ACS Synthetic Biology

Subscriber access provided by CORNELL UNIVERSITY LIBRARY

Article

Engineering modular biosensors to confer metabolite-responsive regulation of transcription

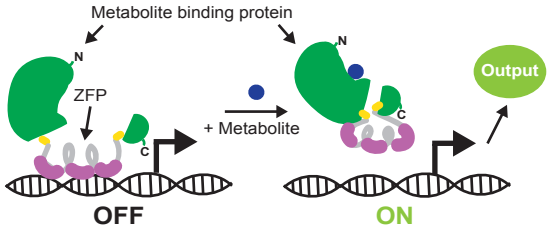
Andrew KD Younger, Neil Chandra Dalvie, Austin G Rottinghaus, and Joshua N. Leonard

ACS Synth. Biol., **Just Accepted Manuscript** • DOI: 10.1021/acssynbio.6b00184 • Publication Date (Web): 17 Oct 2016

Downloaded from http://pubs.acs.org on October 18, 2016

ACS Publications

ACS Synthetic Biology is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036


Metabolite binding protein
N
ZFP
C
+ Metabolite
OFF
N
C
Output
ON

Engineering modular biosensors to confer metabolite-responsive regulation of transcription

Andrew K.D. Younger[1], Neil C. Dalvie[2], Austin G. Rottinghaus[3], and Joshua N. Leonard[2,4,5,6,*]

[1] Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL, 60208, USA
[2] Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, 60208, USA
[3] Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI, 53706, USA
[4] Center for Synthetic Biology, Northwestern University, Evanston, IL, 60208, USA
[5] Chemistry of Life Processes Institute, Northwestern University, Evanston, IL, 60208, USA
[6] Member, Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Evanston, IL, 60208, USA

* To whom correspondence should be addressed. Tel: (847) 491-7455; Fax: (847) 491-3728; Email: j-leonard@northwestern.edu

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.

KEYWORDS

Metabolite biosensor

Protein engineering

Promoter engineering

Maltose binding protein

Zinc finger DNA binding domain

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 [1-3] or protein [4-15]. 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 [3-7, 9, 11, 13-16]. 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 [6]. Similar approaches have been harnessed to screen plasmid libraries to achieve enhanced production of mevalonate [17], triacetic acid lactone [18], and L-lysine [19]. 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 [20], and targeted genome-wide mutagenesis, which has been used to optimize naringenin and glucaric acid production via combinatorial perturbation of endogenous gene regulation [4]. While most investigations to date have applied these methods to bacterial chasses, such approaches may also be extended to yeast and other organisms [9]. 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 [16]. 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) [7]. 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 [3]. The potential utility of biosensor-mediated feedback control is now widely recognized [11, 21, 22], 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 [5]. 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 [10, 23-29]. 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 [24]. 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 [26] or GFP and its derivatives [30]. 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 [10, 31-33]). 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 ([34-39]). 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[40]. 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 [17, 18, 41, 42]. 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 [17, 43-47], although *de novo* design of novel ligand-binding proteins and biosensors is another promising frontier [48, 49]. 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.

RESULTS

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 [50-52]. ZFPs are small (compared to alternative architectures such as TALEs [38, 39]), 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* [36], as well as in eukaryotes [34, 35, 53]. 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 1A) [54], 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 [50, 55]. 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 (Supplementary Figure S2). 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[56, 57]. 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 [36]. 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 (Supplemental Figure S3) and repressibility by BCR-ABL1 (Figure 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 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 1D). Overall, responses were unimodal, such that population-averaged fluorescence measured by flow cytometry corresponded well to comparable metrics obtained by microplate-based assays (Supplemental Figure S4), 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 1

We first examined the impact of promoter design on repressibility by inspection. As depicted in Figure 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 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 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 2C). Adding additional ZFP binding sites downstream of the -10 box did not increase repressibility (Figure 2D). However, adding additional binding sites upstream of the -35 box did effectively increase repressibility (Figure 2E). Decreasing the spacer distance between the -10 box and the first downstream binding site increased repressibility (Figure 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 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

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 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 [58], and these corrected coefficients were used to generate a ranked lists of features (for PLSR coefficients, see Supplementary Figure S5). 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 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 (Supplementary Figure S5). The final method used was Lasso regression, which performs least squares regression with an additional penalty placed on the magnitudes of regression coefficients [59]. 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 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 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.

Figure 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 [60-65]. 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 [66-69], split GFP fragments [28], or context-dependent fluorophores [45], 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 [55, 70]. 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 [55] (see Figure 4A for the proposed mechanism). None of the reporters evaluated exhibited repression upon the induction of SZF biosensor expression (Figure 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” [24], and BCR-ABL1 was inserted between these N and C terminal fragments (see Figure 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 [71]). 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 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 4E). This correspondence suggests that the rules for predicting robust ZFP-mediated repression of reporter output (Figures 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 1D), SP biosensor-mediated repression and alleviation was also unimodal when analyzed by flow cytometry (Supplemental Figure S6).

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 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 (Supplemental Figure S7). Moreover, the addition of maltose did not substantially impact cell growth (Supplemental Figure S8). We next investigated how extracellular maltose concentration impacts biosensor performance (Figure 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 [72] (Figure 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 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 [50], 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 [73], although tetR is understood to achieve exquisite transcriptional repression through contorting the target DNA rather than through high affinity binding alone[74]. 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) [50]. 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 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 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 5

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 [36], while our minimal ZFP (which lacks a transactivation domain) conferred repression even when bound upstream of the +1 site (Figure 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 2G, 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 4D). In contract, the SZF biosensor evaluated conferred no repression of reporter output (Figure 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 [24]; 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 [62, 71]. 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 [60]. 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 [75]. 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 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 [76]. 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 [25, 77-80]. 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[81]. 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[40]. 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.* [5]. 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 5 using a later time point at which cell growth had slowed, to facilitate this comparison (Supplementary Figure S9). 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[82, 83]. By leveraging the modularity conferred by programmable ZFP binding [50, 75], it may also be possible to implement multiple SP biosensors in a single cell. Moreover, high throughput genome engineering approaches such as MAGE [84] 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.

MATERIAL AND METHODS

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.

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) [85]. Custom RBS sequences were designed using the RBS Calculator [86]. 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 [87], BCR-ABL1 [55], and the Zif268 portion of human EGR1 [88] (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. Electronic plasmid sequence files are included as Supplementary Data and representative plasmid maps are included in Supplemental Figure S1.

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.

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.

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 [58]. 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 [59]. 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.

SUPPORTING INFORMATION

Supporting figures, methods, and plasmid sequence files can be found in the supporting information.

ABBREVIATIONS

MBP – Maltose binding protein

ZFP – Zinc finger protein

TALE – Transcription activator-like effector

PLSR – Partial Least squares regression

MSE – Mean squared error

BLA – Beta-lactamase

SZF – Split zinc finger

SP – Split protein

IPTG - Isopropyl β-D-1-thiogalactopyranoside

LB – Lysogeny broth

OD – Optical density at 600 nm

AUTHOR INFORMATION

AY and JNL conceived the study. AY, ND, and AR conducted the experiments. AY, ND, AR, and JNL analyzed the data and wrote the manuscript.

NOTES

The authors declare no competing financial interests.

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.

FUNDING

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).

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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*.

[6] 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.

[7] 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.

[8] 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.

[9] 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*.

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

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

[12] 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*.

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

[14] 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.

[15] 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.

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

[17] 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.

[18] 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.

[19] 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, *Nature communications 4*, 1413.

[20] 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.

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

[22] 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.

[23] 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.
[24] 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.
[25] Guntas, G., and Ostermeier, M. (2004) Creation of an allosteric enzyme by domain insertion, *J Mol Biol 336*, 263-273.
[26] Meister, G. E., and Joshi, N. S. (2013) An engineered calmodulin-based allosteric switch for Peptide biosensing, *Chembiochem 14*, 1460-1467.
[27] 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.
[28] 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.
[29] Wang, B., Barahona, M., Buck, M., and Schumacher, J. (2013) Rewiring cell signalling through chimaeric regulatory protein engineering, *Biochem Soc Trans 41*, 1195-1200.
[30] 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.
[31] Liu, D., Evans, T., and Zhang, F. (2015) Applications and advances of metabolite biosensors for metabolic engineering, *Metab Eng*.
[32] Strianese, M., Staiano, M., Ruggiero, G., Labella, T., Pellecchia, C., and D'Auria, S. (2012) Fluorescence-based biosensors, *Methods Mol Biol 875*, 193-216.
[33] Oldach, L., and Zhang, J. (2014) Genetically encoded fluorescent biosensors for live-cell visualization of protein phosphorylation, *Chem Biol 21*, 186-197.
[34] 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.
[35] 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.
[36] 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.
[37] 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.
[38] Bogdanove, A. J., and Voytas, D. F. (2011) TAL effectors: customizable proteins for DNA targeting, *Science 333*, 1843-1846.
[39] 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.
[40] 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.
[41] 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.
[42] 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.
[43] 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.
[44] 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.

[45] 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.
[46] 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.
[47] 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.
[48] 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.
[49] 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*.
[50] 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.
[51] 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.
[52] 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.
[53] 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.
[54] 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.
[55] 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.
[56] Deaner, M., and Alper, H. S. (2016) Promoter and Terminator Discovery and Engineering, *Adv Biochem Eng Biotechnol*.
[57] Blazeck, J., and Alper, H. S. (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level, *Biotechnol J 8*, 46-58.
[58] 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.
[59] Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso, *J Roy Stat Soc B Met 58*, 267-288.
[60] 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.
[61] 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.
[62] 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.
[63] 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.
[64] 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.

[65] Riggs, P. D. (2012) Engineered Derivatives of Maltose-Binding Protein, In *Protein Engineering* (Kaumaya, P., Ed.), InTech.
[66] 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.
[67] 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.
[68] 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.
[69] 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.
[70] 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.
[71] 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.
[72] 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.
[73] 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.
[74] 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.
[75] 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.
[76] Makhlynets, O. V., Raymond, E. A., and Korendovych, I. V. (2015) Design of allosterically regulated protein catalysts, *Biochemistry 54*, 1444-1456.
[77] Yu, K., Liu, C., Kim, B. G., and Lee, D. Y. (2015) Synthetic fusion protein design and applications, *Biotechnol Adv 33*, 155-164.
[78] 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.
[79] Ostermeier, M. (2005) Engineering allosteric protein switches by domain insertion, *Protein Eng Des Sel 18*, 359-364.
[80] Topilina, N. I., and Mills, K. V. (2014) Recent advances in in vivo applications of intein-mediated protein splicing, *Mob DNA 5*, 5.
[81] 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.
[82] 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.
[83] 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.
[84] 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.
[85] 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.
[86] 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.
[87] 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.

[88] 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.

Figure 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 Supplementary Figure S4. 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.

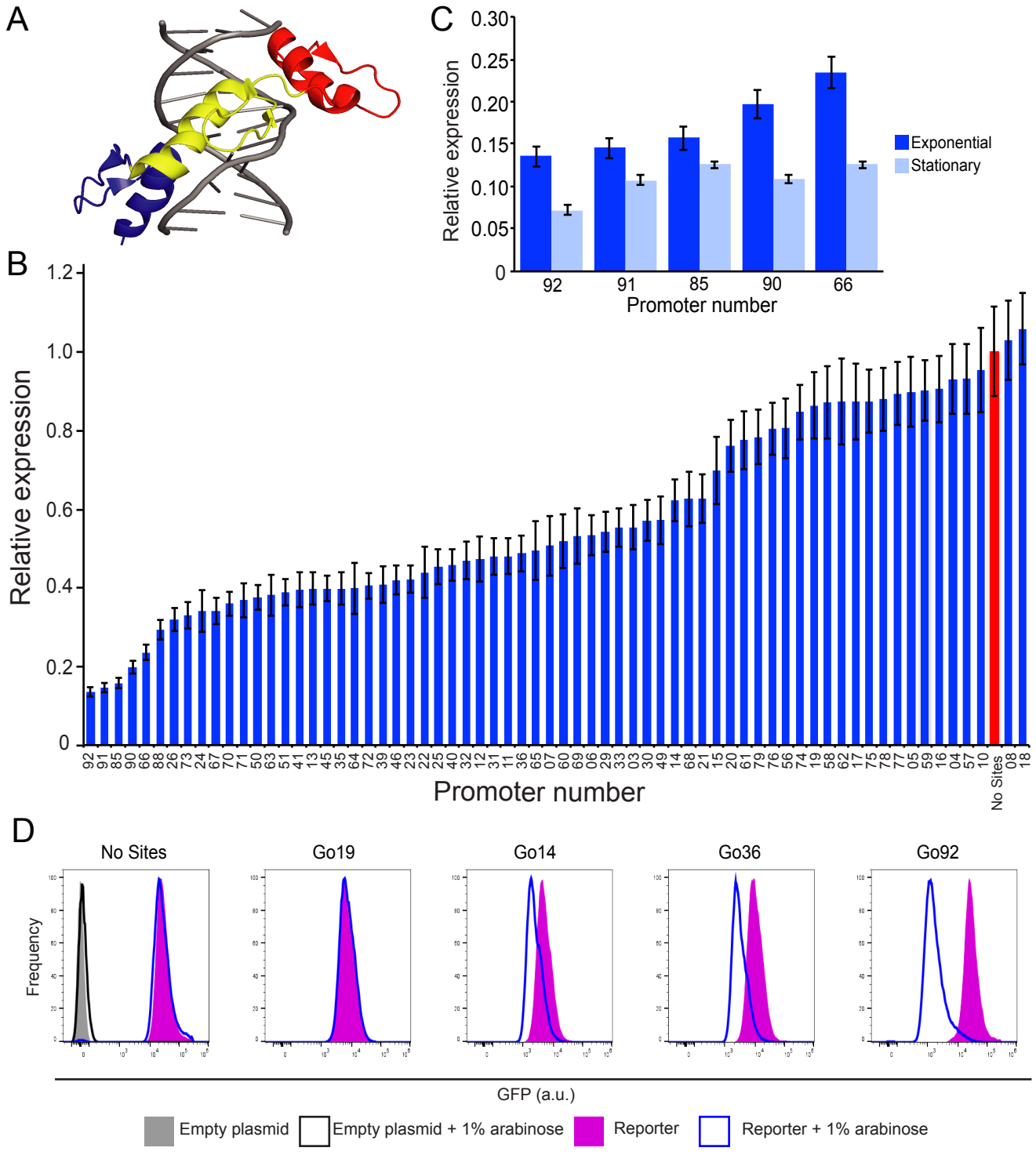
Figure 2. Inspection-based evaluation of promoter design rules. Promoters were manually grouped to represent exploration of design features including (A) presence and location of ZFP binding sites between the -10 box and -35 region, (B) combinatorial effects of having a pair of binding sites between 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 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.

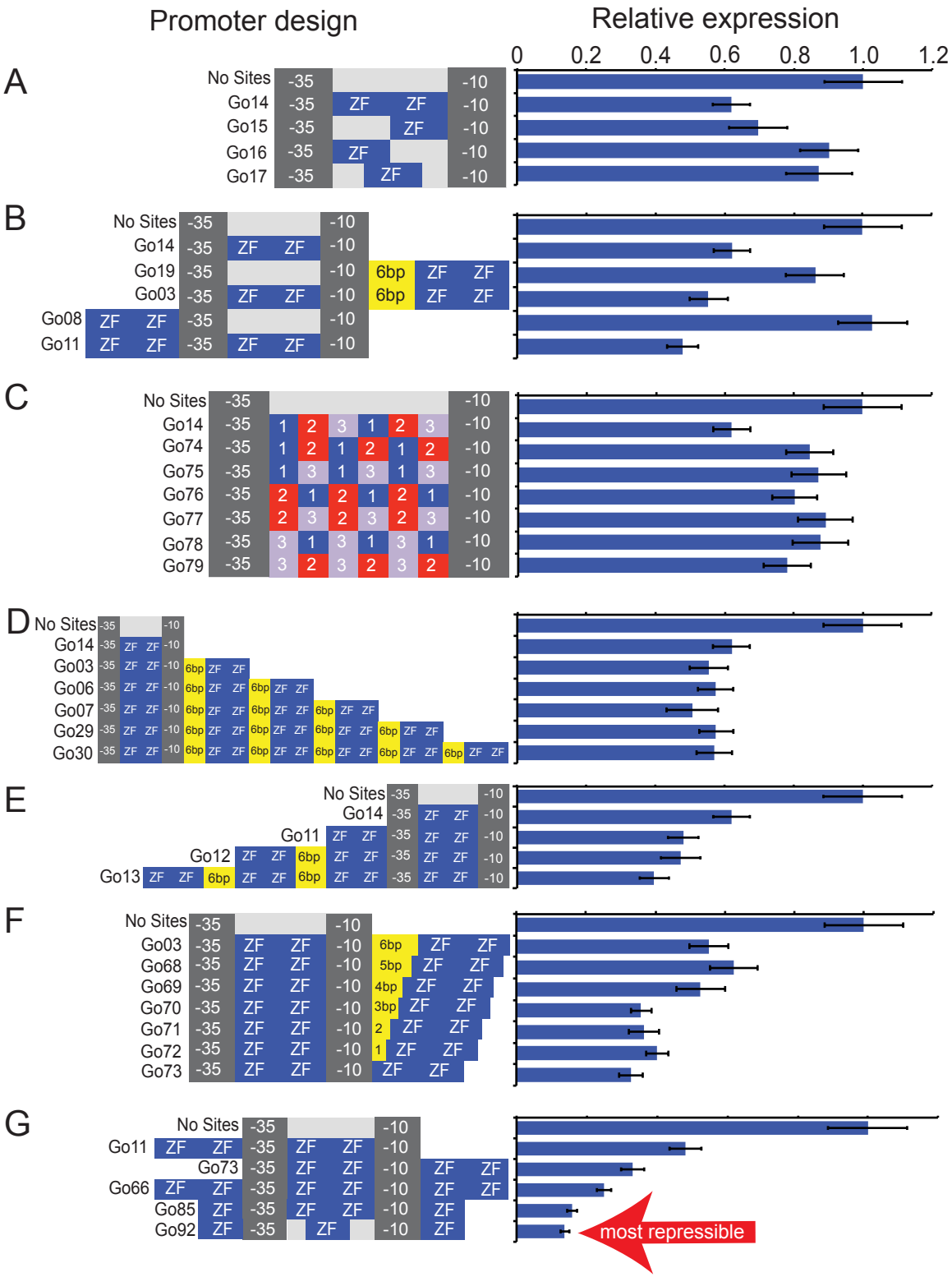
Figure 3. Computational identification of promoter design features conferring ZFP-mediated repressibility (A) Shorthand names and descriptions of the 17 features chosen to describe the promoter library. One binding site is defined as the nine base pair ZFP binding site. One pair is defined as two adjacent 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 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 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 Supplementary Figure S5.

Figure 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 [72], 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 (**$p \leq 0.01$, ***$p \leq 0.001$).

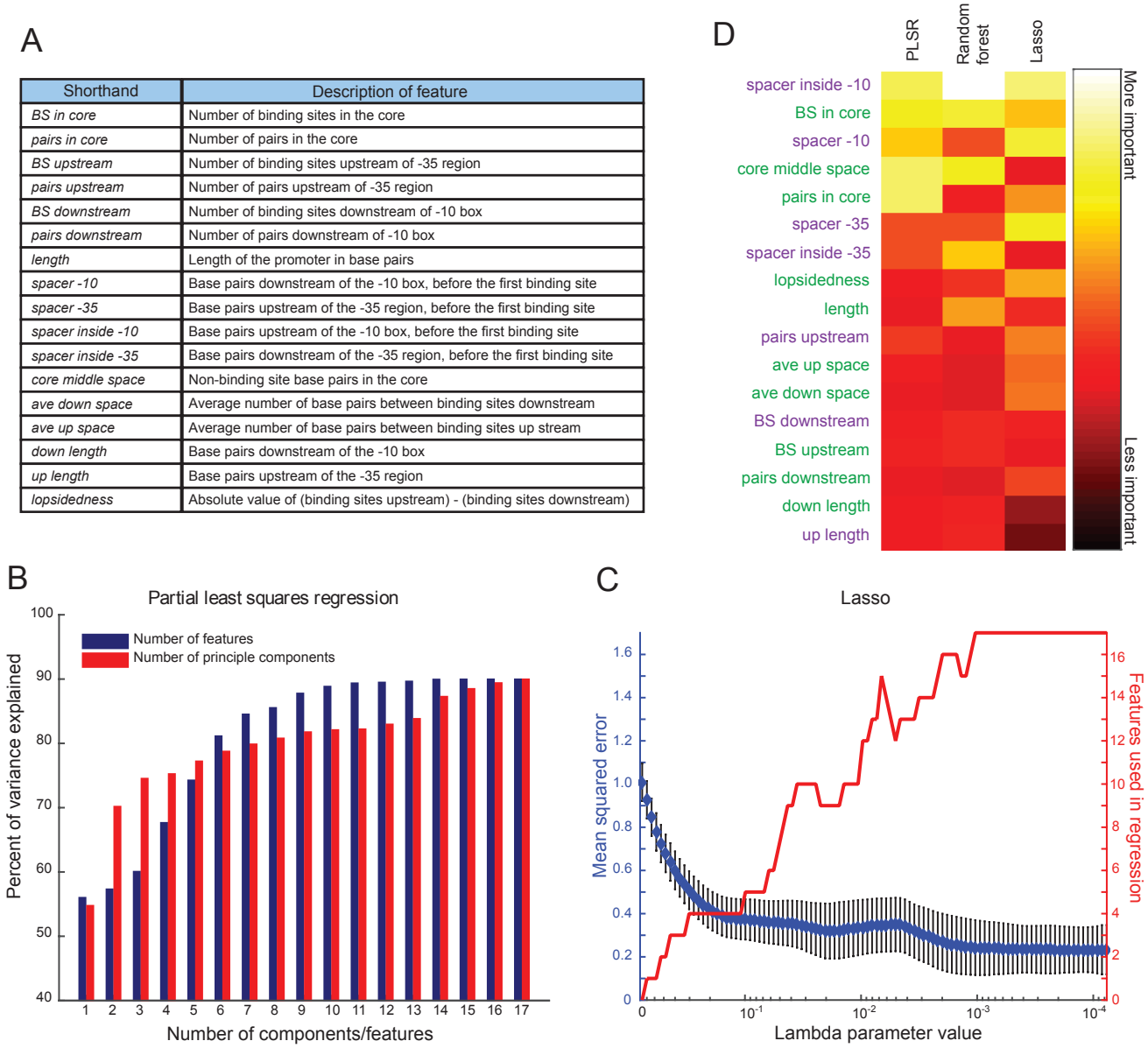
Figure 5. Contributions of biosensor biophysical properties to biosensor performance. (A) At top, the illustration summarizes the biosensor design space explorations described in this figure, using the SP biosensor as a reference case. SP-Zif268 incorporates the tighter binding Zif268 ZFP in place of 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 (*$p \leq 0.05$, *** $p \leq 0.001$).

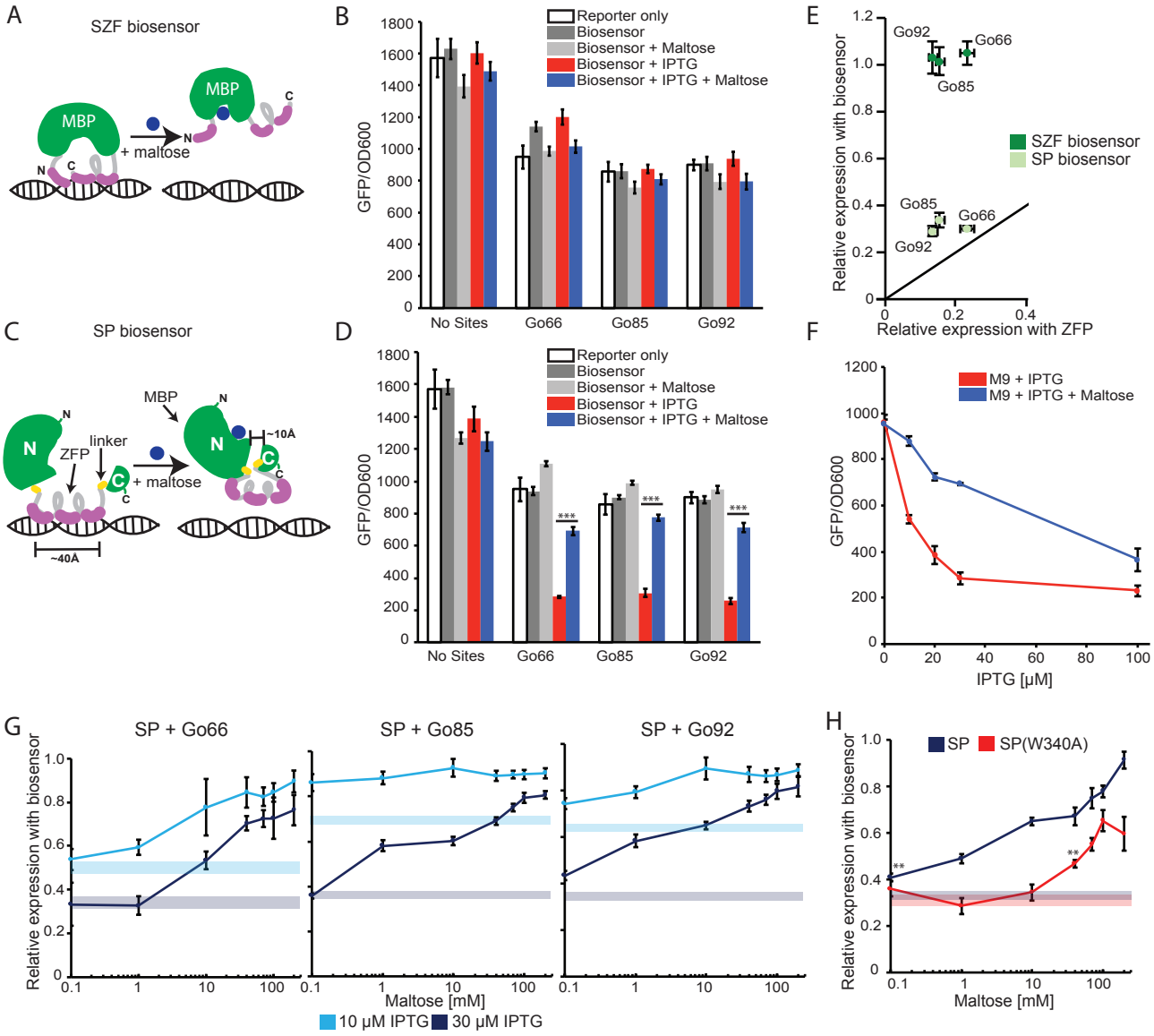

A
C
Relative expression
0.30
0.25
0.20
0.15
0.10
0.05
0
92
91
85
90
66
Promoter number
Exponential
Stationary
B
Relative expression
1.2
1.0
0.8
0.6
0.4
0.2
0
Promoter number
No Sites
D
No Sites
Go19
Go14
Go36
Go92
Frequency
GFP (a.u.)
Empty plasmid
Empty plasmid + 1% arabinose
Reporter
Reporter + 1% arabinose

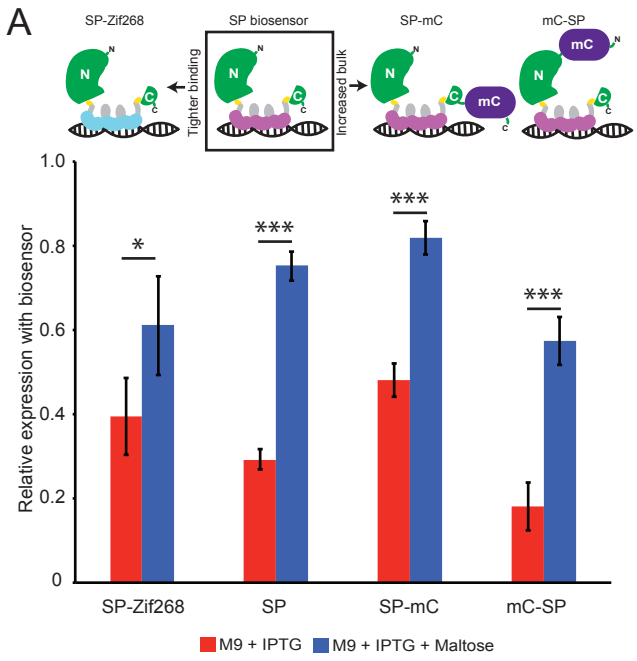

Promoter design
Relative expression
0
0.2
0.4
0.6
0.8
1.0
1.2
A
No Sites -35 -10
Go14 -35 ZF ZF -10
Go15 -35 ZF -10
Go16 -35 ZF -10
Go17 -35 ZF -10
B
No Sites -35 -10
Go14 -35 ZF ZF -10
Go19 -35 -10 6bp ZF ZF
Go03 -35 ZF ZF -10 6bp ZF ZF
Go08 ZF ZF -35 -10
Go11 ZF ZF -35 ZF ZF -10
C
No Sites -35 -10
Go14 -35 1 2 3 1 2 3 -10
Go74 -35 1 2 1 2 1 2 -10
Go75 -35 1 3 1 3 1 3 -10
Go76 -35 2 1 2 1 2 1 -10
Go77 -35 2 3 2 3 2 3 -10
Go78 -35 3 1 3 1 3 1 -10
Go79 -35 3 2 3 2 3 2 -10
D
No Sites -35 -10
Go14 -35 ZF ZF -10
Go03 -35 ZF ZF -10 6bp ZF ZF
Go06 -35 ZF ZF -10 6bp ZF ZF 6bp ZF ZF
Go07 -35 ZF ZF -10 6bp ZF ZF 6bp ZF ZF 6bp ZF ZF
Go29 -35 ZF ZF -10 6bp ZF ZF 6bp ZF ZF 6bp ZF ZF 6bp ZF ZF
Go30 -35 ZF ZF -10 6bp ZF ZF 6bp ZF ZF 6bp ZF ZF 6bp ZF ZF 6bp ZF ZF
E
No Sites -35 -10
Go14 -35 ZF ZF -10
Go11 ZF ZF -35 ZF ZF -10
Go12 ZF ZF 6bp ZF ZF -35 ZF ZF -10
Go13 ZF ZF 6bp ZF ZF 6bp ZF ZF -35 ZF ZF -10
F
No Sites -35 -10
Go03 -35 ZF ZF -10 6bp ZF ZF
Go68 -35 ZF ZF -10 5bp ZF ZF
Go69 -35 ZF ZF -10 4bp ZF ZF
Go70 -35 ZF ZF -10 3bp ZF ZF
Go71 -35 ZF ZF -10 2 ZF ZF
Go72 -35 ZF ZF -10 1 ZF ZF
Go73 -35 ZF ZF -10 ZF ZF
G
No Sites -35 -10
Go11 ZF ZF -35 ZF ZF -10
Go73 -35 ZF ZF -10 ZF ZF
Go66 ZF ZF -35 ZF ZF -10 ZF ZF
Go85 ZF -35 ZF ZF -10 ZF
Go92 ZF -35 ZF -10 ZF
most repressible

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)




A
SZF biosensor
MBP
+ maltose
B
Reporter only
Biosensor
Biosensor + Maltose
Biosensor + IPTG
Biosensor + IPTG + Maltose
GFP/OD600
No Sites
Go66
Go85
Go92
E
Relative expression with biosensor
Relative expression with ZFP
SZF biosensor
SP biosensor
C
SP biosensor
linker
ZFP
~10Å
~40Å
D
F
M9 + IPTG
M9 + IPTG + Maltose
IPTG [μM]
G
SP + Go66
SP + Go85
SP + Go92
Maltose [mM]
10 μM IPTG
30 μM IPTG
H
SP
SP(W340A)


A
SP-Zif268
SP biosensor
SP-mC
mC-SP
Tighter binding
Increased bulk
N
C
mC
Relative expression with biosensor
1.0
0.8
0.6
0.4
0.2
0
*
***
***
***
SP-Zif268
SP
SP-mC
mC-SP
M9 + IPTG
M9 + IPTG + Maltose