter output when the linker lengths on the 335P biosensor are shortened from the wild type 3x (Ala) to 2x Ala, 1x Ala, and 0 AA scarless fusions. The gray curve represents the biosensor and reporter plasmid's GFP expression prior to induction of the biosensor protein. The red curve is the GFP expression with the biosensor induced with 30μM IPTG. The blue curve is the GFP expression with the biosensor induced with 30μM IPTG and 100mM maltose. (C) Mean fluorescence intensity of the four linker variants of the 335P biosensor. The dark gray bars represent the biosensor and reporter plasmid's GFP expression prior to induction of the biosensor protein. The light gray bars represent the biosensor and reporter plasmid's GFP expression prior to induction of the biosensor protein in the presence of 100 mM maltose. The red bars represent the GFP expression with the biosensor induced with 30μM IPTG. The blue bars represent the GFP expression with the biosensor induced with 30μM IPTG and 100mM maltose. Mean fluorescence intensity is averaged from samples run in biological triplicates, and error bars represent one standard deviation.

3.6 Discussion

In this study, we developed and implemented the BERDI method for the generation of maltose-responsive MBP-ZFP fusion proteins in a high-throughput manner to find three new

biosensors. The fact that multiple insertions produced a bi-functional protein is not surprising given that a previous study found twelve functional insertions for a circular permuted GFP into MBP using a similar method [144]. Additionally, another transposon insertion study demonstrated multiple bi-functional insertions of a cytochrome into β-lactamase [57]. A possible explanation for the tolerance of the proteins studied both here and in previous research is that many circularly permuted proteins are able to retain their function, demonstrated in a study that found 15 unique functional circular permutations of an adenylate kinase using transposon mutagenesis [138]. These findings strengthen the need for library based approaches, like the one described here, for developing and screening novel transcription factor biosensors.

Given our goal of identifying possible biosensors using this method, it is important to evaluate the diversity in our naïve library. Although our conservative estimate of coverage from our combined NGS and Sanger sequencing is 13.7%, we have many reasons to believe this number is far from representative of our true diversity. First, the three biosensors that were enriched in the screening process were not detected by either Sanger sequencing or NGS, indicating that our screening method can isolate out rare fusions from the initial library. Furthermore, NGS library preparation requires a PCR-based amplification step and therefore there is a chance of certain insertions providing more favorable priming than others, and the exponential nature of PCR substantially compounds this. The difference in distributions of insertions from the NGS data and the Sanger sequencing of individual colonies further underlines this bias, as does the fact that 10 insertions represented 93% of our total NGS insertion counts and 66% of our insertions identified in our NGS were present at 5 or fewer counts. We believe these combined facts lead to a masking of less frequently amplified insertions in the NGS dataset. This is most apparent in the first window of MBP, where all insertions pooled from forward/reverse insertions and both biological replicates

yielded only 16 distinct insertions, one of which was observed 868,175 times (encompassing 94% of the reads), and out of which 11 were observed less than 10 times each. Were we to repeat this deep sequencing analysis, we would likely explore an alternative method of amplification, such as using blunt-end ligated adapters on all library fragments for primer binding, instead of relying on primers that bind to the MBP and ZFP sequences directly.

In addition to finding three productive new biosensors in our library, we found many variants that were out-of-frame, but that still exhibited mild (less than 2-fold) inducible repression (data not shown). We hypothesize that this is due to non-specific translation, as the start codon of the ZFP remained in the final constructs. Therefore, we recommend that the start codon be removed in subsequent library construction to minimize this issue and further enrich for *productive* biosensors over these false positives.

One of our newly discovered biosensors (270A) had a double zinc-finger insertion, which arises from the fact that the ZFP insertion contains NotI on both the 5' and 3' end to be properly inserted into the palindromic transposon scar, making it possible, however rare, for a double insertion. Our investigation into the single-ZFP variant of that biosensor revealed that the double insertion variant exhibited significantly better repression at the gating threshold used in sorting. It is possible that the presence of two zinc fingers, if correctly folded, increases repression due to the higher local concentration of binding domains. Therefore, this is not inherently problematic or advantageous, but the fact that it originates from the transposon recognition sequences leads us to accept this as a possibility in our library creation. A mutagenesis study on the transposon recognition sequences might reveal an alternative method that removes this possibility, but such a study is outside the scope of this work.

Biosensor dose is critical when evaluating biosensor performance. Increasing biosensor expression increases the repression of the GFP reporter, however it also limits the sensitivity to maltose. This is potentially due maltose transport becoming a limiting factor, and thus the system is overwhelmed with unbound biosensors that are free to bind the reporter. 100 mM maltose was used here as the highest level of maltose that could be added extracellularly before growth was impacted. It is possible if the biosensor and promoter could interact in a purified, *in vitro* system, enough maltose could be added to completely ablate DNA binding. However, if this level is unattainable intracellularly via native maltose transporters then it is of no practical use. Additionally, it is possible that the binding of maltose does not ever completely ablate the ability of the biosensor to bind DNA, therefore, regardless of intracellular maltose concentrations, the GFP reporter may never be completely unbound by biosensors. This performance characteristic is likely to be unique for every biosensor created, necessitating the need for a dose curve of both biosensor and ligand to be performed to find the desired biosensor properties.

Linker composition is vital to biosensor performance. The reference biosensor was not found in the transposon based screen due to the differences in the linkers. However, three novel biosensors were found. This implies that not only does linker composition matter, but that if we had chosen different linkers in the transposon design, we likely still would have found biosensors, albeit potentially an entirely different set. We hypothesized that too long of a linker length would insulate any allosteric interaction upon ligand binding, while no linker may prevent the ZFP from folding in a viable conformation for DNA binding. Therefore, three alanine linkers on would theoretically provide an inert and flexible linker composition. In fact, the variants of the 335P biosensor support this hypothesis: shortening linkers reduced the repression, and changed the maltose responsiveness. Furthermore, why the reference biosensor would be able to bind DNA

with lysine and leucine on the 5' end and asparagine and valine on the 3' end of the ZFP, but not with three alanines implies that potentially both length and composition of linkers are important for performance. As with biosensor expression level, the linkers are likely to impact every biosensor differently.

The 335P biosensor has the performance characteristics that make it capable of distinguishing between a high and low state of maltose that could be utilized for high-throughput screening or feedback control mechanisms. However, the biosensors found by BERDI may not always have the performance characteristic desired for a particular application. Therefore, using BERDI as a starting point to generate functional biosensors, it may be possible to evolve the biosensor by saturating mutagenesis on the three alanine linkers, or even the whole protein, followed by additional rounds of sorting to enrich for different performance characteristics. The three novel biosensors described here were all found by sorting for repressors, then clonal examination for maltose responsiveness. However, if ligand responsive biosensors prove to be exceptionally rare, it would be prudent to use the ligand to sort and enrich for ligand responsive biosensors. Additionally, instead of using GFP and FACS as the screening system for ligand responsive biosensor, GFP could be replaced with the tetA gene encoding the tetracycline/H+ antiporter. Since the ZFP represses transcription, cells that cannot alleviate this repression in the presence of the ligand would be selected against under tetracycline challenge. Whereas biosensor that were ligand responsive would express more tetA. Therefore, growth on tetracycline could be used as another way to enrich for rare, ligand responsive, variants.

Using MBP as a model system for which a biosensor had already been described was useful as it enabled the comparison of new biosensors to the previous reference biosensor. It was also important to design the BERDI method in such a way that did not depend on any previously

described permutation of the protein in addition to being generalizable to any metabolite binding protein of interest. The BERDI method successfully demonstrated that it can generate several novel biosensors from a model metabolite binding protein. Library diversity can be evaluated by restriction enzyme digestion, in addition to sequencing the library members by both Sanger sequencing and NGS. Furthermore, by evaluating biosensor dose along with linker length and composition, the performance of the enriched biosensor can be changed. In summary, the BERDI method is capable of rapidly converting a metabolite binding protein into a metabolite responsive transcription factor without any prior knowledge of permissive sites.

3.7 Acknowledgments

This work was supported by the National Science Foundation (MCB-1341414 to J.N.L., DGE-1324585 to P.Y.S); the Environmental Protection Agency (STAR Fellowship F13A30124 to A.K.D.Y.); National Institute of Health (TRC is supported by 1R01MH103910-01); and Northwestern University (Undergraduate Research Grant to A.J.S.). A.K.D.Y. was supported in part by the Northwestern University Graduate School Cluster in Biotechnology, Systems, and Synthetic Biology, which is affiliated with the Biotechnology Training Program. Flow cytometry experiments were conducted at the Robert H. Lurie Flow Cytometry Core Facility. Traditional sequencing services were performed at the Northwestern University Genomics Core Facility. NGS runs were done courtesy of the UIC Sequencing Core, and special acknowledgements to Stefan Green and Weihua Wang for assisting in the process.

Chapter 4. Development of a screening strain for the inducible overproduction of Farnesyl Pyrophosphate (FPP) to develop a biosensor in the DXP pathway

4.1 Context

In Chapters 2 and 3, evaluating maltose and MBP as a ligand and metabolite binding protein pair comprised an effective model system for a variety of reasons. Maltose monohydrate is an inexpensive sugar, easily soluble in water and media at high concentrations, and non-toxic and readily imported from the media by *E. coli*. MBP is a maltose transporter without catalytic activity on its substrate, and was previously converted into maltose responsive biosensors[52, 53, 95]. Additionally, MBP has many crystal structures and mutants characterized and described. This detailed biophysical understanding of the protein allowed informed hypotheses to be postulated about the potential mechanism of maltose mediated transcriptional regulation of the discovered biosensors. These features made maltose and MBP an attractive model system. However, the utility and applications of a maltose-responsive transcription factor, as described in Chapters 2 and 3, are limited. Therefore, to test the generalizability of the BERDI strategy, developed in Chapter 4, a new ligand and binding protein pair is required.

In comparison to transporters like MBP, there are many metabolite-binding proteins whose primary activity is catalytic. Metabolite binding enzymes are a promising target for biosensor conversion as many have high specificity for their target metabolite. However, metabolite responsive transcription factors are not catalytic, so the target enzymes must have catalytically inactive mutants described. I hypothesize that without this type of mutation available, the residence time of the ligand in the active site would be too short to effectively convey metabolite mediated

regulation of transcription. Additionally, intact catalytic activity may also change the intracellular levels of the metabolite that is being measured, which could confound and complicate the output of the sensor. Therefore, prior to selection of a ligand and enzyme pair I identified a desirable starting point by defining six properties that are most important for this initial investigation: (1) importance of the ligand to be sensed (for basic and applied research), (2) known crystal structure (at least ligand-bound), (3) known kinetic parameters (esp. K_m, the substrate concentration at half-maximal velocity), (4) known active site or mutations that disrupt catalysis without abolishing metabolite binding, (5) knowledge of ligand consumption, production, and, toxicity in *E. coli*, and (6) known method/s for enhancing ligand accumulation to facilitate initial biosensor evaluation.

With this mind, two farnesyl pyrophosphate (FPP) binding enzymes, ispU from *E. coli* and crtM from *Staphylococcus aureus*, were chosen that fulfill all six of the desired properties. FPP is an important precursor molecule for many industrially relevant compounds such as terpenes, terpenoids (isoprenoids), and sterols[146]. FPP, unlike maltose, would not be able to be added exogenously to test biosensor function due to toxicity to *E. coli*[147]. However, *E. coli* natively produces FPP, and the enzymes known to be rate limiting in the pathway have been previously identified[142, 148, 149]. While static overexpression of key rate limiting enzymes has been shown to increase FPP levels, for FACS based screening purposes, the same cell needs to be able to toggle on the overproduction of FPP in an inducible manner. Therefore, to facilitate the testing of an FPP biosensor, a screening strain that inducibly overproduces FPP was developed. I designed and conducted the experiments described in this chapter with the help of three talented undergraduate researchers: Neil Dalvie, Andrea Shepard, and Shreya Udani. This chapter will eventually become an introductory part of a larger paper based on the design and development of FPP-responsive transcription factor biosensors.

4.2 Abstract

Farnesyl pyrophosphate is a valuable metabolic intermediate whose overproduction and accumulation leads to cellular toxicity by an unknown mechanism. Furthermore, general stress responsive promoters have been used as a proxy for FPP toxicity to dynamically regulate production a downstream product of FPP, amorphadiene[12]. However, there currently is no known transcription factor biosensor for FPP directly. Yet there are many naturally occurring FPP binding enzymes that could be converted into transcription factor biosensors using the previously described BERDI method. A screening strain must first be developed to inducibly overproduce FPP to enable single cells the ability to toggle from a low FPP level to a high FPP level, as required for biosensor evaluation by FACS. Since there is not an easy way to monitor FPP production the enzymes that convert FPP into β-carotene, a visible red-orange product that can be quantified via absorption at 450nm, can be introduced as a proxy for FPP. By inducibly expressing the rate limiting enzymes in the pathway upstream of FPP together with enzymes that convert FPP into β-carotene, an inducible FPP production landscape was determined. This strain can inducibly overproduce FPP to physiological, and biotechnologically relevant levels to enable the evaluation of FPP responsive biosensors.

4.3 Introduction

FPP is an intermediate in the biosynthetic pathway for producing industrially relevant terpenes, terpenoids (isoprenoids), and sterols [146]. FPP is also a key branch point for native terpenoid use in cell wall synthesis and redox mediators (quinols). Therefore, careful management of the FPP node controls branching between two native pathways and recombinant synthesis of

higher (>15 carbon) terpenoids. Accumulation of FPP is also toxic by an unknown mechanism [147], and optimization of *E. coli* metabolism to maximize FPP flux while overcoming toxicity has been extensively investigated [142, 148, 149]. Moreover, FPP is readily converted into a visible carotenoid product such as β-carotene, via expression of the *crtEBIY* operon from *Pantoea ananatis*, which provides a useful metric of FPP production [142, 149]. FPP-responsive promoters have been identified and utilized to implement feedback control, although the mechanism of FPP recognition is indirect [12]; thus this case study comprises a useful point of comparison even though it does not utilize or identify an FPP-specific biosensor. Therefore, prior to the development of an FPP responsive transcription factor, a strain that can inducibly overproduce FPP was constructed using β-carotene as a proxy. Using this strain, a range of inducer concentrations and several media formulations were investigated to generate a set of conditions that produce a 13-fold induction of β-carotene over the uninduced strain. The strain was further validated by recapitulating phenotypes associated with high over production of FPP such as cellular toxicity and response to previously described stress responsive promoters.

4.4 Materials and methods

4.4.1 Bacterial strains and culturing

All experiments were conducted in DS941 Z1 Escherichia coli cells generously provided by Sean Colloms (University of Glasgow) (AB1157, recF143, $lacI^q$ lacZ ΔM15, P_{laciq}-LacI, P_{N25}-TetR)[149]. Cells were maintained in Lysogeny Broth (LB) Lennox formulation (10 g/L of tryptone, 5 g/L of yeast extract, 5 g/L of NaCl) supplemented with appropriate antibiotics (Ampicillin 100 μ g/mL, Kanamycin 50 μ g/mL, or Chloramphenicol 15 μ g/mL). All experimental analysis was conducted as noted in either LB media, M9 minimal media (1X M9 salts, 0.2% Casamino Acids,

2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1 mM Thiamine HCl) containing glycerol (0.4%) as the primary carbon source, or R media[150]. Variable amounts of anhydrotetracycline (aTc) and isopropyl β-D-1-thiogalactopyranoside (IPTG) were added as indicated.

4.4.2 Plasmid construction

All plasmids were assembled using standard molecular biology techniques. Plasmid backbones containing "plug-and-play" multiple cloning sites and compatible plasmids containing synthetic parts (mCherry, pTrc2) were generously provided by Jim Collins (MIT)[111]. The low copy number plasmid (pSC101 origin ~1-5 copies per cell) containing a TetR-based aTc-responsive promoter driving the expression of the enzymes for *dxs*, *idi*, and *ispA* with a chloramphenicol resistance cassette was modified from pJKR-L-tetR, a gift from George Church[3] (Addgene #62562). The pA15 medium copy number origin (~10 copies per cell) was obtained from the Registry of Standard Biological Parts, plasmid pSB3K3. The stress-responsive promoters (gadE, rstA) were each PCRed from the *E.* coli genome and cloned into a medium copy number pA15 backbone with the ampicillin resistance cassette, driving the expression of mCherry. The *crtE*, *crtB*, *crtI*, and *crtY* genes from *P. ananatis* that convert FPP into β-carotene were generously provided by Sean Colloms (University of Glasgow) and cloned into a pTrc-based IPTG-inducible expression plasmid in a ColE1 backbone (~ 300 copies per cell) and ampicillin resistance cassette. Custom ribosome binding site (RBS) sequences were developed using the RBS calculator [112]. A summary of the plasmids used in this chapter can be found in Table 4.1.

Table 4.1 Summary of plasmids used in Chapter 4

Plasmid name	Description	Resistance	Origin
pAY449	pLtetO-*dxs-idi-ispA* and constitutive TetR Upstream enzymes.	Cm^R	pSC101
pAY475	pTrc2-*crtEBIY* Downstream enzymes. β-carotene producing.	Kan^R	ColE1
pAY471	pGadE-mCherry. FPP stress responsive reporter.	Amp^R	pA15
pAY472	pRstA-mCherry. FPP stress responsive reporter.	Amp^R	pA15

4.4.3 Microplate-based fluorescence assays and analysis

Cultures were inoculated from single colonies into 2 mL of M9 media and grown overnight to stationary phase. Overnight cultures were diluted 1:10 and grown for 1-2 hours (OD600 ~ 0.5). Cultures were again diluted 1:10, plated in black-walled clear bottom 96-well plates in biological triplicate, and induced with aTc (to drive expression of the upstream enzymes, pAY449) and or IPTG (to drive expression of the β-carotene enzymes pAY475) as indicated. Plates with lids were incubated and shaken in a continuous double orbital pattern at 548 cpm (2 mm) inside a BioTek Synergy H1 plate reader for 16 h with mCherry, and OD600 measurements taken every 15 min. Monochrometer settings were 585/620 nm for mCherry. (mCherry) / (OD600) per well was averaged across biological triplicates and plotted over time. Each error bar represents the standard deviation of the means.

4.4.4 β-carotene extraction and quantification

Cultures were inoculated from single colonies into 2 mL of indicated media and grown overnight to stationary phase. The overnight cultures were diluted 1:10 in biological triplicate and grown for 24 or 48 hours at 30° C while shaking. 2 mL of cultures were then centrifuged for 1 min

at 15000 RCF. The pellets were washed with 500 µl of water, and centrifuged for 1 min at 15000 RCF. The washed cells were re-suspended in 1 mL of dimethyl sulfoxide (DMSO), and placed into a 55°C water bath for 10 minutes. Next, the cells are centrifuged for 2 min at 15000 RCF with the resulting pellet being devoid of any β-carotene. To measure β-carotene levels, the supernatant was pipetted off and the absorbance at 450 nm was quantified using spectrophotometer that had been blanked with plain DMSO. To normalize β-carotene production by cell growth, the optical density at 600 nm wavelength (OD600) was also collected using the spectrophotometer. The cell density measurements were taken by diluting the overnight cultures 1:10 in corresponding media, on a media blanked spectrophotometer. The β-carotene production was defined as (OD450)/(OD600), and was averaged across three biological replicate for comparison between conditions.

4.5 Results

4.5.1 Native DXP pathway and accompanying and inducible overexpression constructs

The first step to develop a strain that could inducibly overexpress FPP was to introduce a plasmid that can convert FPP into the colored pigment β-carotene, because FPP is not readily quantifiable. This was accomplished by expressing four enzymes, *crtE, crtB, crtI,* and *crtY* from *P. ananatis* on a high copy plasmid (~300 copies per cell) from the pTrc2 lacI-based inducible promoter. To ensure that β-carotene was a true proxy for FPP level, these enzymes were placed on a high copy plasmid to maximize expression so that all the FPP is converted into β-carotene (Figure 4.1B). Next, to upregulate the native DXP pathway upstream of FPP (Figure 4.1A), three enzymes that had been previously described (*dxs, idi,* and *ipsA)*[142] to be rate limiting were expressed on a

separate plasmid. The overexpression of these enzymes has the potential to generate FPP at toxic levels. Therefore, these enzymes were expressed from the pLtetO promoter, which has a strong TetR repressed "OFF" state, and placed on a pSC101 based single copy number plasmid (Figure 4.1B). By co-transforming both inducible plasmids into the same cell, the upstream enzymes that overexpress FPP can be induced with anhydrotetracycline (aTc), and the downstream enzymes that convert the FPP into β-carotene can be induced with isopropyl β-D-1 thiogalactopyranoside (IPTG). This way, the expression of both plasmids can be independently tuned with their matching small molecule inducers, within the same cell.

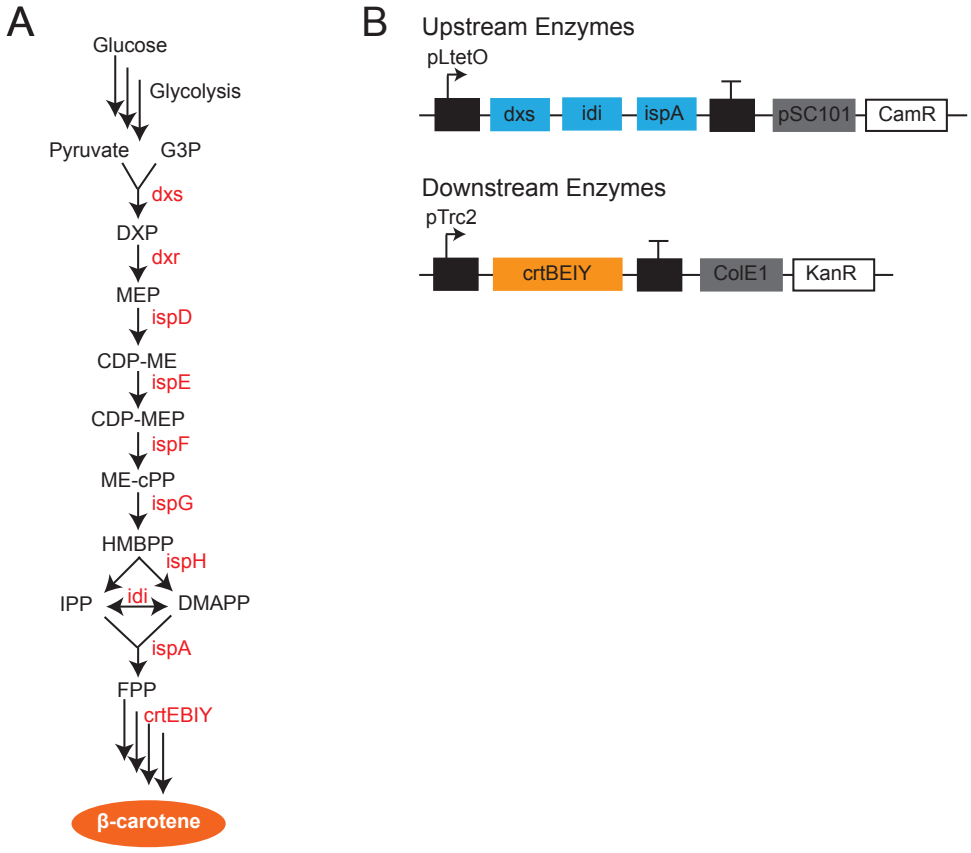


Figure 4.1 Native DXP pathway and inducible overexpression constructs
(A) Native DXP pathway in *E. coli* to generate FPP. Metabolites are shown in black, enzymes are shown in red. crtEBIY is added heterologously on a plasmid to convert FPP into β-carotene. (B) Graphical description of the two inducible expression constructs used in this chapter.

4.5.2 Evaluation of inducer concentrations to assess the β-carotene production landscape

Next, to evaluate the β-carotene production landscape, a combinatorial variation of inducer concentrations was performed. To first evaluate if the downstream enzymes are rate limiting, cells containing only the downstream enzymes were induced with a matrix of inducer concentrations (Figure 4.2A). As the IPTG dose increased, the β-carotene production does not change, until the highest dose of IPTG (100 μM) where the β-carotene production decreases potentially due to the stress associated with the high overproduction of four non-native enzymes. Overall, this indicates that the upstream enzymes are not rate-limiting when FPP is expressed at the basal level. When both the upstream enzymes and downstream enzymes were transformed together, β-carotene production is analyzed using the same range of inducer concentrations (Figure 4.2A). Unlike the downstream plasmid alone, when the levels of FPP are elevated by inducing the upstream pathway enzymes, the downstream enzymes become rate limiting. This can be seen across a row of aTc concentrations; as the IPTG concentration increases, the β-carotene production does as well. The 100 μM IPTG dose together with the 100 ng/mL aTc dose produced the highest level of β-carotene from both plasmids. This combinatorial variation of inducer concentrations was performed in standard LB media as a starting point to determine the optimal concentration of each inducer to use.

Next to evaluate whether β-carotene induction can be further increased by varying the nutrient level the cells receive, three different media types were surveyed. The no inducer case and the 100 μM IPTG + 100 ng/mL aTc case were repeated in standard LB media, in addition to a minimal media (M9) and an extra rich bioreactor media (R). Under these conditions, the β-carotene

production was induced ~ 13-fold from the M9 case without any inducers to the R media case with induction (Figure 4.2B). This indicates that for maximum β-carotene induction, increasing the nutrient level in addition to overexpressing the upstream pathway is an effective strategy for inducible FPP production.

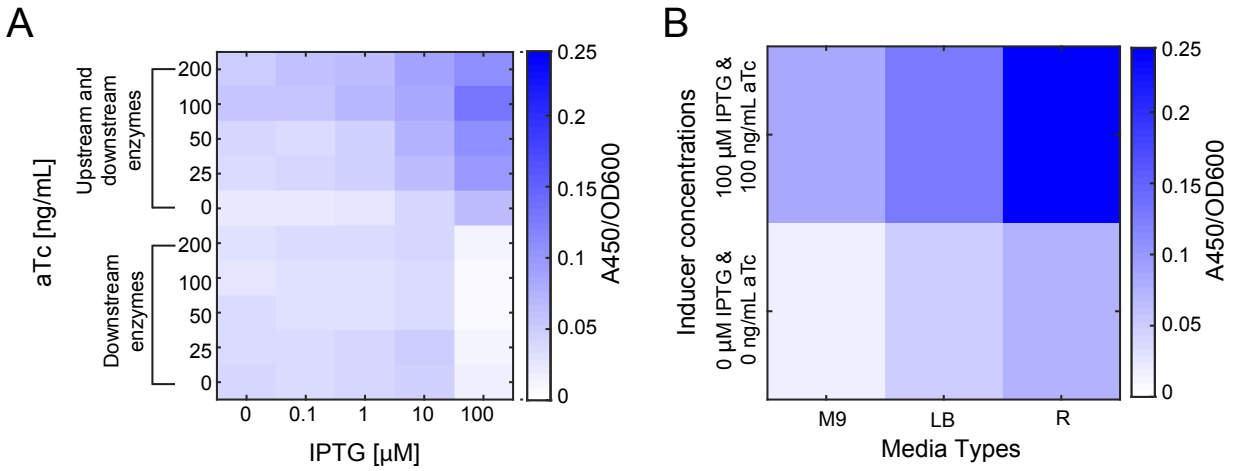


Figure 4.2 Evaluation of inducer concentrations to assess the β-carotene production landscape
(A) Heat map of a matrix of inducer concentrations for both the downstream enzymes alone, or the downstream and upstream enzymes together. (B) β-carotene production is quantified without any inducers or with the indicated inducer concentrations in three different media types. β-carotene production is quantified as absorbance at 450 nm divided by the OD600. Darker blue indicates higher β-carotene production per OD600.

4.5.3 DXP pathway overexpression effect on cell health and stress-responsive reporters

To validate that the inducible system is capable of overexpressing FPP to relevant physiological levels, the ability to induce cellular toxicity was evaluated. High levels of FPP is known to slow down growth through an unknown mechanism, therefore if the inducible upstream pathway can generate high levels of FPP, the growth should be impacted by this induction. To test this, only the upstream pathway enzymes were included so that FPP would build up instead of being converted to β-carotene. Cell growth readings, measured as OD600, were taken every 15 minutes for 16 hours post-induction, with a range of aTc doses (Figure 4.3A). As expected, the

higher the dose of aTc, the larger the growth effect was, confirming that the inducible system can recapitulate the known cellular toxicity phenotype.

Next, two stress responsive promoters, driving the expression of the fluorescent protein mCherry that have been previously shown to respond to FPP induced cellular stress, were also evaluated with the inducible system. The two promoters GadE and RstA were both found to be differentially regulated when the transcriptome from cells with a basal level of FPP was compared to cells that had a static overexpression of FPP[12]. Therefore, to evaluate whether the inducible system can trigger the GadE and RstA stress responsive promoters, mCherry reporters were constructed and transformed with the upstream and downstream plasmids. Natively, GadE is highly expressed under normal conditions and decreases its expression with cellular stress, while RstA has the reverse logic. As previously described, the GadE promoter has high expression without any cell stress, then upon addition of the upstream enzymes, the expression drops, and drops again when the upstream and downstream plasmids are present (Figure 4.3B). Next using the RstA stress responsive promoter, expression was induced when both the upstream and downstream plasmids were present (Figure 4.3C). The behavior of the GadE and RstA reporters is consistent with previously described findings, indicating that this inducible system is producing similar levels of FPP mediated cell stress.

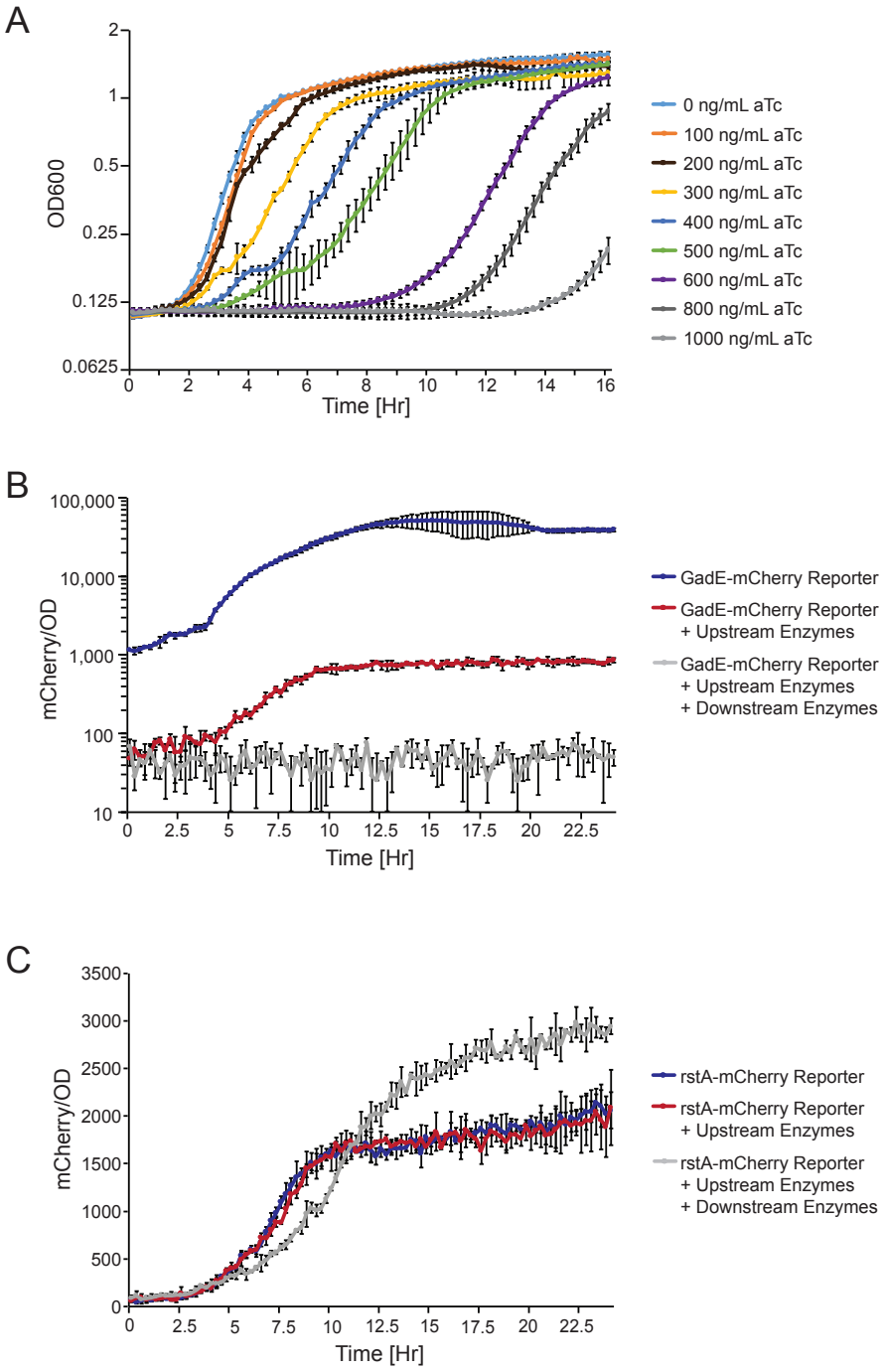


Figure 4.3 DXP pathway overexpression effect on cell health and stress responsive reporters
(A) Cell growth, as measured by OD600, over time of the upstream pathway enzymes being expressed with increasing concentrations of the inducer aTc. (B) mCherry fluorescence normalized by OD600 of the GadE stress responsive promoter plotted over time. (C) mCherry fluorescence normalized by OD600 of the RstA stress responsive promoter plotted over time.

4.6 Discussion

Taken together, these results suggest that the inducible upstream enzymes are capable of inducibly increasing intracellular FPP levels. Using separate plasmids for the upstream enzymes and the downstream enzymes, the 100 μM IPTG dose together with the 100 ng/mL aTc dose produced the highest level of β-carotene from both plasmids. Then by changing the nutrient level with different media formulations, a ~13-fold induction of β-carotene was achieved. Through recapitulating known, high FPP phenotypes, the inducible system was validated.

The most direct way to measure FPP is to use gas chromatography-mass spectrometry (GCMS). However, FPP itself is not volatile enough to be effectively detected, and must first be derivatized into its alcohol farnesol. While there are published methods on this, initial attempts proved challenging to accurately quantify farnesol levels from cellular extracts. The body of evidence, however, for this strains ability to inducibly overproduce FPP is strong in the absence of this direct measurement data.

When analyzing the performance of the stress response promoters, it was unexpected that the addition of the downstream enzymes did not decrease the cell stress associated with FPP levels. Presumably, the conversion of FPP into β-carotene would have relieved the cells FPP related stress, however the opposite was observed. When the downstream plasmid was included, this increased the cell stress that is reported on by both promoters. This is the potential downside of using general stress responsive promoters, even if they respond to high FPP stress, they also respond to other forms of stress. It is possible that having a third plasmid expressing the downstream enzymes in addition to the upstream enzyme plasmid and the reporter plasmid caused more general stress on the cells that was subsequently reported. This underscores the need for a direct way to measure FPP, and the generation of an FPP responsive transcription factor biosensor

could be one possible solution instead of relying on more indirect measurements like β-carotene or the stress responsive promoters.

Chapter 5. Conclusions and Recommendations

5.1 Chapter 2. Engineering modular biosensors to confer metabolite-responsive regulation of transcription

5.1.1 Conclusions

Before the evaluation of several biosensor conversion strategies, zinc finger responsive promoters first needed to be developed. Previously, zinc fingers fused to transcriptional activators or repressors and their cognate promoter had been developed[100-102], however there were no known promoters for the BCR-ABL1 zinc finger, and no investigation into just using the DNA-binding domain without any fusion partners such as transcriptional activator or repressor domains. The latter point was especially important as we intended to use the metabolite binding protein as the transcriptional regulator, and not an additional domain. By rational design, a library of BCR-ABL1 zinc finger-responsive promoters was built, and features important for transcriptional repression by the ZFP were identified by qualitative introspection and computational regression methods. The most repressible promoters were carried forward to be paired with the ZFP based biosensors. In fact, when the repression by the ZFP was compared to the repression by the SP biosensor, the same promoters that were highly repressible by the zinc finger alone were also highly repressible by the biosensor, indicating that the design rules for developing zinc finger responsive promoters also holds true for biosensors built from the same zinc finger.

Prior to the development of the SP biosensor, where a ZFP was inserted internally into MBP at amino acid 316R, all unnatural transcription factor biosensors had relied on the re-engineering, or fusion of a known transcription factor. While a ZFP is a known DNA-binding domain, it is not a transcription factor. Additionally, MBP is natively a periplasmic sugar transporter, so it too has no inherent transcriptional regulation activity. The SP biosensor was the

first example of a way to combine a DNA-binding domain and a metabolite-binding protein into a metabolite responsive transcription factor. Furthermore, the construct did not rely on directed evolution/mutagenesis in order to obtain maltose responsive transcriptional regulation. The SP biosensor did, however, rely on a wealth of previous literature on the development of enzymatic biosensors from MBP fused to other domains. In fact, the 316R split of MBP was previously used to generate a non-transcription factor enzymatic biosensor [53]. Unlike the SP biosensor strategy, the SZF strategy where the zinc finger was split in two and fused to the N and C terminal domains of MBP proved unsuccessful at producing a transcription factor biosensor. This is likely due to a combination of the distance the two halves of the ZFP were held apart by the N and C termini, and their orientation relative to one another, that precluded their ability to bind DNA. Furthermore, this strategy is overall less generalizable as it relies on the N and C termini to be in appropriate orientation to all the two halves of the ZFP to be reconstituted to a degree that enables DNA-binding. The SP strategy on the other hand just relies on the ability to find an internal split site that enables the ZFP to fold in a manner that enables binding, and a position that allows for ligand responsive regulation of the DNA-binding event. While this is by no means trivial, there is not any known inherent structural feature that would always prevent the ability to form a biosensor by the SP method, unlike the SZF method.

5.1.2 Recommendations

The SP biosensor is a successful, maltose responsive transcription factor, however the mechanism of transcriptional regulation is not understood. I hypothesize that in the unbound state, the ZFP portion of the biosensor is folded in a conformation that permits DNA-binding. In the presence of maltose, the two halves of MBP that are split by the ZFP, reconstitute upon ligand

binding. This reconstitution event changes the conformation of the protein and inhibits, at least partially, the ZFPs ability to bind DNA. This hypothesis could be addressed by evaluating a truncation of the SP biosensor that only contains the first half of MBP and the zinc finger. If this protein is no longer able to alleviate transcriptional repression, it could be evidence that the second domain of MBP is required for transcriptional regulation. However, another possible explanation could be that all the maltose binding and transcriptional regulation happens in the front half of MBP and the change in conformation upon ligand binding is transduced into the ZFP and that is what prevents DNA-binding. If that is the mechanism, then this conformational transduction could be dependent on the ZFP being tethered to MBP on both ends, and therefore the truncation biosensor would potentially not be able to differentiate between these two hypotheses. A crystal structure of the protein would answer many of these questions as it would enable a detailed look at the folding of the chimeric biosensor. Additionally, if both the *apo* and *holo* versions of the biosensor were solved, it could provide valuable insights into the mechanism of action of the biosensor.

A potential limitation of the biosensor work presented in chapter 2 is the mild repression exhibited by the biosensor, for example relative to that mediated by tetR. Thus, a fair question is – what could be done to improve the repression and fold change of the signal? From a promoter design aspect, placing the zinc finger binding sites right next to the -10 and -35 regions of the promoter provided strong repression upon ZFP expression. The -10 and -35 regions are where the sigma factor (σ70) binds and recruits RNA polymerase to transcribe the GFP reporter. I hypothesize that this repression was due to the ZFP partially occluding the σ70's ability to bind DNA. Therefore, a way to improve repressibility of the promoter could be to overlap the ZFP binding site with the σ70 binding sites. This could block σ70 binding more dramatically and

potentially decreases the recruitment of RNA polymerase to the promoter. From a biosensor perspective, the length or composition of the linkers between the ZFP and MBP were not thoroughly permutated. As demonstrated in chapter 3, the linkers can dramatically impact overall biosensor performance. Therefore, this could be another way in which to improve the performance characteristics of the biosensor.

5.2 Chapter 3. Development of novel metabolite responsive transcription factors via transposon-mediated, high-throughput protein fusion

5.2.1 Conclusions

Chapter 2 demonstrated that the internal fusion of the ZFP into MBP, at a split of MBP that was previously described to make an enzymatic biosensor, created a maltose responsive transcription factor. However, as previously stated in Chapter 1, this split of MBP is highly unlikely to turn any other protein into a biosensor, even if it was extremely structurally related. MBP is one of the few proteins for which enzymatic biosensors have been developed for, therefore a published split of the protein was available, but most proteins will not have this literature support, or if they do, the split may not result in a functional biosensor. Therefore, to develop a more generalizable method for converting metabolite binding proteins into biosensors using the SP strategy without the reliance on any prior information on potential fusions points, the Biosensor Engineering by Random Domain Insertion (BERDI) method was developed.

By sampling a diverse library of insertions of the ZFP into any metabolite binding protein, functional biosensors can be screen for using the zinc finger responsive GFP reporter developed in Chapter 2. After several rounds of FACS sorting, three new maltose-responsive biosensors were discovered. Including the biosensor generated in Chapter 2, the four insertion positions mapped to

three distinct regions in the three-dimensional structure off MBP, indicating there are several regions where the ZFP can be inserted to generate a functional biosensor. Not surprisingly, the insertions were all on the outside surface of the protein, presumably because internal insertions would destabilize the folding of MBP. Furthermore, a brief investigation into linker length and composition indicated the importance of the linkers in between the ZFP and the MBP on overall biosensor performance.

The success of the BERDI method with MBP is a potentially transformative technology that could enable the conversion of a metabolite binding protein into a metabolite responsive transcription factor with no prior knowledge of previously described splits of the protein.

5.2.2 Recommendations

While our initial application of the BERDI method was successful in generating novel MBP-based biosensors, there are certain aspects of our method that can be improved prior to future applications. The naïve library created by the MuA transposase was analyzed prior to sorting by both Sanger sequencing of individual colonies, and deep sequencing of the entire library pooled together. The two distributions were far from similar, indicating that the PCR amplification of the pool library during NGS preparation introduced a bias into the library. This is potentially due to mis-priming of the amplification primers, or a bias associated with the relative rates of amplification as a function of the insertional position. This could potentially be addressed by ligating on common sequences to each library fragment and then PCR amplify the library using primers that bind to these common sequences. This way all amplification was done by the same primer binding to the same sequence, they could potentially reduce the differences in priming and variable relative amplification biases.

Additional improvement of our method can be accomplished by changing the FACS strategy. After several rounds of sorting for functional repressors, the library was clonally assayed in a microplate for maltose responsive regulation of transcription. These resulted in many colonies that were functional biosensors, however the throughput of this evaluation was much lower compared to the FACS that was done to enrich for repressors. If the biosensors are too rare to be screened with 96-well plates, an alternative strategy could be to use the ligand to sort for functional biosensors. The limitation of this idea is that the entire library must repress the GFP reporter in the presence of biosensor prior to ligand based sorting. Full ligand-induced alleviation of the repression would look just like a library member that could not repress the reporter in the presence of inducer. Therefore, there can be no non-repressors, as these will subsequently be enriched as false positives as candidate ligand-responsive biosensors. An alternative approach would be to use the zinc finger-responsive promoter to drive the expression of the tetA gene. TetA encodes a tetracycline/H+ antiporter that confers resistance to the antibiotic tetracycline. Since the ZFP represses the zinc finger responsive promoter, it could repress the production of tetA, making the cells sensitive to tetracycline in the presence of biosensor. However, if the biosensor is sensitive to the ligand, it will alleviate this repression and survive the challenge with tetracycline. This way, the growth on tetracycline could be a way of enriching for ligand responsive biosensors from a large pool of repressors.

Finally, like the original biosensor from Chapter 2, the new biosensor regulate transcription via an unknown mechanism, furthermore the distinct regions of MBP where the ZFP has been inserted make it very difficult to generalize any one particular mechanism for all successful MBP biosensors. Additionally, the rational changes to the linkers of the 335P biosensor indicate that there is a stepwise change in biosensor performance associated with these linker changes.

Therefore, an in vitro study of the 316R and 335P biosensors, including linker variants, using purified protein could provide new evidence for their mechanism of action. An electrophoretic mobility shift assay (EMSA) using purified promoter DNA sequences and purified biosensors could provide evidence for the promoter occupancy, to determine if more than one biosensor can bind to the promoter, or if the biosensor is prevented from binding DNA in the presence of maltose or if the strength of the interaction is only weakened. By assaying the linker variants of 335P, the impact of the linker changes could be described in more mechanistic detail how a biosensors performance is changed by varying the linker length. These experiments could provide new information on the molecular mechanism of biosensor transcriptional regulation.

5.3 Chapter 4. Development of a screening strain for the inducible overproduction of Farnesyl Pyrophosphate (FPP) to develop a biosensor in the DXP pathway

5.3.1 Conclusions

FPP is a valuable metabolite because it can be enzymatically converted into a wide variety of industrially important, high value compounds. There are many example in the literature of metabolic pathway engineering to overproduce FPP and its downstream molecules. However, in order to evaluate FPP-responsive biosensors, a inducible screening strain needed to be developed. By controlling the overexpression of the upstream rate limiting enzyme on a tightly controlled single copy plasmid, the induction of FPP levels were ~13-fold over baseline after a conversion of FPP into β-carotene. After validation by the recapitulation of FPP toxicity phenotypes, and response to known stress-responsive promoters, an inducible system of FPP production was confirmed. This system can now allow for the evaluation of FPP-responsive biosensors, by inducibly toggling on the overproduction of FPP levels.

5.3.2 Recommendations

FPP was chosen as important model system for the next target BERDI primarily because there is no know transcription factor biosensor for FPP. However, there are many FPP binding enzymes that have been described and crystalized. Two in particular, ispU and crtM, are excellent candidates for biosensor conversion as they have both been crystallized, the FPP binding pocket is known, and mutants that abolish catalytic activity without also abolishing ligand binding are known. Attempting to convert ispU and crtM into FPP-responsive biosensors will address some questions about the generalizability of the BERDI method. Both FPP-binding proteins are enzymes, and therefore it was critical to the selection process that catalytic mutants were available, as catalysis could limit the time the FPP ligand spends bound to the biosensor. How using these mutants will impact biosensor performance, especially in the FPP-bound state, is an interesting and open question. Furthermore, both enzymes are naturally dimers; presumably some insertions cold disrupt dimerization, and it will be interesting to see whether such higher order complex formation helps or hurts biosensor performance. In Chapter 2, fusing non-reactive bulk in the form of the fluorescent protein mCherry increased the repressibility of the 316R/SP biosensor. I hypothesize that the added bulk of a dimeric biosensor will increase its ability to block the sigma factor and RNA polymerase from initiating transcription at the regulated promoter. The development of successful FPP-responsive biosensors, or the failure to do so, will inform the types of proteins that are likely to result in functional biosensors following the BERDI strategy for future investigations.

Chapter 6. References

[1] Dietrich, J. A., McKee, A. E., and Keasling, J. D. (2010) High-throughput metabolic engineering: advances in small-molecule screening and selection, *Annu Rev Biochem 79*, 563-590.

[2] Raman, S., Rogers, J. K., Taylor, N. D., and Church, G. M. (2014) Evolution-guided optimization of biosynthetic pathways, *Proc Natl Acad Sci U S A 111*, 17803-17808.

[3] Rogers, J. K., Guzman, C. D., Taylor, N. D., Raman, S., Anderson, K., and Church, G. M. (2015) Synthetic biosensors for precise gene control and real-time monitoring of metabolites, *Nucleic Acids Res*.

[4] Dietrich, J. A., Shis, D. L., Alikhani, A., and Keasling, J. D. (2013) Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis, *ACS Synth Biol 2*, 47-58.

[5] Zhang, F., Carothers, J. M., and Keasling, J. D. (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids, *Nat Biotechnol 30*, 354-359.

[6] Golynskiy, M. V., Koay, M. S., Vinkenborg, J. L., and Merkx, M. (2011) Engineering protein switches: sensors, regulators, and spare parts for biology and biotechnology, *Chembiochem 12*, 353-361.

[7] Li, S., Si, T., Wang, M., and Zhao, H. (2015) Development of a Synthetic Malonyl-CoA Sensor in Saccharomyces cerevisiae for Intracellular Metabolite Monitoring and Genetic Screening, *ACS Synth Biol*.

[8] Zhao, Y., and Yang, Y. (2015) Profiling metabolic states with genetically encoded fluorescent biosensors for NADH, *Curr Opin Biotechnol 31*, 86-92.

[9] Brockman, I. M., and Prather, K. L. (2015) Dynamic metabolic engineering: New strategies for developing responsive cell factories, *Biotechnol J*.

[10] Watstein, D. M., McNerney, M. P., and Styczynski, M. P. (2015) Precise metabolic engineering of carotenoid biosynthesis in Escherichia coli towards a low-cost biosensor, *Metab Eng*.

[11] Venayak, N., Anesiadis, N., Cluett, W. R., and Mahadevan, R. (2015) Engineering metabolism through dynamic control, *Curr Opin Biotechnol 34C*, 142-152.

[12] Dahl, R. H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T. S., Redding-Johanson, A. M., Petzold, C. J., Mukhopadhyay, A., Lee, T. S., Adams, P. D., and Keasling, J. D. (2013) Engineering dynamic pathway regulation using stress-response promoters, *Nat Biotechnol 31*, 1039-1046.

[13] Liu, D., Xiao, Y., Evans, B. S., and Zhang, F. (2015) Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor-actuator, *ACS Synth Biol 4*, 132-140.

[14] Wang, B., Barahona, M., Buck, M., and Schumacher, J. (2013) Rewiring cell signalling through chimaeric regulatory protein engineering, *Biochem Soc Trans 41*, 1195-1200.

[15] Strianese, M., Staiano, M., Ruggiero, G., Labella, T., Pellecchia, C., and D'Auria, S. (2012) Fluorescence-based biosensors, *Methods Mol Biol 875*, 193-216.

[16] Michener, J. K., Thodey, K., Liang, J. C., and Smolke, C. D. (2012) Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways, *Metab Eng 14*, 212-222.

[17] Kang, Z., Zhang, C., Zhang, J., Jin, P., Zhang, J., Du, G., and Chen, J. (2014) Small RNA regulators in bacteria: powerful tools for metabolic engineering and synthetic biology, *Applied microbiology and biotechnology 98*, 3413-3424.

[18] McLaughlin, K. J., Strain-Damerell, C. M., Xie, K., Brekasis, D., Soares, A. S., Paget, M. S., and Kielkopf, C. L. (2010) Structural basis for NADH/NAD+ redox sensing by a Rex family repressor, *Molecular cell 38*, 563-575.

[19] Hung, Y. P., and Yellen, G. (2014) Live-cell imaging of cytosolic NADH-NAD+ redox state using a genetically encoded fluorescent biosensor, *Methods Mol Biol 1071*, 83-95.

[20] Tsuyama, T., Kishikawa, J., Han, Y. W., Harada, Y., Tsubouchi, A., Noji, H., Kakizuka, A., Yokoyama, K., Uemura, T., and Imamura, H. (2013) In vivo fluorescent adenosine 5'-triphosphate (ATP) imaging of Drosophila melanogaster and Caenorhabditis elegans by using a genetically encoded fluorescent ATP biosensor optimized for low temperatures, *Anal Chem 85*, 7889-7896.

[21] Deuschle, K., Chaudhuri, B., Okumoto, S., Lager, I., Lalonde, S., and Frommer, W. B. (2006) Rapid metabolism of glucose detected with FRET glucose nanosensors in epidermal cells and intact roots of Arabidopsis RNA-silencing mutants, *Plant Cell 18*, 2314-2325.

[22] Fehr, M., Okumoto, S., Deuschle, K., Lager, I., Looger, L. L., Persson, J., Kozhukh, L., Lalonde, S., and Frommer, W. B. (2005) Development and use of fluorescent nanosensors for metabolite imaging in living cells, *Biochem Soc Trans 33*, 287-290.

[23] Deuschle, K., Okumoto, S., Fehr, M., Looger, L. L., Kozhukh, L., and Frommer, W. B. (2005) Construction and optimization of a family of genetically encoded metabolite sensors by semirational protein engineering, *Protein Sci 14*, 2304-2314.

[24] Takanaga, H., Chaudhuri, B., and Frommer, W. B. (2008) GLUT1 and GLUT9 as major contributors to glucose influx in HepG2 cells identified by a high sensitivity intramolecular FRET glucose sensor, *Biochim Biophys Acta 1778*, 1091-1099.

[25] Yang, H., Bogner, M., Stierhof, Y. D., and Ludewig, U. (2010) H-independent glutamine transport in plant root tips, *PLoS One 5*, e8917.

[26] Okumoto, S., Looger, L. L., Micheva, K. D., Reimer, R. J., Smith, S. J., and Frommer, W. B. (2005) Detection of glutamate release from neurons by genetically encoded surface-displayed FRET nanosensors, *Proc Natl Acad Sci U S A 102*, 8740-8745.

[27] Marvin, J. S., Borghuis, B. G., Tian, L., Cichon, J., Harnett, M. T., Akerboom, J., Gordus, A., Renninger, S. L., Chen, T. W., Bargmann, C. I., Orger, M. B., Schreiter, E. R., Demb, J. B., Gan, W. B., Hires, S. A., and Looger, L. L. (2013) An optimized fluorescent probe for visualizing glutamate neurotransmission, *Nat Methods 10*, 162-170.

[28] Mohsin, M., Abdin, M. Z., Nischal, L., Kardam, H., and Ahmad, A. (2013) Genetically encoded FRET-based nanosensor for in vivo measurement of leucine, *Biosens Bioelectron 50*, 72-77.

[29] San Martin, A., Ceballo, S., Baeza-Lehnert, F., Lerchundi, R., Valdebenito, R., Contreras-Baeza, Y., Alegria, K., and Barros, L. F. (2014) Imaging mitochondrial flux in single cells with a FRET sensor for pyruvate, *PLoS One 9*, e85780.

[30] San Martin, A., Ceballo, S., Ruminot, I., Lerchundi, R., Frommer, W. B., and Barros, L. F. (2013) A genetically encoded FRET lactate sensor and its use to detect the Warburg effect in single cancer cells, *PLoS One 8*, e57712.

[31] Monterrubio, R., Baldoma, L., Obradors, N., Aguilar, J., and Badia, J. (2000) A common regulator for the operons encoding the enzymes involved in D-galactarate, D-glucarate, and D-glycerate utilization in Escherichia coli, *J Bacteriol 182*, 2672-2674.

[32] Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M., Ezzati, M., Grantham-McGregor, S., Katz, J., Martorell, R., Uauy, R., Maternal, and Child Nutrition Study, G. (2013) Maternal and child undernutrition and overweight in low-income and middle-income countries, *Lancet 382*, 427-451.

[33] Green, A. A., Silver, P. A., Collins, J. J., and Yin, P. (2014) Toehold switches: de-novo-designed regulators of gene expression, *Cell 159*, 925-939.

[34] Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., Ferrante, T., Ma, D., Donghia, N., Fan, M., Daringer, N. M., Bosch, I., Dudley, D. M., O'Connor, D. H., Gehrke, L., and Collins, J. J. (2016) Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components, *Cell 165*, 1255-1266.

[35] Pardee, K., Green, A. A., Ferrante, T., Cameron, D. E., DaleyKeyser, A., Yin, P., and Collins, J. J. (2014) Paper-based synthetic gene networks, *Cell 159*, 940-954.

[36] Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L., and Gordon, J. I. (2011) Human nutrition, the gut microbiome and the immune system, *Nature 474*, 327-336.

[37] Kotula, J. W., Kerns, S. J., Shaket, L. A., Siraj, L., Collins, J. J., Way, J. C., and Silver, P. A. (2014) Programmable bacteria detect and record an environmental signal in the mammalian gut, *Proc Natl Acad Sci U S A 111*, 4838-4843.

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

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

[40] Binder, S., Schendzielorz, G., Stabler, N., Krumbach, K., Hoffmann, K., Bott, M., and Eggeling, L. (2012) A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level, *Genome biology 13*, R40.

[41] Tang, S. Y., Qian, S., Akinterinwa, O., Frei, C. S., Gredell, J. A., and Cirino, P. C. (2013) Screening for enhanced triacetic acid lactone production by recombinant Escherichia coli expressing a designed triacetic acid lactone reporter, *J Am Chem Soc 135*, 10099-10103.

[42] Brockman, I. M., and Prather, K. L. (2015) Dynamic knockdown of E. coli central metabolism for redirecting fluxes of primary metabolites, *Metab Eng 28*, 104-113.

[43] Liu, D., Evans, T., and Zhang, F. (2015) Applications and advances of metabolite biosensors for metabolic engineering, *Metab Eng*.

[44] Farmer, W. R., and Liao, J. C. (2000) Improving lycopene production in Escherichia coli by engineering metabolic control, *Nat Biotechnol 18*, 533-537.

[45] Xu, P., Li, L., Zhang, F., Stephanopoulos, G., and Koffas, M. (2014) Improving fatty acids production by engineering dynamic pathway regulation and metabolic control, *Proc Natl Acad Sci U S A 111*, 11299-11304.

[46] Solomon, K. V., Sanders, T. M., and Prather, K. L. (2012) A dynamic metabolite valve for the control of central carbon metabolism, *Metab Eng 14*, 661-671.

[47] Soma, Y., Tsuruno, K., Wada, M., Yokota, A., and Hanai, T. (2014) Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch, *Metab Eng 23*, 175-184.

[48] Belousov, V. V., Fradkov, A. F., Lukyanov, K. A., Staroverov, D. B., Shakhbazov, K. S., Terskikh, A. V., and Lukyanov, S. (2006) Genetically encoded fluorescent indicator for intracellular hydrogen peroxide, *Nat Methods 3*, 281-286.

[49] Berg, J., Hung, Y. P., and Yellen, G. (2009) A genetically encoded fluorescent reporter of ATP:ADP ratio, *Nat Methods 6*, 161-166.

[50] Nagai, T., Sawano, A., Park, E. S., and Miyawaki, A. (2001) Circularly permuted green fluorescent proteins engineered to sense Ca2+, *Proc Natl Acad Sci U S A 98*, 3197-3202.

[51] Nausch, L. W., Ledoux, J., Bonev, A. D., Nelson, M. T., and Dostmann, W. R. (2008) Differential patterning of cGMP in vascular smooth muscle cells revealed by single GFP-linked biosensors, *Proc Natl Acad Sci U S A 105*, 365-370.

[52] Jeong, J., Kim, S. K., Ahn, J., Park, K., Jeong, E. J., Kim, M., and Chung, B. H. (2006) Monitoring of conformational change in maltose binding protein using split green fluorescent protein, *Biochem Biophys Res Commun 339*, 647-651.

[53] Guntas, G., Mitchell, S. F., and Ostermeier, M. (2004) A molecular switch created by in vitro recombination of nonhomologous genes, *Chem Biol 11*, 1483-1487.

[54] Guntas, G., and Ostermeier, M. (2004) Creation of an allosteric enzyme by domain insertion, *J Mol Biol 336*, 263-273.

[55] Tullman, J., Guntas, G., Dumont, M., and Ostermeier, M. (2011) Protein switches identified from diverse insertion libraries created using S1 nuclease digestion of supercoiled-form plasmid DNA, *Biotechnol Bioeng 108*, 2535-2543.

[56] Tullman, J., Nicholes, N., Dumont, M. R., Ribeiro, L. F., and Ostermeier, M. (2016) Enzymatic protein switches built from paralogous input domains, *Biotechnol Bioeng 113*, 852-858.

[57] Edwards, W. R., Busse, K., Allemann, R. K., and Jones, D. D. (2008) Linking the functions of unrelated proteins using a novel directed evolution domain insertion method, *Nucleic Acids Res 36*, e78.

[58] Meister, G. E., and Joshi, N. S. (2013) An engineered calmodulin-based allosteric switch for Peptide biosensing, *Chembiochem 14*, 1460-1467.

[59] Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M., and Tsien, R. Y. (1997) Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin, *Nature 388*, 882-887.

[60] Chaudhuri, B., Hormann, F., and Frommer, W. B. (2011) Dynamic imaging of glucose flux impedance using FRET sensors in wild-type Arabidopsis plants, *J Exp Bot 62*, 2411-2417.

[61] Fehr, M., Lalonde, S., Lager, I., Wolff, M. W., and Frommer, W. B. (2003) In vivo imaging of the dynamics of glucose uptake in the cytosol of COS-7 cells by fluorescent nanosensors, *J Biol Chem 278*, 19127-19133.

[62] Fehr, M., Frommer, W. B., and Lalonde, S. (2002) Visualization of maltose uptake in living yeast cells by fluorescent nanosensors, *Proc Natl Acad Sci U S A 99*, 9846-9851.

[63] Kaper, T., Lager, I., Looger, L. L., Chermak, D., and Frommer, W. B. (2008) Fluorescence resonance energy transfer sensors for quantitative monitoring of pentose and disaccharide accumulation in bacteria, *Biotechnol Biofuels 1*, 11.

[64] Lager, I., Fehr, M., Frommer, W. B., and Lalonde, S. (2003) Development of a fluorescent nanosensor for ribose, *FEBS Lett 553*, 85-89.

[65] Tinberg, C. E., Khare, S. D., Dou, J., Doyle, L., Nelson, J. W., Schena, A., Jankowski, W., Kalodimos, C. G., Johnsson, K., Stoddard, B. L., and Baker, D. (2013) Computational design of ligand-binding proteins with high affinity and selectivity, *Nature 501*, 212-216.

[66] Feng, J., Jester, B. W., Tinberg, C. E., Mandell, D. J., Antunes, M. S., Chari, R., Morey, K. J., Rios, X., Medford, J. I., Church, G. M., Fields, S., and Baker, D. (2015) A general strategy to construct small molecule biosensors in eukaryotes, *Elife 4*.

[67] Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A., Perumov, D. A., and Nudler, E. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria, *Cell 111*, 747-756.

[68] Winkler, W., Nahvi, A., and Breaker, R. R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression, *Nature 419*, 952-956.

[69] Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A., and Breaker, R. R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme, *Nature 428*, 281-286.

[70] Ellington, A. D., and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands, *Nature 346*, 818-822.

[71] Jenison, R. D., Gill, S. C., Pardi, A., and Polisky, B. (1994) High-resolution molecular discrimination by RNA, *Science 263*, 1425-1429.

[72] Thompson, K. M., Syrett, H. A., Knudsen, S. M., and Ellington, A. D. (2002) Group I aptazymes as genetic regulatory switches, *BMC Biotechnol 2*, 21.

[73] Bayer, T. S., and Smolke, C. D. (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression, *Nat Biotechnol 23*, 337-343.

[74] Desai, S. K., and Gallivan, J. P. (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation, *J Am Chem Soc 126*, 13247-13254.

[75] Win, M. N., and Smolke, C. D. (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function, *Proc Natl Acad Sci U S A 104*, 14283-14288.

[76] Qi, L., Lucks, J. B., Liu, C. C., Mutalik, V. K., and Arkin, A. P. (2012) Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals, *Nucleic Acids Res 40*, 5775-5786.

[77] Yang, J., Seo, S. W., Jang, S., Shin, S. I., Lim, C. H., Roh, T. Y., and Jung, G. Y. (2013) Synthetic RNA devices to expedite the evolution of metabolite-producing microbes, *Nat Commun 4*, 1413.

[78] Zhou, L. B., and Zeng, A. P. (2015) Exploring Lysine Riboswitch for Metabolic Flux Control and Improvement of l-Lysine Synthesis in Corynebacterium glutamicum, *ACS Synth Biol 4*, 729-734.

[79] Khalil, A. S., and Collins, J. J. (2010) Synthetic biology: applications come of age, *Nat Rev Genet 11*, 367-379.

[80] Khalil, A. S., Lu, T. K., Bashor, C. J., Ramirez, C. L., Pyenson, N. C., Joung, J. K., and Collins, J. J. (2012) A synthetic biology framework for programming eukaryotic transcription functions, *Cell 150*, 647-658.

[81] Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014) A brief history of synthetic biology, *Nat Rev Microbiol 12*, 381-390.

[82] Michalodimitrakis, K., and Isalan, M. (2009) Engineering prokaryotic gene circuits, *FEMS Microbiol Rev 33*, 27-37.

[83] Bertram, R., and Hillen, W. (2008) The application of Tet repressor in prokaryotic gene regulation and expression, *Microb Biotechnol 1*, 2-16.

[84] Hommais, F., Krin, E., Coppee, J. Y., Lacroix, C., Yeramian, E., Danchin, A., and Bertin, P. (2004) GadE (YhiE): a novel activator involved in the response to acid environment in Escherichia coli, *Microbiology 150*, 61-72.

[85] van Sint Fiet, S., van Beilen, J. B., and Witholt, B. (2006) Selection of biocatalysts for chemical synthesis, *Proc Natl Acad Sci U S A 103*, 1693-1698.

[86] Xu, P., Wang, W., Li, L., Bhan, N., Zhang, F., and Koffas, M. A. (2014) Design and kinetic analysis of a hybrid promoter-regulator system for malonyl-CoA sensing in Escherichia coli, *ACS Chem Biol 9*, 451-458.

[87] Meinhardt, S., Manley, M. W., Jr., Becker, N. A., Hessman, J. A., Maher, L. J., 3rd, and Swint-Kruse, L. (2012) Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression, *Nucleic Acids Res 40*, 11139-11154.

[88] Shis, D. L., Hussain, F., Meinhardt, S., Swint-Kruse, L., and Bennett, M. R. (2014) Modular, multi-input transcriptional logic gating with orthogonal LacI/GalR family chimeras, *ACS Synth Biol 3*, 645-651.

[89] Chou, H. H., and Keasling, J. D. (2013) Programming adaptive control to evolve increased metabolite production, *Nat Commun 4*, 2595.

[90] Collins, C. H., Arnold, F. H., and Leadbetter, J. R. (2005) Directed evolution of Vibrio fischeri LuxR for increased sensitivity to a broad spectrum of acyl-homoserine lactones, *Mol Microbiol 55*, 712-723.

[91] Collins, C. H., Leadbetter, J. R., and Arnold, F. H. (2006) Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR, *Nat Biotechnol 24*, 708-712.

[92] Tang, S. Y., Fazelinia, H., and Cirino, P. C. (2008) AraC regulatory protein mutants with altered effector specificity, *J Am Chem Soc 130*, 5267-5271.

[93] Tang, S. Y., and Cirino, P. C. (2011) Design and application of a mevalonate-responsive regulatory protein, *Angew Chem Int Ed Engl 50*, 1084-1086.

[94] Mohn, W. W., Garmendia, J., Galvao, T. C., and de Lorenzo, V. (2006) Surveying biotransformations with a la carte genetic traps: translating dehydrochlorination of lindane (gamma-hexachlorocyclohexane) into lacZ-based phenotypes, *Environ Microbiol 8*, 546-555.

[95] Park, K., Lee, L. H., Shin, Y. B., Yi, S. Y., Kang, Y. W., Sok, D. E., Chung, J. W., Chung, B. H., and Kim, M. (2009) Detection of conformationally changed MBP using intramolecular FRET, *Biochem Biophys Res Commun 388*, 560-564.

[96] Lohmueller, J. J., Armel, T. Z., and Silver, P. A. (2012) A tunable zinc finger-based framework for Boolean logic computation in mammalian cells, *Nucleic Acids Res 40*, 5180-5187.

[97] Younger, A. K., Dalvie, N. C., Rottinghaus, A. G., and Leonard, J. N. (2016) Engineering Modular Biosensors to Confer Metabolite-Responsive Regulation of Transcription, *ACS Synth Biol*.

[98] Zhang, F., and Keasling, J. (2011) Biosensors and their applications in microbial metabolic engineering, *Trends Microbiol 19*, 323-329.

[99] Oldach, L., and Zhang, J. (2014) Genetically encoded fluorescent biosensors for live-cell visualization of protein phosphorylation, *Chem Biol 21*, 186-197.

[100] Park, K. S., Lee, D. K., Lee, H., Lee, Y., Jang, Y. S., Kim, Y. H., Yang, H. Y., Lee, S. I., Seol, W., and Kim, J. S. (2003) Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors, *Nat Biotechnol 21*, 1208-1214.

[101] Bae, K. H., Kwon, Y. D., Shin, H. C., Hwang, M. S., Ryu, E. H., Park, K. S., Yang, H. Y., Lee, D. K., Lee, Y., Park, J., Kwon, H. S., Kim, H. W., Yeh, B. I., Lee, H. W., Sohn, S. H., Yoon, J., Seol, W., and Kim, J. S. (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors, *Nat Biotechnol 21*, 275-280.

[102] Lee, J. Y., Sung, B. H., Yu, B. J., Lee, J. H., Lee, S. H., Kim, M. S., Koob, M. D., and Kim, S. C. (2008) Phenotypic engineering by reprogramming gene transcription using novel artificial transcription factors in Escherichia coli, *Nucleic Acids Res 36*, e102.

[103] Copeland, M. F., Politz, M. C., and Pfleger, B. F. (2014) Application of TALEs, CRISPR/Cas and sRNAs as trans-acting regulators in prokaryotes, *Curr Opin Biotechnol 29*, 46-54.

[104] Bogdanove, A. J., and Voytas, D. F. (2011) TAL effectors: customizable proteins for DNA targeting, *Science 333*, 1843-1846.

[105] Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J., and Voytas, D. F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting, *Nucleic Acids Res 39*, e82.

[106] Oakes, B. L., Nadler, D. C., Flamholz, A., Fellmann, C., Staahl, B. T., Doudna, J. A., and Savage, D. F. (2016) Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch, *Nat Biotechnol 34*, 646-651.

[107] Benson, D. E., Conrad, D. W., de Lorimier, R. M., Trammell, S. A., and Hellinga, H. W. (2001) Design of bioelectronic interfaces by exploiting hinge-bending motions in proteins, *Science 293*, 1641-1644.

[108] Benson, D. E., Haddy, A. E., and Hellinga, H. W. (2002) Converting a maltose receptor into a nascent binuclear copper oxygenase by computational design, *Biochemistry 41*, 3262-3269.

[109] Marvin, J. S., Corcoran, E. E., Hattangadi, N. A., Zhang, J. V., Gere, S. A., and Hellinga, H. W. (1997) The rational design of allosteric interactions in a monomeric protein and its applications to the construction of biosensors, *Proc Natl Acad Sci U S A 94*, 4366-4371.

[110] Marvin, J. S., and Hellinga, H. W. (2001) Conversion of a maltose receptor into a zinc biosensor by computational design, *Proc Natl Acad Sci U S A 98*, 4955-4960.

[111] Litcofsky, K. D., Afeyan, R. B., Krom, R. J., Khalil, A. S., and Collins, J. J. (2012) Iterative plug-and-play methodology for constructing and modifying synthetic gene networks, *Nat Methods 9*, 1077-1080.

[112] Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression, *Nat Biotechnol 27*, 946-950.

[113] Chen, Y. J., Liu, P., Nielsen, A. A., Brophy, J. A., Clancy, K., Peterson, T., and Voigt, C. A. (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints, *Nat Methods 10*, 659-664.

[114] Worringer, K. A., Rand, T. A., Hayashi, Y., Sami, S., Takahashi, K., Tanabe, K., Narita, M., Srivastava, D., and Yamanaka, S. (2014) The let-7/LIN-41 pathway regulates reprogramming to human induced pluripotent stem cells by controlling expression of prodifferentiation genes, *Cell Stem Cell 14*, 40-52.

[115] Janes, K. A., Kelly, J. R., Gaudet, S., Albeck, J. G., Sorger, P. K., and Lauffenburger, D. A. (2004) Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data, *J Comput Biol 11*, 544-561.

[116] Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso, *J Roy Stat Soc B Met 58*, 267-288.

[117] Hurt, J. A., Thibodeau, S. A., Hirsh, A. S., Pabo, C. O., and Joung, J. K. (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection, *Proc Natl Acad Sci U S A 100*, 12271-12276.

[118] Wright, D. A., Thibodeau-Beganny, S., Sander, J. D., Winfrey, R. J., Hirsh, A. S., Eichtinger, M., Fu, F., Porteus, M. H., Dobbs, D., Voytas, D. F., and Joung, J. K. (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly, *Nature protocols 1*, 1637-1652.

[119] Maeder, M. L., Thibodeau-Beganny, S., Osiak, A., Wright, D. A., Anthony, R. M., Eichtinger, M., Jiang, T., Foley, J. E., Winfrey, R. J., Townsend, J. A., Unger-Wallace, E., Sander, J. D., Muller-Lerch, F., Fu, F., Pearlberg, J., Gobel, C., Dassie, J. P., Pruett-Miller, S. M., Porteus, M. H., Sgroi, D. C., Iafrate, A. J., Dobbs, D., McCray, P. B., Jr., Cathomen, T., Voytas, D. F., and Joung, J. K. (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification, *Molecular cell 31*, 294-301.

[120] Hashimoto, H., Olanrewaju, Y. O., Zheng, Y., Wilson, G. G., Zhang, X., and Cheng, X. (2014) Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence, *Genes Dev 28*, 2304-2313.

[121] Deaner, M., and Alper, H. S. (2016) Promoter and Terminator Discovery and Engineering, *Adv Biochem Eng Biotechnol*.

[122] Blazeck, J., and Alper, H. S. (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level, *Biotechnol J 8*, 46-58.

[123] Quiocho, F. A., Spurlino, J. C., and Rodseth, L. E. (1997) Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor, *Structure 5*, 997-1015.

[124] Sharff, A. J., Rodseth, L. E., and Quiocho, F. A. (1993) Refined 1.8-A structure reveals the mode of binding of beta-cyclodextrin to the maltodextrin binding protein, *Biochemistry 32*, 10553-10559.

[125] Sharff, A. J., Rodseth, L. E., Spurlino, J. C., and Quiocho, F. A. (1992) Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis, *Biochemistry 31*, 10657-10663.

[126] Spurlino, J. C., Lu, G. Y., and Quiocho, F. A. (1991) The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis, *J Biol Chem 266*, 5202-5219.

[127] Spurlino, J. C., Rodseth, L. E., and Quiocho, F. A. (1992) Atomic interactions in protein-carbohydrate complexes. Tryptophan residues in the periplasmic maltodextrin receptor for active transport and chemotaxis, *J Mol Biol 226*, 15-22.

[128] Riggs, P. D. (2012) Engineered Derivatives of Maltose-Binding Protein, In *Protein Engineering* (Kaumaya, P., Ed.), InTech.

[129] Ha, J. S., Song, J. J., Lee, Y. M., Kim, S. J., Sohn, J. H., Shin, C. S., and Lee, S. G. (2007) Design and application of highly responsive fluorescence resonance energy transfer biosensors for detection of sugar in living Saccharomyces cerevisiae cells, *Appl Environ Microbiol 73*, 7408-7414.

[130] Meister, G. E., Chandrasegaran, S., and Ostermeier, M. (2008) An engineered split M.HhaI-zinc finger fusion lacks the intended methyltransferase specificity, *Biochem Biophys Res Commun 377*, 226-230.

[131] Ke, W., Laurent, A. H., Armstrong, M. D., Chen, Y., Smith, W. E., Liang, J., Wright, C. M., Ostermeier, M., and van den Akker, F. (2012) Structure of an engineered beta-lactamase maltose binding protein fusion protein: insights into heterotropic allosteric regulation, *PLoS One 7*, e39168.

[132] Martineau, P., Szmelcman, S., Spurlino, J. C., Quiocho, F. A., and Hofnung, M. (1990) Genetic approach to the role of tryptophan residues in the activities and fluorescence of a bacterial periplasmic maltose-binding protein, *J Mol Biol 214*, 337-352.

[133] Kamionka, A., Bogdanska-Urbaniak, J., Scholz, O., and Hillen, W. (2004) Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor, *Nucleic Acids Res 32*, 842-847.

[134] Ramos, J. L., Martinez-Bueno, M., Molina-Henares, A. J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M. T., Brennan, R., and Tobes, R. (2005) The TetR family of transcriptional repressors, *Microbiol Mol Biol Rev 69*, 326-356.

[135] Pabo, C. O., Peisach, E., and Grant, R. A. (2001) Design and selection of novel Cys2His2 zinc finger proteins, *Annu Rev Biochem 70*, 313-340.

[136] Makhlynets, O. V., Raymond, E. A., and Korendovych, I. V. (2015) Design of allosterically regulated protein catalysts, *Biochemistry 54*, 1444-1456.

[137] Yu, K., Liu, C., Kim, B. G., and Lee, D. Y. (2015) Synthetic fusion protein design and applications, *Biotechnol Adv 33*, 155-164.

[138] Mehta, M. M., Liu, S., and Silberg, J. J. (2012) A transposase strategy for creating libraries of circularly permuted proteins, *Nucleic Acids Res 40*, e71.

[139] Ostermeier, M. (2005) Engineering allosteric protein switches by domain insertion, *Protein Eng Des Sel 18*, 359-364.

[140] Topilina, N. I., and Mills, K. V. (2014) Recent advances in in vivo applications of intein-mediated protein splicing, *Mob DNA 5*, 5.

[141] Nistala, G. J., Wu, K., Rao, C. V., and Bhalerao, K. D. (2010) A modular positive feedback-based gene amplifier, *J Biol Eng 4*, 4.

[142] Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. (2009) Programming cells by multiplex genome engineering and accelerated evolution, *Nature 460*, 894-898.

[143] Segall-Shapiro, T. H., Nguyen, P. Q., Dos Santos, E. D., Subedi, S., Judd, J., Suh, J., and Silberg, J. J. (2011) Mesophilic and hyperthermophilic adenylate kinases differ in their tolerance to random fragmentation, *J Mol Biol 406*, 135-148.

[144] Nadler, D. C., Morgan, S. A., Flamholz, A., Kortright, K. E., and Savage, D. F. (2016) Rapid construction of metabolite biosensors using domain-insertion profiling, *Nat Commun 7*, 12266.

[145] Murphy, K. C., Campellone, K. G., and Poteete, A. R. (2000) PCR-mediated gene replacement in Escherichia coli, *Gene 246*, 321-330.

[146] Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S., and Lee, S. Y. (2012) Systems metabolic engineering of microorganisms for natural and non-natural chemicals, *Nat Chem Biol 8*, 536-546.

[147] Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. (2003) Engineering a mevalonate pathway in Escherichia coli for production of terpenoids, *Nat Biotechnol 21*, 796-802.

[148] Kim, S. W., and Keasling, J. D. (2001) Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in Escherichia coli enhances lycopene production, *Biotechnol Bioeng 72*, 408-415.

[149] Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C., Osbourn, A., Keasling, J. D., and Rosser, S. J. (2014) Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination, *Nucleic Acids Res 42*, e23.

[150] Korz, D. J., Rinas, U., Hellmuth, K., Sanders, E. A., and Deckwer, W. D. (1995) Simple fed-batch technique for high cell density cultivation of Escherichia coli, *J Biotechnol 39*, 59-65.

Appendix A. Supplementary information for *Engineering modular biosensors to confer metabolite-responsive regulation of transcription*

A.1 Supplementary methods

A.1.1 Statistical analysis of promoter design features

Statistical analysis of promoter design features

This document provides additional details describing the statistical analysis used to evaluate which promoter design features conferred regulation of the engineered promoters by either DNA-binding proteins or engineered biosensors.

In order to apply computational methods to describe the library of promoters, it was necessary to choose quantitative descriptors, or features, that describe architectural properties of each promoter. We chose 17 features to describe the location of binding sites, relative to the -10 and -35 boxes, and to other binding sites. Two assumptions are associated with our choice of features. First, features were not assumed to be independent. Second, the expression of the reporter gene was assumed to depend solely on the repression of a bound biosensor or zinc finger protein. In order to determine which promoter features are important, three feature selection methods were used. All three methods used the same input and output data to generate a ranking of feature importance. All data were mean-centered and variance-scaled before these methods were applied. Input data consisted of a matrix of promoter indices and feature variables (62 promoter indices in rows, 17 feature variables in columns). Output data are described in the Materials and Methods and Results sections. The “repressibility” value for each promoter was defined as the negative relative expression of the reporter gene (GFP).

The following three feature selection techniques generated rankings of the features in order of importance.

A.1.2 Partial least squares regression

The first feature selection method used was a permutation test using partial least squares regression (PLSR). First, PLSR was executed using the built-in MATLAB function, *plsregress*. This function returns a predictive model for the output values through regression coefficients for each feature. To determine feature importance, a permutation test was used [115]. The output vector was randomly permuted, and PLSR was executed for the meaningless output vector. Regression coefficients were recorded for 1000 permutations, and a mean and standard deviation was calculated for the coefficient for each feature. After many permutations, coefficient means approached zero, as is expected for random permutations, but the standard deviations associated with each coefficient approached a different finite value for each feature, which indicates the degree to which that coefficient fluctuates randomly. Features for which the coefficients were greater in magnitude than the random variance are likely to be more significant. Therefore, the ratio of coefficient magnitude to the standard deviation associated with each coefficient provided a metric by which features were ranked in order of importance. In addition to rankings described in chapter 2 (Figure 2.3), coefficients and standard deviations for each feature are reported Appendix A2.5.

The MATLAB function, *plsregress,* also provides a vector of output variance explained by each feature. Summing these gives the overall variance explained in the output data by the regression. PLSR was executed with one feature removed, and the loss of output variance explained was recorded for each feature. It is important to note that ranking through this loss of

output variance explained, or loss of predictive power, yields the same ranking as the permutation test. Using this ranking, PLSR was executed with an increasing number of features, in order of importance, with the output variance explained recorded each time. The output variance explained was also recorded for an increasing number of principal components used in the regression. These both were plotted to show the contribution of each feature or principal component to the regression model (Figure 2.3B).

A.1.3 Random forest

The second feature selection method utilized was Random Forest. To implement the Random Forest method, we modified a MATLAB script developed by Jaiantilal (https://code.google.com/p/randomforest-matlab/), which was based upon a method originally described by Breiman and Cutler (http://www.stat.berkeley.edu/~breiman/RandomForests/). First, the promoter library was divided into 54 promoters selected randomly to comprise a training set, leaving 8 promoters as a test set. Next, 6 features were "bagged" into a subset by random selection without replacement. The size of this subset is traditionally one third of the total set of features, which in this case rounds to 6 features (http://statweb.stanford.edu/~tibs/ElemStatLearn/). Next, a subset of promoters from the training set was randomly selected with replacement. The size of this subset is similarly one third of the total number of promoters in the training set, yielding an 18 promoter subset. A decision tree was then generated, its predictions were tested against the data from the test set of promoters, and the mean square error was recorded. This process was repeated for a large number of bagged promoter training sets, while retaining the same subset of 6 features. This overall sequence was then repeated for a large number of feature subsets, generating a total of 100 decision trees, each of which used the same test set. Finally, this entire process was repeated

for 100 different random choices of test set, generating a total of 10,000 decision trees. To assess the importance of a feature, the input data within the test set were perturbed such that the feature values associated with each promoter (e.g., number of ZFP binding sites) were randomly permuted by shuffling. Any decision tree that included the feature of interest was then retested using the perturbed input data. The increase in mean squared error (i.e., reduction in predictive power) was averaged over all trees containing this feature. This metric (average increase in mean square error) was thus used to generate a ranking of features by importance, such that features with a greater average increase in mean square error were ranked as more important. (Appendix A2.5).

A.1.4 Lasso regression

The last feature selection method used was Lasso regression, also known as sparse or regularized regression. This type of feature selection is generally considered more robust than a permutation test or random forest, because the selection is built into the model generation, and does not require removing features from a predictive model [116]. Lasso regression uses the least squares method, and is regularized by placing a constraint on the sum of the absolute value of the regression coefficients. Mathematically, the method places a penalty on large coefficient magnitudes by minimizing the following expression:

$$\sum_{i=1}^{N}(y_i - \sum_j \beta_j x_{ij})^2 + \lambda \sum_j |\beta_j|$$

In this expression, y_i represents output data for the ith promoter, β is the regression coefficient for the jth feature, and x_{ij} is input data (feature variable j for promoter i). The value of λ is a tunable parameter that determines the extent of regularization. With this method, coefficients of unimportant features shrink to zero as λ is increased. Using the MATLAB function, *lasso*, Lasso

regression was executed for 100 increasing values of lambda, with the number of features with non-zero coefficients shrinking from 17 to 0. Each regression iteration (corresponding to each value of λ) was tested using 10-fold cross validation, and a mean squared error was recorded for each iteration. For each feature, the number of regression iterations for which it had a non-zero coefficient was recorded. This metric was used to generate a ranking of features in order of importance, with the most important features having non-zero coefficients for larger values of λ. In addition to the feature ranking, the mean squared error and number of features with non-zero coefficients were plotted together versus the value of λ (Figure 2.3C).

A2. Supplementary tables and figures

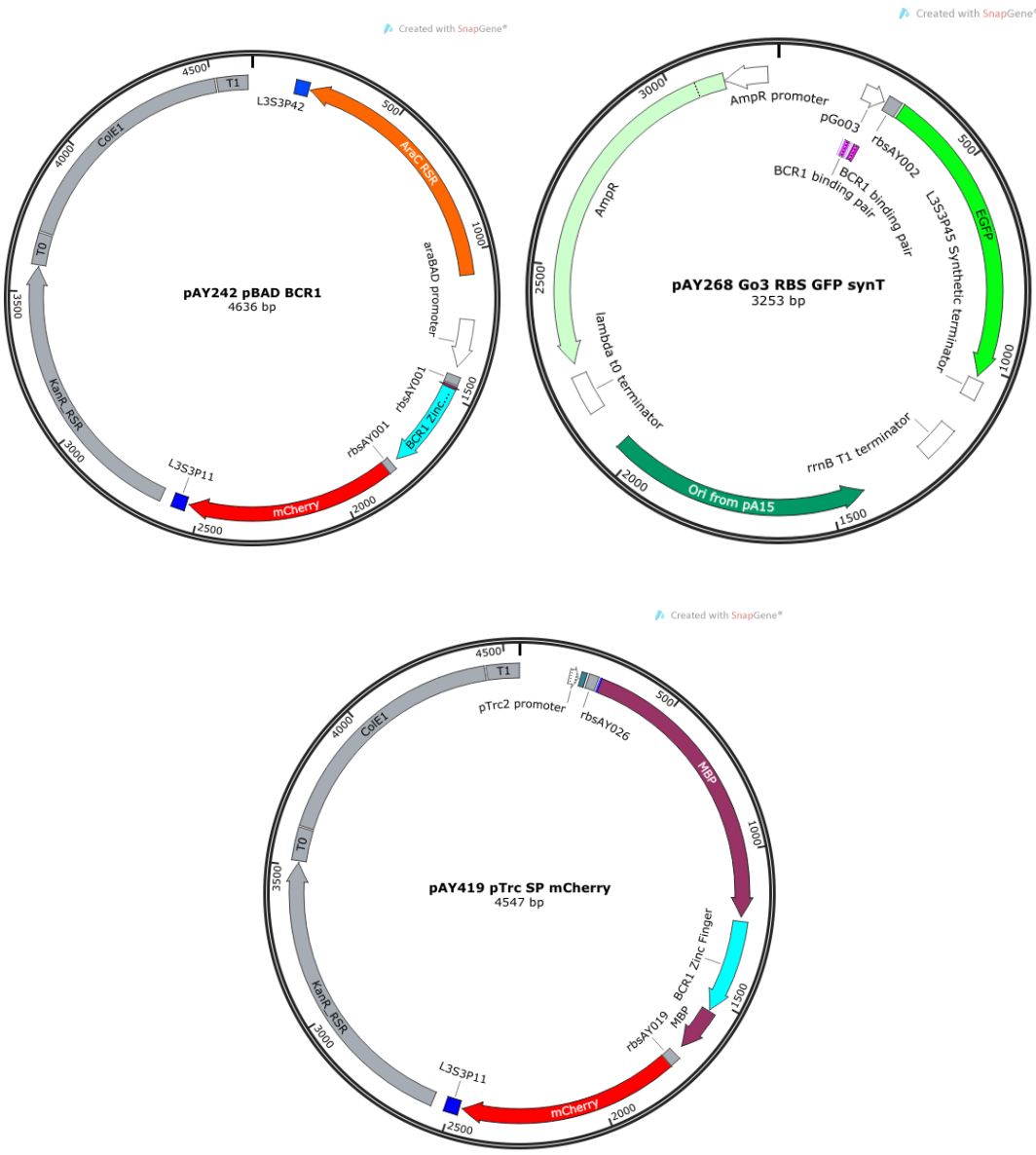


Figure A2.1 Plasmid maps of representative plasmids used in Chapter 2
pAY242 is the pBAD (arabinose-inducible) vector expressing the BCR-ABL1 zinc finger and mCherry (co-cistronically). This is a high copy plasmid (ColE1 origin) and contains the Kan^R resistance marker. pAY268 is a representative reporter plasmid with the Go3 promoter, driving the expression of EGFP. This is a medium copy plasmid (pA15 origin) and contains the Amp^R resistance marker. pAY419 drives expression of the SP biosensor (and mCherry, co-cistronically) from the pTrc2 promoter (which is IPTG-inducible). The site at which MBP is split via the BCR1 insertion is indicated on this map. This is a high copy plasmid (ColE1 origin) and contains the Kan^R resistance marker.

Promoter feature	Sequence and coloring
-10 Box	TATAAT
-35 Box	TTGACA
BCR-ABL1 binding site	gcagaagcc
BCR-ABL1 1st finger	gca
BCR-ABL1 2nd finger	gaa
BCR-ABL1 3rd finger	gcc
Spacer sequence	nnnnn

pGoXX	Promoter sequence (5'-3')
No Sites	TTGACAgatagctcactactatgTATAATgatagcaagactcc
Go03	TTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagcc
Go04	TTGACAgatagctcagtcataggTATAATgatagcgcagaagccgcagaagccactagagcagaagccgcagaagcc
Go05	TTGACAgatagctcagtcataggTATAATgatagcgcagaagccgcagaagccactagagcagaagccgcagaagccactagagcagaagccgcagaagcc
Go06	TTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccactagagcagaagccgcagaagcc
Go07	TTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccactagagcagaagccgcagaagccaggtgagcagaagccgcagaagcc
Go08	gcagaagccgcagaagccTTGACAgatagctcagtcataggTATAATgatagc
Go10	gcagaagccgcagaagccaggtgagcagaagccgcagaagccactagagcagaagccgcagaagccTTGACAgatagctcagtcataggTATAATgatagc
Go11	gcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagc
Go12	gcagaagccgcagaagccactagagcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAAT
Go13	gcagaagccgcagaagccaggtgagcagaagccgcagaagccactagagcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagc
Go14	TTGACAgcagaagccgcagaagccTATAAT
Go15	TTGACAgatagctcagcagaagccTATAATgatagc
Go16	TTGACAgcagaagccagtcataggTATAATgatagc
Go17	TTGACAgatagcagaagcctagaTATAATgatagc
Go18	TTGACAgatagcagaagcctagaTATAATgcagaagccgatagc
Go19	TTGACAgatagctcagtcataggTATAATgatagcgcagaagccgcagaagc
Go20	gcagaagccTTGACAgatagcagaagcctagaTATAATgatagc
Go21	gcagaagccttcgtagcgTTGACAgcagaagccagtcataggTATAATgatagc
Go22	TTGACAgatagctcagcagaagccTATAATgatagcgcagaagcc
Go23	gcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccactagagcagaagccgcagaagccaggtgagcagaagccgcagaagcc
Go24	gcagaagccgcagaagccaggtgagcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccactagagcagaagccgcagaagcc
Go25	gcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccgcagaagccgcagaagcc
Go26	gcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccgcagaagccgcagaagccgcagaagcc
Go29	TTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccactagagcagaagccgcagaagccaggtgagcagaagccgcagaagccactagagcagaagccgcagaagcc
Go30	TTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccactagagcagaagccgcagaagccaggtgagcagaagccgcagaagccactagagcagaagccgcagaagccactagagcagaagccgcagaagcc
Go31	gcagaagccgcagaagccttgTTGACAgcagaagccgcagaagccTATAATgatagcagcgcagaagccgcagaagccggtgcagaagccgcagaagccccctgcagaagccgcagaagcc
Go32	gcagaagccggagcagaagccttgTTGACAgcagaagccgcagaagccTATAATgatagcagcgcagaagccgctgcagaagccggtgcagaagccaatgcagaagccctgcagaagccgtagcagaagcc
Go33	gcagaagccgcagaagccttttggTTGACAgcagaagccgcagaagccTATAATgatagcagggacgcagaagccgcagaagccgaacgtgcagaagccgcagaagccccttaatgcagaagccgcagaagcc
Go35	gcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccactagagcagaagccgcagaagccaggtgagcagaagccgcagaagccactagagcagaagccgcagaagcc
Go36	TTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccaatgcagaagccactagagcagaagcctaagcagaagccaggtgagcagaagccggagcagaagccactagagcagaagcctatgcagaagcc
Go39	gcagaagccgcagaagccgcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccgcagaagccgcagaagcc
Go40	gcagaagccgcagaagccgcagaagccgcagaagccgcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccgcagaagccgcagaagcc
Go41	gcagaagccgcagaagccgttgcagaagccgcagaagccatagcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagcctaagcagaagccgcagaagcc
Go45	TTGACAgcagaagccgcagaagccTATAATgcagaagccgcagaagccactagagcagaagccgcagaagccaggtgagcagaagccgcagaagccactagagcagaagccgcagaagcc
Go46	gcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgcagaagccgcagaagccgcagaagccgcagaagccgcagaagccgcagaagcc
Go49	gcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccgcagaagccgcagaagccgcagaagcc
Go50	gcagaagccagcagaagccTTGACAgcagaagccagcagaagccTATAATgatagcgcagaagccagcagaagccgcagaagccagcagaagccgcagaagccagcagaagcc
Go51	gcagaagccagcagaagccTTGACAgcagaagccagcagaagccTATAATgcagaagccagcagaagccagcagaagccgcagaagccagcagaagcc
Go56	TTGACAgcagcagcagcagcagcaTATAATgatagcgaagccgaagccgaagcc
Go57	TTGACAgcagcagcagcagcagcaTATAATgaagccgaagccgaagcc
Go58	TTGACAgaagccgaagccgaagccTATAATgatagcgcagcagcagcagcagca
Go59	TTGACAgaagccgaagccgaagccTATAATgcagcagcagcagcagca
Go60	TTGACAgcagaagccgcagaagccTATAATgatagcgccgaagcagccgaagca
Go61	TTGACAgccgaagcagccgaagcaTATAATgatagcgcagaagccgcagaagcc
Go62	TTGACAgccgaagcagccgaagcaTATAATgccgaagcagccgaagca
Go63	TTGACAgcagaagccgcagaagccTATAATgatagcgaagcagccgcagaagcc
Go64	TTGACAgaagcagccgcagaagccTATAATgatagcgaagcagccgcagaagcc
Go65	TTGACAgaagcagccgcagaagccTATAATgatagcgcagaagccgcagaagcc
Go66	gcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgcagaagccgcagaagcc
Go67	gcagaagccgcagaagccTTGACAgcagaagccgcagaagccTATAATgatagcgcagaagccgcagaagccgcagaagccgcagaagccgcagaagccgcagaagccgcagaagcc
Go68	TTGACAgcagaagccgcagaagccTATAATgatagagaagccgcagaagcc
Go69	TTGACAgcagaagccgcagaagccTATAATgatagcagaagccgcagaagcc
Go70	TTGACAgcagaagccgcagaagccTATAATgatgcagaagccgcagaagcc
Go71	TTGACAgcagaagccgcagaagccTATAATgagcagaagccgcagaagcc
Go72	TTGACAgcagaagccgcagaagccTATAATggcagaagccgcagaagcc
Go73	TTGACAgcagaagccgcagaagccTATAATgcagaagccgcagaagcc
Go74	TTGACAgcagaagcagaagcagaaTATAATgatagc
Go75	TTGACAgcagccgcagccgcagccTATAATgatagc
Go76	TTGACAgaagcagaagcagaagcaTATAATgatagc
Go77	TTGACAgaagccgaagccgaagccTATAATgatagc
Go78	TTGACAgccgcagccgcagccgcaTATAATgatagc
Go79	TTGACAgccgaagccgaagccgaaTATAATgatagc
Go85	gcagaagccTTGACAgcagaagccgcagaagccTATAATgcagaagcc
Go88	gcagaagccTTGACAgcagaagccgcagaagccTATAATgcagaagcc
Go90	gcagaagccTTGACAgatagctccgcagaagccTATAATgcagaagcc
Go91	gcagaagccTTGACAgcagaagccgatagctccaTATAATgcagaagcc
Go92	gcagaagccTTGACAgatagcgcagaagccctccaTATAATgcagaagcc

Figure A2.2 Engineered promoter library details. Each BCR-ABL1-based promoter used in this study is listed and annotated as per the key at top

All Zif268 promoters (not listed) are identical in every way to their BCR-ABL1 counterparts except that the Zif268 binding site (GCGTGGGCG) replaces each instance of the BCR-ABL1 binding site.

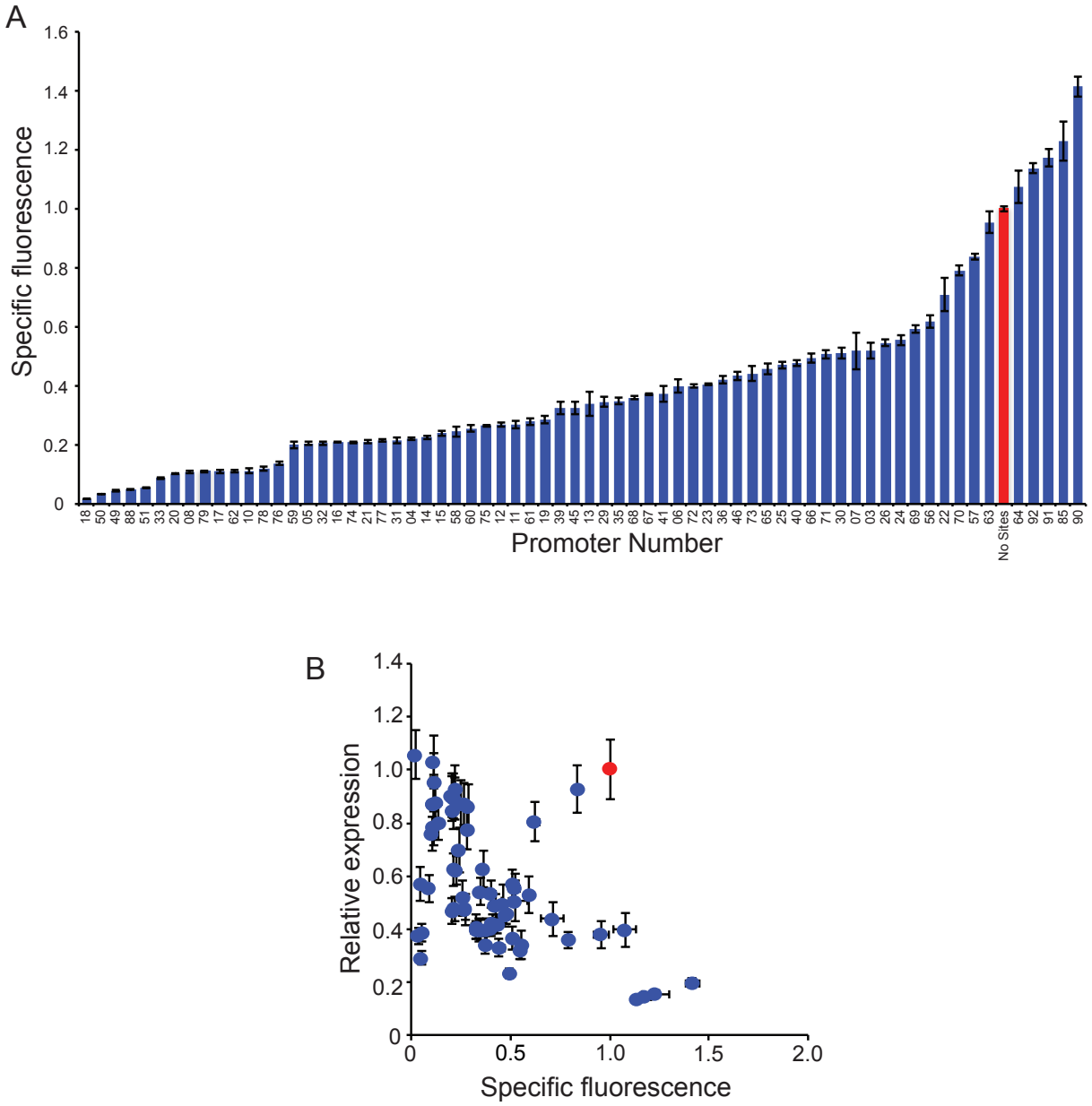


Figure A2.3 Specific fluorescence variation across the promoter library
(A) The impact of BCR-ABL1 expression on GFP reporter output was evaluated using a library of promoters bearing BCR-ABL1 sites at various locations in the promoter. The specific fluorescence (GFP fluorescence per OD_{600}) was measured for each promoter in the library in the absence of BCR-ABL1 and normalized to that of the No Sites control promoter, shown in red. (B) Comparison of specific fluorescence without BCR-ABL1 to BCR-ABL1-mediated repressibility (relative expression); note that each quantity was normalized to value associated with the No Sites control promoter, shown in red. Relative expression is defined as the ratio of GFP/OD600 (for any given promoter) of the induced case relative to that of the uninduced case, divided by this same ratio for the No Sites promoter (a full description and rationale can be found in the Materials and Methods section of the main manuscript). All experiments were run in biological triplicate, and error bars indicate one standard deviation.

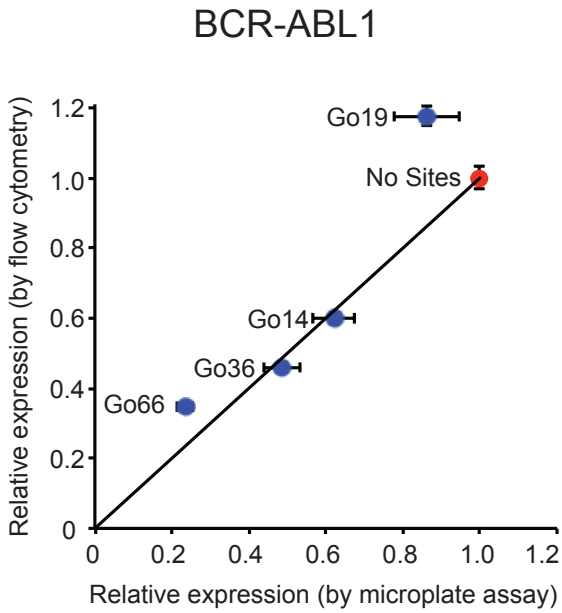
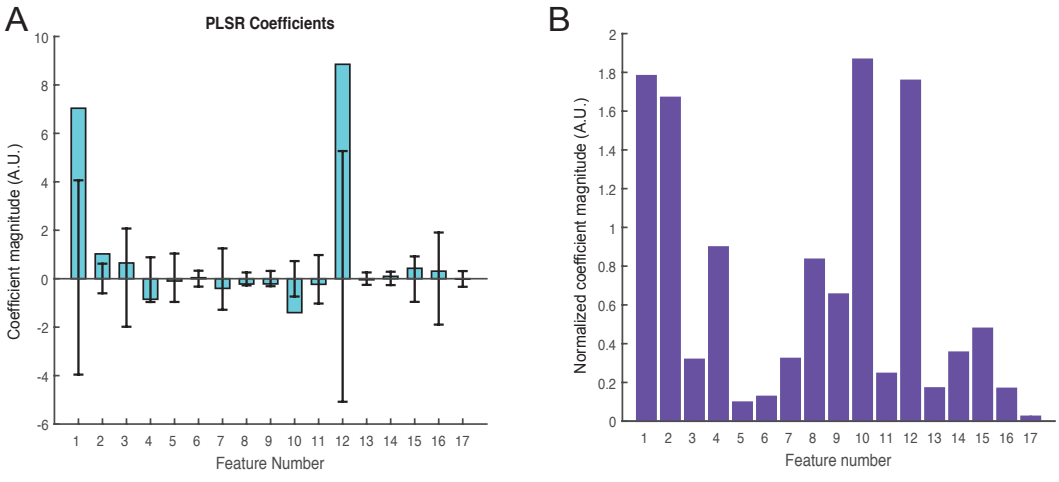


Figure A2.4 Comparison of flow cytometry and microplate assay-based quantification of BCR-ABL1-mediated repressibility

Select promoter constructs were analyzed by both methods. Close association of each point with the diagonal line (y = x) drawn as a visual guide indicates agreement between the two methods of quantifying relative expression, with the possible exception of Go19, which was the least repressible promoter analyzed. All samples were normalized to the No Sites control promoter, shown as a red circle. Experiments were conducted in biological triplicate, and error bars indicate one standard deviation.



Feature Number	Shorthand	Description
1	*BS in core*	Number of binding sites in the core
2	*pairs in core*	Number of pairs in the core
3	*BS upstream*	Number of binding sites upstream of the -35 region
4	*pairs upstream*	Number of pairs upstream of the -35 region
5	*BS downstream*	Number of binding sites downstream of the -10 box
6	*pairs downstream*	Number of pairs downstream of the -10 box
7	*length*	Length of the promoter in base pairs
8	*spacer -10*	Base pairs downstream of the -10 box, before the first binding site
9	*spacer -35*	Base pairs upstream of the -35 region, before the first binding site
10	*spacer inside -10*	Base pairs upstrsam of the -10 box, before the first binding site
11	*spacer inside -35*	Bse pairs downstream of the -35 region, before the first binding site
12	*core middle space*	Non-binding site base pairs in the core
13	*ave down space*	Average number of base pairs between binding sites downstream
14	*ave up space*	Average number of base pairs between binding sites upstream
15	*down length*	Base pairs downstream of the -10 box
16	*up length*	base pairs upstream of the -35 box
17	*lopsidedness*	Absolute value of (binding sites upstream) - (binding sites downstream)

Figure A2.5 Feature selection for BCR-ABL1-mediated repression
(A) Partial least squares repression (PLSR) coefficients associated with each feature (blue bars) are plotted along with corresponding standard deviations (error bars, which indicate the error around the mean coefficient value obtained by iterative permutation; this mean value was 0 in all cases, and thus error bars are plotted as deviations from 0). (B) PLSR coefficients associated with each feature were normalized by dividing each coefficient by its associated standard deviation, as obtained by iterative permutation (see Supplementary Methods). Features with a large normalized coefficient value are most important. (C) Average mean squared errors obtained when each feature was permuted during the Random Forest analysis (see Supplementary Methods). Features associated with a high mean squared error (when permuted) are more important. Feature numbers correspond to those listed in Figure 2.3A in chapter two.

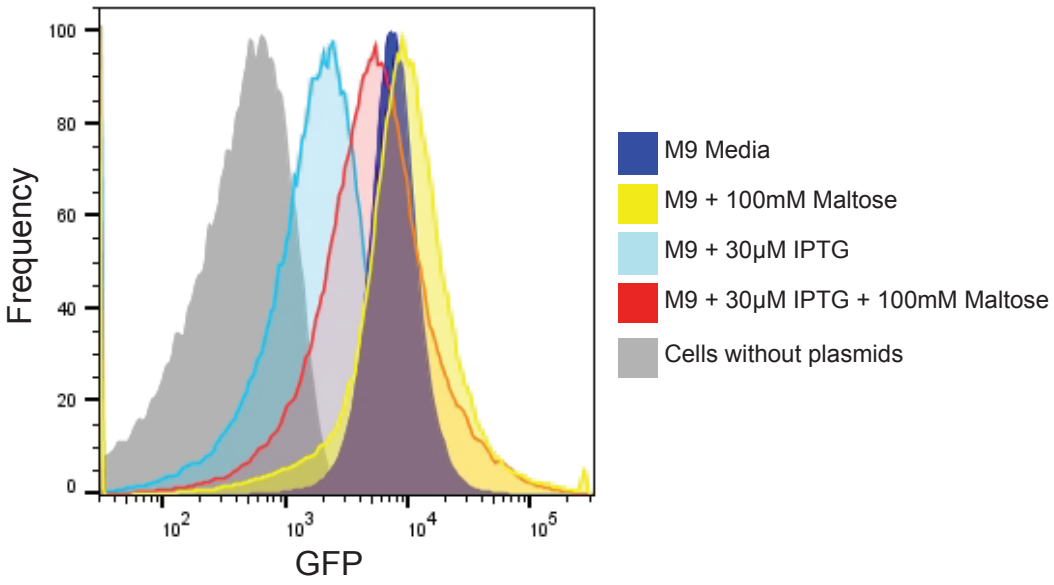


Figure A2.6 Analysis of SP biosensor performance at the individual cell level
Cells containing the SP biosensor and the Go92 reporter were grown and induced in the same manner as was used for microplate reader analysis, and then these cells were analyzed by flow cytometry. Cell were gated using forward and side scatter to exclude debris. The gray histogram represents cells that contain neither the biosensor nor the reporter plasmids. The remaining (colored) histograms represent cells expressing the SP biosensor and the Go92 reporter, cultured under the medium conditions indicated.

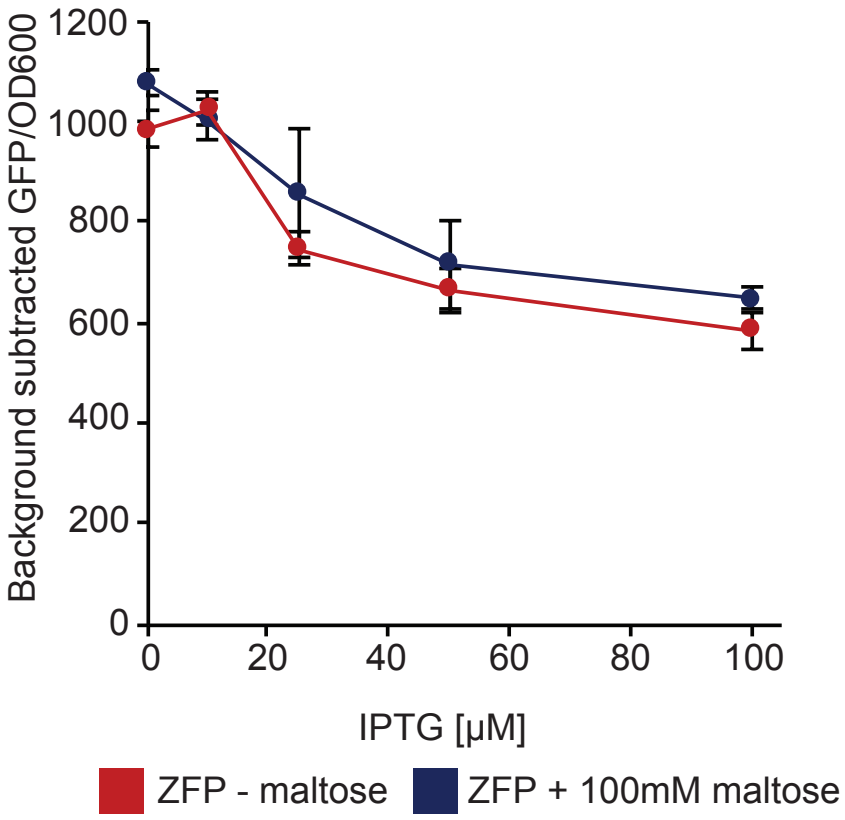


Figure A2.7 The effect of maltose on the BCR-ABL1 zinc finger's repressibility over a range of IPTG induction levels

The BCR-ABL1 zinc finger was expressed with the pTrc2 promoter in combination with the Go66 reporter. Compare to biosensor performance in Figure 2.4F. Note that the differences in repressibility between pTrc-BCR-ABL1 (shown here) and pBAD-BCR-ABL1 (shown in Figure 2.1, Figure 2.2, A2.3, and A2.4) may be attributed to different levels of expression of this ZFP from the aforementioned promoters. Given the potential for catabolite repression of the pBAD promoter in the presence of maltose, pTrc-BCR-ABL1 was constructed to enable testing the effect of maltose on BCR-ABL1-mediated repression of the reporters. All data represent mean values calculated from two independent experiments, each run in biological triplicate, and error bars represent one standard deviation.

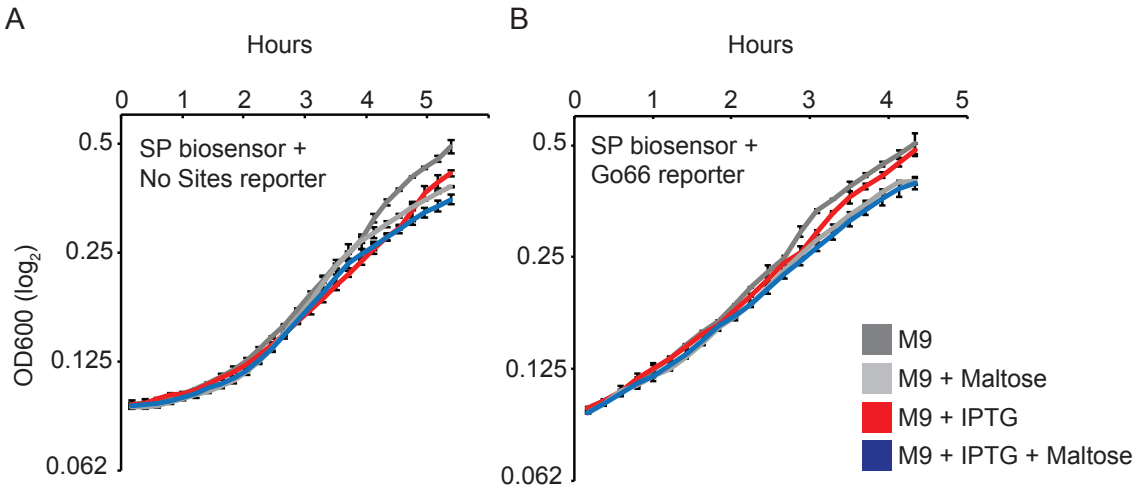


Figure A2.8 Impact of 100mM maltose and IPTG on cell growth
(A, B) Growth curves were collected for cells transformed with either the SP biosensor + the No Sites reporter (A) or the SP biosensor + the Go66 reporter (B). Experiments were conducted in biological triplicate, and error bars indicate one standard deviation.

Biosensor	SP	SP-Zif	SP-mC	mC-SP
Fold repression	4.1 ± 0.1	2.4 ± 0.4	3.5 ± 0.1	6.0 ± 0.4
Fold alleviation	3.0 ± 0.1	1.8 ± 0.3	2.0 ± 0.1	3.8 ± 0.3

Figure A2.9 Fold induction and alleviation calculated using metrics previously applied to natural biosensors
Using the fold induction methods described by Rogers et al. [3], fold repression was calculated by dividing the (maximum, uninduced background subtracted GFP/OD600) by the (induced, background subtracted GFP/OD600, repressed fluorescence). Fold alleviation was similarly calculated by dividing the (maximum, background subtracted GFP/OD600, maltose-induced fluorescence), by the (induced, background subtracted GFP/OD600, repressed fluorescence). GFP/OD600 values were calculated 10 hours after induction. Each range indicated is one standard deviation, in which error was propagated according to the division rule.

Appendix B. Supplementary information for *Development of novel metabolite responsive transcription factors via, transposon mediated, high throughput protein fusion.*

B.1 Supplemental results

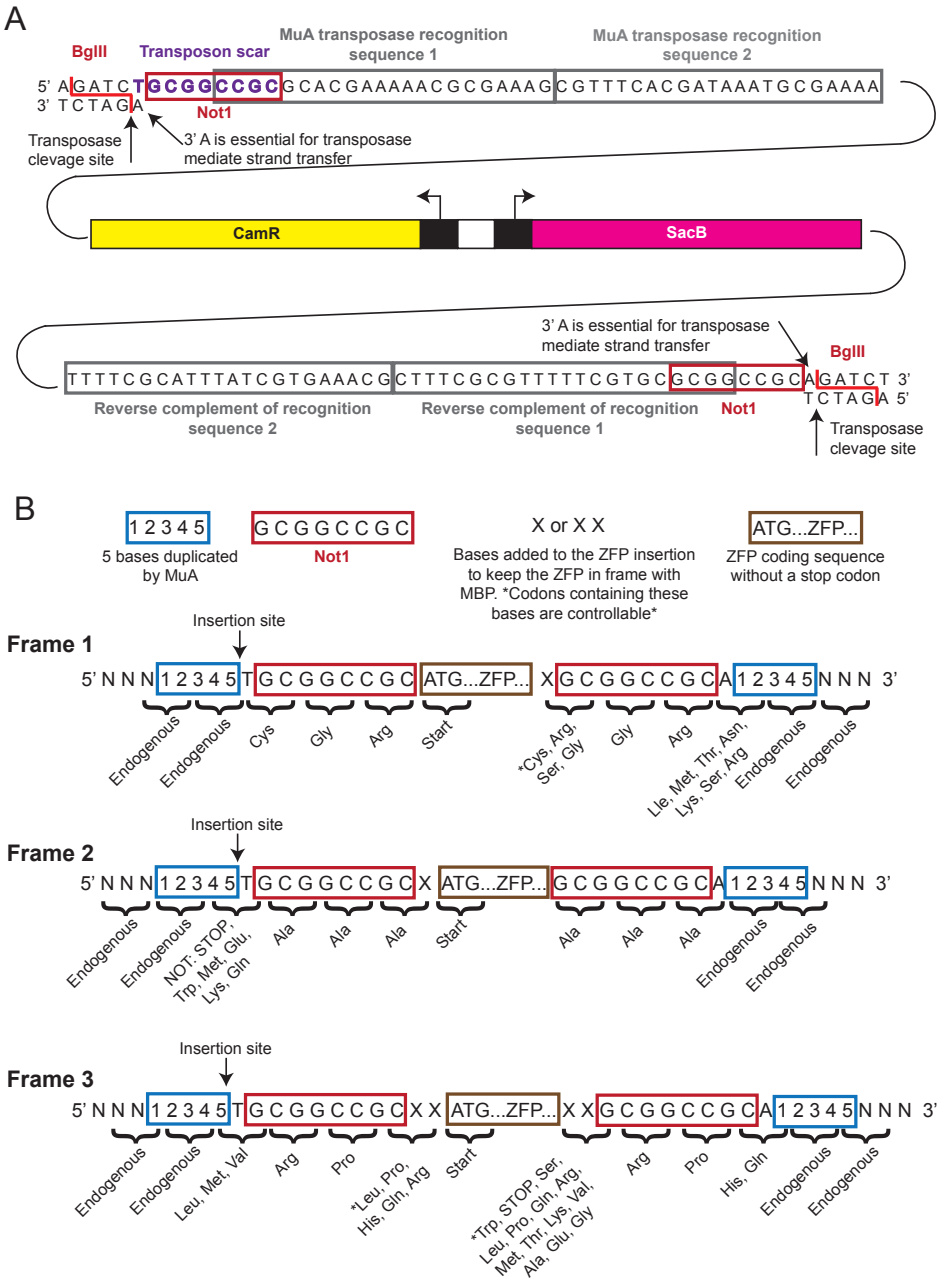


Figure B1.1 Transposon key features and scar options.
(A) Cartoon representation of the transposon. The transposon is digested out of its storage vector pAY438 using the BglII restriction enzymes. This leaves the minimal 3' "A" that is essential for MuA mediated strand transfer during the transposition reaction. All bases outside of this "A" get cleaved from the transposon following transposition. Purple highlighted bases will remain in the gene of interest after the transposon has been excised out. The NotI site is embedded in the MuA recognition site. The MuA recognition sites are show in gray boxes. The CamR and SacB ORFs with their own constitutive promoters are contained within the MuA recognition sequences. (B) The transposon will insert irrespective of frame: therefore, for each of the three frames, we describe the potential bases that need to be added to the ZFP to ensure that both regions of MBP are in frame with the ZFP insertion. By adding a single base to the front of the ZFP for frame 2, all the linkers will be alanines, except for the codon interrupted by transposon, and this cannot become a STOP codon. Frames 1 and 3 yielded poor linker options after the frame of the ZFP and MBP was preserved. Codons containing a "*" are controllable by varying the "X" base identity.

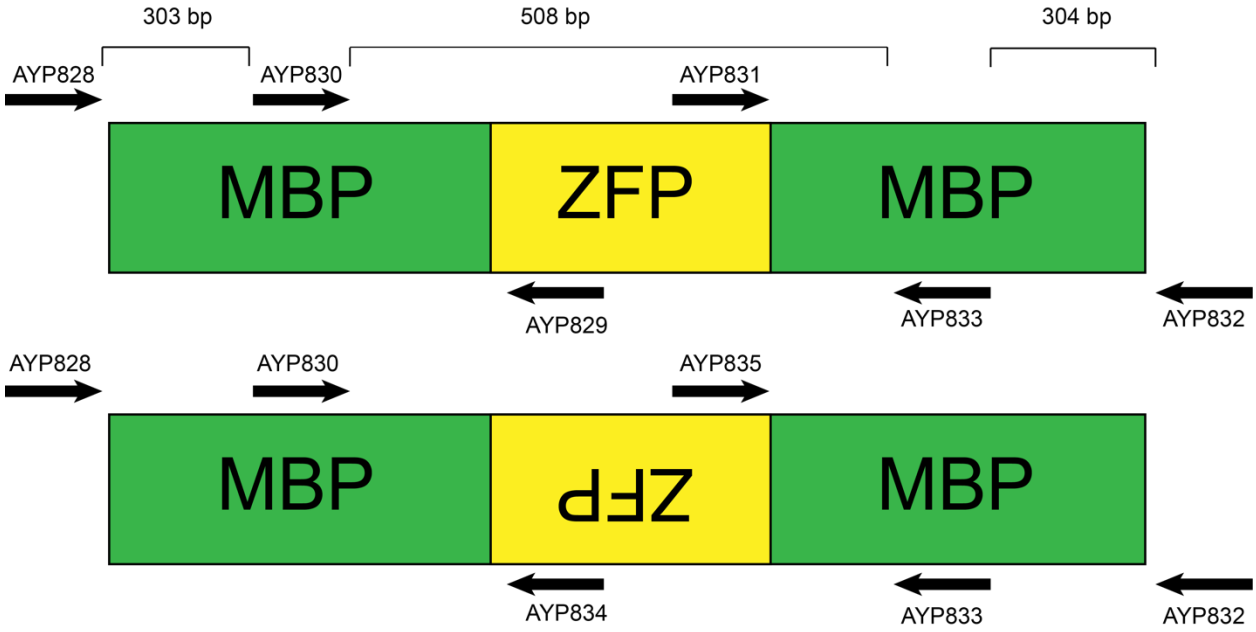


Figure B1.2 Representative graphic of the primers used to prepare library for deep sequencing.
The zinc finger could potentially be anywhere inside MBP and is only drawn in the middle here for ease of visualization. All primers contained a variant of the common sequences necessary for the downstream amplification conducted by the next generation sequencing core. Primers were spaced apart in order to capture insertions in windows of ~300 bp, as the NGS required amplicons < 500 bp in length. The top cartoon describes the primers used for the forward facing ZFP, whereas the bottom cartoon described the reactions that were done to account for possible reverse ZFP insertions due to the palindromic nature of the transposon recognition sequence. Please see table S3 for description of the 8 PCRs.

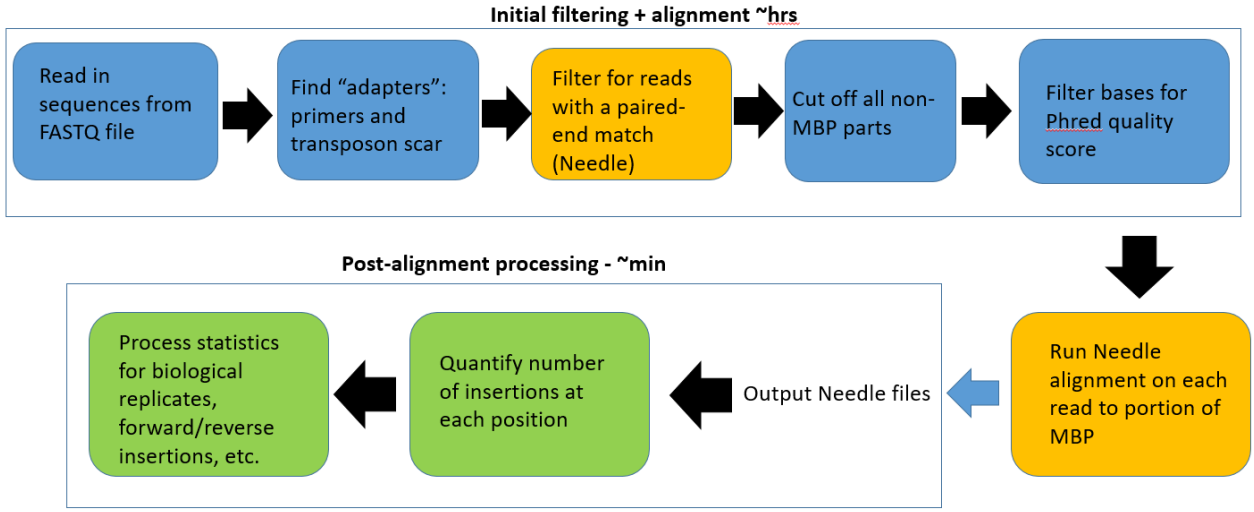


Figure B1.3 Flow Chart of NGS Analysis Pipeline.
Functions were created to conduct each task are outlined in the boxes. The yellow boxes indicate steps where a Needleman-Wunsch alignment. The green boxes indicate post-processing steps that only require the alignment needle file of the insert aligned against the template.

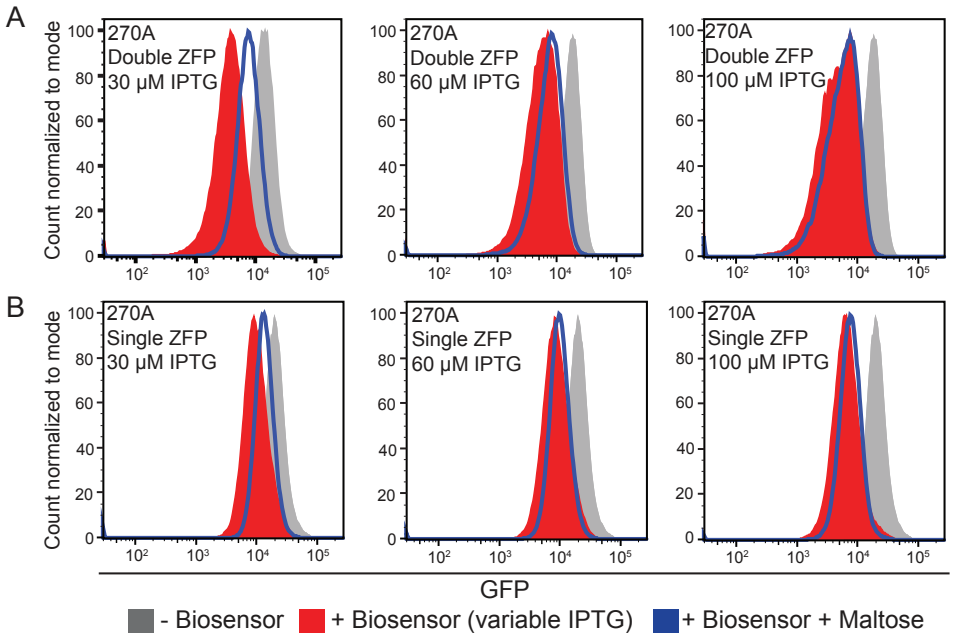


Figure B1.4 270A double ZFP versus 270A single ZFP.
The effects the number of zinc finger inserts into MBP at position 270A on biosensor performance were evaluated using flow cytometry for this construct. IPTG was used to induce biosensor repression, and the maltose addition relieved repression. The plots shown use varying concentrations of IPTG and are grouped by the number of zinc fingers present with (A) two zinc finger insertions, (B) one zinc finger insertion.

Table B1.5 Plasmids used in this study

Plasmid name	Description	Resistance	Origin	Reference
pAY430	pGo92-GFP. Zinc finger repressible promoter driving GFP.	Amp^R	pA15	Younger *et al.*, 2016
pAY447	MBP (no promoter)	Amp^R	ColE1	This work
pAY438	BglII and transposon recognition sequences flanking Cm^R and SacB	Kan^R Cm^R	ColE1	This work
pAY437	BCR-ABL1 zinc finger, without a stop codon or promoter, flanked by NotI sites	Amp^R	ColE1	This work
pAY431	pTrc2 promoter and RBS followed by KpnI and SphI MCS, followed by a terminator	Kan^R	ColE1	This work
pAY419	316R reference biosensor (SP from Younger *et al*. 2016)	Kan^R	ColE1	Younger *et al.*, 2016
pAY470	277A biosensor (in pAY431)	Kan^R	ColE1	This work
pAY450	270A double ZFP biosensor (in pAY431)	Kan^R	ColE1	This work
pAY453	270A single biosensor (in pAY431)	Kan^R	ColE1	This work
pAY451	335P (3AA) biosensor (in pAY431)	Kan^R	ColE1	This work
pAY469	335P (2AA) biosensor (in pAY431)	Kan^R	ColE1	This work
pAY468	335P (1AA) biosensor (in pAY431)	Kan^R	ColE1	This work
pAY460	335P (0AA) biosensor (in pAY431)	Kan^R	ColE1	This work

Table B1.6 Targeted and experimental library sizes

Experimental Step	Theoretical library size	Targeted library size (10x oversampling)	Experimental Results	Experimental Oversampling
Generate library	3144 (bp MBP plasmid) x 2 (forward or reverse) = **6,288**	62,880	~800,000	~125x
Clone transposed gene into expression plasmid	1116 (bp of MBP) x 2 (forward or reverse) = **2,232**	22,320	~30,000	~14x
Exchange transposon for ZFP	1116 (bp of MBP) x 2 (forward or reverse) = **2,232**	22,320	~30,000	~14x

Table B1.7 PCR primer pairs for NGS. CS – Common sequence

PCR Number	Forward primer (common sequence #)	Reverse primer (common sequence #)	ZFP direction
1	AYP828 (CS1)	AYP829 (CS2)	Forward
2	AYP830 (CS1)	AYP829 (CS2)	Forward
3	AYP831 (CS1)	AYP833 (CS2)	Forward
4	AYP831 (CS1)	AYP832 (CS2)	Forward
5	AYP828 (CS1)	AYP835 (CS2)	Reverse
6	AYP830 (CS1)	AYP835 (CS2)	Reverse
7	AYP834 (CS1)	AYP833 (CS2)	Reverse
8	AYP834 (CS1)	AYP832 (CS2)	Reverse

Table B1.8 List of all insertions found by NGS and Sanger sequencing

Insertion site [bp]	Count by NGS	Count by Sanger
1	7	0
2	35358	0
9	2	1
10	1	0
12	1	0
13	1	0
14	13	0
15	868175	1
16	4	0
18	1	0
32	2	0
40	1	0
59	4	0
60	24625	0
74	2	0
152	15	0
201	42	0
285	1	0
286	2708	4
287	2	0
300	3366	0
306	0	1
325	12	0
326	4	0
329	1	0
330	10	0
331	4	0
332	3	0
333	3	0
334	1	0
336	3	0
337	2	0
340	8	0
341	1	0
346	1	0
348	1	0

349	5	0
350	7	0
351	2	0
354	4	0
360	1	0
361	2	0
363	3	0
364	1	0
365	2	0
366	14	0
367	3	0
371	1	0
373	16	0
382	8	0
383	3	0
384	2	0
391	2	0
394	3	0
404	1	0
405	1	0
410	2	0
412	2	0
417	1	0
422	1	0
423	1	0
442	1	0
443	66	1
444	318427	0
445	1	0
448	1	0
494	1	0
550	1	0
561	1	0
563	1	0
564	7093	4
565	130462	0
566	4	0
570	1	0
580	2299	0

581	8766	0
594	27159	2
595	68187	0
614	2642	0
615	2269	0
616	1	0
623	8349	0
637	39957	1
638	5345	0
639	1	0
680	8378	0
681	118	0
670	0	1
706	1	0
711	201312	0
712	1659	0
713	8	0
715	1	0
716	1	0
717	1	0
719	1	0
723	1	0
748	1	0
757	1	0
758	181930	0
759	884	0
762	5	0
771	1	0
779	1	0
785	5	0
786	662975	0
787	666	0
788	1	0
789	1	0
791	4	0
794	1	0
807	66	0
808	2	0
809	262424	0

810	192	0
812	6	0
814	2	0
867	2	4
870	21	0
896	0	1
897	0	1
898	0	1
899	2840	0
900	7824	2
901	41428	15
902	1	0
903	44428	5
904	3	0
907	1	0
938	2	0
953	1	0
976	4	0
978	32589	0
979	14	0
987	3932	0
993	1	0
1004	3219	0
1068	595	0
1080	1	0
1086	72	0
1087	183665	1
1088	3	0
1090	2	0
1099	1	0
1102	1	0
1103	1	0
1104	1	0
1105	1	0
1110	1	0
1111	4	0
1112	236842	0
1113	113	0
1115	1	0

Table B1.9 X^2 statistical test was calculated per the formula:

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

Variable	Value
χ^2	740,018
Degrees of Freedom	1115
P-value	0.00E+00