ACS Synthetic Biology

Freie Universität Berlin

Article

Sort-seq approach to engineering a formaldehyde-inducible promoter for dynamically regulated *Escherichia coli* growth on methanol

Julia Rohlhill, Nicholas Richard Sandoval, and Eleftherios Terry Papoutsakis

ACS Synth. Biol., **Just Accepted Manuscript** • Publication Date (Web): 02 May 2017

ACS Publications

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

Sort-seq approach to engineering a formaldehyde-inducible promoter for dynamically regulated *Escherichia coli* growth on methanol

Julia Rohlhill,[†] Nicholas R. Sandoval,[‡,*] Eleftherios T. Papoutsakis[†,*]

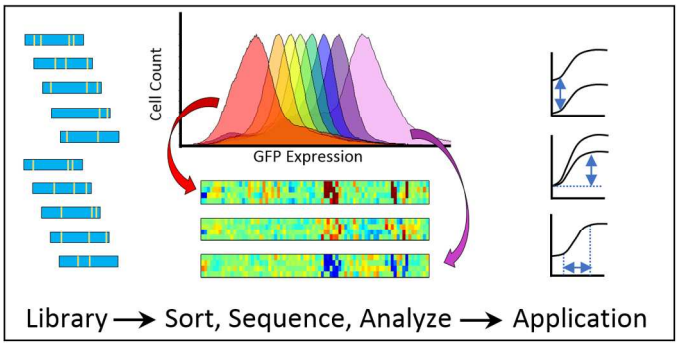
[†]Department of Chemical & Biomolecular Engineering and the Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711, United States

[‡]Department of Chemical & Biomolecular Engineering, Tulane University, New Orleans, Louisiana 70118, United States

For Table of Contents Use Only

Sort-seq approach to engineering a formaldehyde-inducible promoter for dynamically regulated *Escherichia coli* growth on methanol

Julia Rohlhill, Nicholas R. Sandoval, Eleftherios T. Papoutsakis



ABSTRACT

Tight and tunable control of gene expression is a highly-desirable goal in synthetic biology for constructing predictable gene circuits and achieving preferred phenotypes. Elucidating the sequence-function relationship of promoters is crucial for manipulating gene expression at the transcriptional level, particularly for inducible systems dependent on transcriptional regulators. Sort-seq methods employing fluorescence-activated cell sorting (FACS) and high-throughput sequencing allow for the quantitative analysis of sequence-function relationships in a robust and rapid way. Here we utilize a massively parallel sort-seq approach to analyze the formaldehyde-inducible *Escherichia coli* promoter (P_{frm}) at a single-nucleotide resolution. A library of mutated formaldehyde-inducible promoters was cloned upstream of *gfp* on a plasmid. The library was partitioned into bins via FACS based on GFP expression level, and mutated promoters falling into each expression bin were identified with high-throughput sequencing. The resulting analysis identifies two 19-bp repressor binding sites, one upstream of the -35 RNA polymerase (RNAP) binding site and one overlapping with the -10 site, and assesses the relative importance of each position and base therein. Key mutations were identified for tuning expression levels and were used to engineer formaldehyde-inducible promoters with predictable activities. Engineered variants demonstrated up to 14-fold lower basal expression, 13-fold higher induced expression, and a 3.6-fold stronger response as indicated by relative dynamic range. Finally, an engineered formaldehyde-inducible promoter was employed to drive the expression of heterologous methanol-assimilation genes and achieved increased biomass levels on methanol, a non-native substrate for *E. coli*.

KEYWORDS: sort-seq; promoter engineering; biosensor; transcription factor binding; transcriptional fine-tuning.

Precise control of gene expression is a necessity for designing predictable gene circuits in synthetic biology and increasing the yields of products encoded by biosynthetic pathways. Controlling the rate of transcription typically involves controlling the interactions between the RNA polymerase (RNAP), the promoter DNA, and any associated transcriptional regulators. Constitutive expression systems, while often used for heterologous protein production, fail to optimize expression levels of metabolic intermediates, and thus often require the cell to carry high metabolic burdens.[1, 2] Synthetic gene regulatory schemes frequently employ transcription factors such as activators and repressors to introduce various positive and negative feedback control mechanisms.

Complex synthetic regulatory networks require orthogonal transcription factors with preferably modular DNA-binding and sensing domains. When controlling the expression of measurable reporters, such as fluorescent proteins or antibiotic resistance markers, these transcription factor-based biosensors have shown promise by increasing the efficiency of high-throughput screens and selections, and in dynamic pathway regulation by allowing real-time monitoring of intracellular metabolite concentrations.[3, 4] Dynamic regulation, widespread in natural systems, eliminates the need for expensive inducers and offers the potential for optimized schemes which minimize unnecessary metabolic burdens through autonomous pathway balancing.[4] Efforts to expand the synthetic biology toolbox have focused on characterizing a range of biosensors[5] and engineering existing regulators[6] to respond to new effectors.[7, 8] Biosensor development could greatly benefit from additional small molecule sensors and the elucidation of their corresponding operators.

Protein-DNA binding interactions can be investigated and ultimately manipulated by quantifying the sequence-function relationship of promoter DNA. A method for elucidating

sequence-function relationships employs fluorescence-activated cell sorting (FACS) and high-throughput sequencing, generally referred to as 'sort-seq' or 'FACS-seq'.[9] These sort-seq schemes begin with the generation of a library of mutated sequences for the regulatory element or protein of interest, which is large and diverse enough to contain variants displaying a wide range of activities.[9] This library is linked to a genetic reporter and sorted into bins based on fluorescence. Here, GFP expression levels represent the activity of the promoter library cloned upstream of the *gfp* gene (Figure 1). Subsequent sequencing of gated populations enables the use of various analysis methods to quantify the activities of hundreds of thousands of variants. One such technique, originating from information theory, allows the quantification of the relationship between two variables, here the base at each nucleotide position (sequence) and output expression level (function) as determined by discrete sorted bins.[10] This quantification is achieved by calculating the mutual information, or dependence of the two random variables on each other:[11, 12]

$$I(b_i;\mu) \approx \sum_{b_i,\mu} f(b_i,\mu)\ \log_2 \frac{f(b_i,\mu)}{f(b_i)f(\mu)} - c \qquad (1)$$

where b_i is the base at position i, μ is the activity bin, $f()$ represents the joint and marginal frequency distributions and c is a correction factor.[12, 13] If the bases at position i are independent of the resulting expression bin (μ), that position is inconsequential to gene expression. Similarly, mutations with skewed distributions, occurring more frequently in low or high expression bins identify vital nucleotide positions which play a deterministic role in expression level and resulting expression bin. While sort-seq approaches have been used to investigate regulatory sequences and proteins,[14] they have rarely been used in combination with mutual information techniques. Two papers of interest used the approach to analyze mammalian enhancers[11] (termed

massively parallel reporter assay (MPRA)) and the CRP activator binding[12] to the prokaryotic *lac* promoter.

Formaldehyde is a toxic compound but also a common cellular metabolite produced endogenously in all cells at low concentrations from various demethylation reactions.[15] *E. coli* has a native formaldehyde-inducible promoter (P_{frm}), found upstream of the *frmRAB* formaldehyde-detoxification operon. FrmR, the first product of the operon, is a member of the DUF156 family of DNA-binding transcriptional regulators.[16] It binds the *frmRAB* promoter region and is negatively allosterically modulated by formaldehyde.[16, 17] FrmR is specific to formaldehyde, responding to acetaldehyde, methylglyoxal, and glyoxal to a far lesser degree, and not at all to a range of other aldehydes and alcohols tested.[16, 17] The genes *frmA* and *frmB* encode a formaldehyde dehydrogenase and S-formylglutathione hydrolase, respectively, and are responsible for detoxifying formaldehyde to formic acid in a glutathione-dependent pathway.[18] The negative feedback regulation of the *frmRAB* operon is similar to many other prokaryotic operons whereby the transcription factor represses its own transcription.[19] Characterizing P_{frm} and the P_{frm}-FrmR relationship adds another operator-regulator to the synthetic biology toolkit and offers further insight into protein-DNA molecular binding mechanisms.

In addition to its ubiquitous role in all cells, formaldehyde is a central metabolic intermediate for methylotrophs. Synthetic methylotrophy, or the utilization of C_1 compounds such as methanol as a carbon and energy source by non-native methylotrophs, has been pursued in earnest recently as methanol availability increases and its price declines.[20, 21] Previous studies have shown labeling in *E. coli* glycolytic intermediates from ^{13}C-methanol by heterologously expressing three enzymes from *Bacillus methanolicus*.[22] Methanol dehydrogenase (Mdh) converts methanol to formaldehyde, hexulose-6-phosphate synthase (Hps) condenses

formaldehyde with D-ribulose 5-phosphate to form hexulose 6-phosphate (Hu6P), and phospho-3-hexuloisomerase (Phi) isomerizes Hu6P to fructose-6-phosphate (F6P) for entrance into central metabolism. Our group recently utilized a superior Mdh from *B. stearothermophilus*,[23] in addition to the *B. methanolicus* codon-optimized Hps and Phi enzymes, to achieve *E. coli* growth on methanol with a small (1 g/L) yeast extract supplementation, demonstrating extensive ^{13}C labeling from ^{13}C-methanol into glycolytic and TCA intermediates and amino acids, as well as methanol conversion to the specialty chemical naringenin.[24] We have also demonstrated a strategy of scaffoldless enzyme assembly that can be used to achieve superior outcomes in synthetic methylotrophy.[25] Placing formaldehyde assimilation genes under the control of formaldehyde regulation emulates the native regulation of methylotroph *B. methanolicus*, whereby *hps* and *phi* are transcriptionally induced by formaldehyde,[26] and results in autonomous pathway balancing. This dynamically regulated substrate-utilization scheme is particularly beneficial considering the toxicity of formaldehyde, the metabolic burden of constitutively expressing the heterologous methanol assimilation genes at high levels, and the additional burden expected from future strain engineering for the production of valuable chemicals or secondary metabolites.

Here we first characterize the native *E. coli* P_{frm} response and regulation, identifying parameters for improvement. We investigate the influence of P_{frm} architecture on the strength of repressor binding by testing a set of P_{frm} variants and isolate approximate FrmR binding regions. We then describe the deconstruction and analysis of the P_{frm} promoter using a sort-seq approach, obtaining, with single-nucleotide resolution, a map of the importance of each nucleotide position for expression, both with and without formaldehyde induction. The analysis of the resulting rich dataset allowed us to identify mutations capable of changing promoter activity in a directed

manner by manipulating repressor and RNAP binding interactions, and this information was used to design a set of formaldehyde-responsive promoters with tunable basal and induced expression levels. An engineered promoter was further used to implement and improve formaldehyde-controlled *E. coli* consumption of the non-native substrate methanol.

RESULTS AND DISCUSSION

The *E. coli* formaldehyde-inducible promoter is an ideal candidate for engineering. We aimed to construct a formaldehyde reporter plasmid and analyze the response of the native P_{frm}. To construct the reporter plasmid, termed P_{frm}-GFP-P_{lac}-FrmR, P_{frm} was cloned upstream of *gfp* and the FrmR repressor was cloned under the control of the *lac* promoter to limit titration issues. P_{frm} promoter activity was assayed by monitoring GFP expression via flow cytometry. The expression of the reporter plasmid was assayed in the NEB5α and Δ*frmR* strains, representing two different regulatory systems (Figure 2a-b). In the NEB5α strain, FrmR levels were autoregulated by P_{frm} on the *E. coli* chromosome, in addition to being expressed from the reporter plasmid. In the Δ*frmR* strain, FrmR was only expressed from the reporter plasmid.

Time response curves for formaldehyde concentrations from 1-500 μM show maximum activity from 100 to 250 min and up to an 8-fold dynamic range, calculated as the ratio of induced activity to uninduced activity (Figure 2a-b). FrmR expression was expectedly higher in the NEB5α strain as evidenced by 3-fold lower GFP expression at time zero compared to the Δ*frmR* strain. The response curves were modeled with the Hill equation[27, 28]. The Hill equation (equation 2) is used to characterize the induction response and cooperativity of the system, particularly of interest here due to the tetrameric structure[29, 17] of FrmR.

$$P([\mathrm{I}]) = P_{min} + (P_{max} - P_{min})\frac{[\mathrm{I}]^n}{K^n + [\mathrm{I}]^n} \quad (2)$$

where P_{min} and P_{max} represent the basal promoter activity and the maximum promoter activity following induction, respectively; [I] is the formaldehyde inducer concentration; n is the Hill coefficient; and K is the apparent dissociation constant related to the inducer concentration at which promoter activity is half-maximal. The Hill coefficient indicates the cooperativity of the system (i.e. the positive or negative effect a single ligand binding event has on subsequent events), with increasing values >1 corresponding to more sigmoidal response curves and higher cooperativity. The apparent Hill coefficient is 1.18 ± 0.13 when FrmR is present on the chromosome, indicating a noncooperative promoter-repressor-RNAP interaction when FrmR is both expressed from a plasmid and autoregulated by formaldehyde (Table S4). Without chromosomal FrmR expression the apparent Hill coefficient was 0.46 ± 0.02 (Table S4). The response curve for the NEB5α strain without plasmid-expressed FrmR can be seen in Figure S2. Disruption of negative autoregulation typically increases the Hill coefficient, leading to a tighter sigmoidal response curve.[30, 31] The opposite trend is seen here, with derepression continually increasing with higher formaldehyde concentrations, possibly due to the toxicity of formaldehyde and the interruption of autoregulation due to the high levels of plasmid-expressed FrmR.

Compared to similarly characterized operator-regulator biosensors with dynamic ranges up to 210-fold,[5] the 5.5-fold range of the P_{frm}-FrmR system at 100 μM formaldehyde induction has ample room for improvement. It is also highly sensitive, responding to dosed formaldehyde concentrations as low as 1 μM. This initial characterization suggests the *E. coli* P_{frm} is a strong candidate for engineering. Further promoter characterization through the deconstruction and analysis of P_{frm} architecture can identify operator regions with high engineering potential.

Inverted repeats are central to P_{frm} architecture and response. The most distinct feature of the *E. coli* P_{frm} promoter is a 19 bp perfect inverted repeat, originally hypothesized as a FrmR binding site (Figure 3).[32] Operator sequences often contain inverted repeats which can form hairpin loops, an important structural feature for protein-DNA interactions.[33] A similarly regulated FrmR homologue was identified in *Salmonella enterica* under the control of a promoter lacking the large inverted repeat of the *E. coli* P_{frm}.[29] A smaller incomplete 5'-**ATAGTATA**/**TATAGTAT**-3' inverted repeat was noted within the *Salmonella* promoter, disruption of which was shown to ablate FrmR binding.[16] Here, P_{frm}-GFP constructs were generated with different architectures to investigate each side of the large 19-bp inverted repeat, termed here Site A and Site B. Sites were replaced with scrambled and reverse complemented sequences to test the importance of their presence and orientation for FrmR binding.

Replacing site A and site B individually with scrambled sequences resulted in constructs 1 and 3, respectively (Figure 3). Both constructs showed response to formaldehyde, supporting the hypothesis that each site is independently capable of binding FrmR. The presence of only one site leads to higher expression compared to the native two-site P_{frm}, an effect which is amplified with 100 μM formaldehyde induction. The 100 μM formaldehyde concentration was chosen to provide strong induction without inhibiting growth. Site B leads to greater repression compared to site A, as evidenced by the higher expression levels for construct 3 compared to construct 1. The enhanced effect of site B is likely due to its position, which is closer to the transcription start site and overlapping with the -10 site.

Due to the overlap between FrmR binding site B and the -10 site it was difficult to resolve expression differences due to FrmR binding from those due to RNAP binding. In order to investigate this, we shifted site B upstream, decreasing the space between sites A and B from 15

to 10 nucleotides and separating site B from the -10 region in constructs 6-8. The -10 site was optimized to the canonical 'TATAAT' sequence to investigate the effect of FrmR binding with increased basal expression due to RNAP binding. FrmR was still capable of binding to the shifted site B, as shown by the formaldehyde responsiveness of construct 7 which lacks site A. Construct 6 included both fully intact sites and maintained low basal expression while more than doubling induced expression with a 7.3-fold dynamic range, suggesting the ability to increase the dynamic range of the promoter by separating the manipulation of RNAP binding from the transcriptional repressor binding.

Ablation of FrmR binding was achieved in constructs 4 and 5, shown by the lack of formaldehyde response. The expression levels of constructs 4 and 5 should therefore represent only the effect of RNAP binding on transcription. Construct 4 exhibits 1.6-fold higher expression levels than construct 5, possibly due to an effect of the scrambled or reversed site A sequence on RNAP binding. Construct 4 utilizes scrambled sequences for both site A and site B, confirming their necessity for FrmR binding in *E. coli*. Construct 5 has a scrambled site B and a partial reverse complement for site A which was unable to recover binding. The partial reverse complement of site A cannot bind FrmR independently based on construct 5, however it appears to cause stronger repression when site B is also present. Construct 2, with a reversed site A, shows lower basal and induced expression compared to construct 1 which had a scrambled site A. The analogous construct 8 with reversed site A also showed stronger basal and induced repression compared to construct 7 with scrambled site A. This indicates a relationship between the binding sites since the partial reverse complement of site A contributes to repression only when site B is present, but not independently.

High diversity library generation ensures rich information output. We aimed to generate a high diversity P_{frm} library containing variants covering a wide range of basal and induced activities as we cannot analyze the effect of mutations which are not represented. Promoter library construction requires careful consideration for sort-seq experiments to ensure enough diversity to create the variants of interest for downstream analysis. The 200 bp P_{frm} promoter on the P_{frm}-GFP-P_{lac}-frmR reporter plasmid was targeted for mutation with error-prone PCR using primers flanking the region (Table S2). The variability present in the P_{frm}-GFP libraries was assayed using flow cytometry and compared to the unmutated parent P_{frm}-GFP strain. Increasing mutation frequency in the promoter region leads to a wider fluorescence distribution, and therefore a wider range of promoter activity. However, a critical threshold mutation frequency causes an increasing percentage of the population to lose function, usually due to being unable to initiate transcription from an inability to bind RNAP. The goal was to use a highly diverse library containing millions of unique sequences, while retaining function. In this case, the majority of the population should also retain formaldehyde-responsiveness, since the difference in activities among the library clones is essential for identifying repressor binding sites with high precision and accuracy. The final P_{frm}-GFP library was chosen after three successive rounds of error-prone PCR achieved a mutation frequency which maximized the spread of the expressing population while minimizing the relative size of the non-expressing population.

The final library had an average of 6.6 mutations per 200 bps of the promoter, with an average of 2.4 of those mutations located within the first 80 bases. It covered a wider range of GFP expression as evidenced both visually (Figure S3) and by a 2.2-fold increase in the robust percent coefficient of variation (%rCV), a metric for the spread of the fluorescence distribution

resistant to outlier effects (Table S5).[34] Importantly, the fluorescence distribution of the P_{frm} library showed a similar geometric mean as the unmutated P_{frm}, and therefore its wider distribution was due not only to clones with lower expression but also clones with higher GFP expression. Individual colonies were selected and assayed with flow cytometry to confirm the existence of promoter variants resulting in unique fluorescence distributions under both induced and uninduced conditions.

Limiting constraints affecting the size and diversity of the promoter library similarly limit the information output from sort-seq analysis. Mutational bias was minimized by using a blend of polymerases (Methods) instead of the *Taq* polymerase, which has a well-documented preference for mutating A's and T's.[35] Analysis of the final P_{frm} library bias from sequencing results indicated a slight preference for mutating C's and G's, with C→T/G→A being the most common mutation at a mutation rate of 1.7 percent and A→C/T→G being the rarest at a rate of 0.15 percent (Figure S4, Table S6). The per position mutation frequency was 2.8% on average, and ranged from 1% to 8% (Figure S5). Mutations at each nucleotide position along the length of the 200 bp P_{frm} are therefore represented in the final library, and while mutation bias skews the depth of the library by variable representation of mutations, this skew is accounted for in the calculation of mutual information.

High-resolution binding sites and identification of mutations of interest from sort-seq data. We aim to identify nucleotide position targets for engineering and the FrmR binding region with high precision and accuracy through analysis of the sequencing data for each expression gate. Sequencing data were processed to calculate information footprints for each of three experiments with different levels of expressed and bound FrmR (Figure 4, Figure S6). The first two experiments involved the NEB5α strain under induced (100 μM formaldehyde) and

uninduced conditions with FrmR expressed from both the plasmid and the chromosome (Figure 4), and the third experiment involved FrmR expression from the plasmid only in the Δ*frmR* strain, presumably resulting in lower FrmR expression levels (Figure S6). Information footprints visualize the contribution of each nucleotide in the promoter sequence to GFP expression by analyzing the relationship between the mutations at a specific nucleotide position and the bin into which the mutated promoters were sorted.[12] Deleterious mutations reducing transcription would be consistently sorted into low expression bins, identifying the corresponding nucleotide position as one with high 'information content', or high potential for affecting gene expression. Similarly, high information content positions are identified due to mutations at positions causing higher GFP expression, and consistently falling into high expression bins. High information content nucleotide positions within the promoter region are therefore ideal engineering targets for influencing output gene expression.

Heat maps displaying the distribution of mutations across different sequencing bins reveal the effects of mutating any individual nucleotide with exceptional precision (Figure 5). While information footprints communicate the correlation between nucleotide position and gene expression, heat maps include more detailed information about the specific bases at each nucleotide position. Information footprints do not differentiate between positive or deleterious mutations but heat maps visually display sequencing information by showing the distribution of A, T, C, and G at each nucleotide position for each expression gate. High expression (up) mutations of interest were identified by analyzing mutations with a strong pattern of enrichment in high expression sorting bins compared to the unsorted P_{frm} library. Similarly, low expression (down) mutations of interest had extremely low occurrence in high expression bins and were

highly enriched in low expression bins. These trends confirm that the mutations of interest influenced the level of GFP expression, and resulting sorting bin.

Comparing the information footprints for induced and uninduced experiments yields a single-nucleotide resolution binding site for transcriptional repressor FrmR. Within the larger 19-bp inverted repeat, 4-nucleotide inverted repeats can be identified (Figure 4). These 5'-ATAC/GTAT-3' inverted repeats upstream of the -35 site and overlapping with the -10 site have lower information content in the induced state when less FrmR is bound to the region (Figure 4). Within site A, three positions exist with much higher information content in the uninduced state compared to the induced state. Two are within the 4-nucleotide inverted repeats 5'-ATA**C**/**G**TAT-3' and one is directly centered in the 9-nucleotide spacer between site A and site B. Site B shows a similar pattern complicated by the -10 site. The two 5'-ATA**C**/**G**TAT-3' nucleotide positions show higher information content in the uninduced versus induced states. The 3' side of the inverted repeat is entirely within the -10 site. The mutational distribution across expression bins also identifies secondary versions of the FrmR binding site, noting a 3-nt 5'-ATA/TAT-3' inverted repeat and an imperfect 10-nt '**ATATA**GA**ATA**/**TAT**AG**TATAT**-3' inverted repeat directly flanking the 6-nt G and C tracts in site A and site B, respectively (Figure S7). Sequences with mutations extending each of these four inverted repeat structures were sorted into low expression bins, indicating that the DNA hairpin loops adopted multiple possible conformations (Figure S7).

The RNAP binding site, consisting of two six-nucleotide regions centered at the -35 and -10 positions, is easily identifiable in the information footprints. The P_{frm} promoter uses the canonical 'TTGACA' -35 site, and the non-canonical 'TAGTAT' -10 site, with the optimal 17-nucleotide spacer. An alternate 'TATAGT' -10 site two nucleotides upstream was previously

hypothesized[32], however is not supported by the low information content at those two positions. In agreement with the information footprints, the heat maps (Figure 5) show that mutations in the -35 and -10 regions occur frequently in the lowest GFP expression sorting bin. This effect is particularly true for the -35 region which is the consensus sequence, but less so for the -10 region, where three 'up' mutations can be identified. Using these detailed information footprints and binding site information enabled the specific engineering of P_{frm} promoters with predictable activities.

Informed design of tunable formaldehyde-inducible promoters. Mutations identified during sequencing analysis were used in combination to generate variants capable of a wider range of basal and induced promoter activities. Twelve P_{frm} variants were cloned using inverse PCR (Methods, Table S3). Repression mutations were used for construct 14 and 15, up mutations were used for construct 20, and combinations were used for other constructs. Site B down mutations, A→T at position -25 and C→T at position -20, extend the 4-nucleotide inverted repeat to six nucleotides in constructs 14-17 and cause extremely low expression (Figure 6). The induced expression from only one of the four constructs is higher than uninduced expression from the native P_{frm}. The same two down mutations are present in constructs 22-25 however their effects are negated by three up mutations in the -10 region, which essentially scramble the inverted repeat within site B and cause much higher basal and induced expression levels.

Mutations which were highly enriched in high expression bins were used in combination to create high-expression formaldehyde-responsive promoters. Construct 20 (Figure 6) features six up mutations, including three within the -10 region, and had 27-fold higher uninduced GFP expression compared to the native promoter. Construct 20 also retains formaldehyde-responsiveness, with 2-fold higher GFP expression in response to 100 μM formaldehyde than the

native P_{frm}. Construct 24 similarly displays high expression levels with only two essentially nonfunctional down mutations. Increasing the site A inverted repeat from 4 to 5 nucleotides has a significant effect on repression, as seen from the 6-fold lower basal and 3-fold lower induced expression in construct 25 compared to construct 24. Designed constructs exhibited expected expression levels based on the length of site A and site B inverted repeats for down mutations, or the disruption of sites for up mutations.

Quantitative sequence activity models seek to predict the behavior of variants assuming that mutations make additive contributions to activity. These models fail to account for secondary structures in the DNA and sequence features which are particularly important for transcription factor binding. Individual down mutations may cause lower GFP expression, however in combination they silence each other. For example, a G→A mutation at position -39 increases the length of the site A inverted repeat from 4 to 5 nucleotides, as in constructs 14, 17, 18, 19, 20, and 25 (Figure 6). A T→C mutation at position -57 would similarly lengthen the site inverted repeat, however when both mutations occur together the shorter 4-nt site A is maintained as in constructs 15, 19, and 23.

Application of engineered formaldehyde-responsive promoter enables higher methanol growth. The *E. coli* P_{frm} is uniquely qualified to achieve dynamically regulated *E. coli* growth on methanol. Methanol is converted to the toxic intermediate formaldehyde by methanol dehydrogenase (Mdh) in the first step of assimilation, therefore proper pathway balancing is vital to preventing the accumulation of formaldehyde in the cell and associated growth inhibition. Here, with knowledge gained from our previous studies,[24] we pursue a strategy for autonomously sustainable synthetic *E. coli* methylotrophy using formaldehyde-inducible promoters.

P_{frm} and the high-expression P_{frm} construct 20 ($P_{frm}20$) were placed upstream of the methanol assimilation Mdh-Hps-Phi operon in a Δ*frmA* and Δ*pgi* strain. The *frmA* gene, encoding formaldehyde dehydrogenase, was deleted to minimize the loss of formaldehyde to carbon dioxide. Formaldehyde dissimilation in the Δ*frmA* strain still occurs and has been attributed to promiscuous aldehyde dehydrogenase activity, however it is at a much slower rate.[24] Phosphoglucose isomerase (*pgi*) was similarly deleted to force the F6P from methanol assimilation down glycolysis, minimizing the loss of carbon to carbon dioxide during the conversion of F6P to glucose 6-phosphate and eventually ribulose 5-phosphate through the oxidative pentose phosphate pathway. The P_{frm} strain successfully consumed methanol and grew to a higher cell density when media was supplemented with 60 or 240 mM methanol (Figure 7a-b,d-e), demonstrating, for the first time, formaldehyde-induced synthetic methylotrophy. The yield on methanol, calculated as reported[24] by assuming methanol consumption accounted for additional biomass in cultures supplemented with methanol, was similar for the P_{frm} and $P_{frm}20$ strains (Figure 7c,f) however the $P_{frm}20$ strain achieved significantly higher biomass titers than the P_{frm} strain with 60 or 240 mM methanol. We hypothesize the higher formaldehyde-induced expression of key methanol assimilation genes in the $P_{frm}20$ strain enable the more efficient management and assimilation of toxic intracellular formaldehyde.

CONCLUSIONS

We have demonstrated the systematic and quantitative characterization, dissection and analysis of the *E. coli* formaldehyde-inducible promoter at a single-nucleotide resolution. We characterized the native P_{frm} regulation and response, and determined the general FrmR operator region using designed promoter variants. The sort-seq approach and analysis succeeded in not only confirming the FrmR binding site but quantifying the effect of each nucleotide on both

expression and formaldehyde inducibility. Utilizing strategically placed up and down mutations we were able to engineer promoters with a range of basal and induced expression levels in a predictable manner. Application of an engineered formaldehyde-responsive promoter with higher basal and induced expression levels before methanol assimilation genes achieved higher biomass titers than the native *E. coli* P_{frm}, demonstrating not only formaldehyde-controlled synthetic methylotrophy but its improvement through a sort-seq guided engineering approach.

The formaldehyde-inducible *E. coli* promoter is one of dozens of uncharacterized promoters regulated by simple inducible transcriptional regulators. The sort-seq method, analysis, engineering, and application described here can be applied to any transcriptional regulator-operator sequence to be used in synthetic biology or metabolic engineering applications, particularly for the characterization of additional biosensors for gene circuits and dynamic pathway regulation.

METHODS

Strains, plasmids, and growth media. *E. coli* strain NEB5α (New England Biolabs (NEB), Ipswich, MA) was used for plasmid cloning and maintenance. The Δ*frmR* (JW0348-1) and Δ*frmA* (JW0347-1) knockout strains were ordered from the Keio deletion collection.[36] The double deletion Δ*frmA* Δ*pgi* strain was constructed by the method of Datsenko and Wanner on the existing Keio Δ*frmA* strain cured of its kanamycin resistance cassette.[37] The genes on the P_{frm}-Mdh-Hps-Phi plasmid were placed in an operon configuration under the *trc* promoter[38] and the RBS calculator v2.0[39, 40] was used to design synthetic ribosome binding sites for each gene. All strains and plasmids used can be found in Table S1. PCR primers were synthesized by Integrated DNA Technologies (IDT, Coralville, IA) and can be found in Table S2.

For flow cytometry analysis, cells were grown in MOPS minimal medium[41] supplemented with 0.4% w/v xylose, carbenicillin (100 μg/mL), and IPTG (0.1 mM). Minimal media was chosen due to the more defined response curves achieved compared to rich media. Cultures were inoculated from overnights to an OD of 0.05 in 5 mL in 15 mL disposable culture tubes, incubated at 37°C for 2-3 hours, and dosed with 0 to 500 μM formaldehyde (MP Biomedicals, Santa Ana, CA). Inverse PCR was used to create P_{frm} variant constructs 1-8 and 14-25 by amplifying the plasmid backbones and incorporating promoter mutations on the amplification primers (Table S3). Variant plasmids were recircularized with the Q5 Site-Directed Mutagenesis Kit (NEB) and directly transformed into high efficiency chemically competent NEB5α cells (NEB). Variant promoter sequences were confirmed with Sanger sequencing (UD Sequencing and Genotyping Center).

For methanol growth assays, cells were grown in M9 minimal medium[42] supplemented with 1 g/L yeast extract, carbenicillin (100 μg/mL), and 0, 60, or 240 mM methanol (Sigma-Aldrich, St. Louis, MO). Overnight cultures were pelleted, resuspended in M9, and used to inoculate 30 mL fresh media in 250 mL baffled flasks with rubber stoppers to an OD of approximately 0.2. Growth curves were normalized to a starting OD of 0.2. All *E. coli* cultures were grown at 37°C with 250 rpm shaking.

Promoter library generation. Error-prone PCR targeting the 200 bp P_{frm} (Figure S1) was performed using GeneMorph II Random Mutagenesis Kit (Agilent, Santa Clara, CA). The resulting promoter library was purified with a PCR purification kit (QIAGEN, Germantown, MD) and twice used as template to obtain higher mutation rates. The P_{frm}-GFP-P_{lac}-frmR plasmid was amplified omitting the native P_{frm}, treated with DpnI (NEB), and extracted from agarose with a gel extraction kit (QIAGEN). The purified promoter library with unmutated overhang

sequences was inserted back into the plasmid backbone using the NEBuilder HiFi DNA Assembly Master Mix (NEB). Two 20 µL reactions were transformed into a total of 20 aliquots of 50µL chemically competent NEB5α cells. Following a 1 hour recovery in 250 µL SOC medium (NEB), all transformations were combined into two screw-top 125 mL flasks with 30 mL LB supplemented with 1% w/v xylose and 100 µg/mL ampicillin. Cultures were incubated at 37°C for 8 hours, sampled every hour for flow cytometry analysis, and stored frozen at -80°C in 15% v/v glycerol. Plating on solid LB media supplemented with 100 µg/mL carbenicillin after the initial 1 hour recovery indicated a library size of approximately 11 million. For the creation of the Δ*frmR* promoter library, 4 mL NEB5α library frozen stocks were thawed, miniprepped (QIAGEN), and 1 µg was transformed into Δ*frmR* electrocompetent cells. Following a 1 hour recovery in 3 mL SOC, cells were transferred to 15 mL LB supplemented with 100 µg/mL ampicillin for 3 hours, and were stored at -80°C.

Flow Cytometry and Sorting. Cells were analyzed and sorted with a BD FACSAria IIu flow cytometer (Becton Dickinson (BD), Franklin Lakes, NJ). A blue solid state laser (488 nm excitation) and a 530/30 nm filter was used to measure eGFP. FCS files were analyzed using Flowing Software v2.5.1 (Cell Imaging Core, Turku Centre for Biotechnology, Finland). For flow cytometry sampling the geometric mean of the FITC-A fluorescence for 10,000 events was taken as 'promoter activity'. Prior to sorting, the cytometer was calibrated using Accudrop Beads (BD) and SPHERO Rainbow Calibration Particles (Spherotech, Lake Forest, IL).

On the day of sorting, 2-4 mL of library frozen stocks were thawed, centrifuged to remove excess glycerol, and used to inoculate 25 mL MOPS minimal medium supplemented with 0.4% w/v xylose, 100 µg/mL ampicillin, and 0.1 mM IPTG. Cells were monitored hourly until they reached a post-recovery state (~5 hours), as indicated by 85-90% of cells expressing

GFP. Cells were then dosed with 0 or 100 μM formaldehyde, and sorted 2 hours later (Figure S8). The final promoter library in the NEB5α and Δ*frmR* strains was sorted into eight gates with approximately equivalent populations, and 1,000,000 events were collected from each gate directly into LB media. Populations were recovered at 37ºC overnight and mini-prepped for sequencing. Plating indicated approximately 70% of sorted cells survived.

Next-generation sequencing and analysis. Multiplexed sequencing libraries were constructed per manufacturer's instructions with a Nextera DNA Library Preparation Kit (Illumina, San Diego, CA). Pooled libraries were sequenced on a MiSeq desktop sequencer (Illumina) using paired-end sequencing with a read length of 2 x 201 bases at the University of Delaware DNA Sequencing and Genotyping Center. Reads within each experiment and sorted population were processed to remove those under 201 nucleotides and redundant sequences. The number of mismatches between each read and the native sequence, or the hamming distance, was calculated and reads with more than approximately 17 mismatches were discarded. Over 3 million reads met all quality standards and were used for further analysis. Within each bin in an experiment we calculated the frequency of each base at each position from the aligned reads, and divided it by the total number of reads in the experiment to obtain the joint distribution $f(b_i,\mu)$ in equation 1. The marginal distributions, $f(b_i)$ and $f(\mu)$, were calculated by summing the joint distribution along the appropriate dimension. The correction factor[13] in equation 1 was calculated as described[12] with number of possible bases n_b=4 and number of bins n_μ equal to 7 for the Δ*frmR* experiment and 8 for the induced and uninduced NEB5α experiments. Sequence data are available here [https://www.ncbi.nlm.nih.gov/bioproject/383844].

FIGURES

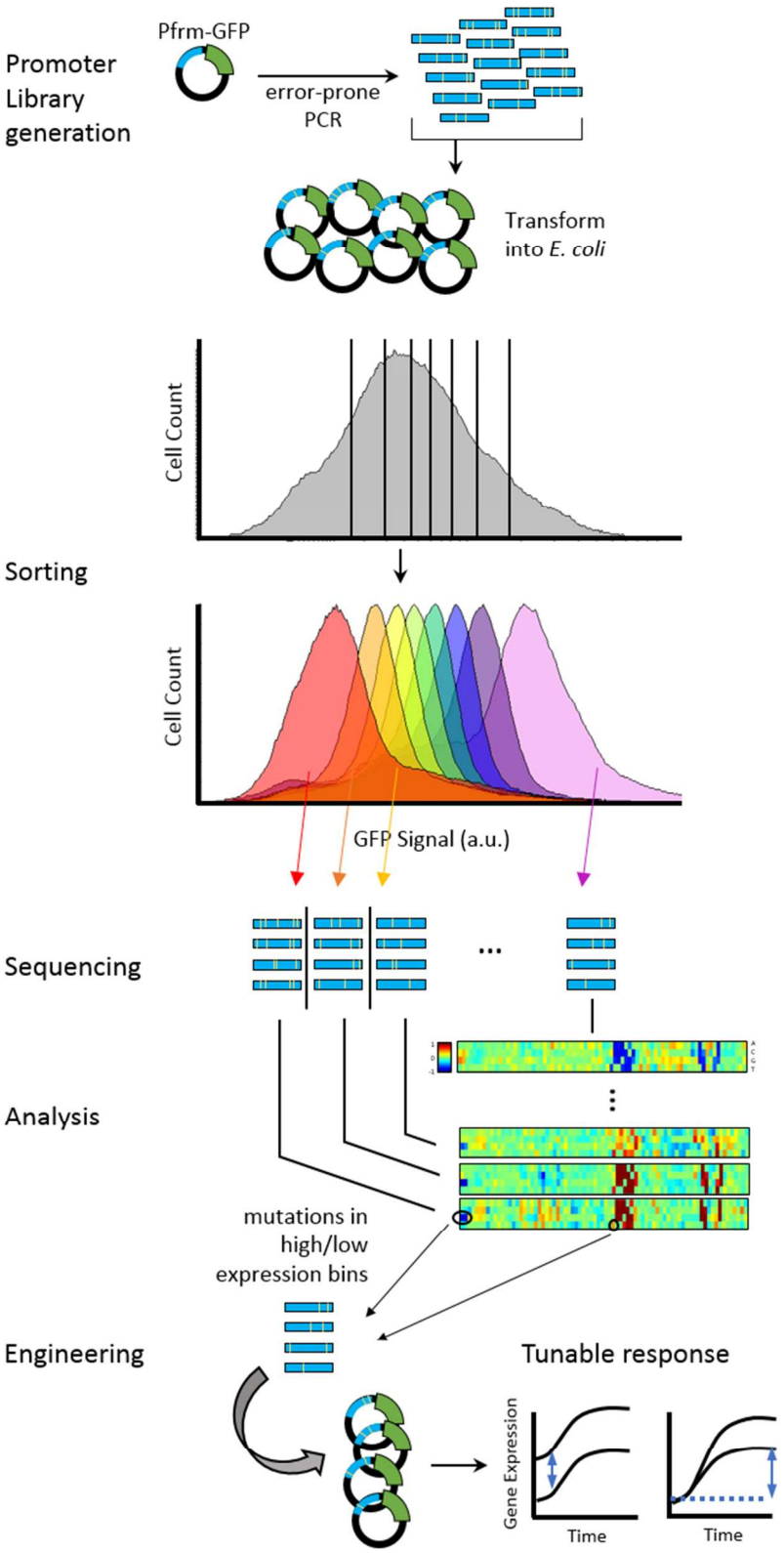

Promoter Library generation
Pfrm-GFP
error-prone PCR
Transform into E. coli
Sorting
Cell Count
Cell Count
GFP Signal (a.u.)
Sequencing
Analysis
mutations in high/low expression bins
Engineering
Tunable response
Gene Expression
Time
Time

Figure 1. Sort-seq experimental method. Promoter library was generated using error-prone PCR and transformed into NEB5α and Δ*frmR* strains. The resulting populations spanned a large range of GFP expression levels and were sorted into 7-8 bins using FACS. Sorted populations were tagged and promoters sequenced, allowing for the identification of mutations leading to higher or lower expression levels. These mutations could then be used to generate inducible promoters with predictable and tunable responses.

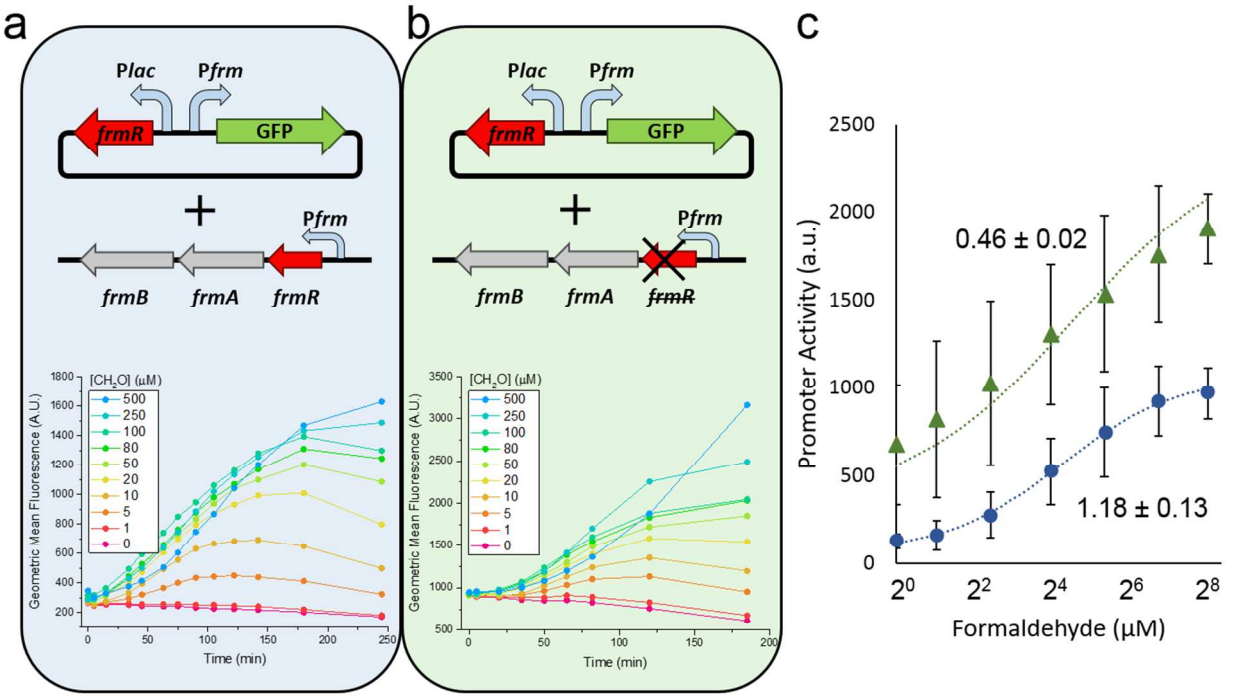


Figure 2. Regulatory mechanisms and response of the *E. coli* formaldehyde-inducible promoter. Two configurations were used to investigate regulation. A P_{frm}-GFP-P_{lac}-frmR reporter plasmid was used with (a) or without (b) without chromosomal expression of FrmR under control of P_{frm}. Representative time response curves for the two configurations after induction with 0-500 μM formaldehyde are shown below their respective regulatory schemes. (c) Response curves fit to the Hill equation for the NEB5α strain with plasmid-expressed FrmR (blue circles), and Δ*frmR* strain with plasmid-expressed FrmR (green triangles). Numbers denote hill coefficients. Error bars represent standard deviation of two replicates tested on different days.

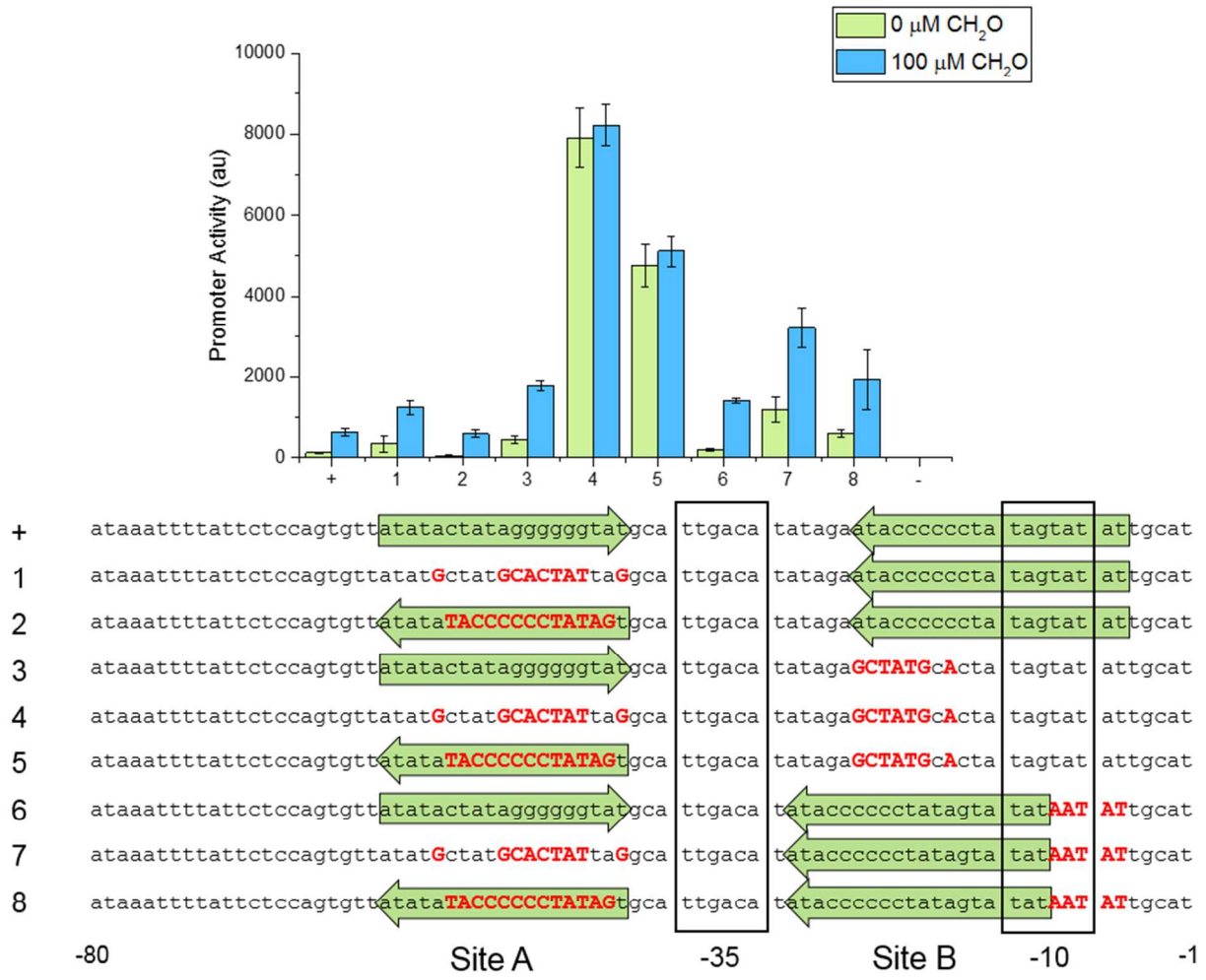


Figure 3. Response of P_{frm} binding site variants 3 hours after dosing with 0 or 100 μM formaldehyde. The 19-bp inverted repeat is shown with two green arrows facing one another to represent their complementary relationship. These two FrmR binding sites were analyzed by removing or reverse complementing the sites in tested constructs. The promoter was deleted to yield the negative control and the positive construct is the native P_{frm} sequence. Error bars represent standard deviation of two replicates tested on different days.

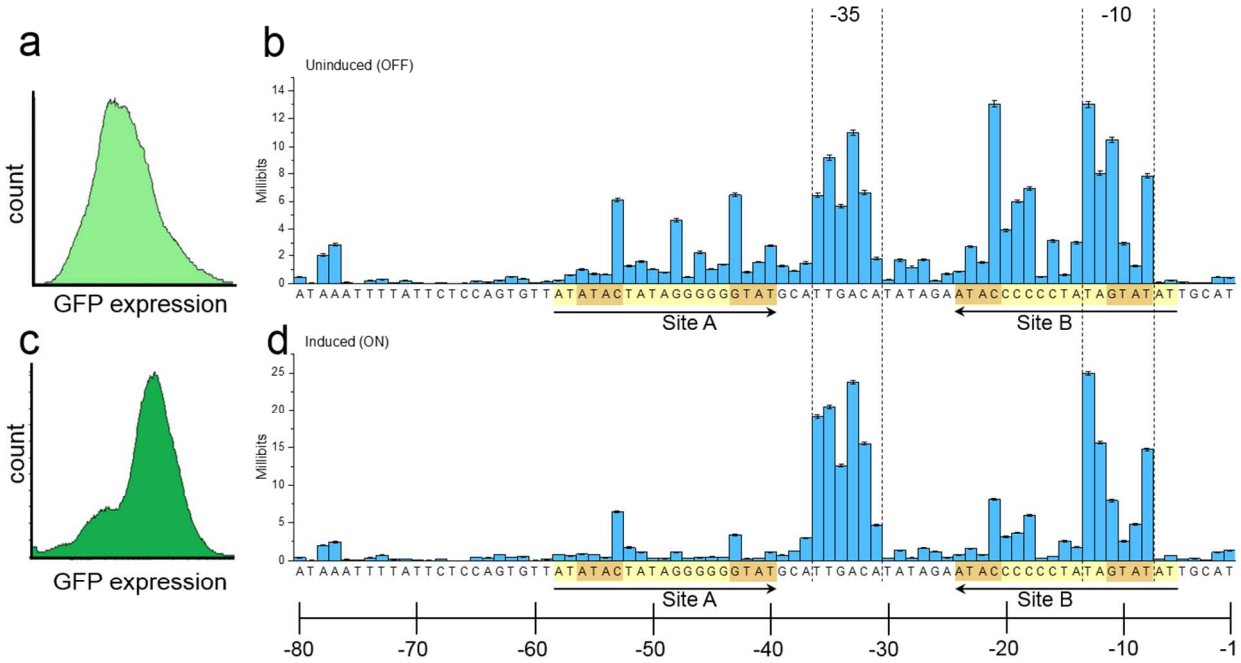


Figure 4. Information footprints of the *E. coli* P_{frm} with different levels of FrmR binding in the NEB5α strain. The distribution of GFP expression from 10,000 cells is shown in the uninduced (a) and induced (c) P_{frm} library. Information footprints for the uninduced (b) and induced with 100 μM formaldehyde (d) experiments illustrate significant nucleotide positions with and without FrmR bound. Yellow highlight shows a large inverted repeat while the two left red sites and two right red sites indicate smaller inverted repeats relevant to FrmR binding. Error bars indicate uncertainties inferred from subsampling.

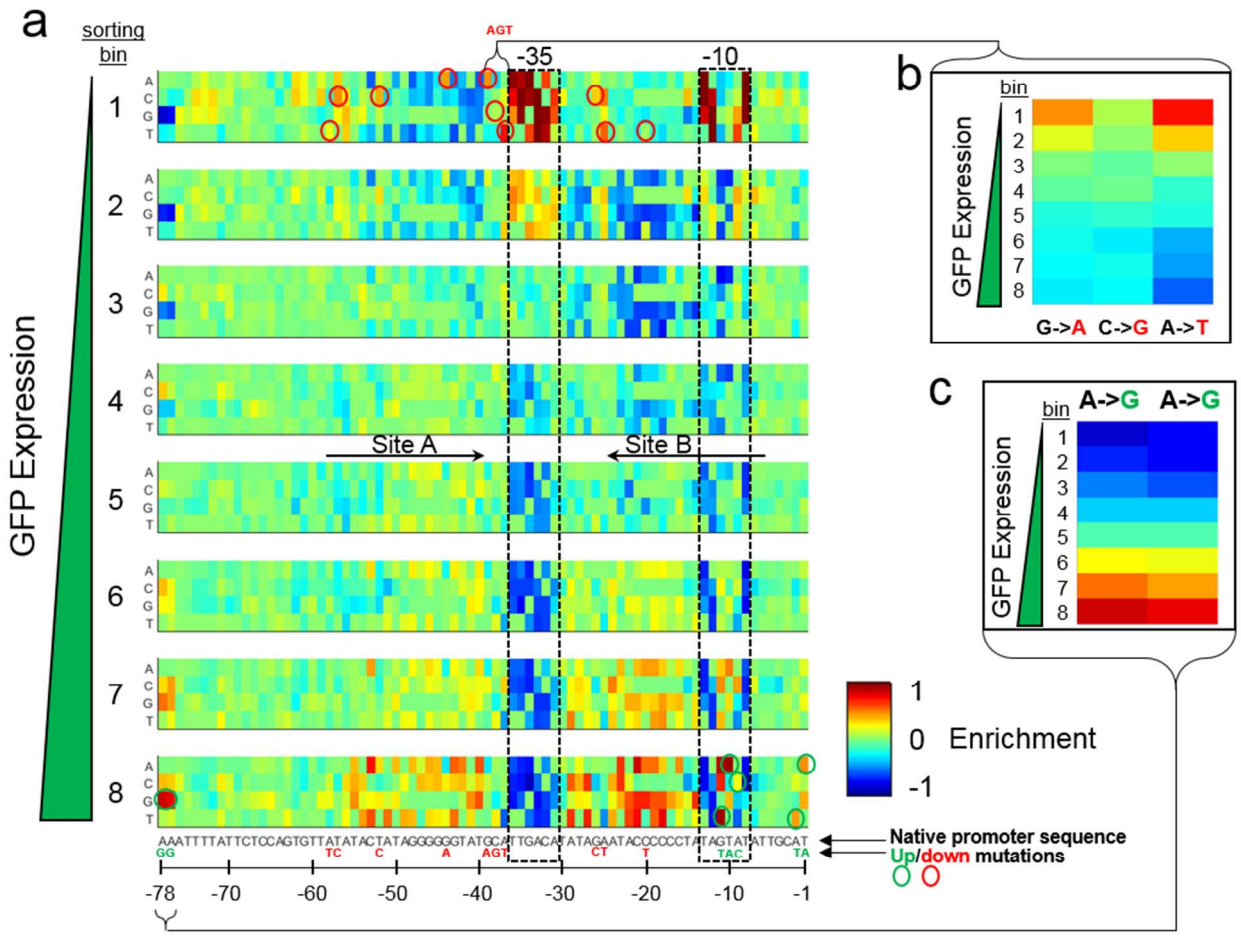


Figure 5. Identification of up and down mutations through sequence analysis. (a) Heat maps of the P_{frm} sequence in each of eight sorting bins in NEB5α strain in uninduced conditions. Mutated promoters isolated from low GFP expression bins are represented in the top heatmaps while those from high GFP bins are represented in the lower heatmaps. Enrichment calculated as log2-fold change of each mutation compared to the unsorted promoter library, where red mutations are highly enriched and blue mutations are rare in each given bin. Native sequence is shown in black below the heat maps and identified up/down mutations are shown at their specified locations in green and red, respectively. (b) Enrichment profiles for three identified down mutations GCA→AGT from directly upstream of the -35 site are shown for each sorting bin. Down mutations are highly enriched (red) in lower expression bins and rare (blue) in higher expression bins. (c) Enrichment profiles for two identified up mutations AA→GG located far upstream. Up mutations are highly enriched (red) in high expression bins and rare (blue) in low expression bins.

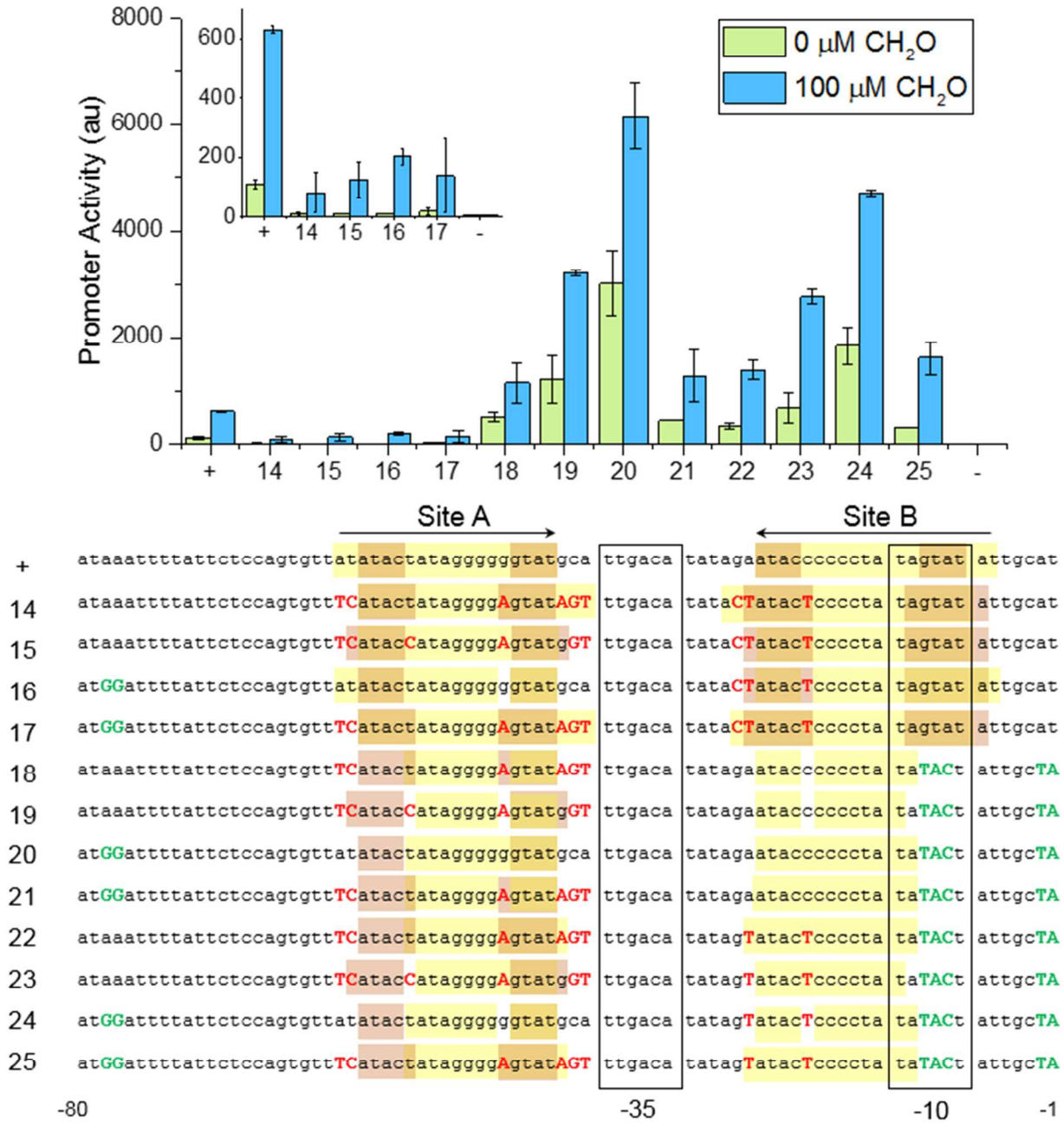


Figure 6. Response of specifically constructed P_{frm} variants 3 hours after dosing with 0 or 100 μM formaldehyde. Variants were constructed with mutations for higher (green) or lower (red) expression levels. The promoter was deleted to yield the negative control and the positive construct is the native P_{frm} sequence. Error bars represent standard deviation of two replicates tested on different days.

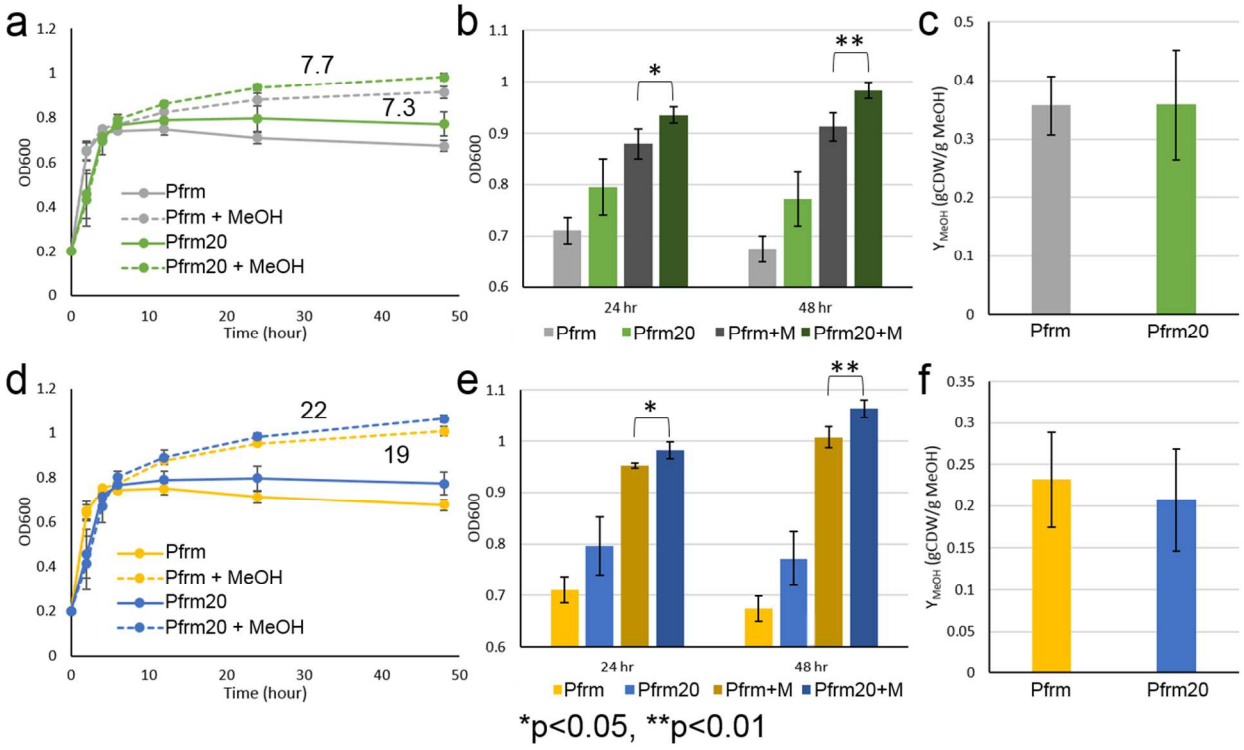


Figure 7. Growth and yield on methanol for P_{frm}-Mdh-Hps-Phi and P_{frm}20-Mdh-Hps-Phi plasmids in the Δ*frmA* Δ*pgi E. coli* strain. Strains were grown with M9 minimal media supplemented with 1 g/L yeast extract and with or without (a,b,c) 60 mM or (d,e,f) 240 mM methanol. (a,d) Growth curves normalized to a starting OD of 0.2. Numbers denote mM methanol consumed for dosed P_{frm} and P_{frm}20 strains. (b,e) OD at 24 and 48 hours for the P_{frm} and P_{frm}20 strains with and without methanol dosing. (c,f) Yield on methanol for the P_{frm} and P_{frm}20 strains in gram cell dry weight (CDW) per gram methanol at 24 hours. Error bars represent standard deviation (n=4).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: XXXXX

Relevant fluorescence distributions; additional response curves, information footprints and heatmaps; clarifying diagrams of Pfrm sequence and architecture; plots and tables characterizing mutation bias and frequency in Pfrm library; tables listing primers, plasmids, and strains from this study; Hill equation fitted parameters (PDF)

AUTHOR INFORMATION

Corresponding Authors

*N.R.S. E-mail: nsandova@tulane.edu

*E.T.P. E-mail: papoutsakis@dbi.udel.edu

Present Address

†Department of Chemical & Biomolecular Engineering and the Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711, United States

‡Department of Chemical & Biomolecular Engineering, Tulane University, New Orleans, Louisiana 70118, United States

ORCID
Julia Rohlhill: 0000-0003-2975-3922
Nicholas R. Sandoval: 0000-0002-1743-3975
Eleftherios T. Papoutsakis: 0000-0002-1077-1277

Author Contributions

JR and NRS designed the experiments. JR and NRS performed flow cytometry analysis and sorting, and next generation sequencing and analysis. JR designed and performed methanol consumption experiments. JR wrote the manuscript with aid from NRS and ETP. ETP conceived and suggested the concept for this study. All authors edited the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Funding was provided by the ARPA-E REMOTE Contract DE-AR0000432. J. R. was partially supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under award number R01GM085232. N.R.S. was supported by the NIH through the National Research Service Award F32GM109617-01A1. We thank R. Kyle Bennett for the construction and provision of the Mdh-Hps-Phi operon and Δ*frmA* Δ*pgi* double knockout *E. coli* strain.

REFERENCES

(1) Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., and Koffas, M. A. G. (2016) Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol. 34*, 652-664.

(2) Glick, B. R. (1995) Metabolic load and heterogenous gene expression. *Biotechnol. Adv. 13*, 247-261.

(3) Rogers, J. K., Taylor, N. D., and Church, G. M. (2016) Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol. 42*, 84-91.

(4) Cress, B. F., Trantas, E. A., Ververidis, F., Linhardt, R. J., and Koffas, M. A. G. (2015) Sensitive cells: Enabling tools for static and dynamic control of microbial metabolic pathways. *Curr. Opin. Biotechnol. 36*, 205-214.

(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. 43*, 7648-7660.

(6) Taylor, N. D., Garruss, A. S., Moretti, R., Chan, S., Arbing, M., Cascio, D., Rogers, J. K., Isaacs, F. J., Kosuri, S., Baker, D., Fields, S., Church, G. M., and Raman, S. (2016) Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods 13*, 177-183.

(7) Mahr, R., and Frunzke, J. (2016) Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl. Microbiol. Biotechnol. 100*, 79-90.

(8) Libis, V., Delépine, B., and Faulon, J.-L. (2016) Sensing new chemicals with bacterial transcription factors. *Curr. Opin. Microbiol. 33*, 105-112.

(9) Peterman, N., and Levine, E. (2016) Sort-seq under the hood : implications of design choices on large-scale characterization of sequence-function relations. *BMC Genomics 17*, 206.

(10) Atwal, G. S., and Kinney, J. B. (2016) Learning Quantitative Sequence–Function Relationships from Massively Parallel Experiments. *J. Stat. Phys. 162*, 1203-1243.
(11) Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C. G., Kinney, J. B., Kellis, M., Lander, E. S., and Mikkelsen, T. S. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol. 30*, 271-277.
(12) Kinney, J. B., Murugan, A., Callan, C. G., and Cox, E. C. (2010) Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc. Natl. Acad. Sci. U.S.A. 107*, 9158-9163.
(13) Treves, A., and Panzeri, S. (1995) The Upward Bias in Measures of Information Derived from Limited Data Samples. *Neural Comput. 7*, 399-407.
(14) Gasperini, M., Starita, L., and Shendure, J. (2016) The power of multiplexed functional analysis of genetic variants. *Nat. Protoc. 11*, 1782-1787.
(15) Chen, N. H., Djoko, K. Y., Veyrier, F. d. r. J., and McEwan, A. G. (2016) Formaldehyde stress responses in bacterial pathogens. *Front. Microbiol. 7*, 257.
(16) Osman, D., Piergentili, C., Chen, J., Sayer, L. N., Usón, I., Huggins, T. G., Robinson, N. J., and Pohl, E. (2016) The Effectors and Sensory Sites of Formaldehyde-Responsive Regulator FrmR and Metal-Sensing Variant. *J. Biol. Chem. 291*, 19502-19516.
(17) Denby, K. J., Iwig, J., Bisson, C., Westwood, J., Rolfe, M. D., Sedelnikova, S. E., Higgins, K., Maroney, M. J., Baker, P. J., Chivers, P. T., and Green, J. (2016) The mechanism of a formaldehyde-sensing transcriptional regulator. *Sci. Rep. 6*, 38879.
(18) Gonzalez, C. F., Proudfoot, M., Brown, G., Korniyenko, Y., Mori, H., Savchenko, A. V., and Yakunin, A. F. (2006) Molecular basis of formaldehyde detoxification: Characterization of two S-formylglutathione hydrolases from Escherichia coli, FrmB and YeiG. *J. Biol. Chem. 281*, 14514-14522.
(19) Thieffry, D., Huerta, a. M., Pérez-Rueda, E., and Collado-Vides, J. (1998) From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. *Bioessays 20*, 433-440.
(20) Whitaker, W. B., Sandoval, N. R., Bennett, R. K., Fast, A. G., and Papoutsakis, E. T. (2015) Synthetic methylotrophy: engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization. *Curr. Opin. Biotechnol. 33*, 165-175.
(21) Olah, G. A. (2013) Towards Oil Independence Through Renewable Methanol Chemistry. *Angew. Chem. Int. Ed. 52*, 104-107.
(22) Müller, J. E. N., Meyer, F., Litsanov, B., Kiefer, P., Potthoff, E., Heux, S., Quax, W. J., Wendisch, V. F., Brautaset, T., Portais, J.-C., and Vorholt, J. A. (2015) Engineering Escherichia coli for methanol conversion. *Metab. Eng. 28*, 190-201.
(23) Sheehan, M. C., Bailey, C. J., Dowds, B. C., and McConnell, D. J. (1988) A new alcohol dehydrogenase, reactive towards methanol, from Bacillus stearothermophilus. *Biochem. J. 252*, 661-666.
(24) Whitaker, W. B., Jones, J. A., Bennett, K., Gonzalez, J., Vernacchio, V. R., Collins, S. M., Palmer, M. A., Schmidt, S., Antoniewicz, M. R., Koffas, M. A., and Papoutsakis, E. T. (2017) Engineering the Biological Conversion of Methanol to Specialty Chemicals in Escherichia coli. *Metab. Eng. 39*, 49-59.

(25) Price, J. V., Chen, L., Whitaker, W. B., Papoutsakis, E., and Chen, W. (2016) Scaffoldless engineered enzyme assembly for enhanced methanol utilization. *Proc. Natl. Acad. Sci. U.S.A. 113*, 12691-12696.
(26) Jakobsen, Ø. M., Benichou, A., Flickinger, M. C., Valla, S., Ellingsen, T. E., and Brautaset, T. (2006) Upregulated transcription of plasmid and chromosomal ribulose monophosphate pathway genes is critical for methanol assimilation rate and methanol tolerance in the methylotrophic bacterium Bacillus methanolicus. *J. Bacteriol. 188*, 3063-3072.
(27) Wang, B., Kitney, R. I., Joly, N., and Buck, M. (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat. Commun. 2*, 508.
(28) Xu, P., Vansiri, A., Bhan, N., and Koffas, M. A. G. (2012) EPathBrick: A synthetic biology platform for engineering metabolic pathways in E. coli. *ACS Synth. Biol. 1*, 256-266.
(29) Osman, D., Piergentili, C., Chen, J., Chakrabarti, B., Foster, A. W., Lurie-Luke, E., Huggins, T. G., and Robinson, N. J. (2015) Generating a Metal-Responsive Transcriptional Regulator to Test What Confers Metal-Sensing in Cells. *J. Biol. Chem. 290*, 19806-19822.
(30) Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K., and Balázsi, G. (2009) Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proc. Natl. Acad. Sci. U.S.A. 106*, 5123-5128.
(31) Madar, D., Dekel, E., Bren, A., and Alon, U. (2011) Negative auto-regulation increases the input dynamic-range of the arabinose system of Escherichia coli. *BMC Syst. Biol. 5*, 111.
(32) Herring, C. D., and Blattner, F. R. (2004) Global transcriptional effects of a suppressor tRNA and the inactivation of the regulator frmR. *J. Bacteriol. 186*, 6714-6720.
(33) Stanton, B. C., Nielsen, A. a. K., Tamsir, A., Clancy, K., Peterson, T., and Voigt, C. a. (2014) Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol. 10*, 99-105.
(34) Huber, P. J. (1981) *Robust Statistics*, John Wiley & Sons.
(35) Drummond, D. A., Iverson, B. L., Georgiou, G., and Arnold, F. H. (2005) Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. *J. Mol. Biol. 350*, 806-816.
(36) Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol. 2*, 2006.0008.
(37) Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A. 97*, 6640-6645.
(38) Brosius, J., Erfle, M., and Storella, J. (1985) Spacing of the -10 and -35 regions in the tac promoter. *J. Biol. Chem. 260*, 3539-3541.
(39) Espah Borujeni, A., Channarasappa, A. S., and Salis, H. M. (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res. 42*, 2646-2659.

(40) 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.
(41) Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974) Culture medium for enterobacteria. *J. Bacteriol. 119*, 736-747.
(42) Sambrook, J. F., E. F., and Maniatis, T. (1989) *Molecular cloning: A laboratory manual*. 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.