https://utilitasmathematicapublishing.com
Journal of Synthetic Computability and Biology (JSCB)
Vol: 1, ISSUE: 1, 2025
Pg No: 1-12

Statistical modeling aids in the development of high-performance whole cell biosensors

Ronnie, Frankie

Roskilde University, Roskilde, Denmark

ABSTRACT

Whole cell biosensors are genetic systems that, for sensing and control applications, correlate the presence of a chemical or other stimulus with a user-defined gene expression output. However, the degree of gene expression of biosensor regulatory components needed for optimal performance is not intuitive, and multidimensional experimental space is not effectively explored by standard iterative procedures. In order to overcome these obstacles, we effectively mapped gene expression levels and improved the performance of biosensors using a design of experiments (DoE) methodology. This approach was used on two biosensors that react to protocatechuic acid and ferulic acid, byproducts of the catabolic degradation of lignin biomass. In order to create biosensor designs with both digital and analog dose-response behavior, we systematically changed the biosensor's dose-response behavior using DoE. We did this by increasing the maximum signal output by up to 30 times, improving the dynamic range by more than 500 times, increasing the sensing range by approximately 4 orders of magnitude, increasing the sensitivity by more than 1500 times, and modulating the slope of the curve. This DoE approach has potential for optimizing metabolic pathways and regulatory systems built from new, poorly understood components.

KEYWORDS: Design of experiments, Definitive screening design, Whole cell biosensors, Protocatechuic acid, Ferulic acid

INTRODUCTION

A biosynthetic system of biological components called a whole cell biosensor is made to transform a stimulus—such as the presence of a chemical, a change in osmolality, or a redox state—into a quantifiable physiological response. By measuring a somewhat basic quantifiable output, biosensors allow for the quick and straightforward detection of tiny molecule effectors. It is possible to build biosensors that enable simple, high-throughput evaluation of the relevant stimuli, in contrast to the low or medium throughput of conventional chemical analytical methods. In order to detect a variety of stimuli, allosteric transcription factors (aTFs) have been widely used for biosensing applications.1–4 In the absence of a particular effector, repression-based aTFs bind their homologous promoter-operator, preventing transcription. A conformational shift brought on by the effector molecule's attachment to the aTF results in a loss of DNA binding and derepression, which in turn triggers the expression of a reporter gene like gfp. Biosensors have found use in biotechnological and biomedical sensing, diagnostic devices, monitoring metabolism, controlling protein expression, and identifying novel genes in metagenomics libraries.5–7 A biosensor system must frequently undergo additional iterative tuning after initial creation in order to function well in its new genetic setting. For optimal biosensor performance, a number of parameters need to be optimized: dynamic range, which is the ratio of the system's ON and OFF states, should be high to allow more confident "hit" identification due to a high signal-to-noise ratio; output in the ON-state (reporter expression level) should be

maximized to allow signal detection in the presence of background noise and to achieve high levels of gene expression for sensing and control applications; and output in the OFF-state (leakiness) should be minimized to allow accurate measurements at low signal levels. Additionally, a biosensor's sensitivity, sensing range, and specificity should be taken into account for specific applications. In primary screening applications, biosensors should have high sensitivity to enable binary (yes/no) classification of positive hits and analyte detection at low levels (<μM). Biosensors that react to a broad range of inducer concentrations would enable initial hits to cluster and be separated into distinct subgroups according to analyte concentration for subsequent screening applications. In order to reduce false detection from analytes with nearly comparable chemical structures or properties, the biosensor must also be extremely selective for biotransformation and other diagnostic applications. The current absence of generic biosensor engineering and design guidelines restricts the wider use of biosensors in sensing and control applications.2. Mechanistic modeling, rational engineering, and directed evolution have all been used to optimize biosensors.8–14 Although these strategies have been effective in clarifying biosensors, they frequently need either the utilization of well-characterized sensory elements or resource-intensive, iteratively controlled evolution using technically difficult selection techniques. Although statistical modeling and structured, multivariate experimentation have not yet been used to address this issue, they may provide quick, cost-effective ways to improve biosensors and clarify general design principles.

Numerous engineering and process sectors make extensive use of statistical modeling and structured multivariate experimentation.15–17 Design of experiments (DoE), a statistical technique that combines modeling and experimentation, is used to methodically explore multidimensional experimental space with the fewest possible experimental runs (Figure 1A). Through a series of statistically structured experiments, it enables researchers to interpret nonintuitive relationships and optimize poorly understood processes.18 During the development of bioprocesses, DoE is frequently used to optimize environmental parameters including temperature, duration, and concentration. Because these factors are continuous, it is easy to conduct systematic experimental exploration under a variety of multivariate settings. In recent times, DoE has also been effectively employed in an increasing variety of applications to optimize genetic variables for metabolic engineering of biosynthetic pathways;19–22 its use is a potent technique to significantly improve metabolic pathways' performance. DoE has not yet been used in more complicated genetic systems with numerous protein-protein and protein-DNA interactions, which are more likely to exhibit nonlinear effects, even if these studies demonstrate its usefulness for linear biosynthesis routes. The difficulty of transforming various closely related but distinct genetic designs into so-called continuous factors complicates this. Without this conversion, the possibility of fewer necessary experiments would be limited because evaluation of all experimental circumstances would have to be done again for every genetic design. We illustrate the usefulness of this method for regulatory system optimization by using a contemporary DoE framework23 in the development of biosensors.

Here, we used DoE to overcome the drawbacks of conventional iterative design-build-test-learn experimental methods for genetic system optimization, which can be expensive in terms of time and resources. We want to investigate and enhance the functionality of several aTF-based small-molecule-responsive biosensors in order to evaluate this technology. Three libraries of regulatory components—two promoters and one RBS—were created, and their expression performance was evaluated. A very effective,

structured coarse-grained map of experimental space was explored using fractional sampling and linear regression modeling. To maximize the performance of a two-gene protocatechuic acid (PCA) sensitive biosensor, this process was used.24 As a key intermediary of lignin catabolic pathways in microbes and an aromatic compound generated from lignocellulosic biomass, PCA is of biotechnological interest for the conversion of lignin into high-value compounds.25, 26 The widely used E. coli recombinant expression systems were then compared to an improved PCA biosensor. The pcaK transporter was then added to the PCA biosensor to boost its sensitivity. By putting pcaK under the direction of a PCA-responsive inverter, biosensors functioning under an analog dosage response modality were subsequently constructed using the DoE idea and regulatory components. Lastly, a more sophisticated enzyme-coupled biosensor that includes three functional genes for the detection of ferulic acid27—a significant aromatic chemical building block produced from lignin—was used to evaluate the DoE idea and regulatory components.26 The outcomes of this optimization effort indicate that other biosensors' performance could be improved by using this strategy as well. All things considered, the method showed that DoE could effectively map experimental space and create genetic systems with much improved output signal, basal control, dynamic range (signal-to-noise), and sensitivity.

RESULTS AND DISCUSSION

PCA biosensor design. In a prior study, we developed a two-plasmid PCA biosensor (PAB)24 that was made up of the PcaV repressible PPV promoter upstream from a reporter gene (GFP) on one plasmid and the PCA-responsive allosteric transcription factor (aTF), PcaV from Streptomyces coelicolor, controlled by a constitutive PlacI promoter on the other plasmid. The PAB was merged into a single plasmid (pPPV-GFP-pcaV) to make deployment easier. The dynamic range of this single plasmid PAB was good (ON/OFF = 417; Table 1). Nevertheless, in contrast to other widely utilized
E. systems of expression for E. Coli.28
We aimed to maximize signal output and dynamic range by redesigning the PAB and methodically altering the genetic components that comprise this biosensor, using DoE to direct the process, in order to investigate whether biosensor performance might be enhanced (Figure 2A). We sought to develop a statistical model explaining the interaction of the genetic components by employing DoE to optimize the PAB and enhance the performance of different iterations. This model might be used as a guide for future attempts to design, optimize, and modify biosensors.
PCA Biosensor Refactoring Under DoE Guidance. In order to determine which parameters are most crucial to the process being studied, a screening phase is conducted first. This is followed by an optimization phase in which the components are modified to achieve the intended optimum. The low level is coded as -1, the intermediate level as 0 and the high level as +1. These factors are set at distinct "levels" that fall within a predetermined range. In order to apply DoE and remodel the PAB, we had to first determine which factors were likely to affect biosensor performance and then translate those factors into levels appropriate for DoE. The constitutive proB promoter (henceforth Preg), which controls pcaV expression; the PcaV-repressible PPV promoter (henceforth Pout); and the G10 RBS (henceforth RBSout), which controls the expression of the sensor output sf GFP (Figure 2A), were the three genetic regulatory components that were chosen and altered to control the transcription and translation of the components that make up the PAB. These three elements were chosen for a methodical research using Design of Experiments since it has been demonstrated that they are all

significant for a biosensor's response1. Since RNAP binding and translation rate, which are determined by the promoter and RBS, respectively, may have distinct effects on the system's response curve, we chose to modify these separately.1 Throughout the first set of experiments, we maintained the same transcriptional terminators, gene orientation, antibiotic selection marker, and plasmid copy number; however, in subsequent experiments, we changed the copy number by converting a stable multicopy system to a single-copy system (see below).

Following the selection of three elements for analysis, we generated, screened, and ranked the performance of libraries for each of these components (Preg, Pout, and RBSout) in order to transform them into continuous variables. By transforming categorical data—in this case, a specific promoter or RBS—into continuous variables with a broad expression range, this step makes statistical-model based optimization easier. In order to increase the confidence of the component performance within the genetic context of the biosensor and to guarantee that the expression level of the library was finely resolved and spanned a wide range, we chose to create new libraries rather than use libraries that had already been published. The libraries were built in the pSEVA131 vector with mCherry acting as a stand-in for pcaV to represent regulator expression and sf GFP as the biosensor output. In order to reconstitute a functioning biosensor after library building and performance evaluation, pcaV was used in place of mCherry (see below). To avoid transcriptional read-through, the genes producing sf GFP and mCherry were positioned differently and separated by a spacer of around 150 bp. RBSout and Pout were employed to

• X = 0, P = Pmax + Pmin

Since logarithmic variables more accurately reflect the cellular biophysics of transcription and translation, the linlog transformation was previously proven to be crucial for the effective application of a DoE-based optimization process,19 which is why it was used here. For each library, the strongest member was recoded as +1, the worst as -1, and the midpoint level 0 was the geometric average of levels +1 and -1. The members were ranked from -1 to +1 (Figure 2C,D,E). To completely explore the gene expression space, 10 648 (22 × 22 × 22) combinations would be required due to the size of the screened libraries. By using structured screening designs, DoE seeks to minimize the number of combinations required to properly explore an experimental space (Figure 1A) and determine the relative importance of various factors.18 While a variety of screening designs are available in a DoE methodology, the definitive screening design (DSD) was chosen because it avoids confounding pairs of second-order effects and allows for the estimation of curvature (nonlinearity) in a factor-response relationship, which is likely to be found in biological systems.23 DSD designs use three levels instead of two levels, allowing for some estimation of curvature (nonlinearity) in a factor-response relationship. Here, a strong RBS (gaaataaggaggtaatacaa) was used to regulate expression of mCherry, while DSD was used to lower control expression of sf GFP and Preg. This resulted in a construct called p131B, which was used as the starting point for library production. We decided to randomly insert the nucleotides at the following locations in order to create the individual libraries: (i) 3 Ns were added to the −10 hex-box for Preg to create Preg-lib (Figure 2B); (ii) 3 Ns were added to both the −10 and −35 hex-boxes for Pout to create Pout-lib (Figure 2B); and (iii) 6 Ns were added to the core RBS binding region for RBSout to create RBSout-lib (Figure 2B). For Preg-lib, Pout-lib, and RBSout-lib, this resulted in total theoretical library sizes of 64 (43), 4096 (46), and 4096 (46), respectively. E was used to screen the mutant libraries for sfGFP fluorescence. coli for Pout-lib and RBSout-lib, and for Preg-lib's mCherry fluorescence. 22

members were chosen from each library after the first screening in order to cover a broad range of fluorescence values. The following equation was used to calculate promoter and RBS activity by calculating the rate of fluorescent protein (FP) production in accordance with previously published work29,30:

Pout-lib had a 46-fold range (maximum of 17 340 and minimum 377), RBSout-lib had a 160-fold range (maximum of 15 860 and minimum of 99), and Preg-lib had a 46-fold range (maximum of 8101.3 and minimum of 177). These libraries covered a broad range of FP synthesis rates (Figure 2C,D,E). The following equation was used to rescale the expression data from libraries created using a linlog transformation previously mentioned19:

Since logarithmic variables more accurately reflect the cellular biophysics of transcription and translation, the linlog transformation was previously proven to be crucial for the effective application of a DoE-based optimization process,19 which is why it was used here. For each library, the strongest member was recoded as +1, the worst as -1, and the midpoint level 0 was the geometric average of levels +1 and -1. The members were ranked from -1 to +1 (Figure 2C,D,E).
To completely explore the gene expression space, 10 648 (22 × 22 × 22) combinations would be required due to the size of the screened libraries. By using structured screening designs, DoE seeks to minimize the number of combinations required to properly explore an experimental space (Figure 1A) and determine the relative importance of various factors.18 While a variety of screening designs are available in a DoE methodology, the definitive screening design (DSD) was chosen because it avoids confounding pairs of second-order effects and allows for the estimation of curvature (nonlinearity) in a factor-response relationship, which is likely to be found in biological systems.23 DSD designs use three levels instead of two levels, allowing for some estimation of curvature (nonlinearity) in a factor-response relationship. Here, a strong RBS (gaaataaggaggtaatacaa) was used to regulate expression of mCherry, while DSD was used to lower control expression of sf GFP and Preg. This resulted in a construct called p131B, which was used as the starting point for library production. The nucleotides were randomly inserted at the following locations to create the individual libraries: (i) 3 Ns were added to the −10 hex-box for Preg to create Preg-lib (Figure 2B); (ii) 3 Ns were added to both the −10 and −35 hex-boxes for Pout to create Pout-lib (Figure 2B); and (iii) 6 Ns were added to the core RBS binding region for RBSout to create RBSout-lib (Figure 2B). For Preg-lib, Pout-lib, and RBSout-lib, this resulted in total theoretical library sizes of 64 (43), 4096 (46), and 4096 (46), respectively. E was used to screen the mutant libraries for sfGFP fluorescence. coli for Pout-lib and RBSout-lib, and for Preg-lib's mCherry fluorescence. 22 members were chosen from each library after the first screening in order to cover a broad range of fluorescence values. The following equation was used to calculate promoter and RBS activity by calculating the rate of fluorescent protein (FP) production in accordance with previously published work29,30:
Ratio of synthesis X = FP(x)tp2 − FP(x)tp1 a compression ratio of 819:1 for all testing setups from 10 648 to 13 (Table 1).
Variants of PCA biosensors are statistically modeled. The constructs were created in accordance with the DSD, which was produced using statistical software (Materials and Methods) and is displayed in Table 1 (for the raw data, see Supplementary Table S1). To accommodate for the anticipated high number of statistically significant factors and interactions, extra runs were added above the bare minimum needed (2n + 1 = total run number, n (number of factors) = 3). After substituting pcaV for mCherry, all 13 constructs (Figure 3 and Table 1) were successfully put together and changed into E. Coli. Next, we determined the biosensor

dynamic range (ON/OFF) and measured end-point sfGFP fluorescence when uninduced (OFF) and induced with 1 mM PCA (ON) to evaluate the performance of each of the several PAB permutations. A wide range of values for the measured responses are provided by the results of these trials, which are displayed in Figure 4A, B, and Table 1. While the worst performer (pD8) produced a negligible output signal (50 RFU/OD) and was hardly responsive, with a dynamic range of 1.2-fold, the best performing candidate (pD2) demonstrated an excellent maximum signal (62 071 RFU/OD) and good dynamic range (156-fold) while maintaining tight basal control. This underscores the significance of a library-based optimization approach.

The significance of each major influence and its interactions was evaluated using factor screening analysis. Half-normal plots, which enable the interpretation of factor effect on each of the three responses (OFF, ON, and ON/OFF), were used to pick significant factors. For dynamic range (ON/OFF), factor screening analysis showed that Pout ($p < 0.0001$), Pout × RBSout ($p < 0.0001$), and Preg ($p = 0.0004$) had the largest significant effects. Pout, RBSout, and Pout × RBSout exhibited the most significant impact for both the ON and OFF biosensor output responses (Supplementary Figure S1). We used a standard least-squares regression (SLSR) model and analysis of variance (ANOVA) to statistically model the data utilizing those parameters that were found to be significant ($p < 0.05$) for biosensor performance (Materials and Methods).

The factors and interactions that had the most effects on the three responses are shown by comparing the effect sizes in the SLSR (Supplementary Figure S1, Supplementary Table S2, and S3). Figure 4C displays the model's prediction profile. As anticipated, Pout and RBSout are predicted to have the biggest impact for maximum output (ON) and should be set at +1 for maximum signal output, whereas RBSout is the strongest determinant for basal output (OFF) and should be decreased if the basal output signal is too high. ON/OFF is improved by increasing Pout and RBSout, although this improvement is nonlinear when the expression level is changed from middle to maximal (0 versus +1). When examining the effect of RBSout at −1 vs. 0, the model reveals an intriguing trade-off: both OFF and ON are scaled proportionally, meaning that ON/OFF does not significantly change. The OFF level shows a plateau at high expression levels of RBSout (+1), while the ON level rises, causing vertical extension and an increase in ON/OFF. Remarkably, altering the amount of Preg that regulates the expression of pcaV also has a nonlinear influence on the ON/OFF response (Figure 4C). The system functions with a lower dynamic range at high (+1) and low levels (−1) for Preg, whereas the ideal PcaV level to attain the maximum dynamic range is close to the medium level (0). It is not surprising that the biosensor's dynamic range is smaller at low aTF expression levels because there inadequate transcription factor in the system to completely disrupt the development of the RNAP-promoter complex.31 The unexpected increase in output signal at high PcaV levels indicates that too much PcaV disrupts the creation of stable regulator-promoter complexes. Together, these results demonstrate the value of a three-level DSD because, in a two-level design with only a high and low level, the nonlinear effect of the Preg level would have been missed.19,20 By using the DSD, we were able to reliably identify nonlinear effects in the design space and allocate the nonlinear effects to the RBSout and Preg levels. Due to significant aliasing across nonlinear effect terms inside the specified data structure, classic DoE screening designs are unable to assign nonlinear effects in this way. This means that even while nonlinear effects can be detected by typical screening designs, they cannot be linked to a causative factor without supplementing the DoE design with more experimental data, which would take more time and money. This demonstrates a key benefit of the final screening strategy used in this case. The

necessity for further experimental rounds to determine the origin of the nonlinear response is eliminated by the trustworthy resolution of this nonlinear effect.

After nonlinearity was found in the expression space under investigation, we attempted to further resolve the curvature in the Preg promoter activity landscape. In order to accomplish this, we conducted extra experiments in which the level of Preg was set at four distinct levels (−0.56, −0.28, 0.36, and 0.67), while Pout and RBSout were set at the highest level (+1) in order to investigate the terrain surrounding the Preg midpoint (Figure 5A). The dynamic range of the measured responses for these PAB iterations is shown. The final construct, p131C− B10, was developed with Preg set at 0.14 and Pout and RBSout maintained at +1 since the data indicated an optimal dynamic range between levels 0 and 0.36. With a dynamic range of 276 times, this build performed the best out of all the tests (Figure 5A and Supplementary Table S14).

The model was adjusted based on the validation trial results to produce a new prediction profile that describes the data (Figure 5B). Some design guidelines that should apply to other repression-based aTF biosensor systems can be deduced from this model: (i) build the most potent chimeric promoter-operator and RBS combination feasible, followed by (ii) fine-tune the level of regulator with a broad range of expression levels. (iii) Reduce the strength of the RBS driving signal output if, after adjusting the regulator, an acceptable dynamic range cannot be achieved. Crucially, in just eighteen experimental trials, we were able to locate the optima and create this statistical model. Additionally, we can be certain that the PAB has been configured to its best potential because the experimental space has been effectively mapped using this DoE approach. We titrated the original PAB with full induction at 4 and the best variation (p131C−B10). mM PCA (Supplemental Table S5 and Figure 5C), which demonstrated that we had increased the dynamic range by 25% (417- to 521-fold) and the output signal by more than 30 times (3121 to 97 099 RFU/OD).

The impact of copy number on biosensor performance. We next transferred the plasmid-based multicopy biosensor system to a single-copy system on the chromosome to examine the impact of copy number on the PAB's performance. The biosensor was cloned into a pKIKO vector and placed into the arsB locus in several permutations (Materials and Methods). As previously, the Preg level was changed while Pout and RBSout were set at the highest level. The chromosomal PABs' reactions were evaluated. In accordance with the copy number reduction from the pBBR1 origin, which is estimated to have 5–10 copies per E, we discovered that the maximum intensity of output signal was lowered by approximately ten times from the plasmid-based biosensor (Supplementary Table S6). coli cell.32,33 As illustrated in Figure 5D, the system's overall dynamic range was decreased (276-fold to 42-fold; Supplementary Table S6) and the quantity of Preg required for the ideal dynamic range was raised from 0.14 to 0.61 (Figure 5A,D). It is well known that expression correlates proportionately with gene-dosage;34,35; however, this relationship is complex and nonlinear,36 and copy reduction is thought to upset the equilibrium of aTF-based systems by lowering the steady state aTF concentration.37 Mechanistic approaches have tried to explain these findings and suggest that a stronger promoter for pcaV is necessary to reduce basal expression of sfGFP from the biosensor when implemented as a single-copy system. We were able to swiftly rework the biosensor to fine-tune the genome-integrated PAB's biosensor performance because of the broad range of expression space spanned by the calibrated regulatory component libraries.

METHODS

supplies. Cloning, in vivo DNA assembly, 51–54 plasmid propagation, promoter/RBS

characterisation, and biosensor experiments were all performed using Escherichia coli DH5α (NEB, #C2987U). Different non-T7 expression methods were utilized in benchmarking tests using E. Coli BL21 (NEB, #C2530H), while T7 RNAP-based expression systems were employed with E. Coli BL21 (DE3) (NEB, #C2527H) that had a genomic copy of T7 RNAP. The revised ferulic acid-responsive biosensor was tested on E. coli BW25113. All investigations involving E. coli strains were conducted in Luria-Bertani (LB) media, with the exception of the promoter/RBS library characterisation, which employed EZ rich (EZ rich defined medium kit, Teknova, #M2105). Ampicillin (100 µg/mL), kanamycin (50 µg/mL for plasmid selection and 25 µg/mL for genome integration), or hygromycin (100 µg/mL) were added to LB and EZ rich, unless otherwise specified. The water came from a Millipore Milli-Q filtration device. Stock solutions of protocatechuic acid, isopropyl β-D-1-thiogalactopyranoside (IPTG), L-arabinose, and L-mannose were made in sterile water, whereas a stock solution of ferulic acid was made in dimethyl sulfoxide (DMSO). Antibiotics and chemicals were bought from Formedium, Sigma, or Fisher. Synthetic genes and DNA oligos were acquired from GeneArt and/or IDT.

Cloning molecules. Supplementary Tables S12 and S13 include primer sequences and a list of plasmids, respectively. Digestions were performed in accordance with accepted procedures, and restriction enzymes were acquired from NEB. DNA fragments were created for cloning using Q5 polymerase (NEB, #M0491S), and genomic insertion genotyping was done using Phire II (Thermo Fisher, #F126S). NEBuilder was used to complete isothermal assembly55 (NEB, #E2621S). DpnI (NEB) was applied to the fragments produced by PCR. To confirm sequence identity, Sanger sequencing was performed on each construct. The PCA biosensor was pSEVA131,56,57, which contained the ampicillin selection marker and the BBR1 origin. For the inverter system, pSEVA 261 with the p15A origin and kanamycin selection marker was employed. For the ferulic acid biosensor, pET28a (Novagen), which has the pBR322 origin and kanamycin selection marker, was utilized. See the Supporting Information for complete details on the molecular cloning.

RBS Library/Promoter Screening. In two rounds, clones from the Preg-lib, Pout-lib, and RBSout-lib libraries were chosen and described. In the initial round, 960 distinct

Using a Hamilton Star robotic platform, clones from each library were selected from transformation plates and arranged into a square-welled 96 deep-well plate (DWP) with breathable seals and media (0.5 mL LB with ampicillin). In a shaker-incubator (Infors HT), the plates were cultivated for 16 hours at 30 °C, 950 rpm, and 75% humidity. A microtiter plate shaker (Stuart) was used to incubate 2 µL of the cultures for 3 hours at 37 °C and 1000 rpm after they had been subcultured into 198 µL of EZ rich media plus ampicillin in black, clear, flat-bottomed 96-well microtiter plates (MTP; Grenier). To get an end-point measurement, fluorescence and optical density (OD λ = 700 nm) were measured in a ClarioStar microplate reader (BMG). GFP fluorescence and mCherry fluorescence were assessed at λEx/λEm = 488/520 and 570/620 nm, respectively. In order to prevent bleed-through from mCherry fluorescence, OD700 was measured rather than OD600.58 22 clones covering a broad range of RFU/OD were chosen from each library using the normalized value of fluorescence to optical density. The background signal for autofluorescence was adjusted for in all RFU/OD measurements by deducting the RFU/OD value of the empty vector negative control.

Using a shaker-incubator (Infors HT), individual colonies were arrayed in triplicate in DWPs containing 0.5 mL LB plus ampicillin (with breathable seals) and grown for 16 hours at 30 °C at 950 rpm, 75% humidity. Selected clones from each library were streaked onto ampicillin plates and grown

overnight for the second round. 75ul of the overnight cultures were transferred to a black microtiter plate with 50 µL of 50% glycerol in order to create "one-shot" stock solutions for cryopreservation. These plates were stored at -80 °C after being well mixed in an MTP shaker (1000 rpm, 1 min). A DWP containing 495 µL of LB plus ampicillin was inoculated with 5 µL of each well after the cryopreserved MTPs had been thawed in an MTP shaker (37 °C, 1000 rpm) for 30 minutes in order to measure promoter activity. In a shaker incubator, the DWP plates were cultivated for 16 hours at 30 °C, 950 rpm, and 75% humidity. The main culture was created using the overnight precultures by transferring 4 µL of cells into a 96-well MTP that had a clear, black bottom and contained 196 µL of EZ rich plus ampicillin. The cells were then incubated in an MTP shaker at 1000 rpm at 37 °C. At 2 and 3 hours, which were previously identified as the times when E. coli grows at its fastest rate under our experimental conditions, OD700 and FP (fluorescent protein) fluorescence were measured. The calculation of promoter activity was based on the literature29,30 (see eq 1). According to the literature19, promoter activity was converted into a logarithmic dimensionless variable (see eq 2). Genomic Integrations. Using lambda red recombination, genomic cassettes were introduced into the E. coli DH5α chromosome.A Qiagen PCR purification column was used to clean up the integration cassette after 59 specific PAB variants were transferred to the pKIKOarsBKM integration vector and 60 amplified using primers AB 39/40. To induce the production of the λ Red recombinase proteins, E. coli DH5α was transformed using the pSIM18 vector, cultured to an OD600 of approximately 0.3, and then heat-shocked for 15 minutes at 42 °C. The cells were electroporated with 300 ng of the PCR product and the transformants chosen on LB plates supplemented with kanamycin at 37 °C after being washed five times in ice-cold sterile water. Using primers AB 34/61, colony PCR was used to confirm that the cassette was inserted at the right locus. In order to cure the pSIM18 strains, clones were restreaked onto LB plates after being subcultured for a whole night at 42 °C on LB + kanamycin.

containing either hygromycin (no growth) or kanamycin (growth) to verify pSIM18 deletion. DoE trials. On LB plus ampicillin or LB plus kanamycin plates, E. Coli DH5α cultures with the plasmid or chromosome-based PAB variations were streaked, respectively. As previously mentioned, "one-shot" stock solutions for cryopreservation were created using individual clones of each strain. Before 10 µL of each well was inoculated into a DWP containing 190 µL of LB + ampicillin (for plasmid-based variations) or LB only (for chromosomally integrated variants), the cryopreserved MTPs were thawed in an MTP shaker set to 37 °C and 1000 rpm for 30 minutes for DoE trials. 5 µL of cells were transferred into a DWP containing 445 µL of LB with ampicillin to create the main culture. The precultures were created by growing the DWP plates for 16 hours at 30 °C, 950 rpm, and 75% humidity in a shaker incubator. DWPs were incubated for two hours at 37 °C and 1000 rpm in an MTP shaker. After that, 50 µL of PCA (10 mM) was added to the clones to induce them to a final concentration of 1 mM PCA, and they were allowed to develop for an additional three hours. To pellet the cells, the DWPs were centrifuged at 2250g for 10 minutes. The spent media was then pipetted with 500 µL PBS. After pelleting and re-washing the cells, 50 µL of the cell suspension was put into a clear-bottomed black MTP with 150 µL of PBS, and OD700 and GFP fluorescence were quantified. Benchmarking of Expression Systems and Biosensor Titrations. Individual clones of each strain were used to inoculate 5 mL of LB plus ampicillin in a 50 mL conical tube, which was then grown for 16 hours at 37 °C and 180 rpm in a shaking incubator (New Brunswick I26). Strains with inducible expression systems were streaked onto LB plates supplemented with ampicillin. After diluting the overnight

cultures 1:100 in LB + ampicillin in a DWP, they were incubated for 2 hours at 1000 rpm and 37 °C in a Stuart MTP shaker. They were then induced by adding the proper inducer and let to grow for an additional 3 hours. The titration included the following final inducer concentrations: 4000, 1000, 250, 62.5, 15.6, 3.9 μM, and no inducer. The FA With the exception of concentrations of 1000, 200, 40, 8, 1.6, and 0.32 μM and the absence of an inducer, biosensor titration was performed in the same manner. As previously mentioned, the cells were pelleted, cleaned, and their OD700 and fluorescence were assessed.
Extension of Dose Response. On solid LB medium supplemented with ampicillin and kanamycin (25 μg/mL and 12.5 μg/mL, respectively) and 1 mM PCA, E. coli DH5α with the p131C−B10 biosensor and p261-lacI-pcaK variations were plated. In a 50 mL conical tube, single isolated colonies were inoculated into 5 mL of LB supplemented with the necessary antibiotics. The tubes were then shaken at 180 rpm for 16 hours at 37 °C. Following a 100-fold dilution in new LB media including antibiotics, the cells were moved to a 96-well DWP and incubated for two hours at 37 °C at 1000 rpm in an MTP shaker. The final culture volume reached 500 μL after the proper inducer concentration was introduced after this outgrowth period. The cells were cultured for a further twenty-four hours. Since the two plasmid inverter systems and the expression of the system components had a major negative impact on the growth, a longer induction period was necessary. The following were the final inducer concentrations: 1000, 200, 40, 8, 1.6, 0.32, 0.064, 0.0128, and 0 μM PCA.
Modeling and data processing. JMP Pro 12 (SAS Institute Inc.) was used for all data processing and statistical analysis, including standard least-squares regression, factor screening, and experiment design. Response information was trans-

created to log10. The DoE definitive screening tool was used to create the DSD data table, and factors were chosen using Lenth's t-ratio, Half-Normal plot analysis of the factor contrast, and Lenth's pseudo standard error (PSE). SLSR fitting included factor contrasts that were considered significant for the model and that did not follow the half-normal distribution. By analyzing simultaneous p-values, factor significance was evaluated, enabling the model's factor importance to be evaluated. Because effect heredity was preserved, if a factor was included in a significant interaction term with another factor but was not considered significant on its own, both terms from this interaction were included in the model fitting. The PSE, which is produced from an estimation of the residual standard error utilizing inactive terms inside the model, was used to generate simultaneous p-values. A Monte Carlo simulation with 10,000 runs was performed from this PSE in order to estimate the p-value. PRISM 7 (GraphPad Software) was used to create the graphs, which were then fitted using a Hill fit.

REFERENCES

1. Ang, J., Harris, E., Hussey, B. J., Kil, R., and McMillen, D. R.
2. (2013) Tuning response curves for synthetic biology. *ACS Synth. Biol. 2*, 547−567.
3. De Paepe, B., Peters, G., Coussement, P., Maertens, J., and De Mey, M. (2017) Tailor-made transcriptional biosensors for optimizing microbial cell factories. *J. Ind. Microbiol. Biotechnol. 44*, 623−645.
4. Lim, H. G., Jang, S., Jang, S., Seo, S. W., and Jung, G. Y. (2018) Design and optimization of genetically encoded biosensors for high-throughput screening of chemicals. *Curr. Opin. Biotechnol. 54*, 18−25.
5. Shi, S., Ang, E. L., and Zhao, H. (2018) In vivo biosensors: mechanisms, development, and applications. *J. Ind. Microbiol. Biotechnol. 45*, 491−516.
6. Liu, D., Evans, T., and Zhang, F. (2015) Applications and advances of metabolite biosensors for metabolic engineering. *Metab. Eng. 31*, 35−43.
7. Koch, M., Pandi, A., Borkowski, O., Cardoso Batista, A., and Faulon, J. L. (2019) Custom-made transcriptional biosensors for metabolic engineering. *Curr. Opin. Biotechnol. 59*, 78−84.
8. Mahr, R., and Frunzke, J. (2016) Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl. Microbiol. Biotechnol. 100*, 79−90.
9. Lutz, R., and Bujard, H. (1997) Independent and

tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res. 25*, 1203– 1210.

10. Gatti-Lafranconi, P., Dijkman, W. P., Devenish, S. R., and Hollfelder, F. (2013) A single mutation in the core domain of the lac repressor reduces leakiness. *Microb. Cell Fact. 12*, 67.

11. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005) Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev. 15*, 116–124.

12. Tashiro, Y., Kimura, Y., Furubayashi, M., Tanaka, A., Terakubo, K., Saito, K., Kawai-Noma, S., and Umeno, D. (2016) Directed evolution of the autoinducer selectivity of *Vibrio f ischeri* LuxR. *J. Gen. Appl. Microbiol. 62*, 240–247.

13. Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J., and Voigt, C. A. (2019) *Escherichia coli* "Marionette" strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol. 15*, 196–204.

14. Ellefson, J. W., Ledbetter, M. P., and Ellington, A. D. (2018) Directed evolution of a synthetic phylogeny of programmable Trp repressors. *Nat. Chem. Biol. 14*, 361–367.

15. Wang, B., Barahona, M., and Buck, M. (2015) Amplification of small molecule-inducible gene expression via tuning of intracellular receptor densities. *Nucleic Acids Res. 43*, 1955–1964.

16. Mandenius, C. F., and Brundin, A. (2008) Bioprocess optimization using Design-of-Experiments methodology. *Biotechnol. Prog. 24*, 1191–1203.

17. Kumar, V., Bhalla, A., and Rathore, A. S. (2014) Design of Experiments applications in bioprocessing: Concepts and approach. *Biotechnol. Prog. 30*, 86–99.

18. Franceschini, G., and Macchietto, S. (2008) Model-based design of experiments for parameter precision: State of the art. *Chem. Eng. Sci. 63*, 4846–4872.

19. Lendrem, D. W., Lendrem, B. C., Woods, D., Rowland-Jones, R., Burke, M., Chatfield, M., Isaacs, J. D., and Owen, M. R. (2015) Lost in space: design of experiments and scientific exploration in a Hogarth Universe. *Drug Discovery Today 20*, 1365–1371.

20. Xu, P., Rizzoni, E. A., Sul, S. Y., and Stephanopoulos, G. (2017) Improving Metabolic Pathway Efficiency by Statistical Model-Based Multivariate Regulatory Metabolic Engineering. *ACS Synth. Biol. 6*, 148–158.

21. Zhou, H., Vonk, B., Roubos, J. A., Bovenberg, R. A., and Voigt, C. A. (2015) Algorithmic co-optimization of genetic constructs and growth conditions: application to 6-ACA, a potential nylon-6 precursor. *Nucleic Acids Res. 43*, 10560–10570.

22. Carbonell, P., Jervis, A. J., Robinson, C. J., Yan, C., Dunstan, M., Swainston, N., Vinaixa, M., Hollywood, K. A., Currin, A., Rattray, N. J. W., Taylor, S., Spiess, R., Sung, R., Williams, A. R., Fellows, D., Stanford, N. J., Mulherin, P., Le Feuvre, R., Barran, P., Goodacre, R., Turner, N. J., Goble, C., Chen, G. G., Kell, D. B., Micklefield, J., Breitling, R., Takano, E., Faulon, J. L., and Scrutton, N. S. (2018) An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Commun. Biol. 1*, 66.

23. Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J., and Dueber, J. E. (2013) Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res. 41*, 10668– 10678.

24. Jones, B., and Nachtsheim, C. J. (2011) A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects. *J. Qual Technol. 43*, 1–15.

25. Machado, L. F. M., Currin, A., and Dixon, N. (2019) Directed evolution of the PcaV allosteric transcription factor to generate a biosensor for aromatic aldehydes. *J. Biol. Eng. 13*, 91.

26. Xu, Z., Lei, P., Zhai, R., Wen, Z., and Jin, M. (2019) Recent advances in lignin valorization with bacterial cultures: micro-organisms, metabolic pathways, and bio-products. *Biotechnol. Biofuels 12*, 32.

27. Beckham, G. T., Johnson, C. W., Karp, E. M., Salvachua, D., and Vardon, D. R. (2016) Opportunities and challenges in biological lignin valorization. *Curr. Opin. Biotechnol. 42*, 40–53.

28. Machado, L. F., and Dixon, N. (2016) Development and substrate specificity screening of an in vivo biosensor for the detection of biomass derived aromatic chemical building blocks. *Chem. Commun. (Cambridge, U. K.) 52*, 11402–11405.

29. Morra, R., Shankar, J., Robinson, C. J., Halliwell, S., Butler, L., Upton, M., Hay, S., Micklefield, J., and Dixon, N. (2016) Dual transcriptional-translational cascade permits cellular level tuneable expression control. *Nucleic Acids Res. 44*, No. e21.

30. Davis, J. H., Rubin, A. J., and Sauer, R. T. (2011) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res. 39*, 1131–1141.

31. Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., de Mora, K., Glieberman, A. L., Monie, D. D., and Endy, D. (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng. 3*, 4.

32. Lanzer, M., and Bujard, H. (1988) Promoters largely determine the efficiency of repressor action. *Proc. Natl. Acad. Sci. U. S. A. 85*, 8973–8977.

33. Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., 2nd, and Peterson, K. M. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-

resistance cassettes. *Gene 166*, 175−176.

34. Jahn, M., Vorpahl, C., Hubschmann, T., Harms, H., and Muller, S. (2016) Copy number variability of expression plasmids determined by cell sorting and droplet digital PCR. *Microb. Cell Fact. 15*, 211.

35. Carrier, T., Jones, K. L., and Keasling, J. D. (1998) mRNA stability and plasmid copy number effects on gene expression from an inducible promoter system. *Biotechnol. Bioeng. 59*, 666−672.

36. Kittleson, J. T., Cheung, S., and Anderson, J. C. (2011) Rapid optimization of gene dosage in *E. coli* using DIAL strains. *J. Biol. Eng. 5*, 10.

37. Mileyko, Y., Joh, R. I., and Weitz, J. S. (2008) Small-scale copy number variation and large-scale changes in gene expression. *Proc. Natl. Acad. Sci. U. S. A. 105*, 16659−16664.