Synthetic Biology **Biotechnology and Bioengineering**
DOI 10.1002/bit.26455

Development of a Formaldehyde Biosensor with Application to Synthetic Methylotrophy†

Benjamin M. Woolston*
Department of Chemical Engineering, Massachusetts Institute of Technology

Timothy Roth*
Department of Chemistry and Chemical Biology, Harvard University
The Broad Institute of MIT and Harvard
Howard Hughes Medical Institute

Ishwar Kohale
Department of Biological Engineering, Massachusetts Institute of Technology

David Liu
Department of Chemistry and Chemical Biology, Harvard University
The Broad Institute of MIT and Harvard
Howard Hughes Medical Institute

Gregory Stephanopoulos#
Department of Chemical Engineering, Massachusetts Institute of Technology
25 Ames Street, MIT 56-469C
Cambridge, Massachusetts 02139
gregstep@mit.edu
(617) 253-4583

*These authors contributed equally to the work
#Corresponding author

†This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/bit.26455]

Additional Supporting Information may be found in the online version of this article.

Received May 26, 2017; Revision Received September 1, 2017; Accepted September 14, 2017

Abstract

Formaldehyde is a prevalent environmental toxin and a key intermediate in single carbon metabolism. The ability to monitor formaldehyde concentration is therefore of interest for both environmental monitoring and for metabolic engineering of native and synthetic methylotrophs, but current methods suffer from low sensitivity, complex workflows, or require expensive analytical equipment. Here we develop a formaldehyde biosensor based on the FrmR repressor protein and cognate promoter of *E. coli*. Optimization of the native repressor binding site and regulatory architecture enabled detection at levels as low as 1 μM. We then used the sensor to benchmark the *in vivo* activity of several NAD-dependent methanol dehydrogenase (Mdh) variants, the rate-limiting enzyme that catalyzes the first step of methanol assimilation. In order to use this biosensor to distinguish individuals in a mixed population of Mdh variants, we developed a strategy to prevent cross-talk by using glutathione as a formaldehyde sink to minimize intercellular formaldehyde diffusion. Finally, we apply this biosensor to balance expression of *mdh* and the formaldehyde assimilation enzymes *hps* and *phi* in an engineered *E. coli* strain to minimize formaldehyde build-up while also reducing the burden of heterologous expression. This biosensor offers a quick and simple method for sensitively detecting formaldehyde, and has the potential to be used as the basis for directed evolution of Mdh and dynamic formaldehyde control strategies for establishing synthetic methylotrophy.

Keywords: Biosensor, Synthetic Methylotrophy, Directed Evolution, Synthetic Biology, Metabolic Engineering

Introduction

Formaldehyde is a highly toxic chemical, and is classified as a human carcinogen by the International Agency for Research of Cancer (IARC, 2006). As a potent electrophile, its toxicity stems from its ability to react rapidly with nucleophilic components of DNA, RNA and proteins, leading to protein and DNA damage in the form of crosslinking (Salthammer, 2013). Somewhat paradoxically, formaldehyde is also a ubiquitous intermediate in one-carbon metabolism across the tree of life. In methanotrophs, methane is oxidized to methanol by methane monooxygenase (Mmo), and then to formaldehyde by a PQQ-dependent methanol dehydrogenase (Mdh) (Chistoserdova et al., 2009). Methylotrophic yeasts such as *Pichia pastoris* convert methanol to formaldehyde using an FAD-linked alcohol oxidase (AOX) (Cregg et al., 1989), and Gram-positive methylotrophs typified by *Bacillus methanolicus* perform the same conversion using an NAD-linked Mdh (Müller et al., 2015a).

In all these organisms, formaldehyde acts as a branch point between further oxidization to CO_2 for energy conservation, and incorporation into biomass via the serine cycle, ribulose monophosphate (RuMP) pathway, or the xylulose-5-phosphate (Xu5P) pathway (Chistoserdova et al., 2009; Müller et al., 2015a; Yurimoto et al., 2011). These pathways are of growing interest in the field of metabolic engineering, where researchers seek to convert relatively cheap methanol feedstocks into higher value commodity chemicals with either native or ‘synthetic’ methylotrophs (Kalyuzhnaya et al., 2015; Whitaker et al., 2015). Besides methylotrophs, formaldehyde is present at low levels in all organisms as a result of demethylation reactions (Kalasz, 2003). Because of

its cytotoxicity, the intracellular formaldehyde concentration must be tightly controlled, which has led to the evolution of a variety of highly coordinated metabolic strategies for detoxifying formaldehyde (Yurimoto et al., 2005). The need to keep the concentration of formaldehyde low while supporting high flux places an even more stringent burden on methylotrophs that rely on formaldehyde metabolism for growth.

The ability to easily measure intracellular formaldehyde could therefore provide basic insights into the regulation of formaldehyde metabolism in native methylotrophs, as well as aid in the development of synthetic methylotrophy. However, typical methods are limited by low sensitivity, cumbersome workflows, or the requirement for expensive HPLC instrumentation. In the work conducted thus far in *E. coli* on synthetic methylotrophy, formaldehyde has been measured in culture supernatants using the Nash assay (Nash, 1953), taking advantage of the fact that formaldehyde can diffuse rapidly across the cell membrane (Müller et al., 2015b; Whitaker et al., 2016; Wu et al., 2016). Due to the low assay sensitivity (Limit of detection, LOD 1 μM), evaluation of the synthetic methanol assimilation pathway activity required elimination of the native detoxification pathway. The assay is also limited to small numbers of samples in a kinetic experiment due to the need to separate cells from supernatant before analysis. The gold standard for environmental formaldehyde quantification involves derivatization with 2,4-dinitrophenyl hydrazine (2,4-DNPH), followed by HPLC to separate the various carbonyl derivatives before quantification via UV (Salthammer, 2013). While highly sensitive (LOD 0.2 μM) (EPA, 1996), this technique suffers from the same bottleneck of requiring cell separation before analysis, and the additional challenges of low throughput and high cost due to the requirement for HPLC separation.

Recently, there has been tremendous interest in developing genetically encoded biosensors for monitoring the concentration of a multitude of different compounds (Zhang et al., 2015; Zhang et al., 2016a; Zhang et al., 2016b). These sensors offer several advantages compared to traditional methodologies: Since the signal is often a fluorescent protein such as GFP, sensor read-out can be determined easily using widely available instrumentation without the need for separating cells from their media, and is therefore adaptable for high-throughput sampling. In addition, because GFP is stable, the fluorescence signal represents an integral of the substrate concentration over time, allowing for significantly increased sensitivity compared to single time-point measurements. Biosensors are also powerful tools for directed evolution (Packer and Liu, 2015; Raman et al., 2014), thus the development of a formaldehyde sensor would aid ongoing efforts to evolve more active variants of Mdh (Wu et al., 2016). Finally, biosensors are gaining interest in metabolic engineering for their ability to actuate a dynamic metabolic response to the presence of the target analyte (Brockman and Prather, 2015; Woolston et al., 2013). Such a regulatory strategy could mitigate the toxicity of formaldehyde in an engineered methylotrophy.

The native glutathione-dependent formaldehyde detoxification pathway in *E. coli* provides a convenient architecture for a formaldehyde biosensor. In this pathway, formaldehyde reacts spontaneously with the nucleophilic cysteine residue of glutathione to form the hemiacetal *S*-(hydroxymethyl)-glutathione. This adduct is enzymatically oxidized to *S*-formylglutathione by FrmA. Hydrolysis of this species by FrmB liberates glutathione and produces formate, which is much less toxic than formaldehyde (Gonzalez et al., 2006). Expression of *frmA* and *frmB* is controlled by a repressor

protein FrmR, expressed in the same operon. In the absence of formaldehdye, FrmR binds to the promoter region, preventing transcription. In the presence of formaldehyde, the nucleophilic Cys36 of the FrmR reacts with formaldehyde and causes a conformational change that results in dissociation from the promoter (Denby et al., 2016; Law, 2012; Osman et al., 2016). Addition of 0.25 mM formaldehyde is sufficient to induce an approximately 100-fold increase in *frmA* transcript after 30 minutes (Herring and Blattner, 2004). Previously, Tralau and coworkers employed a GFP-linked biosensor based on this regulatory system to detect formaldehyde produced during the oxidation of dimethylglycine in *E. coli* (Tralau et al., 2009). In a more recent publication, this biosensor was improved through a sort-seq approach to screen multiple mutations of the *frm* promoter region to improve dynamic range and elucidate the repressor bindings sites (Rohlhill et al., 2017). The authors further showed that the evolved stronger promoter could enhance the benefit of methanol on biomass formation during growth on yeast extract.

Here we report further characterization of the biosensor, rational mutation to the promoter to increase sensitivity at low formaldehyde concentrations, and significant advances in the application of the sensor to the ongoing efforts to engineer synthetic methylotrophy. First, we compared the response of an autoregulated version of the sensor, where *frmR* was expressed from its own promoter in an operon with the signal, to a variant where *frmR* was expressed under the orthogonal p_{tet} promoter to remove the negative feedback of the original oncstruct. Second, we tuned to physiological relevant formaldehyde concentrations (1-40 µM in wild-type cells) (Müller et al., 2015b) by systematic mutation of the *frm* promoter region. To demonstrate the utility of the

refined biosensor, we then showed that the reporter could detect formaldehyde production in engineered *E. coli* strains without deletion of *frmA*. To examine the potential of the assay for Mdh evolution, we verified that our reporter could distinguish between the varying activities of various Mdh genes, and then developed methods to distinguish between individuals within a mixed population by the addition of glutathione to minimize interceullar formaldehyde diffusion. Finally, as a proof-of-concept and first step toward dynamic regulation we examined the ability of our biosensor to report formaldehyde levels in strains engineered for methanol assimilation, facilitating high-throughput optimization of induction conditions for Mdh and the formaldehyde assimilation enzymes Hexulose phosphate synthase (Hps) and Phosphohexulose isomerase (Phi*)*.

Materials and Methods

Reagents

Unless specified, all chemical reagents were purchased in the highest grade available from Sigma. For luciferase reporter assays, M9 salts and LB, Agar, and casamino acids were purchased from US Biological Life Sciences. For all other assays, DIFCO M9 salts and LB, BACTO Agar, and casamino acids were purchased from BD. Methanol-free formaldehyde (16% v/v) was purchased as 1 mL ampules from ThermoFisher (Catalog # 28906), and dilutions for cellular assays were prepared fresh daily. Trace Elements (MD-TMS) and Vitamin Solution (MD-VS) were purchased from ATCC.

Strains and Plasmids

Bacterial strains and plasmids used in this study are listed in Supplemental Table 1. *E. coli* DH5α was used as a cloning host. Luciferase-based assays and MDH comparisons

were carried out using *E. coli* S1030 (Carlson et al., 2014) with or without *frmA* knocked out, as described in the text. Pathway optimization assays were conducted in *E. coli* MG1655(DE3) (Tseng et al., 2009). pBbS2k-RFP was a gift from Jay Keasling (Addgene plasmid #35330). pETM6-mCherry was a gift from Mattheos Koffas (Addgene plasmid #66534).

Cloning

Plasmids were constructed using USER cloning, Gibson Assembly Master Mix, or KLD Enzyme Mix (NEB). DNA fragments were generated by PCR with the primers listed in Supplemental Table 2. Phusion U Hot Start DNA Polymerase (Life Technologies) was used with all USER cloning primers, whereas Q5 Polymerase (NEB) was used for all other primers. All amplicons were digested with DpnI before PCR purification for Gibson Assemblies and KLD ligations, or during PCR fragment assembly for USER cloning. All restriction endonucleases were purchased from NEB. Assembled vectors were transformed into either DH5α chemically competent cells (NEB) or Mach 1 T1R chemically competent cells (Invitrogen), and verified by Sanger sequencing. The *frmA* deletion in S1030 was made using the protocol described by Datsenko and Wanner (Datsenko and Wanner, 2000). Cell growth for cloning purposes was carried out using DIFCO LB media supplemented with appropriate antibiotics (Kanamycin, 50 μg mL^{-1}; Carbenicillin, 50 μg mL^{-1}; Spectinomycin, 50 μg mL^{-1}).

Luciferase-Based Reporter Assays

For all reporter assays, a minimum of three individual colonies from transformation plates were grown overnight in $M9^{+}$ medium (M9 supplemented with 0.2% glucose, 0.1% casamino acids, and 1% (v/v) each of trace mineral and vitamin solution). Cultures

were diluted 100-fold into fresh medium and grown until early exponential phase (OD approximately 0.4) supplemented with anhydrotetracycline (ATc) to induce *frmR* expression. For direct testing of induction, formaldehyde was added to 200 µL of sample in a black, clear-bottom 96-well plate (Corning) and analyzed for optical density at 600nm and luciferase activity at 37°C on an Infinite Pro M1000 plate reader (Tecan).

GFP-Based Formaldehyde Reporter and MDH Assays

GFP reporter assays with formaldehyde were conducted almost identically to the luciferase-based ones, but cell growth and fluorescence were measured on a SpectraMax M2e (Molecular Devices) spectrophotometer at 37°C with continuous shaking. For Mdh-linked assays, cells were induced with 10 mM arabinose and immediately transferred to 96-well plates (200 µL per well), where 50 µL methanol was added to a final concentration of 0-500 mM. Plates were covered with Breathe-Easy sealing membranes (Sigma), and growth continued at 37°C with maximum agitation on a Jitterbug 2.0 (Boekel). Pathway optimization experiments were conducted analogously, except that induction was mediated by IPTG and ATc. GFP fluorescence was assessed hourly to minimize temperature fluctuations that would affect growth and Mdh activity. Excitation and emission wavelengths were 488 nm and 525 nm respectively.

Formaldehyde Measurement by Nash Assay

Formaldehyde concentration in culture supernatants was assayed by a modification of the Nash reaction (Nash, 1953) for 96-well plate format. 200 µL of cells were centrifuged for 1 minute at 13,000 RPM, and 125 µL supernatant transferred to the plate. 125 µL Nash reagent (5 M ammonium acetate, 50 mM acetylacetone) was added

to each well, the plate was incubated at 37°C for one hour, and absorbance was read at 412 nm. A standard curve was prepared in culture medium in the range from 100 µM to 0 µM. The plate was kept on ice until all samples had been collected.

Flow Cytometry

Cells from GFP assays were diluted 500-fold into PBS buffer for FC analysis, which was performed using a BD FACS LSRII HTS-2. GFP fluorescence was measured using the 488nm laser and 530/30 filter, and mCherry fluoresecence was measured using the 561nm laser and 575/26 bandpass filter. Cells were gated based on FSC-H and SSC-H, and 10,000 events falling into this window were recorded. GFP vs mCherry plots were converted into tsv files using the freeware software Cyflogic, and analyzed and plotted using in-house Python scripts.

Model Regression

Response curves from both luciferase and GFP assays were fitted to the Hill equation, and confidence intervals around the model parameters (S_{max}, n, and K_a) were calculated using the MATLAB functions nlinfit() and nlparci(), respectively. The form of the Hill equation used was:

$$S = S_{min} + (S_{max} - S_{min}) \left(\frac{[F]^n}{K_a^n + [F]^n} \right)$$

where S is the measured signal, [F] is the concentration of formaldehyde, S_{max} the signal at saturation S_{min} the signal at 0 formaldehyde, n the Hill coefficient, and K_a the dissociation constant.

Results and Discussion

Auto-regulated negative-feedback promoters are a means of maintaining steady-state levels of gene-expression across varying conditions, providing greater stability to genetic networks (Becskei and Serrano, 2000). However, this feedback regulation also attenuates the dynamic range of the reporter system, obscuring differences in signal across broad input ranges. Previous work with the *frm* promoter showed relatively limited signal gain in response to toxic concentrations of formaldehyde (Tralau et al., 2009). We hypothesized that we could achieve a stronger, dose-dependent response to formaldehyde by taking the promoter out of its auto-regulatory context. To test this, we placed *frmR* expression under inducible control of the p_{tet} promoter (construct pTR47). We then compared this construct to an analogous, auto-regulated synthetic plasmid (construct pTR47auto) (Figure 1). Both constructs express the luciferase reporter *luxAB* from the frm promoter, the latter of which was subcloned out of the *E. coli* genome. Interestingly, there was a SNP in this region in our strain (see Supplementary Information), but it did not occur in a sequence predicted to be important in controlling DNA geometry or facilitating protein-DNA interactions (Law, 2012). With construct pTR47, in the absence of ATc induction, significant background signal was observed, which rose further upon addition of formaldehyde. The most likely explanation for the high background signal is that the amount of FrmR that arises from chromosomal expression is insufficient to fully repress the multiple-copy plasmid-borne *frm* promoter. That the signal increased further upon formaldehyde addition suggests that this low level of *frmR* expression is still sufficient to partially repress the promoter. Addition of 40ng mL^{-1} ATc to induce *frmR* expression lowered background signal by 100-fold without significantly lowering observed reporter values for the ‘on' state, enabling

detection of formaldehyde between 1-100 μM with three orders of magnitude difference in signal, and a dissociation constant (K_a) of 37±1.5 μM. Comparing this de-coupled repressor to an analogous auto-regulated system (pTR47auto), we saw an order of magnitude improvement in signal-to-background gene expression (Figure 1). In addition, the response in the decoupled system showed significantly higher cooperativity, with Hill coefficient n = 2.7±0.2 compared to n=1.2±0.1.

Despite the improvement derived from removing auto-regulation, we sought better separation of signal response at lower formaldehyde concentrations more relevant in cells engineered for methanol assimilation (1-40 μM in wild-type cells) (Müller et al., 2015c). Literature precedent in the RncR repressor protein of *E. coli* suggested that repeated cytosine and guanine tracts can induce A-form DNA geometry, likely playing an important role in DNA-protein interactions at binding sites at or around these tracts (Iwig and Chivers, 2009). In the native p_{frm} sequence, two repeated stretches of guanine or cytosine nucleotides have been implicated as a contributing factor to the DNA geometry around the position that would typically bind to recruit the σ70 portion of the *E. coli* RNA polymerase to initiate transcription of downstream genes (Law, 2012). Changing two cytosine nucleotides in one tract (construct pTR47m4) allowed us to observe higher signal at lower formaldehyde doses compared to the wild-type binding sequence (construct pTR47), reducing the K_a from 37±1.5 to 11±0.9 μM, and increasing the Hill coefficient from 2.7±0.2 to 5.0±2.5 (Figure 2). To ensure formaldehyde specificity, the reporter was tested with acetaldehyde, propionaldehyde, and methanol. None of these substrates elicited luciferase expression below mM concentration, and are therefore not significant activators of FrmR at physiological

concentrations (Supplemental Figure 1). The mutations tested in our m4 variant were examined individually by Rohlhill et al.; however none of the variants examined in that work had mutations at both positions within the C- tract without otherwise completely removing the tract. Our results suggest that this repeated C-tract may represent a sequence feature in which the type and combination of mutated bases can synergistically affect the responsiveness of the promoter/repressor system.

Because quantification based on fluorescent proteins is more widely used in biosensor applications, we additionally constructed a GFP version of the biosensor encoding the superfolder GFP (Pédelacq et al., 2006) in place of luciferase in pTR47m4 (construct pTR47m4-gfp). This construct showed a similar response to formaldehyde, but with a lower signal-to-noise ratio and higher limit of detection (LOD) of 10 µM (Figure 3). The Hill coefficient was reduced to 1.8±1.5, and the K_a increased to 21±11 µM. We further assessed this construct in the context of a *ΔfrmA* strain to see if removal of the detoxification system would improve sensitivity. In this strain, the reporter was much more sensitive at low formaldehyde concentration, with a K_a of 6.3±2.8 µM, resulting in a reduced LOD of 5 µM. Due to the higher baseline promoter activity in this strain, the overall dynamic range did not increase substantially. Interestingly, there was also considerably higher variability in the signal in the *ΔfrmA* background, possibly because the absence of a detoxification system amplifies any initial noise in the formaldehyde spike in that strain.

Comparison of Mdh Variants

Prior work in synthetic methylotrophy has shown the kinetic properties of Mdh to be much less favorable than those of the formaldehyde assimilation enzymes Hps and Phi,

which has led to a search for catalytically more active homologs (Whitaker et al., 2016; Wu et al., 2016). The in *vivo* comparison of different Mdh candidates has typically involved a labor-intensive Nash assay, which requires pelleting and washing cultures in order to chemically measure formaldehyde concentrations (Müller et al., 2015b; Whitaker et al., 2016). The experiment is conducted in a *ΔfrmA* strain otherwise detoxification lowers the formaldehyde concentration to virtually undetectable levels. This process is low-throughput and does not lend itself to the simultaneous comparison of many variants. In their efforts to evolve Mdh, Wu and coworkers developed a 96-well assay where the Nash reagent is added directly to the culture upon methanol addition. However, this procedure still required separation of the supernatant (Wu et al., 2016). Rapid alternatives to assay *in vivo* Mdh activity with minimal user intervention would therefore be valuable.

To test whether our formaldehyde reporter could meet this need, we cloned the evolved variant of *mdh2* from *Cupriavidus necator* (Wu et al., 2016) into pTR48, and transformed this plasmid into S1030 and S1030 *ΔfrmA* cells containing pTR47m4-gfp. Figure 4 shows the time course of GFP expression after simultaneous addition of methanol and *L*-arabinose (to induce *mdh* expression). A clear methanol-dependent signal was observed in both strain backgrounds as early as 30 minutes into the assay, and the response saturated after two hours. These results clearly demonstrate the ability of the sensor to rapidly detect Mdh activity, even without deletion of endogenous formaldehyde detoxification pathways. The dynamic range, defined as the ratio of final GFP signal to initial signal, was higher in the WT cells (10.4±1.4 vs. 4.0±0.7), primarily due to the high background in the *ΔfrmA* strain, therefore we chose the WT background

for further studies. In a parallel experiment with the same cells, we also measured formaldehyde concentration in the media using the Nash assay, to compare the sensitivity of the two methods (Figure 4). While methanol-dependent formaldehyde production was clearly measurable in both strains, there was only a 1.6-fold increase in concentration over the same time course, highlighting the superiority of the biosensor for *in vivo* measurement of Mdh activity. Given the similarity in LOD for both the Nash assay and our reporter with spiked formaldehyde, the increased dynamic range is most likely due to the amplification of the signal afforded by continual GFP expression in the presence of formaldehyde.

With a functional assay in hand, we set about comparing several candidate Mdhs. *mdh1* and *mdh2* from *B. methanolicus* were chosen as the major isoenzyme expressed during methylotrophic growth (Irla et al., 2015; Müller et al., 2014) and the most active candidate in *E. coli*, respectively (Müller et al., 2015b). To assess the importance of the activator protein (Act), we generated a construct that co-expressed *mdh2* and *act*. This protein is a Nudix hydrolase which, at least *in vitro*, hydrolyzes the nicotinamide mononucleotide moiety of the NADH cofactor of MDH, leading to a drastic reduction in the K_m for methanol (Arfman et al., 1991; Kloosterman et al., 2002; Krog et al., 2013). We also included the alcohol dehydrogenase *adhA* from *Corynebacterium glutamicum*, which has a very low reported K_M (Kotrbova-Kozak et al., 2007), as well as the *mdh* from *Geobacillus stearothermophilus* (Whitaker et al., 2016), and both the WT and evolved variants of *mdh2* from *Cupriavidus necator* (Wu et al., 2016).

After two hours, signal was detected for all variants except Mdh1 (Figure 5). The strongest signal, from the evolved variant of *C. necator* Mdh2, showed over a 20-fold

increase in OD-normalized GFP signal compared to an mCherry control. In general, the differences in fluorescence intensity between variants matched trends in previous reports: Mdh2 from *B. methanolicus* outperforms Mdh1 (30 mU mg^{-1} vs. 1.7 mU mg^{-1}), and co-expression of Act in *E. coli* does not change the *in vivo* activity (Müller et al., 2015b). This is reflected in our data by the 9-fold difference in signal between Mdh2 and Mdh1 at the highest methanol concentration, and the absence of any difference between Mdh2 and Mdh2+Act. It should be noted that although the Mdh1 strain showed no GFP signal after 2 hours, fluorescent signal above the control could be seen upon further incubation for a total of 8 hours (Supplementary Figure 2), long after signals from the other variants had reached saturation, reflecting the extremely slow but still detectable rate of methanol oxidation. Whitaker et al showed that the Mdh from *G. stearothermophilus* outperforms the Mdh2 from *B. methanolicus* at low concentrations (60 mM) of methanol (Whitaker et al., 2016). This is also reflected in our data, where at 50 mM methanol the former shows a fluorescent intensity 2.8-fold higher than the latter. As expected, the evolved variant of the *C. necator* enzyme outperformed the WT version. Given the reportedly low K_M of AdhA from *C. glutamicum*, we were surprised to see no improvement over Mdh2 from *B. methanolicus*, which supports more recently reported K_M data for this enzyme (Wu et al., 2016). Our assay identified the evolved variant from *C. necator* and the Mdh from *G. stearothermophilus* as the most promising candidates for engineering synthetic methylotrophy, but could not distinguish between the two. Taken together, the data presented here show that our assay can be used to detect *in vivo* formaldehyde production by Mdh in a WT background and faithfully reconfirms trends reported in previous literature. From inoculation to assay completion,

the procedure requires a total of 5 hours, and only one intervention to add the inducer and methanol. This a significant simplification over the current state of the art, and should enable comparison of different Mdh candidates in a high-throughput manner. It should be pointed out that, since Mdh expression was induced at the same time methanol was added, the comparisons presented here between the different Mdh variants also include variations in their expression and folding kinetics.

Formaldehyde Biosensor as the Basis for Directed Evolution of MDH Variants

Having demonstrated that the formaldehyde biosensor could be used to discriminate between Mdh candidates, we were interested in evaluating the potential of the sensor for the directed evolution of novel variants using high-throughput technologies such as FACS or PACE (Esvelt et al., 2011). In these approaches, the assay must be able to discriminate between candidates within a mixed population. Since formaldehyde diffuses rapidly across the cell membrane, we were concerned that the lack of spatial segregation of high- and low- activity mutants in a library could lead to the enrichment of cheaters, where the formaldehyde produced by a high-activity variant could diffuse into a cell with a low-activity variant and activate the reporter. To assess this possibility, we co-inoculated a culture with two strains: one carrying pTR48mdh2, and one carrying pTR48mCherry, which expresses *mCherry* instead of *mdh2*. Both strains contained the reporter plasmid pTR47m4GFP, but only the one with *mdh2* should be able to produce formaldehyde from methanol and activate the reporter. If formaldehyde diffusion can activate the reporter in cells not expressing *mdh*, we would expect to see GFP signal in cells expressing *mCherry* after treatment with methanol. If not, the mCherry cells should show no GFP fluorescence. Cells were grown under the same conditions as before, and

the population was analyzed by flow cytometry two hours after induction and methanol addition. As shown in Figure 5, the mCherry+ cells showed mean GFP fluorescence similar to the mCherry- cells, indicating that formaldehyde produced in one cell was able to activate the reporter in another.

In order to prevent this cross-talk, we devised two strategies to reduce the extracellular formaldehyde concentration: 1) addition of NAD-dependent formaldehyde dehydrogenase (FaDH) from *Pseudomonas sp.* and NAD^+, to enzymatically oxidize the formaldehyde to formate; 2) addition of glutathione as a formaldehyde scavenger. Addition of the FaDH and NAD^+ was unsuccessful, possibly due to the breakdown of the enzyme in the supernatant or unfavorable reaction conditions. In contrast, addition of increasing concentrations of glutathione, from 2 to 10 mM, increased the difference in GFP fluorescence between the $mCherry^+$ and $mCherry^-$ cells (Figure 6). At 10 mM, the $mCherry^+$ cells showed only background levels of GFP, indicating complete sequestration of formaldehyde in the supernatant as the glutathione adduct. The addition of glutathione negatively impacted cell growth, with the OD at analysis of the 10 mM culture roughly half that of the 0 mM control. Despite the growth defect, the GFP fluorescence was roughly the same, indicating that the glutathione had no impact on the activation of the reporter or the activity of Mdh. These results clearly show the potential for cheaters in this assay, and establish that glutathione should be added as a formaldehyde sink if this assay is employed for directed evolution of Mdh activity.

Biosensor-Based Pathway Tuning

Since formaldehyde is highly toxic, strains engineered to metabolize methanol must carefully balance the production of formaldehyde by Mdh with its consumption by Hps

and Phi. This balance can be engineered at the transcriptional level by varying the induction of the upstream and downstream pathways, and assessing the formaldehyde concentration. Here we used the formaldehyde reporter to facilitate high-throughput testing of different induction concentrations. The evolved *mdh2* from *C. necator* was placed under the control of the T7 promoter, and *hps* and *phi* as an operon under control of the p_{tet} promoter (construct pETMEOH560). Formaldehyde was measured using the biosensor plasmid pTR59gfp, which is derived from pTR47m4 but with constitutive expression of *frmR* to avoid cross-talk with the tet promoter in the methanol plasmid. Cells were grown on xylose to provide non-limiting levels of Ru5P. Without induction of either *mdh* or *hps/phi*, only background GFP signal was detected (Figure 7). Upon induction with IPTG, the GFP signal rose to a maximum 14-fold increase at 50 μM, before falling at higher induction levels. This fall in signal is likely attributable to the low solubility of Mdh at high induction levels, which leads to slower growth and reduced GFP synthesis (Supplemental Figure 3). At 25 μM IPTG, low-level induction of *hps/phi* with 10 ng mL^{-1} is sufficient to reduce the formaldehyde concentration to background levels. At 50 μM IPTG, the higher Mdh activity requires an increase in Hps/Phi induction to compensate. At 100 μM IPTG and above, even full induction of the downstream pathway with 200 ng mL^{-1} ATc was insufficient to fully consume formaldehyde. This insufficiency may arise because high level IPTG induction reduces the expression level from the p_{tet} promoter, as shown in experiments where IPTG and ATc controlled GFP and RFP respectively (Supplemental Figure 4), such that the Hps/Phi activity is too low to balance formaldehyde generation.

Taken together, the results here demonstrate the high sensitivity of the reporter through its ability to detect formaldehyde even in strains engineered for its consumption, and without deletion of the endogenous detoxification system. They also demonstrate how the reporter can be used to optimize the relative expression level of *mdh* and *hps/phi* to maximize flux while minimizing expression burden.

Conclusions

In this work, we made significant improvements to a formaldehyde biosensor, and demonstrated its utility in ongoing efforts to establish synthetic methylotrophy. Engineering of the promoter binding site and regulatory architecture led to dramatically improved sensitivity over the previous version. Our luciferase-based system could detect the addition of 1 μM exogenous formaldehyde with a 10-fold increase in signal. That we were able to achieve this substantial improvement to sensitivity with a rational approach reinforces the continued utility of rational design in promoter engineering.

We further showed our reporter could discriminate between Mdh candidates, and with the addition of glutathione as a formaldehyde sink could detect differences within a mixed population. In addition, we showed that the dynamic range of the sensor matches well with physiological formaldehyde concentrations in strains engineered for methanol assimilation, allowing high-throughput optimization of induction conditions for the upstream and downstream pathway. These proof-of-concept experiments demonstrate the utility of the biosensor, and pave the way for directed evolution of Mdh, as well as the implementation of dynamic control strategies in synthetic methylotrophy. Finally, levels of formaldehyde in drinking water are routinely 50 μg L^{-1} (1.7 μM) (WHO, 2005),

therefore this assay has the additional potential beyond the synthetic biology applications to be used in environmental sensing and waste management

Acknowledgements

This work was supported by funding from ARPA-E under the award DE-AR0000433, and partially by Cancer Center Support Grant P30-CA14051 from the NCI. This work was also supported by funding from the U.S. National Institutes of Health (NIH) under grants R01 EB022376 and NIH R35 GM118062, as well as the Howard Hughes Medical Institute. BMW and TBR are supported by the National Science Foundation Graduate Research Fellowship under Grant Nos. 1122374 and DGE-1144152, respectively.

References

Arfman N, Van Beeumen J, De Vries GE, Harder W, Dijkhuizen L. 1991. Purification and characterization of an activator protein for methanol dehydrogenase from thermotolerant Bacillus spp. *J. Biol. Chem.* **266**:3955–60. http://www.ncbi.nlm.nih.gov/pubmed/1995643.

Becskei A, Serrano L. 2000. Engineering stability in gene networks by autoregulation. *Nature* **405**:590–593.

Brockman IM, Prather KLJ. 2015. Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnol. J* **10**:1360–1369.

Carlson JC, Badran AH, Guggiana-Nilo DA, Liu DR. 2014. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat Chem Biol* **10**:216–222. http://www.ncbi.nlm.nih.gov/pubmed/24487694.

Chistoserdova L, Kalyuzhnaya MG, Lidstrom ME. 2009. The Expanding World of Methylotrophic Metabolism. *Annu. Rev. Microbiol.* **63**:477–499.

Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA. 1989. Functional Characterization of the Two Alcohol Oxidase Genes from the Yeast Pichia pastoris **9**.

Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**:6640–5. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=18686&tool=pmcentrez&rendertype=abstract.

Denby KJ, Iwig J, Bisson C, Westwood J, Rolfe MD, Sedelnikova SE, Higgins K,

Maroney MJ, Baker PJ, Chivers PT, Green J. 2016. The mechanism of a formaldehyde-sensing transcriptional regulator. *Sci. Rep.* **6**:38879. http://www.nature.com/articles/srep38879.

EPA. 1996. Method 8315A. Determination of carbonyl compounds by high performance liquid chromatography (HPLC).

Esvelt KM, Carlson JC, Liu DR. 2011. A system for the continuous directed evolution of biomolecules. *Nature* **472**:499–503. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3084352&tool=pmcentrez&rendertype=abstract.

Gonzalez CF, Proudfoot M, Brown G, Korniyenko Y, Mori H, Savchenko A V, Yakunin AF. 2006. Molecular basis of formaldehyde detoxification. Characterization of two S-formylglutathione hydrolases from Escherichia coli, FrmB and YeiG. *J. Biol. Chem.* **281**:14514–22. http://www.ncbi.nlm.nih.gov/pubmed/16567800.

Herring CD, Blattner FR. 2004. Global transcriptional effects of a suppressor tRNA and the inactivation of the regulator frmR. *J. Bacteriol.* **186**:6714–6720.

IARC. 2006. Monographs on the evaluation of carcinogenic risk to humans: Formaldehyde, 2-Butoxyethanol and 1-tert-Butoxypropan-2-ol. *Int. Agency Res. Cancer* **88**:1–390. http://monographs.iarc.fr/ENG/Monographs/vol83/mono83-1.pdf.

Irla M, Neshat A, Brautaset T, Rückert C, Kalinowski J, Wendisch VF. 2015. Transcriptome analysis of thermophilic methylotrophic Bacillus methanolicus MGA3 using RNA-sequencing provides detailed insights into its previously uncharted transcriptional landscape. *BMC Genomics* **16**:73. http://www.biomedcentral.com/1471-2164/16/73.

Iwig JS, Chivers PT. 2009. DNA Recognition and Wrapping by Escherichia coli RcnR. *J. Mol. Biol.* **393**:514–526. http://dx.doi.org/10.1016/j.jmb.2009.08.038.

Kalasz H. 2003. Biological Role of Formaldehyde, and Cycles Related to Methylation, Demethylation, and Formaldehyde Production. *Mini-Reviews Med. Chem.* **3**:175–192.

Kalyuzhnaya M, Puri AW, Lidstrom ME. 2015. Metabolic engineering in methanotrophic bacteria. *Metab. Eng.* **29**:142–152. http://www.ncbi.nlm.nih.gov/pubmed/25825038.

Kloosterman H, Vrijbloed JW, Dijkhuizen L. 2002. Molecular, biochemical, and functional characterization of a Nudix hydrolase protein that stimulates the activity of a nicotinoprotein alcohol dehydrogenase. *J. Biol. Chem.* **277**:34785–92. http://www.ncbi.nlm.nih.gov/pubmed/12089158.

Kotrbova-Kozak A, Kotrba P, Inui M, Sajdok J, Yukawa H. 2007. Transcriptionally regulated adhA gene encodes alcohol dehydrogenase required for ethanol and n-propanol utilization in Corynebacterium glutamicum R. *Appl. Microbiol. Biotechnol.* **76**:1347–56. http://www.ncbi.nlm.nih.gov/pubmed/17646983.

Krog A, Heggeset TMB, Müller JEN, Kupper CE, Schneider O, Vorholt J a, Ellingsen TE, Brautaset T. 2013. Methylotrophic Bacillus methanolicus Encodes Two Chromosomal and One Plasmid Born NAD(+) Dependent Methanol Dehydrogenase Paralogs with Different Catalytic and Biochemical Properties. *PLoS One* **8**:e59188. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3602061&tool=pmcentrez&rendertype=abstract.

Law JR. 2012. Molecular Basis of Bacterial Formaldehyde Sensing; University of

Manchester.

Müller JEN, Heggeset TMB, Wendisch VF, Vorholt J a., Brautaset T. 2015a. Methylotrophy in the thermophilic Bacillus methanolicus, basic insights and application for commodity production from methanol. *Appl. Microbiol. Biotechnol.* **99**:535–551. http://link.springer.com/10.1007/s00253-014-6224-3.

Müller JEN, Meyer F, Litsanov B, Kiefer P, Potthoff E, Heux S, Quax WJ, Wendisch VF, Brautaset T, Portais J-C, Vorholt J a. 2015b. Engineering Escherichia coli for methanol conversion. *Metab. Eng.* **28**:190–201. http://linkinghub.elsevier.com/retrieve/pii/S1096717615000026.

Müller JEN, Litsanov B, Bortfeld-Miller M, Trachsel C, Grossmann J, Brautaset T, Vorholt JA. 2014. Proteomic analysis of the thermophilic methylotroph Bacillus methanolicus MGA3. *Proteomics* **14**:725–737.

Müller JEN, Meyer F, Litsanov B, Kiefer P, Potthoff E, Heux S, Quax WJ, Wendisch VF, Brautaset T. 2015c. Engineering Escherichia coli for methanol conversion. *Metab. Eng.* **28**:190–201. http://dx.doi.org/10.1016/j.ymben.2014.12.008.

Nash T. 1953. The colorimetric estimation of formaldehyde by means of the Hantzsch reaction. *Biochem. J.* **55**:416–421.

Osman D, Piergentili C, Chen J, Sayer LN, Usón I, Huggins TG, Robinson NJ, Pohl E. 2016. The Effectors and Sensory Sites of Formaldehyde-responsive Regulator FrmR and Metal-sensing Variant. *J. Biol. Chem.* **291**:19502–19516.

Packer MS, Liu DR. 2015. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**:379–394. http://dx.doi.org/10.1038/nrg3927.

Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS. 2006. Engineering and

characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**:79–88. http://www.ncbi.nlm.nih.gov/pubmed/16369541.

Raman S, Rogers JK, Taylor ND, Church GM. 2014. Evolution-guided optimization of biosynthetic pathways. *Proc. Natl. Acad. Sci.* **111**:201409523. http://www.pnas.org/content/111/50/17803.abstract.

Rohlhill J, Sandoval NR, Papoutsakis ET. 2017. Sort-seq approach to engineering a formaldehyde-inducible promoter for dynamically regulated Escherichia coli growth on methanol.

Salthammer T. 2013. Formaldehyde in the ambient atmosphere: From an indoor pollutant to an outdoor pollutant? *Angew. Chemie - Int. Ed.* **52**:3320–3327.

Tralau T, Lafite P, Levy C, Combe JP, Scrutton NS, Leys D. 2009. An internal reaction chamber in dimethylglycine oxidase provides efficient protection from exposure to toxic formaldehyde. *J. Biol. Chem.* **284**:17826–34. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2719421&tool=pmcentrez&rendertype=abstract.

Tseng H-C, Martin CH, Nielsen DR, Prather KLJ. 2009. Metabolic engineering of Escherichia coli for enhanced production of (R)- and (S)-3-hydroxybutyrate. *Appl. Environ. Microbiol.* **75**:3137–45. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2681625&tool=pmcentrez&rendertype=abstract.

Whitaker WB, Jones JA, Bennett K, Gonzalez J, Vernacchio VR, Collins SM, Palmer MA, Schmidt S, Antoniewicz MR, Koffas MA, Papoutsakis ET. 2016. Engineering the Biological Conversion of Methanol to Specialty Chemicals in Escherichia coli.

Metab. Eng.:1–11.

Whitaker WB, Sandoval NR, Bennett RK, Fast AG, Papoutsakis ET. 2015. Synthetic methylotrophy: Engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization. *Curr. Opin. Biotechnol.* Elsevier Ltd. http://dx.doi.org/10.1016/j.copbio.2015.01.007.

WHO. 2005. Formaldehyde in Drinking Water.

Woolston BM, Edgar S, Stephanopoulos G. 2013. Metabolic Engineering: Past and Future. *Annu. Rev. Chem. Biomol. Eng.*:259–288. http://www.ncbi.nlm.nih.gov/pubmed/23540289.

Wu TY, Chen CT, Liu JTJ, Bogorad IW, Damoiseaux R, Liao JC. 2016. Characterization and evolution of an activator-independent methanol dehydrogenase from Cupriavidus necator N-1. *Appl. Microbiol. Biotechnol.* **100**:4969–4983. http://dx.doi.org/10.1007/s00253-016-7320-3.

Yurimoto H, Kato N, Sakai Y. 2005. Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. *Chem. Rec.* **5**:367–375.

Yurimoto H, Oku M, Sakai Y. 2011. Yeast methylotrophy: Metabolism, gene regulation and peroxisome homeostasis. *Int. J. Microbiol.* **2011**.

Zhang J, Jensen MK, Keasling JD. 2015. Development of biosensors and their application in metabolic engineering. *Curr. Opin. Chem. Biol.* **28**:1–8. http://dx.doi.org/10.1016/j.cbpa.2015.05.013.

Zhang J, Sonnenschein N, Pihl TPB, Pedersen KR, Jensen MK, Keasling JD. 2016a. Engineering an NADPH/NADP+ redox biosensor in yeast. *ACS Synth.*

Biol.:acssynbio.6b00135. http://pubs.acs.org/doi/abs/10.1021/acssynbio.6b00135.

Zhang J, Barajas JF, Burdu M, Ruegg TL, Dias B, Keasling JD. 2016b. Development of a Transcription Factor Based Lactam Biosensor. *ACS Synth. Biol.*:acssynbio.6b00136. http://pubs.acs.org/doi/abs/10.1021/acssynbio.6b00136.

Figure Legends

Figure 1 - Reporter construct design and formaldehyde response

Top: The de-coupled plasmid pTR47 controls *frmR* expression through the P_{tet} promoter, while the pTR47auto plasmid mirrors the native context of the frm operon of *E. coli*. Bottom: Comparison of the coupled (pTR47auto) and de-coupled (pTR47) reporter constructs, with or without anhydrotetracycline (ATc). Error bars show the standard deviation of three biological replicates. RLU, Relative Luminescence Units; OD, Optical Density at 600nm. Breaks in the lines for the 40ng mL^{-1} ATc graph at low concentrations of formaldehyde are due to negative values for luminescence at those time-points.

Figure 2 – Rational promoter engineering increases sensitivity and dynamic range

Top: Schematic showing the specific differences in the pTR47 and pTR47m4 constructs. Nucleotides in the binding sequence that are underlined show predicted FrmR binding regions. Blue nucleotides are putative σ70 binding sequences. Nucleotides in purple represent the cytosine repeat tract, and red nucleotides show the mutations introduced in pTR47m4. Bottom: Comparison of the formaldehyde-based response of cells containing the wild-type (pTR47, Black) and mutant (pTR47m4, Red) binding sites in the de-coupled circuit architecture. Error bars show the standard deviation across three biological replicates. RLU, Relative Luminescence Units; OD, Optical Density at 600nm.

Figure 3 – Evaluation of a GFP reporter variant in WT and ΔfrmA Strains

Left: Time course of the dose-dependent response to formaldehyde in WT cells. Right: Fitting of a Hill plot to the final GFP signal in WT (red) or *ΔfrmA* (black) background. Dashed vertical line represents the K_a of the reporter (21 μM or 6.2 μM). Error bars represent standard deviation of three biological replicates. RFU, raw fluorescence signal (AU).

Figure 4 –GFP Biosensor Enables Sensitive Mdh Activity Measurements

Left: Timecourse of GFP expression from reporter pTR47m4-gfp in response to formaldehyde generated *in vivo* by cells expressing Mdh. S1030 WT (black) or ΔfrmA (blue) cells expressing the evolved *mdh2* from *C. necator* (pTR48cnmdh4-1) were incubated with without 500 mM methanol (squares and triangles, respectively), and fluorescence was monitored for 150 minutes. Right: Time course of formaldehyde production in the supernatant of the same cells, measured by the Nash assay. Error bars represent standard deviation of three technical replicates.

Figure 5 – in vivo Comparison of Mdh variants using GFP formaldehyde biosensor

GFP fluorescence (RFU) measured two hours after induction of *mdh* expression and addition of various concentrations of methanol, and normalized to culture density (OD). Error bars represent standard deviations for three individual colonies. The control is cells carrying pTR48mCherry.

Figure 6 – Analysis and elimination of cheaters in a mixed population

Left: Schematic showing potential false activation of reporter in cells with no *mdh* (bottom) by diffusion from cells with active *mdh* (top) in a mixed culture. Cells that are only GFP+ are true formaldehyde producers, whereas cells that are both mCherry+ and GFP+ are ‘cheaters’ activated by diffusion. Right: Analysis of GFP fluorescence in *mdh*-

containing cells (blue bars) and mCherry-containing cells (pink bars) by FC with varying concentrations of glutathione added as an extracellular formaldehyde sink to prevent diffusion.

Figure 7 – High-throughput pathway balancing monitored by the formaldehyde reporter

Normalized GFP fluorescence in cells carrying pETMEOH560 2 hours after induction with various concentrations of ATc and IPTG, and treatment with 500 mM methanol. IPTG induction controls *mdh* expression, while ATc induction controls *hps* and *phi* (top right). Bars represent means of three biological replicates. RFU, raw fluorescence units; OD, optical density at 600 nm.

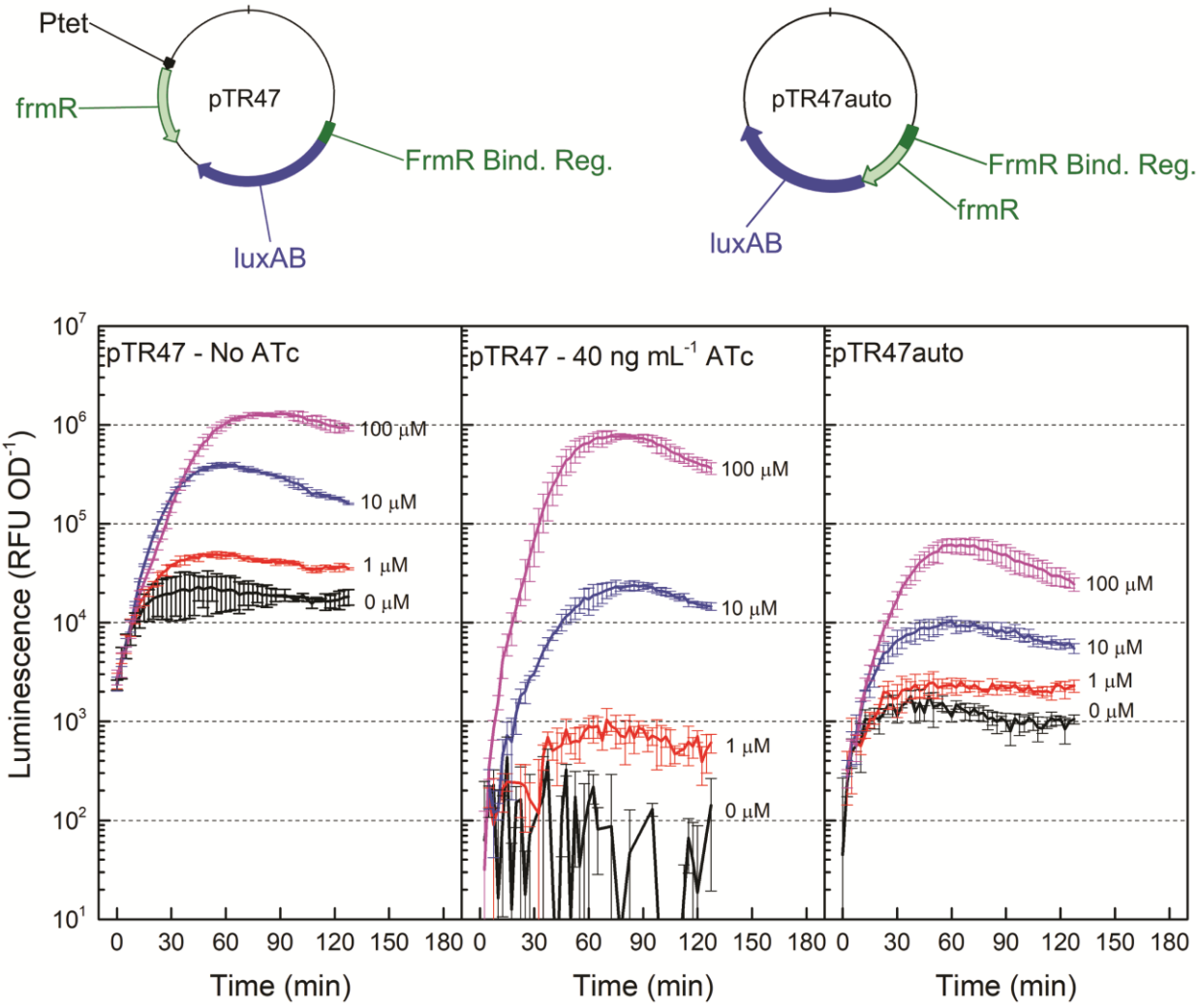

Figure 1

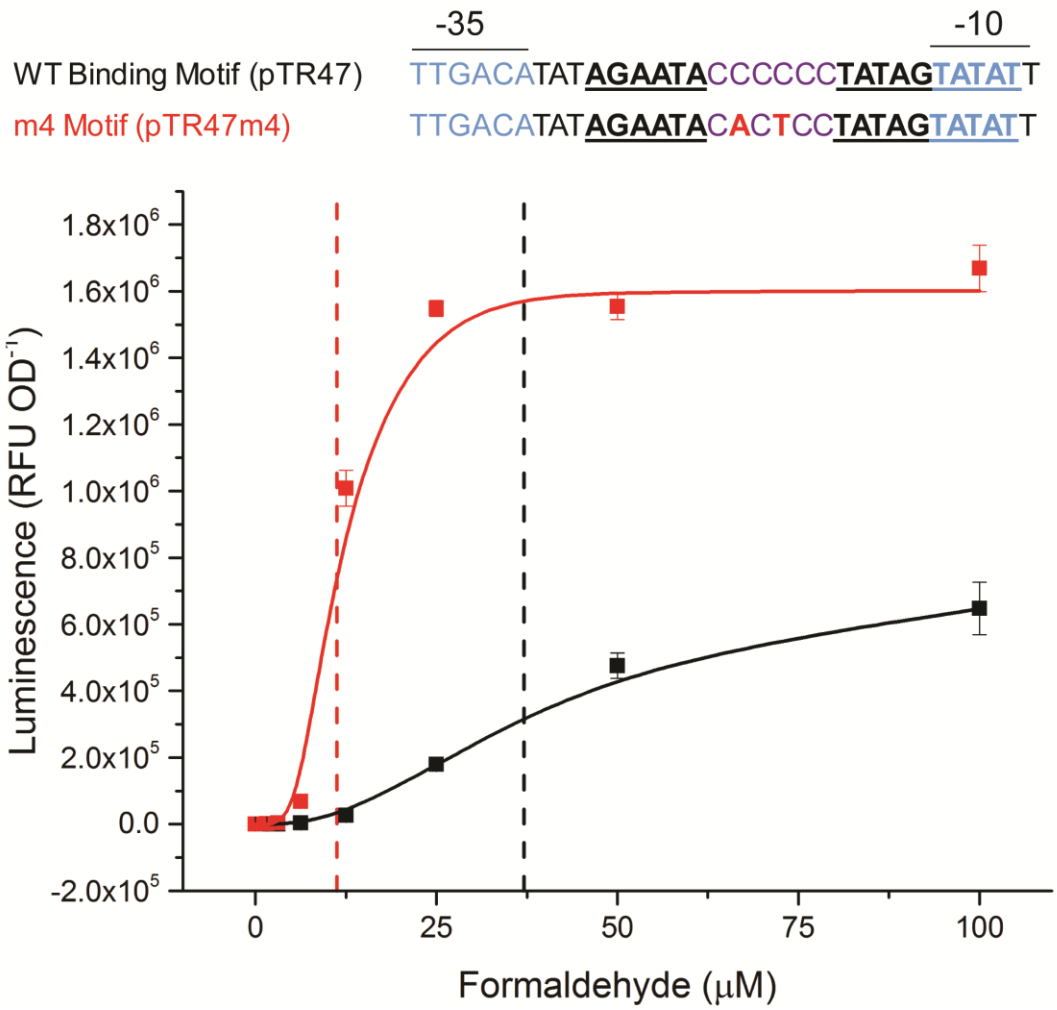

Figure 2

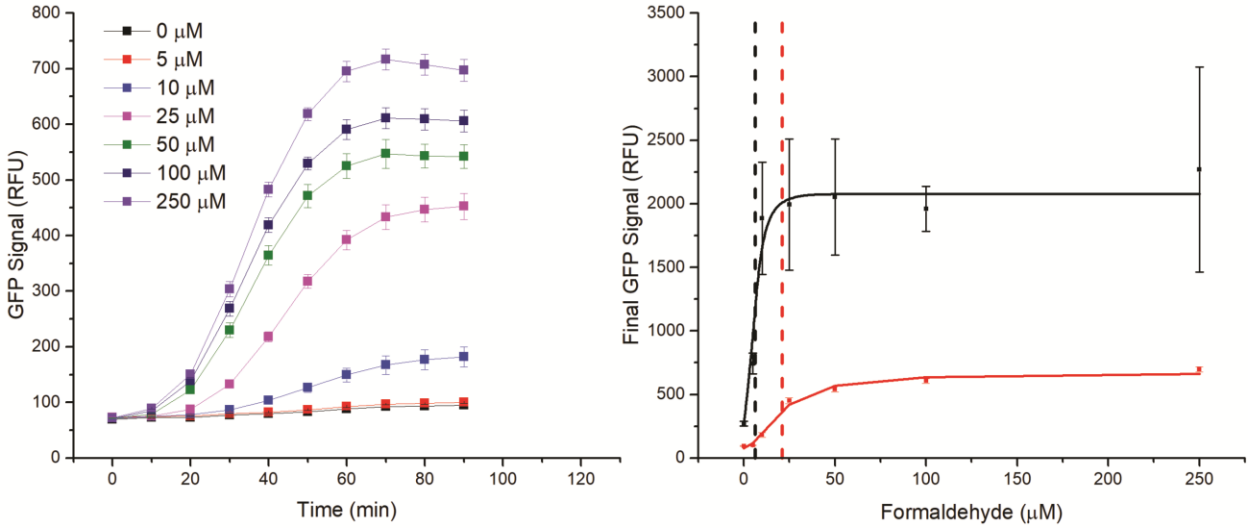

Figure 3

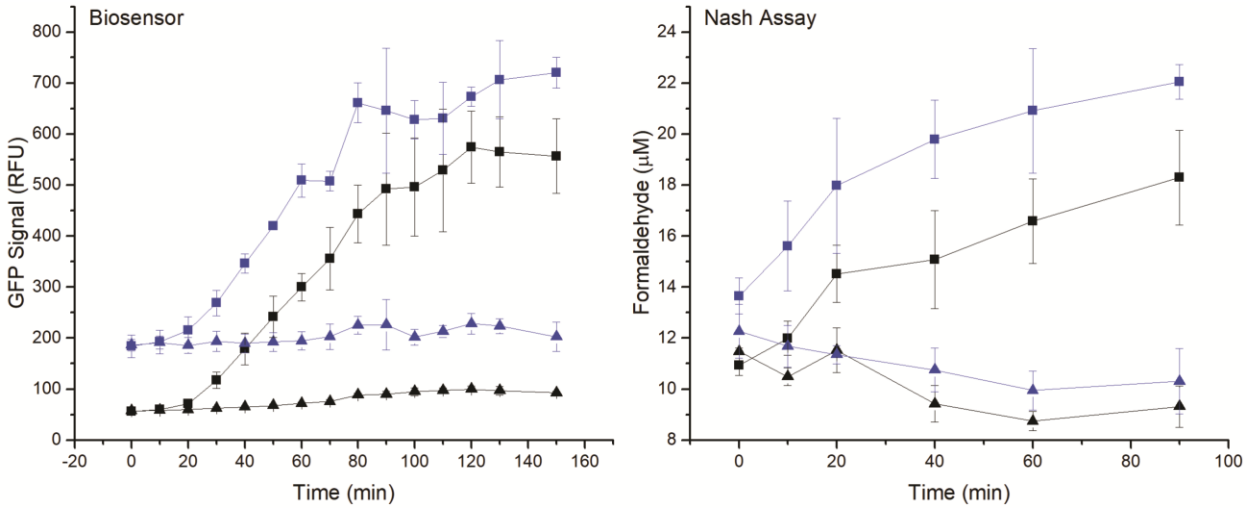

Figure 4

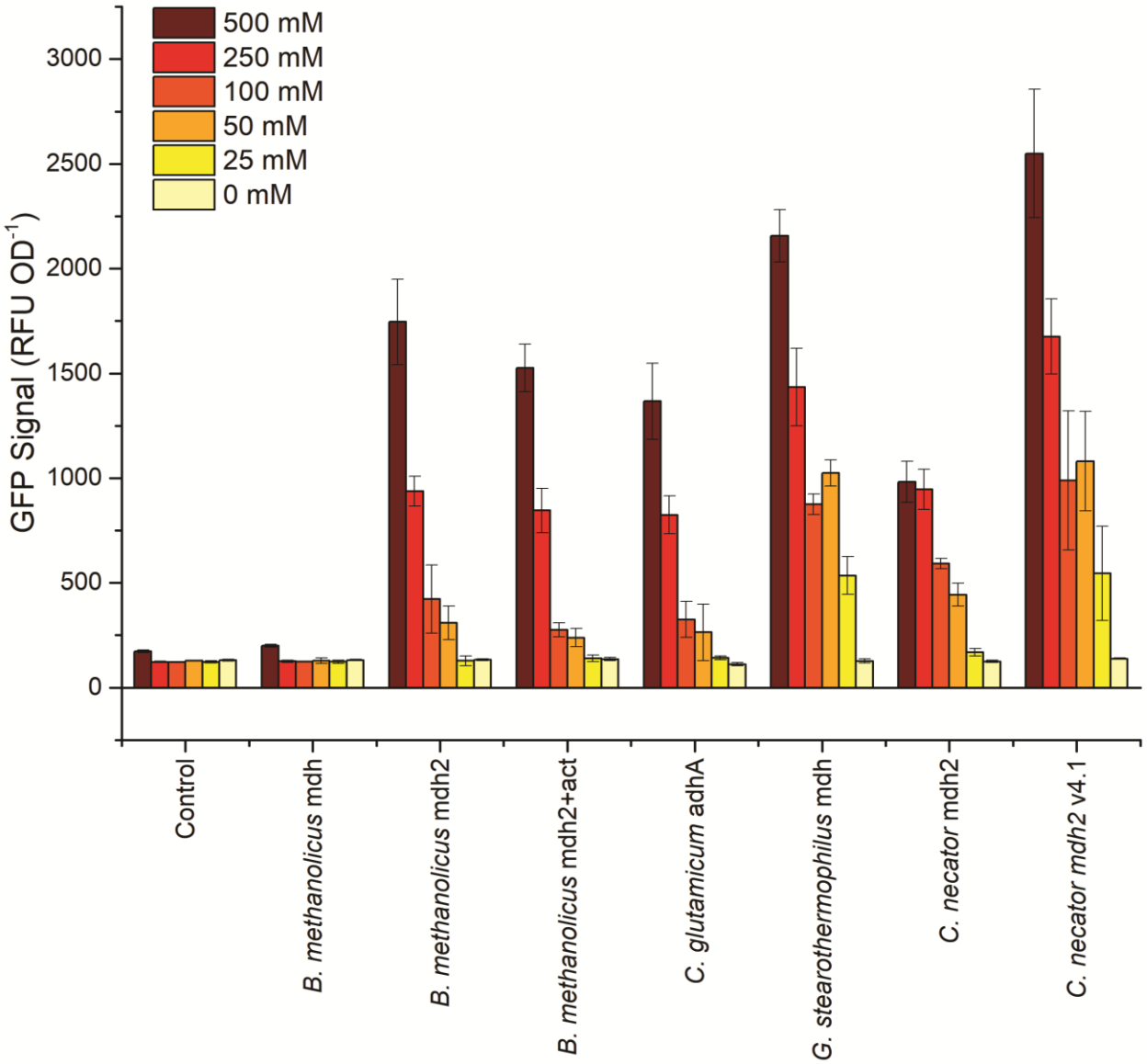

Figure 5

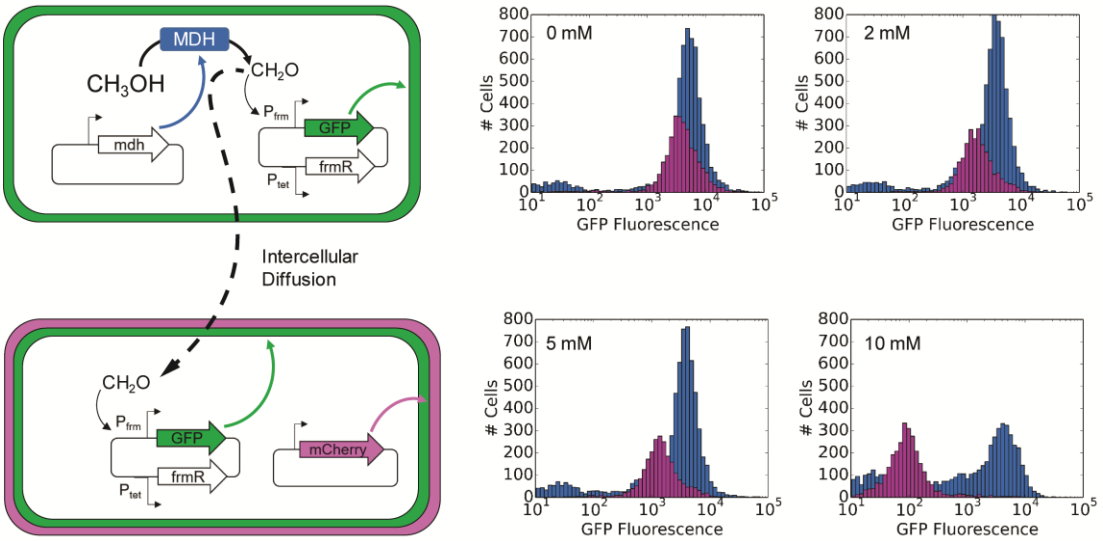

Figure 6

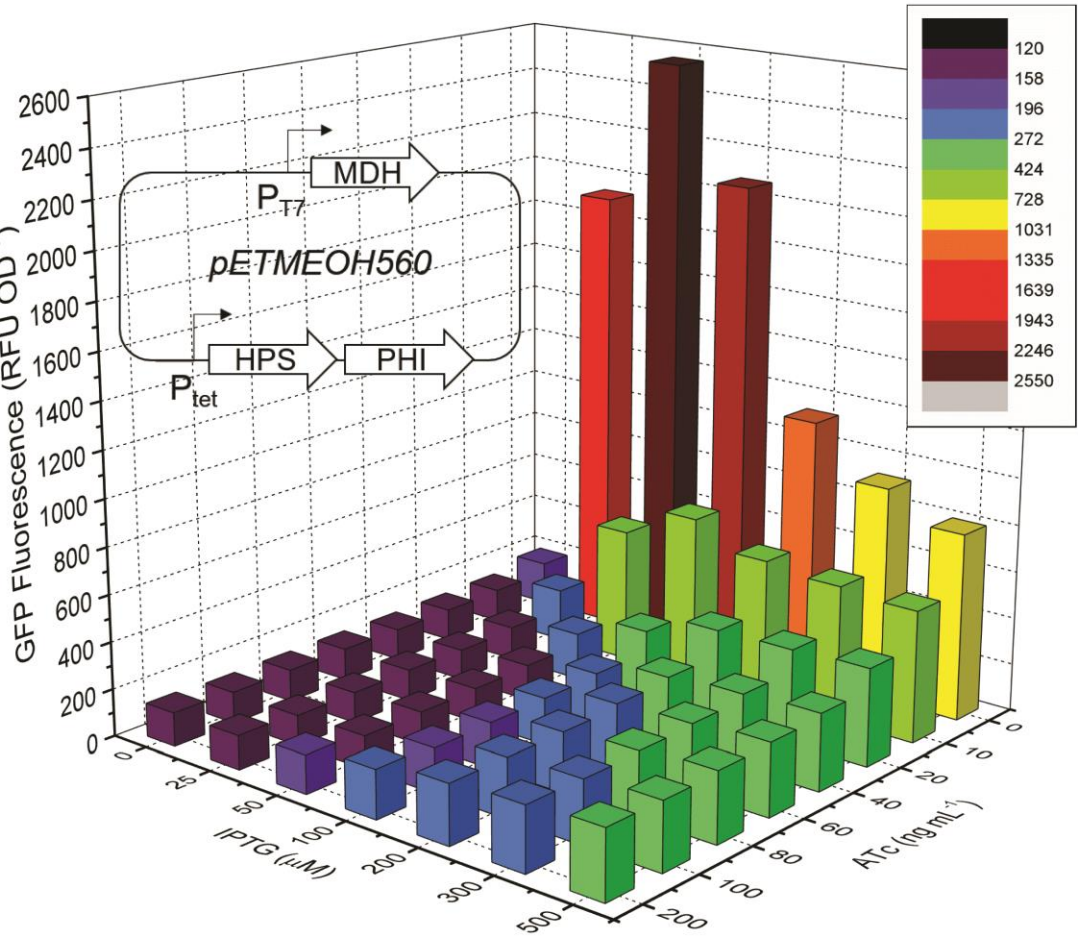

Figure 7