ENGINEERING *Escherichia coli* FOR GROWTH ON METHANOL THROUGH DYNAMIC REGULATION AND PROTEIN ENGINEERING

by

Julia Rohlhill

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemical Engineering

Fall 2019

ENGINEERING *Escherichia coli* FOR GROWTH ON METHANOL THROUGH DYNAMIC REGULATION AND PROTEIN ENGINEERING

by

Julia Rohlhill

Approved: _____
Eric M. Furst, Ph.D.
Chair of the Department of Chemical and Biomolecular Engineering

Approved: _____
Levi T. Thompson, Ph.D.
Dean of the College of Engineering

Approved: _____
Douglas J. Doren, Ph.D.
Interim Vice Provost for Graduate and Professional Education and Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Eleftherios T. Papoutsakis, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Wilfred Chen, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Maciek R. Antoniewicz, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Elizabeth Clarke, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

Throughout my PhD journey there have been many people who have provided me with meaningful technical, material, and emotional support. I would first like to thank my advisor, Terry Papoutsakis, for being an excellent mentor. My professional development has been solidly placed on a foundation he has provided. My writing, critical thinking, and experimental approach has benefited greatly from both his advice and through observation of his engineering perspective and success. Beyond the lab, I want to thank Terry for his ability to alleviate stress when we feel the pressures of graduate school, and for appreciating the importance of a full work-life balance. I want to thank my committee members Wilfred Chen, Maciek Antoniewicz, and Liz Clarke, for their collaboration and advice. They have challenged me to think critically about my research findings and to view those findings from new perspectives.

I want to thank my lab mates for being able to bring levity to the disappointment that can accompany repeated failure. Working and joking with them on a daily basis will be sorely missed. From the beginning thank you to Alan for being so generous with his time and genuinely caring so much for others, Elli for making me feel at home in lab from the beginning and being a good friend, and Nich for being an incredible mentor to me and training me when I joined lab. Thank you to Brian, Vanessa, and Lisa, for their willingness to share their knowledge and experience. Thank you to Stephanie for her comfort and reassurance during stressful times, Hannah for her camaraderie, Kamil for his sense of humor, Samik for his fabulous jokes, and Bryan for frequently lending a helping hand. Thanks to the undergrads I've

trained over the years: Alec, Evan, Logan, and Vicky for helping me to develop as a mentor. And thank you to our new members Jess, Will, and Jonathan, who will be continuing the ETP lab. Finally, I'd like to thank my immediate seniors Kyle and Chen-Yuan for their invaluable professional and personal advice, keeping me sane towards the end, and being sincerely decent human beings whom I respect at the highest level. Every single one of my labmates, as well as our other undergraduates, masters students, and collaborators, has contributed to a truly fantastic lab culture over the years and I am deeply appreciative of it.

I want to thank my centennial classmates for their help and camaraderie in our first year and throughout my PhD, in particular Lauren, Camil, John, Andrew, Kat, Jannat, and Wesley. Conversations over meals with John especially helped me think through many experimental problems and reminded me of the wider impact of and motivation for the work we do on a daily basis.

I want to thank my family, nuclear and extended, for their support through this journey. I will never be able to thank my parents enough for giving me the freedom to pursue my interests, and teaching me to value my own happiness. Thank you to my aunt who is my second mother, for being present for every major event in my life. And thank you to my in-laws for giving me family in Delaware and enthusiastically celebrating my accomplishments.

Thank you to Anna and Dominic for forcing me to see a bigger picture, and reinforcing my motivation for completing this degree. And absolutely most of all, I want to thank my husband Willem. I am eternally grateful for his love, support, and penchant for making bread. Life goes on during a PhD, and we have gone from dating, to engaged, to married, to parents twice over. Thank you, for truly everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Industrial microbial fermentation is frequently used to produce a wide array of desirable products, including amino acids, vitamins, recombinant proteins, pharmaceuticals, and alternative renewable fuels. Production of these compounds is most commonly accomplished utilizing glucose or other sugar substrates. Methanol has emerged in recent years as an attractive non-food feedstock option due to the increased production of natural gas. Methanol is produced from methane, the main component of natural gas, or can be produced from renewable sources at a higher cost. Methanol offers a high degree of reduction, or more electrons per carbon than sugar substrates, which translates to improved product titers, as well as a low contamination risk for large-scale fermentations. Native methylotrophic bacteria, capable of utilizing single-carbon substrates for their carbon and energy needs, have limited genetic tools to engineer the production of heterologous products. *Escherichia coli* is a model organism that has been extensively researched, has a well-developed genetic toolbox, and has been previously engineered to produce a wide array of products. For these reasons, *E. coli* is an appropriate host organism for the implementation of synthetic methylotrophy.

Attempts to engineer synthetic methylotrophy in *E. coli* have struggled with various bottlenecks. Most implementations rely on the expression of at least three

heterologous enzymes for methanol assimilation. An NAD-dependent methanol dehydrogenase (Mdh) oxidizes methanol to formaldehyde, a cytotoxic compound, which is fixed to ribulose 5-phosphate (Ru5P) by 3-hexulose-6-phosphate synthase (Hps) to form hexulose 6-phosphate. Subsequent isomerization by 6-phospho-3-hexuloisomerase (Phi) generates fructose 6-phosphate, an intermediate in central carbon pathways. We target several points of weakness to improve methanol assimilation, focusing on instituting formaldehyde-based regulation in *E. coli* to avoid accumulation of the toxic intermediate, and driving methanol oxidation through redox perturbation and protein engineering. Characterization of an *E. coli* formaldehyde-inducible promoter, P_{frm}, allows for its utilization in implementation dynamic formaldehyde regulation, driving the expression of genes involved in methanol and formaldehyde assimilation directly in response to intracellular formaldehyde concentrations. Specific genes from native methylotroph *Bacillus methanolicus* involved with the regeneration of Ru5P, necessary for formaldehyde fixation, were also expressed and placed under formaldehyde control to improve formaldehyde assimilation. Redox perturbation achieved by knocking out malate dehydrogenase, an NAD-dependent enzyme, to drive NAD-dependent methanol oxidation, resulted in dramatically higher growth benefit with methanol on a yeast extract co-substrate. Lastly, the Mdh sourced from *Geobacillus stearothermophilus* was engineered for higher methanol activity and selectivity to drive methanol oxidation.

Together, we show these approaches to significantly improve the utilization of methanol carbon and energy in *E. coli*. Future attempts to create a pure synthetic

methylotrophy will likely need to continue focusing on regulation and the adequate expression of relevant enzymes for sustained methanol assimilation and growth.

Chapter 1

INTRODUCTION

1.1 Methanol as a substrate for microbial fermentations

Natural gas production has risen over 25% in the United States from 2013 through 2018, due to the discovery of additional sources of recoverable natural gas and technological advances increasing the efficiency of extraction.[1] This increase has motivated interest in utilizing methane, the primary component of natural gas, and its oxidized product methanol, as a substrate for microbial fermentations. Traditional microbial fermentations utilize sugar substrates, which can compete with the food supply.[2] One approach for alternative substrates is lignocellulosic biomass, a complex mixture of polysaccharides and lignin which does not compete with the food supply.[3] Current utilization remains difficult however due to the energy and cost intensive methods such as enzymatic or chemical digestion required to pretreat the biomass and convert it into fermentable sugars.[3] Methanol offers several advantages over traditional sugar substrates. It has a degree of reduction of 6 compared to 4 for glucose, which means more electrons per carbon to improve product yields. Many microorganisms are sensitive to methanol toxicity, leading to a lower risk of contamination for industrial fermentations.[4] Methanol also has high availability and a decreasing price due to its chemical conversion from methane derived from natural gas.[5]

While the majority of methanol is currently produced from natural gas, methane from bio-gas can also be used, and the production of renewable methanol, or bio-methanol, has been investigated as well. Bio-methanol production typically involves reacting carbon dioxide from renewable or waste sources with hydrogen produced via electrolysis. A renewable methanol plant built in Iceland reports a production of 4000 metric tons of methanol per year using carbon dioxide from geothermal steam, and hydrogen produced via electrolysis from renewable sources of energy.[6, 7]

Carbon dioxide is an industrial waste, and thus can be sourced from biogas, municipal solid waste, or from the atmosphere. The electrolysis of hydrogen is the most expensive and energy-intensive portion of this process, accounting for approximately 65% of all costs.[8] Coupling of renewable energies to hydrogen production is essential for an economical process and to reduce net CO_2 generation during methanol production.[9] One technoeconomic analysis focused on coupling bio-methanol synthesis with hydrogen electrolysis powered by wind energy,[10] and a life-cycle assessment of bio-methanol production from fermentation-sourced CO_2 and hydrogen from wind powered electrolysis projected 82-86% lower greenhouse gas emissions.[11] Others have focused on modeling other forms of renewable energy, such as the coupling of solar energy to traditional or more environmentally friendly methanol production processes.[12] Active research in bio-methanol production focuses on reducing the cost of water electrolysis for hydrogen production, investigating efficient carbon capture technologies, and incorporating renewable energy into traditional methanol production processes.

1.1.1 Current progress engineering biological methane to methanol conversion

Native methanotrophs utilize enzymes to convert methane to methanol *in vivo*. Methanotrophs have few genetic tools necessary to produce high titers of products of interest. Previous attempts to express heterologous soluble methane monooxygenases (MMOs) from methanotrophs have been met with limited success. The β-subunit of particulate MMO (pMMO) was expressed in *Escherichia coli* with detectable activity, and protein engineering of the P450 monooxygenase has been attempted as well.[13]

Most recently, a modified version of the particulate methane monooxygenase (pMMO) from *Methylococcus capsulatus* (Bath) was successfully expressed in *E. coli* and purified, however functionality *in vivo* was not tested.[14] The significant challenges involved with expression of MMO enzymes in *E. coli*, as well as obtaining reasonable MMO activity, motivate our focus on implementing methanol assimilation, a potential future foundation for attempts at synthetic methanotrophy.

1.1.2 Choosing a host for methanol utilization

Native methylotrophs capable of consuming single-carbon substrates such as methane and methanol as their sole source of carbon and energy can include yeasts, bacteria, fungi, and archaea. Methanol utilization in particular can be found in certain bacteria and yeasts. Progress has been made in developing genetic tools for some native methylotrophic bacteria. For example, plasmid transformation protocols have been developed for *Bacillus methanolicus* and there are a few expression vectors and inducible promoter systems which have been used with success in the organism.[15, 16] *B. methanolicus* has also been used for the production of lysine, glutamate, and cadaverine.[16, 17] Other methylotrophic bacteria have also been used or engineered for

the production of amino acids including serine, threonine, glutamate, and lysine, however they have rarely been engineered for the production of non-native products.[18]

Despite the progress made in recent years, the genetic toolbox for native methylotrophs is relatively undeveloped, and the choices for product synthesis are limited. *E. coli* is one of the most studied microorganisms, with a well-developed set of genetic tools for chromosomal integrations and deletions, heterologous gene expression, and controlled inducible expression systems. *E. coli* has also been engineered to produce a wide array of heterologous products, and is a frequently used industrial host. Given these considerations, *E. coli* is an appropriate host for the development of synthetic methylotrophy, and success in this endeavor would allow for the methanol-driven production of dozens of industrially relevant products.

1.2 *E. coli* synthetic methylotrophy

Native methylotrophs capable of consuming single-carbon substrates such as methane and methanol as their sole source of carbon and energy can include yeasts, bacteria, fungi, and archaea. Methanol utilization in particular can be found in certain bacteria and yeasts. Progress has been made in developing genetic tools for some native methylotrophic bacteria. For example, plasmid transformation protocols have been developed for *Bacillus methanolicus* and there are a few expression vectors and inducible promoter systems which have been used with success in the organism.[15, 16] *B. methanolicus* has also been used for the production of lysine, glutamate, and cadaverine.[16, 17] Other methylotrophic bacteria have also been used or engineered for the production of amino acids including serine, threonine, glutamate, and lysine, however they have rarely been engineered for the production of non-native products.[18]

Despite the progress made in recent years, the genetic toolbox for native methylotrophs is relatively undeveloped, and the choices for product synthesis are limited. *E. coli* is one of the most studied microorganisms, with a well-developed set of genetic tools for chromosomal integrations and deletions, heterologous gene expression, and controlled inducible expression systems. *E. coli* has also been engineered to produce a wide array of heterologous products, and is a frequently used industrial host. Given these considerations, *E. coli* is an appropriate host for the development of synthetic methylotrophy, and success in this endeavor would allow for the methanol-driven production of dozens of industrially relevant products.

1.2.1 Enzymes for methanol oxidation

Native methylotrophs have developed three major classes of methanol oxidation enzymes, distinguished by their electron acceptor. Alcohol oxidases (AOXs), used by methylotrophic yeasts, convert methanol and O_2 to formaldehyde and hydrogen peroxide.[19] Methylotrophic bacteria utilize either NAD^+ or pyrroloquinoline quinone (PQQ)-dependent methanol dehydrogenases (Mdhs). *E. coli* produces and utilizes NAD^+ as a cofactor but is unable to produce PQQ without the heterologous expression of multiple proteins. While the thermodynamics are far more favorable for PQQ-dependent Mdh and AOX catalyzed reactions at physiological conditions, NAD-dependent methanol oxidation is the most technically feasible with regard to the transfer and expression of necessary heterologous enzymes.[5] NAD-dependent Mdhs also allow for the possibility of operating under anaerobic conditions, which may be desired for the production of some metabolites.

The first assessment of identifying suitable NAD-dependent Mdhs in *E. coli* was undertaken by Müller et al., where the *in vivo* and *in vitro* activities were assayed of Mdhs sourced from *B. methanolicus*, *B. coagulans*, *Desulfitobacterium hafniense*, *Lysinibacillus fusiformis*, *L. sphaericus*, and *Desulfotomaculum kuznetsovii*.[20] The Mdh2 from *B. methanolicus* MGA3 displayed the highest *in vivo* activity, but the Mdh2 of *B. methanolicus* PB1 showed higher activity during a dynamic ^{13}C-methanol labeling experiment.[20] Attempts to engineer NAD-dependent Mdhs for improved methanol assimilation in *E. coli* have been undertaken for the Mdh2 from *Cupriavidus necator* N-1,[21] and the *B. methanolicus* Mdh2.[22] The *C. necator* N-1 Mdh was improved through directed evolution, with one variant, denoted CT4-1, achieving an 83% lower K_m towards methanol of 22 mM.[21] The *B. methanolicus* Mdh2 was engineered with a phage-assisted evolution approach, and achieved a 48% lower K_m for methanol of 329 mM.[22]

Whitaker et al. compared the *B. methanolicus* Mdh2 with the *Geobacillus stearothermophilus* Mdh, which has a reported K_m for methanol of 20 mM.[23, 24] During growth on methanol and yeast extract the *G. stearothermophilus* Mdh was determined to be far superior both in biomass titer, formaldehyde production, and ^{13}C-methanol incorporation.[23] Engineering of the superior *G. stearothermophilus* Mdh for improved methanol activity and selectivity will be one focus of this dissertation.

1.2.2 Enzymes for formaldehyde fixation

Formaldehyde assimilation can occur through multiple pathways, including the ribulose monophosphate (RuMP) pathway, the serine cycle, and the ribulose bisphosphate pathway. Of these, the RuMP is the simplest to implement in *E. coli*,

requiring the expression of two heterologous enzymes: 3-hexulose-6-phosphate synthase (encoded by *hps*) and 6-phospho-3-hexuloisomerase (encoded by *phi*). The RuMP pathway is also more favorable energetically compared to the serine cycle, producing one NADH for every three formaldehyde molecules assimilated.[23]

Müller et al. initially tested the *in vivo* and *in vitro* activity of Hps enzymes expressed either in an operon or fusion configuration with Phi, sourced from *B. methanolicus* MGA3 and *Methylobacillus flagellatus* KT.[20] The authors also tested the *in vitro* activity of Phi enzymes sourced from *B. methanolicus* and *M. flagellatus* KT. Of these, they determined the *B. methanolicus hps* and *phi* displayed the highest activity.[20] Subsequent attempts at synthetic methylotrophy have continued using the *B. methanolicus* Hps and Phi,[23, 25] the Hps and Phi sourced from *M. flagellatus*, the *B. methanolicus* Hps with the *M. capsulatus* Phi,[26] or the *B. methanolicus* Hps with the *M. flagellatus* Phi.[27]

An earlier publication seeking to establish formaldehyde resistance also showed functional expression of a fusion of *hps* and *phi* sourced from *Mycobacterium gastri* MB19, and observed higher activity for the fusion compared to when they were expressed separately.[28] A protein fusion approach combining *B. methanolicus* MGA3 Mdh3 and the previously published *M. gastri* Hps-Phi fusion showed dramatic improvements for *in vitro* rates of fructose 6-phosphate production and methanol consumption, as well as improved methanol consumption *in vivo*.[29]

1.2.3 Strain engineering approaches

Strain engineering approaches for improving methanol assimilation by *E. coli* include any gene integrations, deletions, or modifications, implemented by rational design,

adaptive evolution, or random mutagenesis. The first main strain modification found to be highly beneficial was the deletion of the *E. coli* formaldehyde dehydrogenase enzyme, encoded by *frmA*. This was implemented by rational design,[20, 23] to decrease formaldehyde loss to carbon dioxide versus assimilation through central carbon pathways, but its importance was also verified during evolution studies which would repeatedly find deleterious mutations in the formaldehyde detoxification pathway.[27, 30]

Several publications have made various improvements to methanol assimilation rates through rational and evolution-based strain engineering approaches intended to force methanol carbon to be consumed by *E. coli*. These have had success in improving methanol consumption and are likely the best approach for using methanol as a co-substrate to improve product yields, however by design the strains are reliant on a co-substrate and cannot achieve growth on methanol alone. Meyer et al. was the first to publish this approach, and used a series of software models with an iterative approach to identify gene knock-outs which would render *E. coli* unable to grow on a multi-carbon source alone, but allow it to grow in combination with methanol.[30] Utilizing this information, the authors knocked out *edd*, encoding phosphogluconate dehydratase, and *rpiAB*, encoding ribose-5-phosphate isomerases, provided plasmids encoding *B. methanolicus* PB1 *mdh2* and *hps* and *phi* from *M. flagellatus* to generate the strain MeSV1. After approximately 35 generations of growth on methanol, gluconate, and pyruvate, the strain exhibited methanol-dependent growth on gluconate with a small amount of yeast extract (0.1 g/L). The authors then knocked out *maldh* encoding malate dehydrogenase, resulting in dramatically improved methanol-dependent growth after only three passages with pyruvate.[30] The

maldh deletion was intended to address a hypothesis of redox imbalance, which will be discussed in more depth later.

Chen et al. designed two strains for methanol auxotrophy, relying on either ribose or xylose as a co-substrate.[27] The deletion of *rpe* encoding ribulose-phosphate 3-epimerase, rendered an *E. coli* strain unable to grow on ribose, while a deletion *rpiAB* encoding ribose-5-phosphate isomerase A and B, rendered an *E. coli* strain unable to grow on xylose. In both cases, the provision of methanol and expression of an engineered *mdh2* from *C. necator*, *hps* from *B. methanolicus*, and *phi* from *M. flagellatus*, rescued growth. RBS libraries upstream of the *mdh*, *hps*, and *phi* genes and laboratory evolution led to strains with improved growth, and butanol and ethanol production was also demonstrated with the auxotrophic strains.[27]

Strain engineering to improve methanol assimilation on yeast extract or amino acid substrates has also been an active area of research. Methanol contribution towards biomass titer and methanol carbon incorporation during growth, measured via ^{13}C-labeling, was first demonstrated by Whitaker et al. in a Δ*frmA* background strain with *G. stearothermophilus* Mdh and *B. methanolicus* Hps and Phi.[23] The production of the flavonoid product naringenin was also engineered in the same strain, demonstrating methanol carbon conversion to a valuable product.[23] Another study analyzed individual amino acid co-substrates for methanol utilization and showed that deletion of the leucine-responsive regulatory protein (encoded by *lrp*) led to improved methanol utilization.[31] Bennett et al. utilized a deletion of phosphoglucose isomerase (encoded by *pgi*) to improve methanol incorporation during growth with glucose, as glucose flux is forced to the oxidative pentose phosphate pathway (PPP), producing the ribulose 5-phosphate intermediate needed for formaldehyde fixation.[25] The Δ*frmA*

Δ*pgi* strain also demonstrated improved acetone titers and methanol utilization, and small labeling improvements in several amino acids during growth with yeast extract and methanol. Most relevant to this dissertation, Bennett et al. integrated five heterologous non-oxidative PPP genes (*rpe*, *tkt*, *fba*, *glpX*, *pfk*) from *B. methanolicus* into the Δ*frmA* strain with Mdh, Hps, and Phi.[25] Expression of these genes involved in ribulose 5-phosphate regeneration resulted in increased methanol assimilation into a range of intracellular metabolites and amino acids, emphasizing the importance of ribulose 5-phosphate pools for continuous methanol and formaldehyde assimilation.[25]

1.2.4 Regulation challenges

Most, if not all, native *E. coli* substrates trigger regulatory responses when they enter the cell or are added to the medium via intracellular and extracellular sensor-regulators. These responses induce the expression of genes responsible for catabolizing the target substrates, as well as the repression of unrelated genes, or those capable of catabolizing alternative substrates. When a preferred substrate such as glucose is added to *E. coli* growth medium, it is transported into the cell and phosphorylated via the phosphotransferase (PTS) system. One component of this system, EIIA, exists predominantly in its unphosphorylated form upon rapid glucose uptake, and in this form will bind to transporters for alternate carbon sources such as lactose and maltose, preventing their uptake and leading to carbon catabolite repression (CCR).[32] For non-glucose sugar substrates, the global transcriptional regulator CRP (cAMP receptor protein) in combination with cAMP (cyclic AMP), acts as an activator or a repressor for hundreds of genes.[33] These complex regulatory responses are responsible for efficient resource allocation, and expression of catabolic

enzymes at appropriate levels allows for complete utilization of the recognized substrate.

Similarly, native methylotrophs often have tight regulatory schemes for strongly upregulating the transcription of genes involved in the consumption of methanol when methanol is present in the medium. Methylotrophic yeast *Pichia pastoris* has the most well-known methanol induction system. The *P. pastoris* promoter upstream of its alcohol oxidase 1 gene (P_{AOX1}) is strongly induced in the presence of methanol, and repressed in the presence of glucose, glycerol, or ethanol.[34, 35] Methylotrophic bacteria are no exception to this pattern. Methylotroph *M. smegmatis* appears to use a two-component system to regulate expression of its methanol dehydrogenase,[36] *Methylobacterium organophilum* has a methanol sensing domain MxcQ,[37] and *Paracoccus denitrificans* relies on the sensor-regulator FlhRS for regulating formaldehyde and methanol oxidation,[38] among others.[39, 40]

B. methanolicus displays 3- to 40-fold transcriptional upregulation of genes involved in methanol assimilation during methanol growth.[41] This includes 2.9-fold upregulation of *mdh*, 6.8-fold and 6.0-fold upregulation of *hps* and *phi*, respectively, and 6- to 40-fold upregulation of genes involved in ribulose 5-phosphate regeneration (*rpe*, *tkt*, *fba*, *pfk*, and *glpX*). Furthermore, the specific inducer for the observed upregulation was determined to be formaldehyde for *hps* and *phi*.[41] Formaldehyde-induced gene expression of *hps* and *phi* has been observed in non-methylotrophs as well for the proposed purpose of formaldehyde dissimilation.[42, 43] A separate RNA-sequencing analysis for *B. methanolicus* measured 'very high' transcript abundance for *hps*, *phi*, *pfk*, and *rpi*, and hypothesized the high abundance to be due to methanol in most of the growth media.[44]

There have been some attempts to develop a methanol-sensing *E. coli* strain through the fusion of methanol-sensing domains from native methylotrophs with *E. coli* transmitter domains. The MxcQZ methanol-sensing domain from *M. organophilum* XX was fused with *E. coli* EnvZ to create a two component system capable of sensing methanol consequently phosphorylating OmpR to induce gene expression downstream of the *ompC* promoter.[37] The same group engineered two other systems for *E. coli* with EnvZ, but fused it with either the FlhS methanol-sensing domain,[45] or the MxaY methanol-sensing domain,[46] both sourced from *P. denitrificans*. All of these methods sense exogenous methanol rather than intracellular methanol, and require a two-step process to achieve the expression of a target gene.

As a non-native substrate, methanol is not recognized by *E. coli*, and thus does not trigger any known regulatory response upon supplementation, at non-toxic levels, into the medium. Gonzalez et al. demonstrated improved methanol utilization in a methylotrophic *E. coli* strain with a deletion of the leucine-responsive regulatory protein, highlighting the importance of regulation for methanol assimilation.[31] *E. coli* does have a native formaldehyde-responsive promoter, P_{frm}, upstream of its formaldehyde detoxification operon *frmRAB*.[47] The high toxicity of formaldehyde makes it an ideal inducer for regulating synthetic methylotrophy. Utilization of dynamic formaldehyde control in this dissertation is inspired by native methylotrophic regulation and driving gene expression directly in response to cell needs.

1.2.5 Redox considerations

Using a NAD-dependent methanol dehydrogenase in *E. coli* may disrupt the cellular redox balance, particularly if the high flux desired for methanol assimilation is

achieved. During an effort to engineer a gluconate- and methanol-dependent growth phenotype in *E. coli* discussed previously, flux balance calculations suggested that most of the cellular NADH was produced via the Mdh.[30] In that case, the authors hypothesized a surplus of NADH was leading to stalled growth. They knocked out the NAD-dependent malate dehydrogenase (encoded by *maldh*) within the TCA cycle to reduce cellular NADH, a decision motivated by thermodynamic considerations, as lower NADH concentrations would decrease the Gibbs energy of the methanol oxidation reaction, as well as the genetics and gene expression patterns of native methylotrophs.[30]

Native methylotrophs that utilize the RuMP formaldehyde fixation pathway often have incomplete or less active TCA cycles. Obligate methylotroph *Methylobacillus flagellatus* has been shown with genomic and proteomic studies to be missing the α–ketoglutarate, malate, and succinate dehydrogenase enzymes.[48, 49] Analysis of enzyme levels in *B. methanolicus* MGA3 suggested a less active TCA cycle,[50] in agreement with previous observations of downregulated TCA cycle enzymes during growth on methanol compared to mannitol.[51] Another study involving a genome-scale reconstruction for native methylotroph *M. extorquens* AM1 claimed the TCA cycle to be operating in an incomplete mode during methanol growth.[52]

Meyer et al. demonstrated significantly improved gluconate- and methanol-dependent growth after only three passages for their stain containing the *maldh* knock-out (in addition to knock-outs of *edd* and *rpiAB*).[30] After continued passaging to improve growth, they also observed mutations in *nadR*, encoding the DNA-binding transcriptional repressor and nicotinamide mononucleotide adenylyltransferase NadR. These mutations were hypothesized to reduce the function of the NAD repressor,

theoretically leading to higher levels of NAD^+ and enabling higher methanol oxidation.[30] In a parallel approach implementing xylose- and methanol-dependent growth, a mutation in *cyaA* was found to contribute highly to improved growth after laboratory evolution.[27] The *cyaA* gene encodes an enzyme that produces cyclic AMP (cAMP) from ATP, and the authors hypothesized that a deleterious mutation in their evolved strain led to lower transcription of TCA cycle enzymes, a favorable condition for NAD-dependent methanol oxidation.[27] Direct assaying of NADH levels in an *E. coli* methylotroph also measured higher NADH levels with methanol compared to the no methanol control.[53] The authors took advantage of this by engineering higher NADPH conversion from the surplus NADH and producing lysine.[53] These findings motivate our work in characterizing and applying a malate dehydrogenase knock-out *E. coli* strain for improved synthetic methylotrophy.

1.3 Dissertation aims and outline

Here, we approach the challenge of *E. coli* synthetic methylotrophy with the goal of improving methanol assimilation, as measured by increased growth with methanol supplementation and increased methanol carbon incorporation as measured with ^{13}C-labeling. We use several methods to improve methanol utilization, all based on regulation and gene expression patterns present in native methylotrophic bacteria discussed previously. First, we characterize and engineer the native *E. coli* formaldehyde-responsive promoter P_{frm} to manipulate its dynamic range utilizing a sorting and sequencing approach to identify specific nucleotide mutations which strengthen or weaken the binding of transcriptional repressor FrmR or RNA

polymerase (RNAP) to the promoter region. Development of a tunable formaldehyde biosensor has applications beyond synthetic methylotrophy, expanding the toolbox of genetic parts for use in environmental applications or in combination with other orthogonal biosensors. Second, we apply the P_{frm} promoter to non-oxidative pentose phosphate pathway genes from *B. methanolicus* for formaldehyde-controlled expression of enzymes responsible for ribulose 5-phosphate regeneration. In combination with this approach we delete the *malate dehydrogenase* (*maldh*) gene, theoretically disrupting the NAD^+/NADH ratio in the cell to drive methanol oxidation with the NAD-dependent Mdh. Third, we engineered the *G. stearothermophilus* Mdh informed through a sorting and sequencing approach, and tested identified variants of interest with *in vitro*, *in vivo*, and production assays. Together, these approaches improve synthetic methylotrophy by targeting initial methanol oxidation, ribulose 5-phosphate regeneration and therefore continuous formaldehyde fixation, and regulation and redox states to implement important characteristics of native methylotrophs into an *E. coli* synthetic methylotroph.

Chapter 2

SORT-SEQ APPROACH TO ENGINEERING A FORMALDEHYDE-INDUCIBLE PROMOTER FOR DYNAMICALLY REGULATED *Escherichia coli* GROWTH ON METHANOL

Reprinted with permission from: Rohlhill J, Sandoval NR, Papoutsakis ET (2017) Sort-Seq Approach to Engineering a Formaldehyde-Inducible Promoter for Dynamically Regulated *Escherichia coli* Growth on Methanol, *ACS Synthetic Biology* 6(8): 1584-1595 (https://pubs.acs.org/doi/abs/10.1021/acssynbio.7b00114)

2.1 Background

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.[54, 55] 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.[56, 57] 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.[57] Efforts to expand the synthetic biology toolbox have focused on characterizing a range of biosensors[58] and engineering existing regulators[59] to respond to new effectors.[60, 61] 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'.[62] 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.[62] 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 2.1). 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.[63] This quantification is achieved by calculating the mutual information, or dependence of the two random variables on each other:[64, 65]

$$I(b_i;\mu) \approx \sum_{b_i,\mu} f(b_i,\mu)\ \log_2 \frac{f(b_i,\mu)}{f(b_i)f(\mu)} - c \qquad (2.1)$$

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.[65, 66] 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,[67] they have rarely been used in combination with mutual information techniques. Two papers of interest used the approach to analyze mammalian enhancers[64] (termed massively parallel reporter assay (MPRA)) and the CRP activator binding[65] 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.[68] *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.[69] It binds the *frmRAB* promoter region and is negatively allosterically modulated by formaldehyde.[69, 70] 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.[69, 70] 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.[71] The negative feedback regulation of the *frmRAB* operon is similar to many other prokaryotic operons whereby the transcription factor represses its own transcription.[72] 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.[5, 6] Previous studies have shown labeling in *E. coli* glycolytic intermediates from ^{13}C-methanol by heterologously expressing three enzymes from *Bacillus methanolicus*.[20] 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*,[73] 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.[23] We have also demonstrated a strategy of scaffoldless enzyme assembly that can be used to achieve superior outcomes in synthetic methylotrophy.[29] Placing formaldehyde assimilation genes under the control of formaldehyde regulation emulates the native methylotrophic regulation of methylotroph *B. methanolicus*, whereby *hps* and *phi* are transcriptionally induced by formaldehyde,[74] 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.

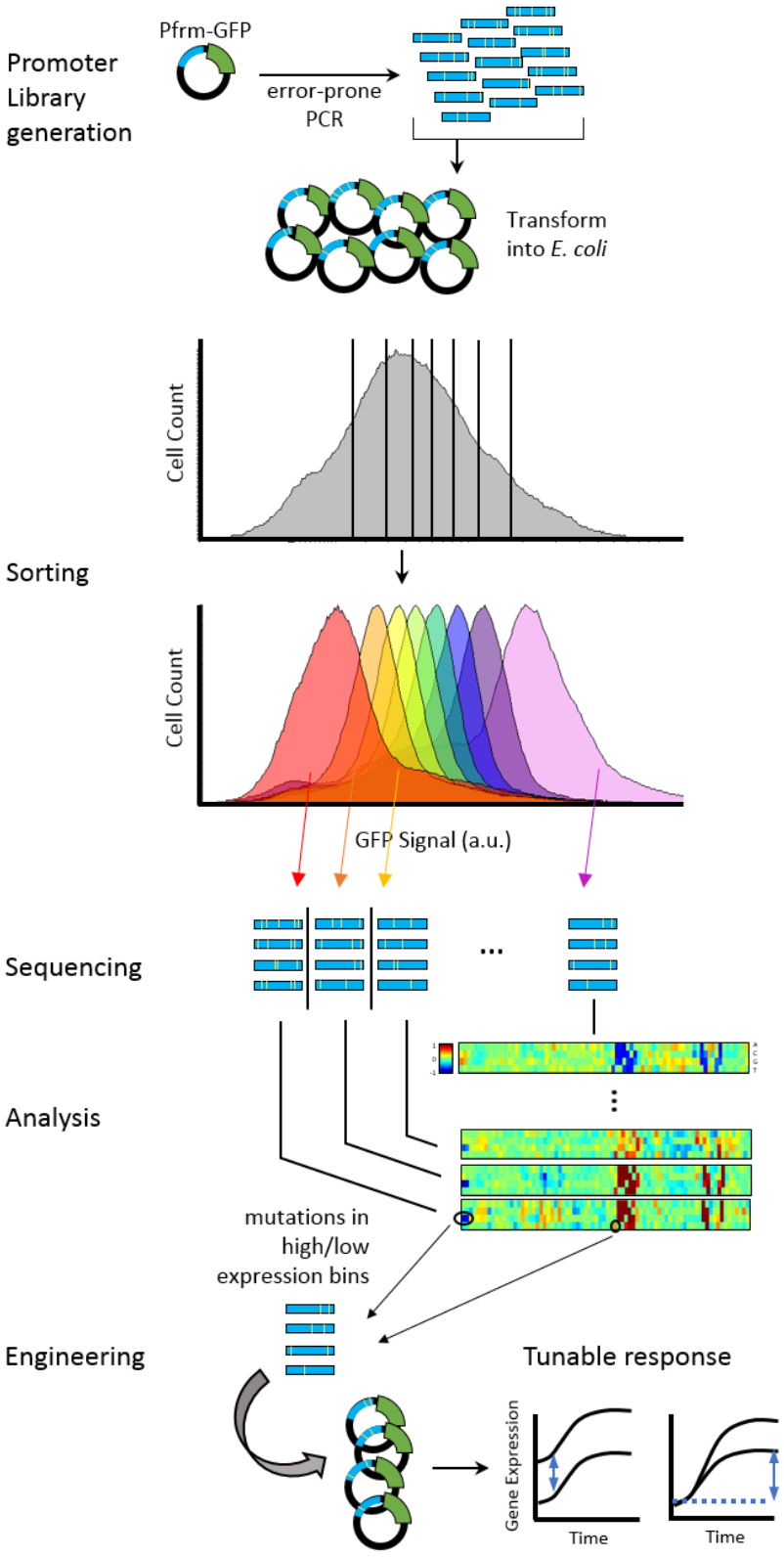
Promoter Library generation
Pfrm-GFP
error-prone PCR
Transform into E. coli
Sorting
Cell Count
Cell Count
GFP Signal (a.u.)
Sequencing
Analysis
mutations in high/low expression bins
Engineering
Tunable response
Gene Expression
Time
Time

Figure 2.1 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.

2.2 Materials and Methods

2.2.1 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.[75] 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.[76] The genes on the P_{frm}-Mdh-Hps-Phi plasmid were placed in an operon configuration under the *trc* promoter[77] and the RBS calculator v2.0[78, 79] was used to design synthetic ribosome binding sites for each gene. All strains and plasmids used can be found in Table A.1. PCR primers were synthesized by Integrated DNA Technologies (IDT, Coralville, IA) and can be found in Table A.2.

For flow cytometry analysis, cells were grown in MOPS minimal medium[80] 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 A.3). 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[81] 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.

2.2.2 Promoter library generation

Error-prone PCR targeting the 200 bp P_{frm} (Figure A.1) 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.

2.2.3 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 A.8). 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.

2.2.4 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 2.1. The marginal distributions $f(b_i)$ and $f(\mu)$, were calculated by summing the joint distribution along the appropriate dimension. The correction factor[66] in equation 2.1 was calculated as described[65] 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 at https://www.ncbi.nlm.nih.gov/bioproject/383844.

2.3 Results

2.3.1 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 2.2A-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 2.2A-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.[82, 83] The Hill equation (equation 2.2) is used to characterize the induction response and cooperativity of the system, particularly of interest here due to the tetrameric structure[84, 70] of FrmR.

$$P([\mathrm{I}]) = P_{min} + (P_{max} - P_{min})\frac{[\mathrm{I}]^n}{K^n + [\mathrm{I}]^n} \tag{2.2}$$

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 A.4). Without chromosomal FrmR expression the apparent Hill coefficient was 0.46 ± 0.02 (Table A.4). The response curve for the NEB5α strain without plasmid-expressed FrmR can be seen in Figure A.2. Disruption of negative autoregulation typically increases the Hill coefficient, leading to a tighter sigmoidal response curve.[85, 86] 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,[58] 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.

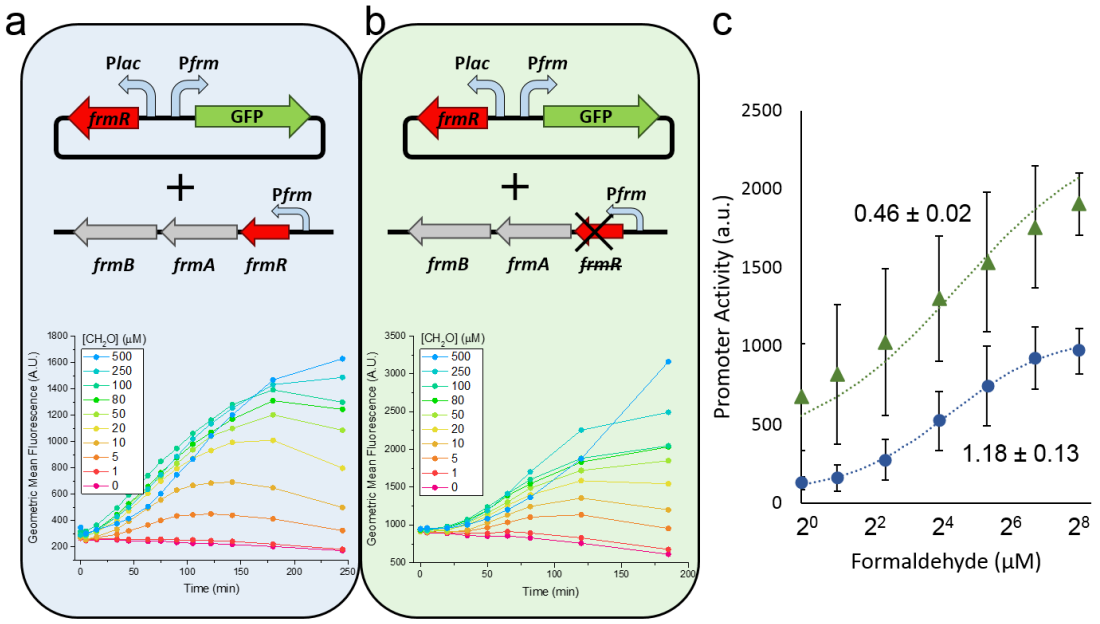

Figure 2.2 Regulatory mechanisms and response of the *E. coli* formaldehyde-inducible promoter. Two configurations were used to investigate regulation. A P_{frm}-GFP-P_{lac}-frmR reporter plasmid was used with (a) or without (b) without chromosomal expression of FrmR under control of P_{frm}. Representative 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.

2.3.2 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 2.3).[47] Operator sequences often contain inverted repeats which can form hairpin loops, an important structural

feature for protein-DNA interactions.[87] 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}.[84] A smaller incomplete 5'-**ATAGTATA/TATAGTAT**-3' inverted repeat was noted within the *Salmonella* promoter, disruption of which was shown to ablate FrmR binding.[69] 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 2.3). 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.

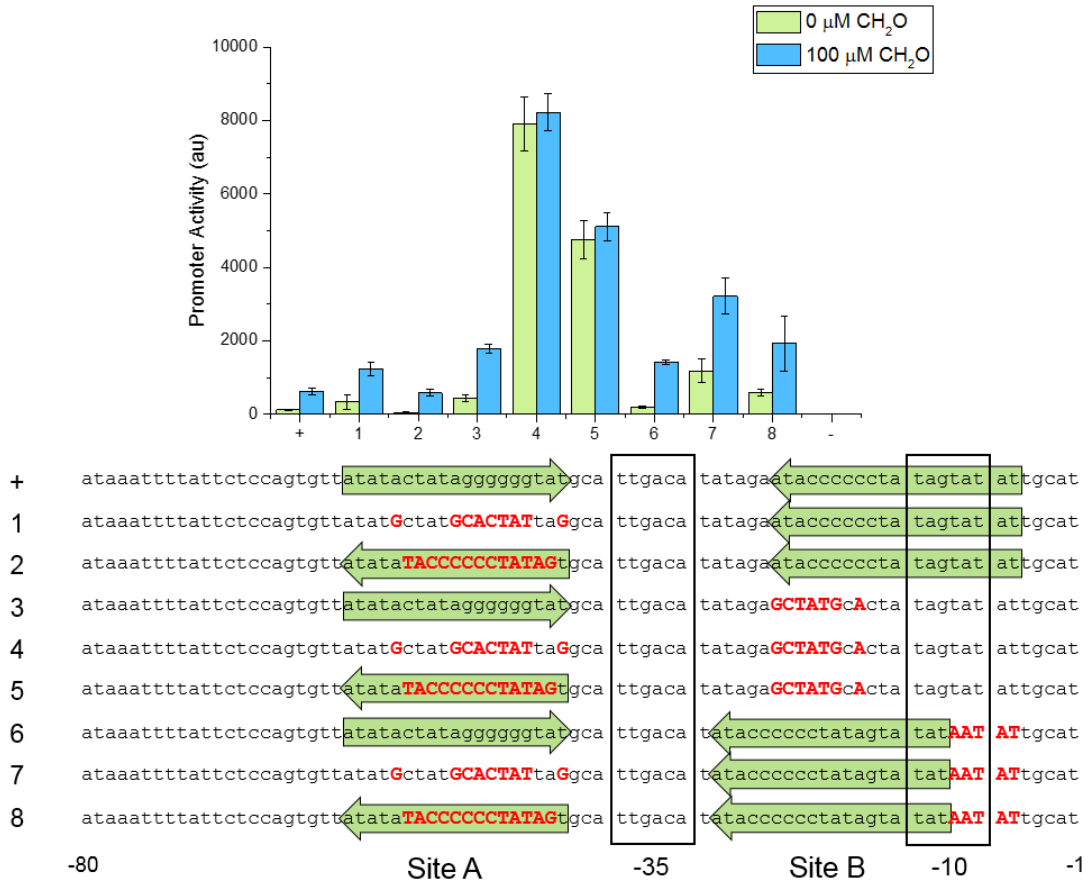

Figure 2.3 Response of P_{frm} binding site variants 3 hours after dosing with 0 or 100 µM formaldehyde. The 19-bp inverted repeat is shown with two green arrows facing one another to represent their complementary relationship. These two FrmR binding sites were analyzed by removing or reverse complementing 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.

2.3.3 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 A.2). 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 A.3) 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 A.5).[88] 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.[89] 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 A.4, Table A.6). The per position mutation frequency was 2.8% on average, and ranged from 1% to 8% (Figure A.5). 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.

2.3.4 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 2.4, Figure A.6). 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 2.4), and the third experiment involved FrmR expression from the plasmid only in the Δ*frmR* strain, presumably resulting in lower FrmR expression levels (Figure A.6). 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.[65] 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 2.5). 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 2.4). 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 2.4). 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 A.7). 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 A.7).

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[47], however is not supported by the low information content at those two positions. In agreement with the information footprints, the heat maps (Figure 2.5) 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.

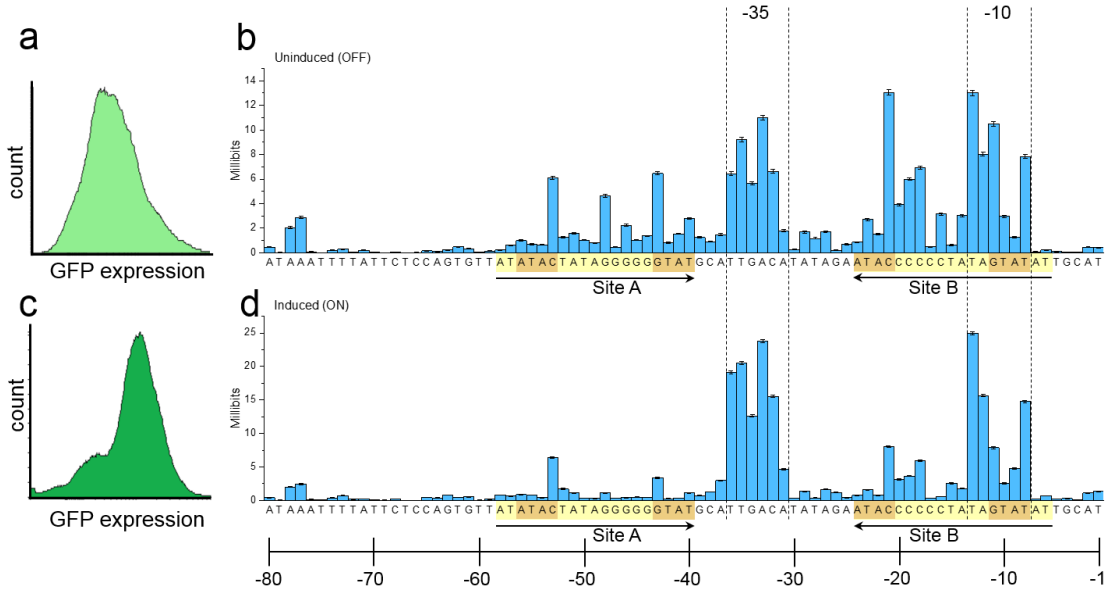

Figure 2.4 Information footprints of the *E. coli* P_{frm} with different levels of FrmR binding in the NEB5α strain. The distribution of GFP expression from 10,000 cells is shown in the uninduced (a) and induced (c) P_{frm} library. Information footprints for the uninduced (b) and induced with 100 μM 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.

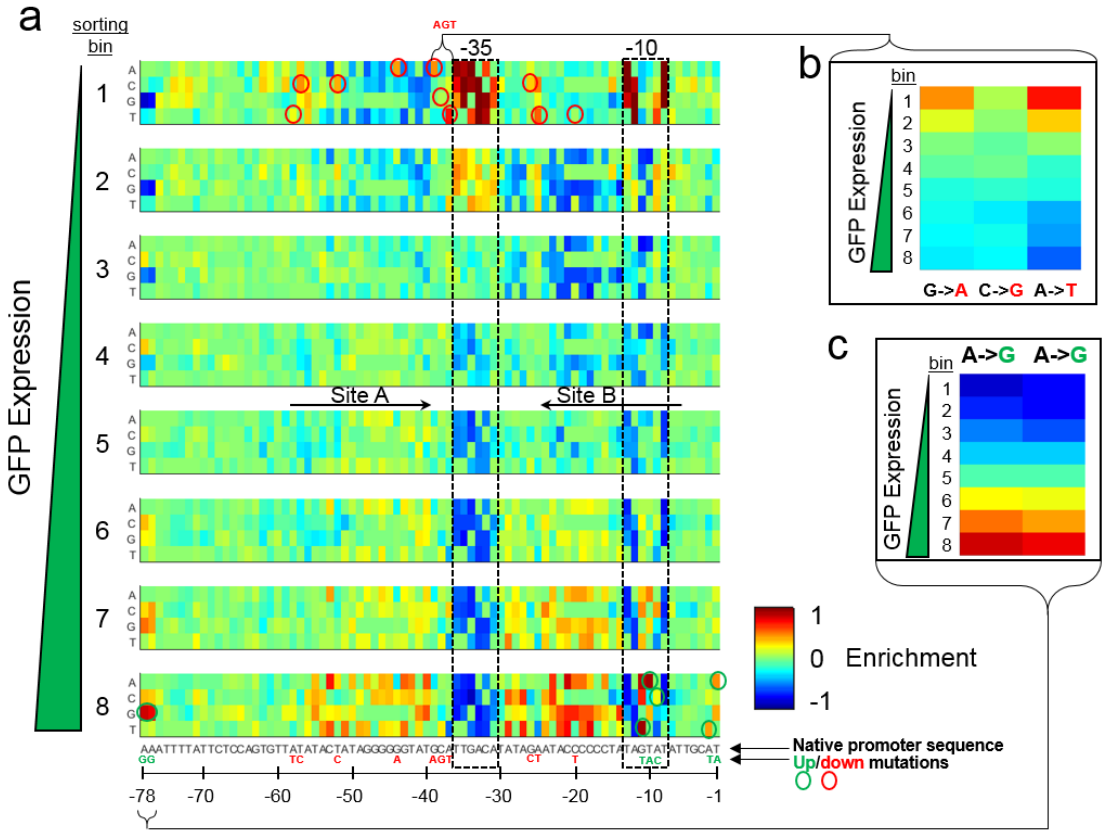

Figure 2.5 Identification of up and down mutations through sequence analysis. (a) Heat maps of the P_{frm} sequence in each of eight sorting bins in NEB5α strain in uninduced conditions. Mutated promoters isolated from low GFP expression bins are represented in the top heatmaps while those from high 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.

2.3.5 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 A.3). 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 2.6). 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 2.6) 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 2.6). 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.

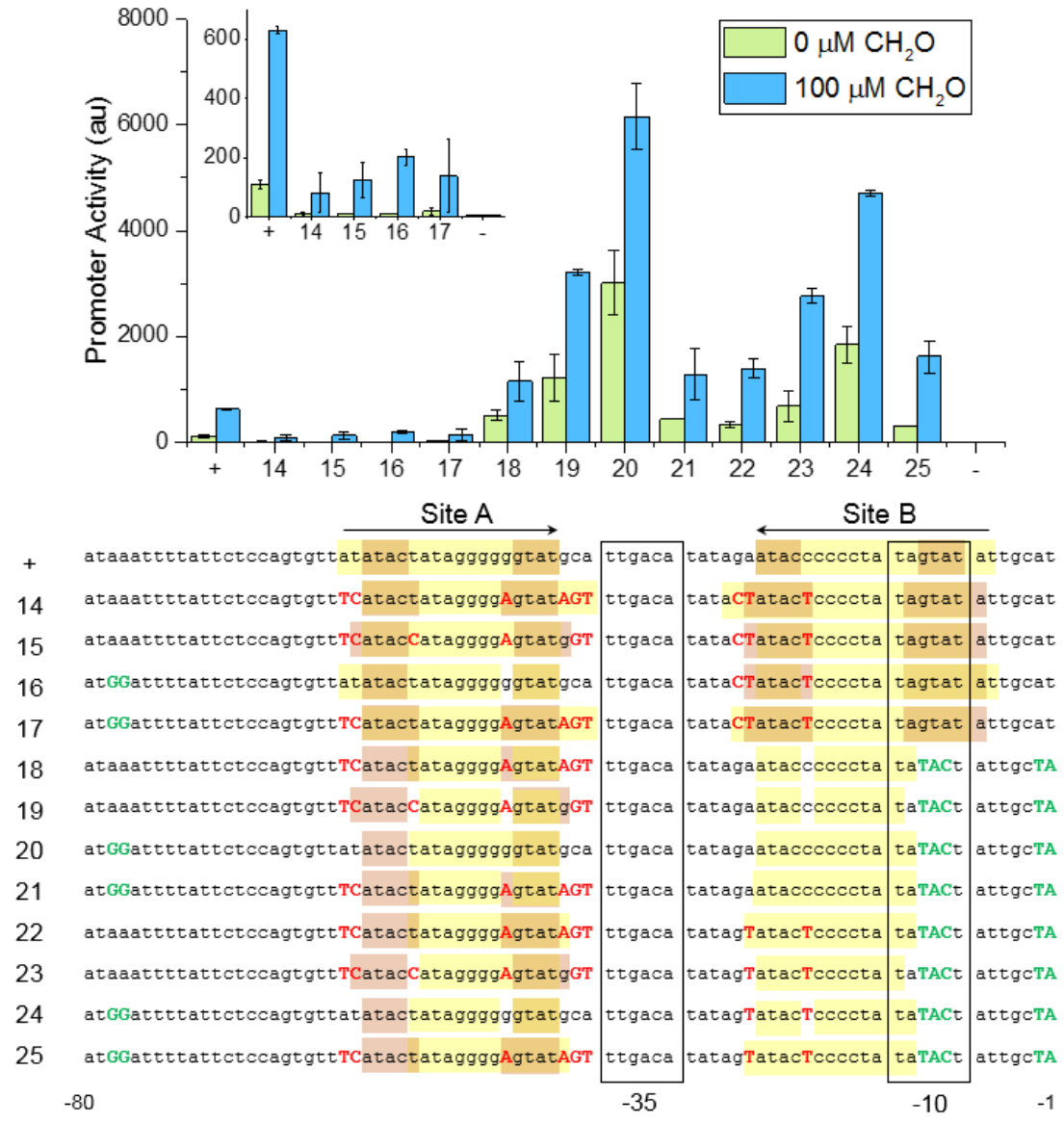

Figure 2.6 Response of specifically constructed P_{frm} variants 3 hours after dosing with 0 or 100 µM formaldehyde. Variants were constructed with mutations for higher (green) or lower (red) expression levels. The promoter was deleted to yield the negative control and the positive construct is the native P_{frm} sequence. Error bars represent standard deviation of two replicates tested on different days.

2.3.6 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,[23] 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.[23] 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 2.7A-B, D-E), demonstrating, for the first time, formaldehyde-induced synthetic methylotrophy. The yield on methanol, calculated as reported[23] 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 2.7C, 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.

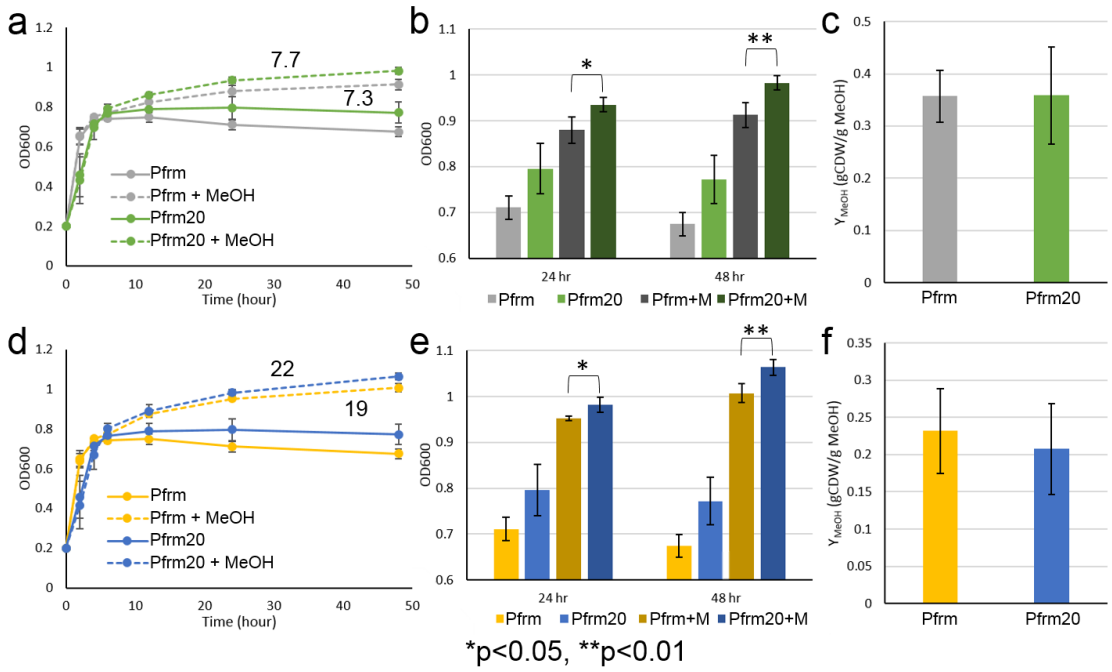

Figure 2.7 Growth and yield on methanol for P_{frm}-Mdh-Hps-Phi and P_{frm}20-Mdh-Hps-Phi plasmids in the Δ*frmA* Δ*pgi E. coli* strain. Strains were grown with M9 minimal media supplemented with 1 g/L yeast extract and with or without (a,b,c) 60 mM or (d,e,f) 240 mM methanol. (a,d) Growth curves normalized 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).

2.4 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.

Acknowledgements

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.

Chapter 3

IMPROVING SYNTHETIC METHYLOTROPHY VIA DYNAMIC FORMALDEHYDE REGULATION OF PENTOSE PHOSPHATE PATHWAY GENES AND REDOX PERTURBATION

3.1 Background

Synthetic methylotrophy, or the engineering of methanol utilization for carbon and energy needs in non-native methylotrophs, has been a highly sought-after goal in metabolic engineering in recent years. The increasing availability and decreasing price of methane from natural gas has renewed interest in utilizing methane, and its reduced product methanol, as a substrate for microbial fermentations.[5] Methanol has also been argued for as an attractive non-food substrate due to its higher degree of reduction compared to conventional sugar alternatives. Renewable methanol, or bio-methanol, sources are being explored by reacting carbon dioxide from geothermal steam, biogas, or municipal solid waste, with hydrogen produced via electrolysis.[6] Improving renewable methanol production is an active area of research, including a technoeconomic analysis coupling bio-methanol synthesis with electrolytic hydrogen production powered by wind energy,[10] investigating the incorporation of renewable energy into carbon recycling, and reducing the cost of electrolysis for hydrogen production.[8]

Native methylotrophs, which are capable of consuming single-carbon substrates such as methane and methanol to fulfill their carbon and energy needs, can include yeasts, bacteria, fungi, and archaea. Methanol utilization specifically can be found in bacteria and yeasts. Despite progress in the development of protocols for engineering native methylotrophic bacteria, the genetic toolbox remains relatively undeveloped and the number of potential products limited.[15] Methylotrophic bacteria such as *Bacillus methanolicus* and *Geobacillus stearothermophilus* are valuable resources for the implementation of methylotrophy in the platform organism *Escherichia coli* (*E. coli*).[23] Successfully engineering *E. coli* for methanol utilization would effectively allow for the methanol-driven production of dozens of industrially relevant products.

Attempts to create an *E. coli*-based synthetic methylotroph have achieved significant progress in the past 5 years.[90] Groups have demonstrated ^{13}C-labeled methanol incorporation in resting cells,[20] methanol incorporation into biomass using a yeast extract co-substrate,[23] and significantly improved methanol carbon utilization in combination with co-substrates through methanol auxotrophy strain engineering.[27 30] These advances in *E. coli*-based synthetic methylotrophy have relied on NAD-dependent methanol dehydrogenases for initial methanol oxidation and the ribulose monophosphate (RuMP) pathway for subsequent formaldehyde fixation. Expression of three genes encoding native methylotrophic enzymes are utilized to convert methanol to the central carbon intermediate fructose 6-phosphate. An NAD-dependent methanol dehydrogenase, Mdh, converts methanol to formaldehyde, which is fixed with ribulose 5-phosphate (Ru5P) via 3-hexulose-6-phosphate synthase, Hps, to form hexulose-6-

phosphate. Hexulose-6-phosphate is subsequently isomerized by 6-phospho-3-hexuloisomerase, Phi, to form fructose 6-phosphate (Figure 3.1A). Research on the Mdh, Hps, and Phi enzymes includes the investigation of enzymes from different methylotrophs,[20] enzyme evolution,[21] and protein fusion approaches.[91] However significant challenges need to be overcome for *E. coli* to grow on methanol as a sole carbon and energy source. We focus here on maximizing methanol carbon incorporation and the methanol growth benefit with three approaches: (1) implementing dynamic regulation machinery in response to formaldehyde, (2) expressing heterologous non-oxidative pentose phosphate pathway genes for the regeneration of Ru5P, and (3) strain engineering to increase reliance on methanol for biomass production.

Implementing dynamic regulation mechanisms in the cell should allow *E. coli* to effectively recognize methanol as a substrate. In this respect, the *E. coli* formaldehyde-inducible promoter, P_{frm}, is particularly useful. The P_{frm} promoter, found upstream of the formaldehyde-detoxification *frmRAB* operon in *E. coli*, has been found to respond with high specificity to formaldehyde, which disrupts the DNA binding of the transcriptional repressor FrmR (Figure 3.1B).[70] Using the P_{frm} promoter, the expression of genes involved in continued methanol and formaldehyde assimilation can be regulated directly in response to cell needs. Formaldehyde induction also prevents the accumulation of the toxic intermediate, which can lead to slowed growth or cell death.[92]

In addition to *mdh*, *hps*, and *phi*, heterologous non-oxidative pentose phosphate pathway (PPP) genes from the methylotroph *B. methanolicus* MGA3 have also been investigated for their ability to bypass or alleviate metabolic bottlenecks. The expression of five *B. methanolicus* genes: *rpe*, *tkt*, *fba*, *glpX*, and *pfk*, which were integrated into the chromosome of *E. coli* and constitutively expressed, were shown to significantly improve ^{13}C-labeled methanol incorporation in a range of intracellular metabolites and amino acids, including a 41% improvement in labeled glycine and a 17% improvement in labeled 3-phosphoglycerate (3PG).[25] We hypothesize that further investigation of these five genes and optimization of their expression can lead to further improvements in methanol utilization.

In an effort to generate gluconate- and methanol-essential growth in *E. coli*, flux balance calculations predicted that most cellular NADH was produced from the heterologous Mdh,[30] and it was thus hypothesized that lower TCA cycle activity, and therefore lower NADH production via the TCA cycle, would improve growth on methanol and gluconate. This hypothesis is supported by metabolic flux analysis on the Δ*maldh* strain during growth on glucose, which predicts significantly reduced TCA cycle activity leading to high acetate overflow production.[93] Native methylotrophs utilizing the RuMP pathway often have incomplete TCA cycles, including the obligate methylotroph *Methylobacillus flagellatus*, which is missing the α–ketoglutarate, malate, and succinate dehydrogenase enzymes.[48] In an attempt to mimic methylotrophs and re-calibrate the redox balance, Meyer et al. generated a knock-out of the NAD-dependent malate dehydrogenase, encoded by *maldh*, which

led to major improvements in growth with methanol and gluconate.[30] Manipulation of the redox balance to make the oxidation of methanol more thermodynamically favorable can increase methanol assimilation and its proportional contribution to cellular reducing equivalents.

3.2 Materials and Methods

3.2.1 Chemicals and reagents

Methanol and glacial acetic acid were purchased from Fisher Scientific (Hampton, NH). Ammonium acetate, acetylacetone, and 1-Bromo-3-chloropropane were purchased from Sigma-Aldrich (St. Louis, MO). ^{13}C-methanol (99% ^{13}C) was purchased from Cambridge Isotope Laboratories (Tewksbury, MA). *E. coli* NEB5α competent cells, Q5 High-Fidelity 2X Master Mix, Q5 Site-Directed Mutagenesis Kit, and NEBuilder HiFi DNA Assembly Master Mix were purchased from NEB (Ipswich, MA). Ambion DNA-*free* kit was purchased from Thermo Fisher Scientific (Waltham, MA), the iScript cDNA synthesis kit was purchased from Bio-Rad (Hercules, CA), and PerfeCTa SYBR Green SuperMix for IQ was purchased from QuantaBio (Beverly, MA).

3.2.2 Strains and plasmids

Strains and plasmids used in this study are listed in Table 1. Primers used for Q-RT-PCR and for the construction of plasmids are listed in Table B.1. *E. coli* NEB5α was used to propagate all plasmids during plasmid construction, and *E. coli* BW25113 was

used as the base strain for characterizing growth and methanol assimilation. The *E. coli* BW25113 Δ*frmA* strain was originally obtained from the Keio collection.[94] The removal of the kanamycin cassette from the *frmA* locus via pCP20 and subsequent deletion of *maldh* was performed as described.[95] Plasmids pRT and pFGP were constructed utilizing Gibson cloning methods and NEBuilder HiFi DNA Assembly Master Mix. The *rpe-tkt* and *fba-glpX-pfk* operons,[25] were cloned into pACM4[83] under the control of the 200 bp P_{frm} promoter[96] to create plasmids pRT and pFGP, respectively. Plasmids pRTF, pRTG, pRTP, were also constructed utilizing Gibson cloning methods and NEBuilder HiFi DNA Assembly Master Mix, and used plasmids pRT and pFGP as templates. Plasmids pR and pT were generated via the Q5 Site-Directed Mutagenesis Kit (NEB) using pRT as a template. Details of construction and primers used can be found in Table B.1. The core methanol assimilation plasmid pM with the *G. stearothermophilus mdh*, and *B. methanolicus hps* and *phi* genes, was constructed previously.[25]

Table 3.1 Strains and plasmids used in this study

Strain or plasmid	Relevant characteristics[a]	Source or reference
E. coli strains		
BW25113	*Δ(araD-araB)567 ΔlacZ4787(::rrnB-3) λ- rph-1 Δ(rhaD-rhaB)568 hsdR514*	Datsenko and Wanner, 2000
ΔA	BW25113 Δ*frmA::FRT*	Baba et al., 2006
ΔAΔm	BW25113 Δ*frmA::FRT* Δ*maldh::FRT*	This study

ΔAΔm evol.	See methods	This study
Plasmids		
pM	pETM6-P_{trc}-*Gs mdh -Bm hps -Bm phi*, Amp^R	Bennett et al., 2018
pRT	pAC-P_{frm}-*Bm rpe -Bm tkt*, Cm^R	This study
pFGP	pAC- P_{frm} -*Bm fba -Bm glpX -Bm pfk*, Cm^R	This study
pR	pAC- P_{frm} -*Bm rpe*, Cm^R	This study
pT	pAC- P_{frm} -*Bm tkt*, Cm^R	This study
pRTF	pAC- P_{frm} -*Bm rpe -Bm tkt -Bm fba*, Cm^R	This study
pRTG	pAC- P_{frm} -*Bm rpe -Bm tkt -Bm glpX*, Cm^R	This study
pRTP	pAC- P_{frm} -*Bm rpe -Bm tkt -Bm pfk*, Cm^R	This study

[a] Amp^R, ampicillin resistance gene; Cm^R, chloramphenicol resistance gene; *Bm*, coding sequence sourced from *Bacillus methanolicus*; *Gs*, coding sequence sourced from *Geobacillus stearothermophilus*; *mdh*, methanol dehydrogenase gene; *hps*, 3-hexulose-6-phosphate synthase gene; *phi*, 6-phospho-3-hexuloisomerase gene; *rpe*, ribulose-5-phosphate 3-epimerase gene; *tkt*, transketolase gene; *fba*, fructose-1,6-bisphosphate aldolase gene; *glpX*, fructose-1,6-bisphosphatase gene; *pfk*, phosphofructokinase gene.

3.2.3 Media and growth conditions

For ^{13}C-methanol labeling experiments, colonies were used to inoculate 30 mL of M9 minimal media supplemented with 0.5 g/L yeast extract and the appropriate antibiotics with and without 100 mM ^{13}C-methanol in 250 mL baffled flasks enclosed with tinfoil. Cultures were grown at 37°C at 250 rpm for two days, then centrifuged (4000 rpm, 10 min), washed with 500 μL of M9 minimal medium, and pellets were kept at -20°C until processing. Growth assays were also inoculated from colony and were

performed with either 0.5 or 1 g/L yeast extract with and without 100 mM methanol in 250 mL baffled flasks or 50 mL conical tubes. Cell growth was measured using a DU370 spectrophotometer from Beckman Coulter (Brea, CA) at 600 nm wavelength. Antibiotics were added at the following final concentrations when appropriate: carbenicillin at 100 μg/mL for maintenance of Amp^R plasmids, and chloramphenicol (dissolved in water) at 25 μg/mL for maintenance of Cm^R plasmids.

For the evolution of the ΔAΔm(pM) strain, two 1 L Eppendorf BioFlo 120 bioreactors (Eppendorf, Hamburg, Germany) were run in parallel with 0.5 L working volume. Colonies were used to inoculate 10 mL overnight cultures (37°C, 250 rpm) in M9 minimal medium with 1 g/L yeast extract. Overnight cultures were used to inoculate 500 mL of M9 minimal media with 0.5 g/L yeast extract and 100 mM methanol in reactors to an OD of 0.005. After 18 h, feeds with fresh media of the same composition were pumped into the reactors at a rate of 0.1 mL/min, while overflow was pumped to waste. Over the course of 70 days the concentration of yeast extract was gradually reduced from 0.5 to 0.1 g/L while maintaining the concentration of methanol at 100 mM.

3.2.4 Formaldehyde assay

Formaldehyde was assayed during growth through the colorimetric Nash assay.[97] Samples were collected during growth on 1 g/L yeast extract with 100 mM methanol through centrifugation (2 min, 13000 rpm) followed by the transfer of 400 μL supernatant to a microcentrifuge tube. Formaldehyde concentration was determined

through the addition of 800 µL Nash reagent (7.5 g ammonium acetate, 150 µL glacial acetic acid, and 100 µL acetylacetone brought to 50 mL with water) to each sample, incubation for 30 min at 37°C, and measurement using a DU370 spectrophotometer from Beckman Coulter (Brea, CA) at 412 nm wavelength.

3.2.5 RNA extraction and Q-RT-PCR analysis

Samples for RNA extraction were isolated from *E. coli* Δ*frmA*(pM)(pRT), and Δ*frmA*(pM)(pFGP) cultures which were inoculated from colonies and grown in 50 mL 1xM9 with 0.5 g/L yeast extract with and without 100 mM methanol in 250 mL baffled flasks for two days. Approximately 20 mL of each culture was centrifuged (3500 rpm, 30 min), and the pellets were resuspended in 1 mL cold Trizol, followed by the addition of 200 µL cold 1-Bromo-3-chloropropane. Samples were then vortexed and centrifuged at 4°C (13,000 rpm, 15 min), and 900 µL isopropanol was added to the collected supernatant for a 3 hour precipitation at -20°C. Precipitated RNA was pelleted through centrifugation (13,000 rpm, 30 min, 4°C), washed twice with 500 µL cold 75% ethanol, centrifuged again (13,000 rpm, 15 min), and the RNA was resuspended in 100 µL DEPC-treated water. A second precipitation was performed by adding 250 µL of 100% ethanol and 35 µL 3M sodium acetate in DEPC-treated water to samples, mixing through inverting, and precipitating overnight at -20°C. Similar to the first precipitation, RNA was pelleted through centrifugation (13,000 rpm, 30 min, 4°C), washed twice with 250 µL cold 80% ethanol, centrifuged

again (13,000 rpm, 15 min), and resuspended in 50 μL DEPC-treated water. Purified RNA was DNAse treated using the Ambion DNA-*free* kit, and cDNA was synthesized using the Bio-Rad iScript cDNA synthesis kit, both following manufacturer's protocols. The cDNA was then used for quantitative RT-PCR (Q-RT-PCR) using the Quantabio PerfeCTa SYBR Green SuperMix for IQ reagent and a Bio-Rad CFX96 Touch Real-Time PCR Detection System. The primers used to amplify *rpe*, *tkt*, *fba*, *glpX*, *pfk*, and the reference gene *ihfB* are listed in Table B.1. Experiments included two biological replicates and three technical replicates. Expression fold changes of the 100 mM methanol samples were normalized to the undosed (no methanol) samples using *ihfB* as the housekeeping gene.

3.2.6 ^{13}C labeling analysis

Intracellular metabolites and amino acids were extracted and analyzed as previously described[23] with minor modifications. Briefly, cells were pelleted and stored in microcentrifuge tubes at -20°C until extraction. For extraction, 1 mL of 70% (v/v) ethanol was heated to 70°C, added to each frozen cell pellet and vortexed for 30 s. The suspension was then heated at 95°C for 5 min, and cooled on ice for 5 min. Thereafter, the suspension was separated by centrifugation at 14,000 rpm for 5 min. The supernatant was transferred to a clean microcentrifuge tube and dried under nitrogen at 37°C. To derivatize the samples for GC-MS analysis, 50 μL of 2 wt% methoxyamine in pyridine solution was first added to the sample, and the sample was incubated at 37°C for 1 h. Thereafter, 50 μL of MTBSTFA+1% tBDMS was added to the sample

and incubated at 60°C for 30 min. The derivatized sample was then centrifuged for 5 min at 14,000 rpm, and 60 µL of the supernatant was transferred to a GC injection vial.

GC-MS analysis for intracellular metabolites was performed as described.[98] GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30m, 0.25 mm i.d., 0.25 µm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass spectrometer operating under ionization by electron impact (EI) at 70 eV. ^{13}C-labeling was determined from the measured mass isotopomer data. First, the mass isotopomer distributions were corrected for natural isotope abundance.[99] Next, average ^{13}C-labeling was determined as follows: Average ^{13}C-labeling = sum (Mi*i)/n, where n is the number of carbon atoms for the measured MS fragment less carbon atoms from derivatization reagent, Mi is the corrected mass isotopomer abundance, and i represents the number of ^{13}C-atoms making up that fragment.

3.3 Results

3.3.1 Select non-oxidative pentose phosphate pathway (PPP) genes expressed under the Pfrm promoter in the *E. coli* strain Δ*frmA* are strongly upregulated during growth with 100 mM methanol

Native methylotroph *Bacillus methanolicus MGA3* contains the large plasmid pBM19, which is necessary for the organism's growth on methanol.[100] Five genes of the non-

oxidative PPP are contained on this plasmid, and each has been shown to be 6- to 40-fold transcriptionally upregulated when *B. methanolicus* is grown on 200 mM methanol.[41] The *E. coli* formaldehyde-inducible promoter P_{frm} (Figure 3.1B) can achieve similar dynamic regulation of target genes in a synthetic methylotrophic strain.[96] Using a methylotrophic *E. coli* strain capable of utilizing methanol by expressing the minimal set of the three enzymes Mdh, Hps, and Phi (Figure 3.1A), we have previously shown that expression of these five PPP genes under the control of the P_{trc} promoter is capable of increasing methanol assimilation into intracellular metabolites by 12 and 41% in malate and glycine, respectively.[25] Here, we hypothesized that if these five PPP genes are expressed in *E. coli* from a methanol- or formaldehyde-responsive promoter, it would enable improved growth on and utilization of methanol. P_{frm} was placed upstream of the five codon-optimized *B. methanolicus* PPP genes capable of regenerating the ribulose 5-phosphate (Ru5P) necessary for formaldehyde fixation. The five genes were organized in two operon configurations, with each operon placed on a separate plasmid (Figure 3.1D). The first plasmid, pRT, bore genes *rpe* and *tkt*, encoding ribulose 5-phosphate 3-epimerase and transketolase, respectively. The second plasmid, pFGP, included *fba*, *glpX*, and *pfk*, encoding fructose 1,6-bisphosphate aldolase, fructose/sedoheptulose bisphosphatase, and phosphofructokinase, respectively. The Δ*frmA* strain with *mdh*, *hps*, and *phi*, will hereafter be referred to as strain ΔA(pM), and additional plasmids or strains will be referred to as listed in Table 1.

Formaldehyde is produced endogenously at low levels in all living cells due to routine cell functions such as demethylation reactions.[68] For both native and synthetic methylotrophs, when cells utilize methanol, formaldehyde, its oxidation product, accumulates in the cells and is frequently secreted in the medium. During growth of *E. coli* ΔA(pM) on 100 mM methanol with 1 g/L yeast extract as co-substrate to support protein synthesis, the formaldehyde concentration rises over the course of 24 hours to 25 μM, 5-fold higher than the 5 μM formaldehyde concentration measured under no methanol conditions (Figure 3.1C). Previous work has shown higher expression of GFP under P_{frm} control at concentrations as low as 5 μM in minimal growth media conditions.[96] Thus, we hypothesized that intracellular formaldehyde levels reached during growth with methanol would be sufficient to induce the expression of genes under P_{frm} control. To assess this hypothesis, we used quantitative RT-PCR (Q-RT-PCR) to confirm gene expression and measure expression differences with and without methanol in the medium. All five target gene transcripts in strains ΔA(pM)(pRT) and ΔA(pM)(pFGP) showed significantly higher expression during methanol growth (Figure 3.1D). When the two strains were grown in a medium with 100 mM methanol and 0.5 g/L yeast extract, the five genes were upregulated between 8- to 30-fold compared to control conditions (no methanol in the medium). These levels are similar to the 6- to 40-fold upregulation shown for the same genes in *B. methanolicus* during methanol growth.[41] These data show that the P_{frm} promoter is functional and induces the expression of the five genes in a formaldehyde-dependent fashion.

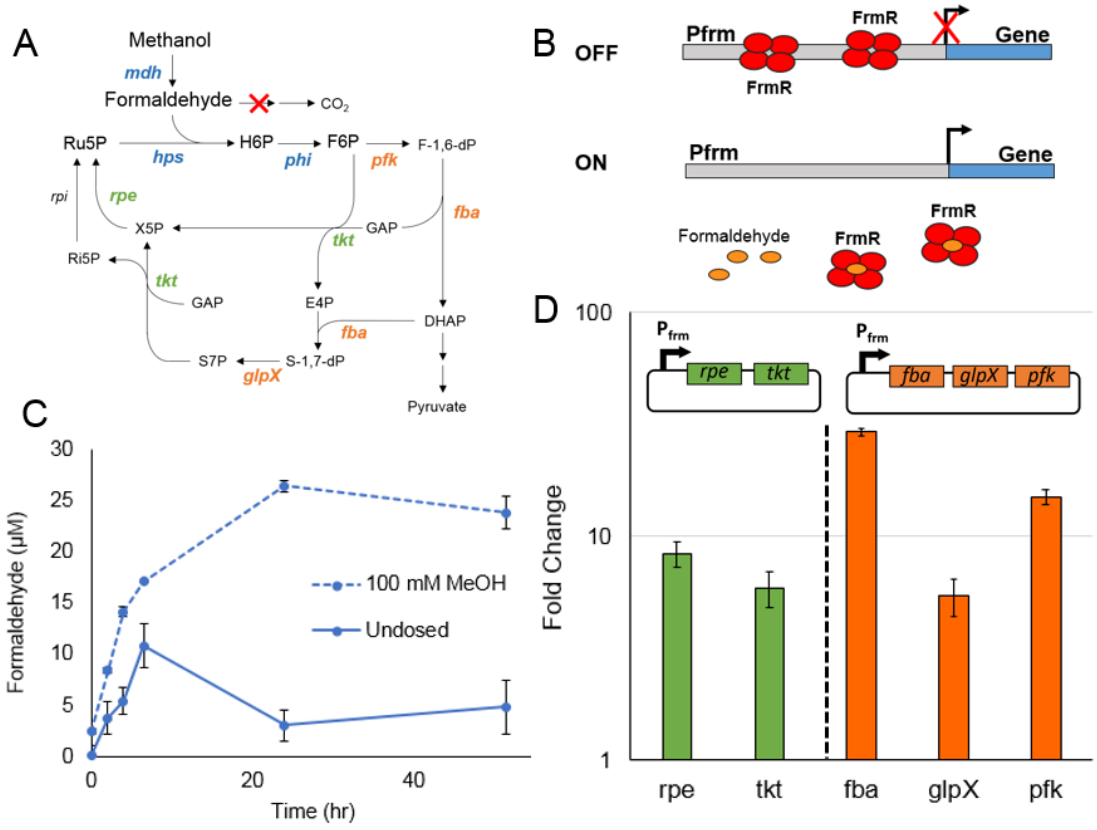
A
Methanol
mdh
Formaldehyde
CO2
Ru5P
hps
H6P
phi
F6P
pfk
F-1,6-dP
rpi
rpe
X5P
GAP
fba
tkt
Ri5P
tkt
GAP
E4P
DHAP
fba
S7P
S-1,7-dP
glpX
Pyruvate
B
OFF
Pfrm
FrmR
Gene
FrmR
ON
Pfrm
Gene
Formaldehyde
FrmR
FrmR
C
Formaldehyde (μM)
30
25
20
15
10
5
0
0
20
40
Time (hr)
100 mM MeOH
Undosed
D
Fold Change
100
10
1
Pfrm
rpe
tkt
Pfrm
fba
glpX
pfk
rpe
tkt
fba
glpX
pfk

Figure 3.1 Expression of genes under the control of the formaldehyde-responsive promoter (P_{frm}) in the methylotrophic *E. coli* strains ΔA(pM)(pRT) and ΔA(pM)(pFGP), in response to 100 mM methanol in a medium containing 0.5 g/L of yeast extract. (A) Methanol assimilation utilizing methanol dehydrogenase (encoded by *mdh*) for the conversion of methanol to formaldehyde, and the ribulose monophosphate (RuMP) pathway for fixation of formaldehyde to ribulose 5-phosphate (Ru5P) to form fructose 6-phosphate (F6P) via 3-hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase, encoded by *hps* and *phi*, respectively. Non-oxidative pentose phosphate genes *rpe*, *tkt*, *fba*, *glpX*, and *pfk*, are involved in regenerating Ru5P from F6P. H6P, hexulose 6-phosphate; F-1,6-dP, fructose 1,6-bisphosphate; GAP, glyceraldehyde 3-phosphate; DHAP, dihydroxyacetone phosphate; E4P, erythrose 4-phosphate; X5P, xylulose 5-phosphate; S-1,7-dP, sedoheptulose 1,7-bisphosphate; S7P, sedoheptulose 7-phosphate; Ri5P, ribose 5-phosphate. (B) Schematic depicting the mechanism by which the *E. coli* P_{frm} promoter achieves the repressed "off" and induced "on" states. (C) Formaldehyde accumulation during growth of the methylotrophic *E. coli* strain ΔA(pM) on 1 g/L yeast extract with and without 100 mM methanol. Error bars indicate standard deviation, n=2. (D) Fold change of mRNA levels in cultures with 0.5 g/L yeast extract and 100 mM methanol compared to the no methanol condition at 18 hours, determined using Q-RT-PCR data for genes placed under the P_{frm} promoter.

3.3.2 Co-expression of two PPP proteins, ribulose 5-phosphate 3-epimerase and transketolase, leads to higher biomass concentration due to methanol utilization, and higher intracellular labeling from ^{13}C-methanol

We then examined the effect of formaldehyde-regulated expression of the five *B. methanolicus* PPP genes on growth and methanol assimilation for the ΔA(pM)(pRT) and ΔA(pM)(pFGP) strains. While the integration of all five *B. methanolicus* PPP genes previously led to increased methanol assimilation,[25] stable episomal expression of the five genes simultaneously was difficult to achieve (unpublished work) necessitating the separation of the five genes into the two operons discussed

previously. The methanol growth benefit (MGB), as defined in equation 3.1, was used as one metric to assess the effectiveness of methanol utilization by the *E. coli* strains.

$$\%MGB = \frac{X^{+MeOH} - X^{-MeOH}}{X^{+MeOH}} * 100 \quad (3.1)$$

where X is the maximum biomass concentration reached. Plasmid expression of *B. methanolicus rpe* and *tkt* genes enhanced the methanol growth benefit with 100 mM methanol by 35%, from 46% in the ΔA(pM) control strain to 62% in the ΔA(pM)(pRT) strain (Figure 3.2A-B). When the ΔA(pM)(pRT) strain was grown on 100 mM methanol and 0.5 g/L yeast extract, it achieved a higher biomass concentration compared to both the control ΔA(pM) and the ΔA(pM)(pFGP) strains (Figure 3.2A). Expression of the *B. methanolicus rpe* and *tkt* genes also improved incorporation of ^{13}C-methanol into intracellular metabolites (Figure 3.2C). Intracellular glutamate, TCA cycle intermediate fumarate, and lower glycolytic intermediate 3-phosphoglycerate (3PG) for strain ΔA(pM)(pRT) displayed significantly higher methanol incorporation ranging from 10 to 35% higher than the ΔA(pM) control strain (Figure 3.2C). In contrast, expression of the *B. methanolicus fba*, *glpX*, and *pfk* genes in strain ΔA(pM)(pFGP) led to a lower growth benefit than the ΔA(pM) control and lower methanol incorporation into intracellular metabolites, including a significant 25% decrease in labeling for the lower glycolytic intermediates 3PG and phosphoenolpyruvate (PEP) (Figure 3.2C, Figure B.1B). While the metabolic

burden of expressing three heterologous enzymes may explain the lower biomass titer for the ΔA(pM)(pFGP) strain, the differences in labeling suggest these three genes (*fba*, *glpX*, and *pfk*) in combination, and as expressed here, are not capable of improving Ru5P regeneration.

Further testing showed that, individually, expression of the *B. methanolicus* *rpe* and *tkt* genes were also unable to generate the improved methanol growth and labeling results achieved by the combined expression of *rpe* and *tkt* (Figure 3.2D-E). Together, *rpe* and *tkt* form a pathway for the biosynthesis of Ru5P from fructose 6-phosphate, and a complete cycle in combination with the *hps* and *phi* genes for regenerating Ru5P upon formaldehyde incorporation (Figure 3.1A). *E. coli*'s native *rpe* and *tktA/tktB* genes are apparently unable to produce enough Ru5P for methylotrophic growth under these conditions, either due to enzymatic differences or issues related to regulation and expression. Adding *fba*, *glpX*, and *pfk* individually to the synthetic *rpe* and *tkt* operon also failed to achieve a strain with higher growth or labeling (Figure 3.2D-E). This is surprising, as *fba* and *glpX* encode enzymes which produce intermediates that the transketolase enzyme can utilize (Figure 3.1A). Overexpression of the native *E. coli glpX* encoded enzyme also previously led to 4-fold increased concentration of Ru5P.[26]

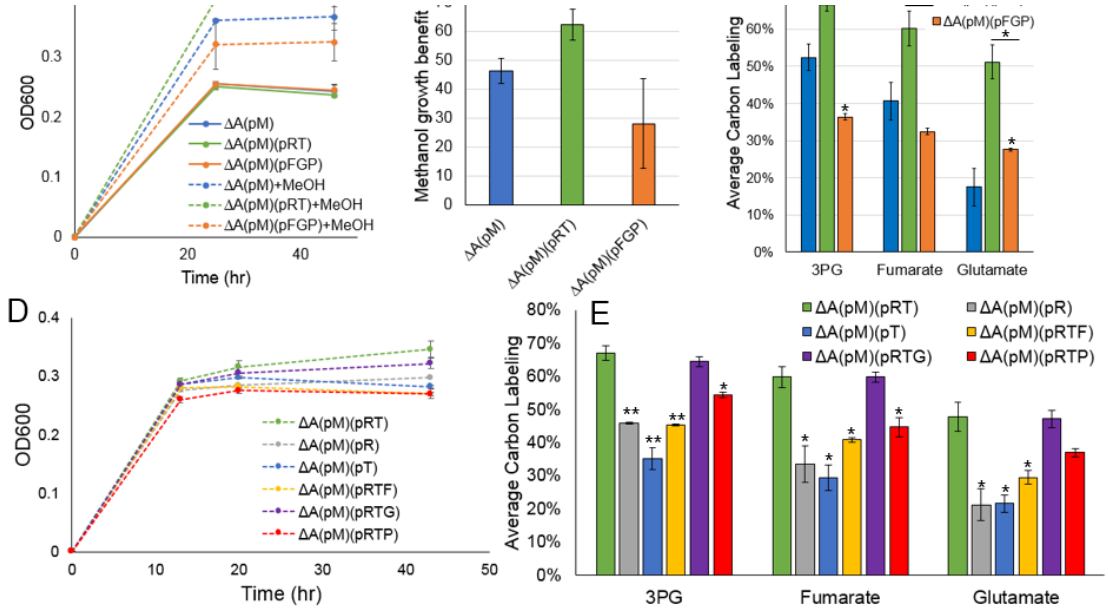

Figure 3.2 Growth profiles and methanol assimilation from ^{13}C-labeled methanol into intracellular metabolites for *E. coli* Δ*frmA* strains with episomal expression of combinations of *B. methanolicus* PPP genes *rpe*, *tkt*, *fba*, *glpX*, and *pfk*. (A) Growth profiles in 0.5 g/L yeast extract with (dashed lines) and without (solid lines) 100 mM methanol for ΔA(pM), ΔA(pM)(pRT), and ΔA(pM)(pFGP) *E. coli* strains. (B) Methanol growth benefit calculated from growth profiles in (A). (C) Methanol assimilation of 100 mM ^{13}C-methanol into intracellular metabolites 3-phosphoglycerate (3PG), fumarate, and glutamate in the ΔA(pM), ΔA(pM)(pRT), and ΔA(pM)(pFGP) strains after two days. Two-tailed *t*-tests indicate significance compared to the ΔA(pM) strain unless indicated otherwise. Growth curves with 100 mM ^{13}C-methanol and 0.5 g/L yeast extract (D), and labeling in 3PG, fumarate, and glutamate from ^{13}C methanol (E) after two days for *E. coli* strains ΔA(pM)(pRT), ΔA(pM)(pR), ΔA(pM)(pT), ΔA(pM)(pRTF), ΔA(pM)(pRTG), and ΔA(pM)(pRTP). Error bars indicate standard deviation, n=2. Two-tailed *t*-tests indicate significance compared to ΔA(pM)(pRT). * P<0.05, ** P<0.01.

3.3.3 Deletion of the malate dehydrogenase gene (*maldh*) enhances methanol growth benefit

Successfully achieving synthetic methylotrophy relies on the utilization of methanol to fulfill both carbon and energy requirements. Improvements in labeling indicate higher levels of methanol carbon being incorporated into central metabolism, however increasing the methanol growth benefit is also a desired goal as an indication of the methanol contribution to biomass accumulation. It has been previously hypothesized that manipulating the redox balance in the cell by decreasing the concentration of cellular NADH will make NAD-dependent methanol oxidation more thermodynamically favorable.[30] Deletion of the NAD-dependent malate dehydrogenase gene (*maldh*) effectively severs the TCA cycle in *E. coli*, theoretically reducing cellular NADH levels, and emulating native methylotrophs with incomplete or less active TCA cycles (Figure 3.3A).[50] To test this hypothesis, we deleted the *maldh* gene in the Δ*frmA* base strain with *mdh*, *hps*, and *phi*, denoted as ΔAΔm(pM), and characterized the strain for growth on methanol and intracellular labeling from ^{13}C-labeled methanol. Growth of the ΔAΔm(pM) strain was initially slower than the ΔA(pM) control strain on 1 g/L yeast extract both with and without methanol, a result which could be predicted from the diminished TCA cycle activity (Figure 3.3B). However, the methanol growth benefit with 100 mM methanol in the medium was over 3-fold higher than the ΔA(pM) control strain, indicating that methanol enables the formation of biomass to a much larger degree for the ΔAΔm(pM) strain under these conditions (Figure 3.3B). Despite this growth benefit, the methanol carbon being

assimilated into intracellular metabolites was significantly lower in the ΔAΔm(pM) strain compared to the ΔA(pM) control (Figure 3.3C). As expected, TCA cycle intermediates such as malate and fumarate were most severely affected, both with 80% lower average carbon labeling, from 50% in ΔA(pM) to 10% in ΔAΔm(pM) (Figure 3.3C). Decreased labeling was also found in the lower glycolytic intermediates, with a 30% reduction for 3PG and PEP, as well as a 40% decrease in alanine labeling and over 80% decrease in glutamate labeling (Figure 3.3C). While alanine can be produced from pyruvate, glutamate is synthesized from α-ketoglutarate within the TCA cycle, explaining the sharply reduced glutamate labeling impact of the *maldh* deletion. The lack of correlation between methanol growth benefit and methanol carbon assimilation indicates the biomass generated with the addition of methanol is not solely due to methanol carbon. We hypothesize the methanol growth benefit is largely due to the increased reliance on reducing equivalents provided through the generation of NADH during methanol oxidation.

In an attempt to improve methanol carbon assimilation in the ΔAΔm(pM) strain, adaptive laboratory evolution (ALE) experiments were performed starting with the parent ΔAΔm(pM) strain in two chemostats with 100 mM methanol and decreasing concentrations of yeast extract over the course of 70 days. This resulted in a mutant strain with a faster initial growth rate, but which maintained the large methanol growth benefit characteristic of the ΔAΔm(pM) strain (Figure 3.4A). Faster initial growth allowed the evolved strain to reach significantly higher biomass concentrations at 18 and 24 hours compared to the parent strain (Figure 3.4A). The

evolved strain also achieved significantly higher ^{13}C-labeling from methanol into intracellular metabolites compared to the parent ΔAΔm(pM) strain, particularly within TCA cycle intermediates (Figure 3.4B). These results may suggest some compensation within the evolved strain leading to higher TCA cycle fluxes while maintaining a large MGB. *E. coli* has another malate dehydrogenase, the malate:quinone-oxidoreductase Mqo, encoded by the gene *mqo*, which has been found to partially compensate for the conversion of malate to oxaloacetate when *maldh* is knocked out.[101] Mqo utilizes quinone rather than NAD as an election acceptor, and thus may improve the TCA cycle flux while maintaining the redox impact of the *maldh* gene knock-out.

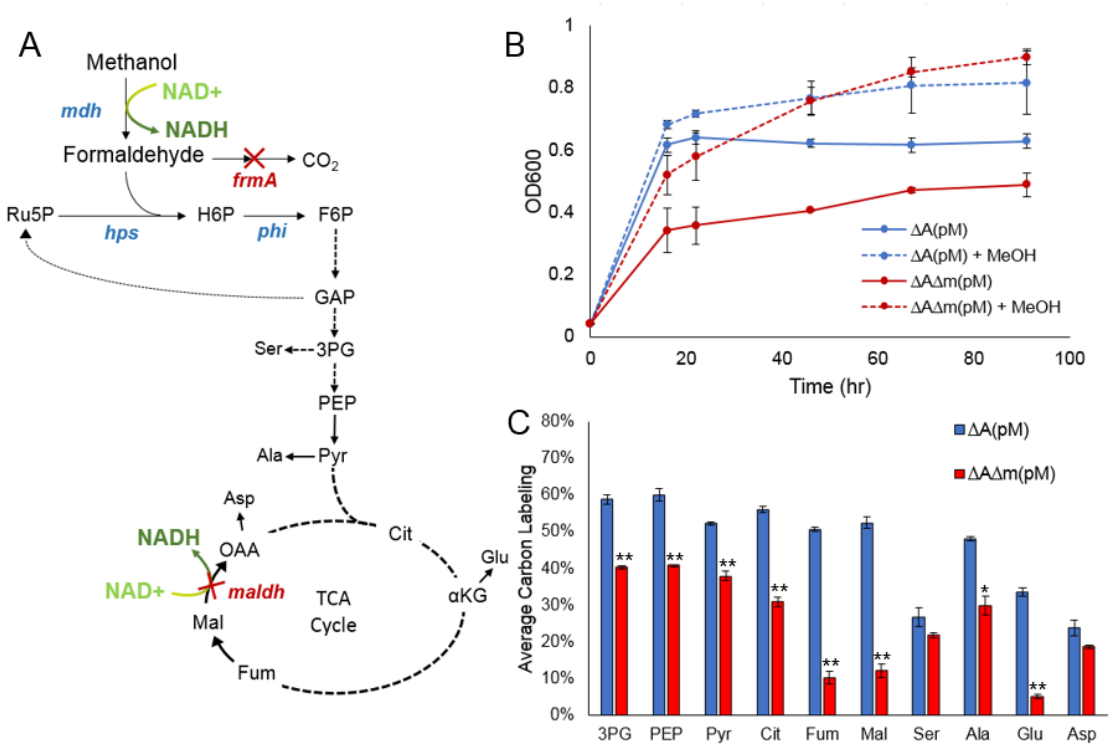

Figure 3.3 Deletion of the NAD-dependent malate dehydrogenase gene (*maldh*) results in a higher methanol growth benefit but lower assimilation from ^{13}C-methanol when grown in 100 mM ^{13}C-methanol and a small amount of yeast extract. (A) Methanol dehydrogenase (*mdh* gene) and malate dehydrogenase (*maldh* gene) are both NAD-dependent enzymes capable of affecting the cellular NAD/NADH balance. 3PG, 3-phosphoglycerate; PEP, phosphoenolpyruvate; Pyr, pyruvate; Cit, citrate; αKG, α-ketoglutarate; Fum, fumarate; Mal, malate; Ser, serine; Ala, alanine; Glu, glutamate; Asp, aspartate; OAA, oxaloacetate. (B) Growth curves of the ΔA(pM) and ΔAΔm(pM) strains in medium with 1 g/L yeast extract with (dashed lines) and without (solid lines) 100 mM methanol. (C) Average carbon labeling from ^{13}C-labeled methanol of intracellular metabolites after two days of growth in 0.5 g/L yeast extract and 100 mM ^{13}C-methanol for the ΔA(pM) and ΔAΔm(pM) strains. Error bars indicate standard deviation, n=2. Two-tailed *t*-tests indicate significance compared to ΔA(pM). * P<0.05, ** P<0.01.

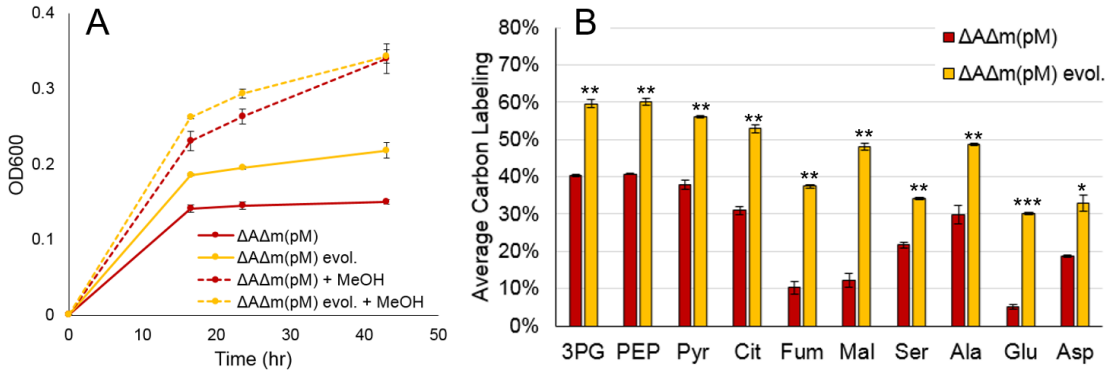

Figure 3.4 Evolution of the ΔAΔm(pM) *E. coli* strain results in faster biomass accumulation and higher ^{13}C-methanol assimilation to intracellular metabolites during growth on 100 mM ^{13}C-methanol and 0.5 g/L yeast extract. (A) Growth profiles for the parent and evolved ΔAΔm(pM) strains in 0.5 g/L yeast extract with (dashed lines) and without (solid lines) 100 mM ^{13}C-methanol. (B) Average carbon labeling of intracellular metabolites from ^{13}C-methanol for the parent and evolved ΔAΔm(pM) strains. Error bars indicate standard deviation, n=2. Two-tailed *t*-tests indicate significance compared to the parent ΔAΔm(pM) strain. * P<0.05, ** P<0.01, *** P<0.001.

3.3.4 Highest methanol growth benefit and ^{13}C-methanol incorporation in intracellular metabolites achieved in the Δ*frmA* Δ*maldh* strain upon formaldehyde-responsive expression of the *B. methanolicus rpe* and *tkt* genes

Plasmid-borne expression of the *B. methanolicus rpe* and *tkt* significantly increased methanol carbon incorporation into intracellular metabolites in the ΔA(pM)(pRT) strain, while deletion of the NAD-dependent malate dehydrogenase gene *maldh* led to

an improved methanol growth benefit with yeast extract as a co-substrate. To examine if combining the two strategies would lead to an improved methanol-utilization phenotype, we expressed *rpe* and *tkt* under formaldehyde control in the ΔAΔm(pM) strain, creating the ΔAΔm(pM)(pRT) strain (Figure 3.5A). Growth curves for the ΔAΔm(pM)(pRT) strain with 100 mM methanol and 1 g/L yeast extract supplementation indicate it reaches higher biomass concentrations than both the parent and evolved ΔAΔm(pM) strains (Figure 3.5B). To probe the effects of formaldehyde-regulated *rpe* and *tkt* expression in the Δ*frmA* Δ*maldh* strain, a short-term resting cell ^{13}C-methanol labeling assay was utilized. Cells were grown on 1 g/L yeast extract and 100 mM methanol, and at 21 and 44 hours post-inoculation, half the culture was centrifuged, washed, and resuspended in M9 minimal media with 500 mM ^{13}C-methanol. Cells were then incubated at 37°C for an hour to allow the ^{13}C-methanol to be incorporated into intracellular metabolites based on the proteome and availability of co-factors and intermediates at the time of sampling. Analysis of average carbon labeling after one hour provides valuable information regarding the methanol assimilation potential of the cell. The ΔAΔm(pM)(pRT) strain exhibited dramatically higher methanol assimilation after 21 hours of methanol growth, including 2-fold higher labeling in glycolytic intermediate 3PG, TCA cycle intermediates citrate, fumarate, and malate, and over 3- and 5-fold higher labeling in glutamate and aspartate, respectively, compared to the ΔAΔm(pM) evol. strain (Figure 3.5C). The ΔAΔm(pM) evol. strain showed moderate improvements over the ΔAΔm(pM) parent strain after 21 hours of methanol growth, including a 59% labeling improvement in

3PG, 79% improvement in both fumarate and malate, and an over 2-fold improvement in glutamate labeling (Figure 3.5C). After 44 hours of growth, the ΔAΔm(pM)(pRT) strain shows slightly lower labeling, while both the ΔAΔm(pM) and ΔAΔm(pM) evol. strains show higher labeling compared to day one, particularly in the TCA cycle intermediates citrate, fumarate, and malate (Figure 3.5D). Based on these data, the ΔAΔm(pM)(pRT) strain is more capable of assimilating methanol carbon at earlier timepoints compared to the parent and evolved ΔAΔm(pM) strains. Similar trends were also observed when collecting samples directly after two days of growth in 100 mM ^{13}C-methanol and 0.5 g/L yeast extract (Figure B.2).

The ΔAΔm(pM)(pRT) strain also compares favorably to the ΔA(pM)(pRT) strain. When grown on 100 mM methanol with 0.5 g/L yeast extract, the ΔAΔm(pM)(pRT) strain maintained an elevated MGB of 103%, which is significantly higher than both the 46% growth benefit of the ΔA(pM) and 62% growth benefit of the ΔA(pM)(pRT) strains (Figure 3.6A). The ΔAΔm(pM)(pRT) strain also displayed similar or higher levels of ^{13}C-labeled methanol carbon incorporation in intracellular metabolites as the ΔA(pM)(pRT) strain, which is particularly surprising considering the extremely low labeling patterns in the ΔAΔm(pM) base strain (Figure 3.6B, 3.3C). The *B. methanolicus* ribulose 5-phosphate 3-epimerase and transketolase enzymes improved labeling in the ΔAΔm(pM)(pRT) strain from 40% to 75% in the lower glycolytic metabolites 3PG and PEP compared to the ΔAΔm(pM) control (Figure 3.6B, 3.3C). The formaldehyde-responsive expression of the *rpe* and *tkt* genes was also explored in the evolved ΔAΔm(pM) strain, however it exhibited a similar

phenotype to the ΔAΔm(pM)(pRT) strain for both methanol growth benefit and labeling in intracellular metabolites in 0.5 g/L yeast extract and 100 mM methanol (Figure B.3).

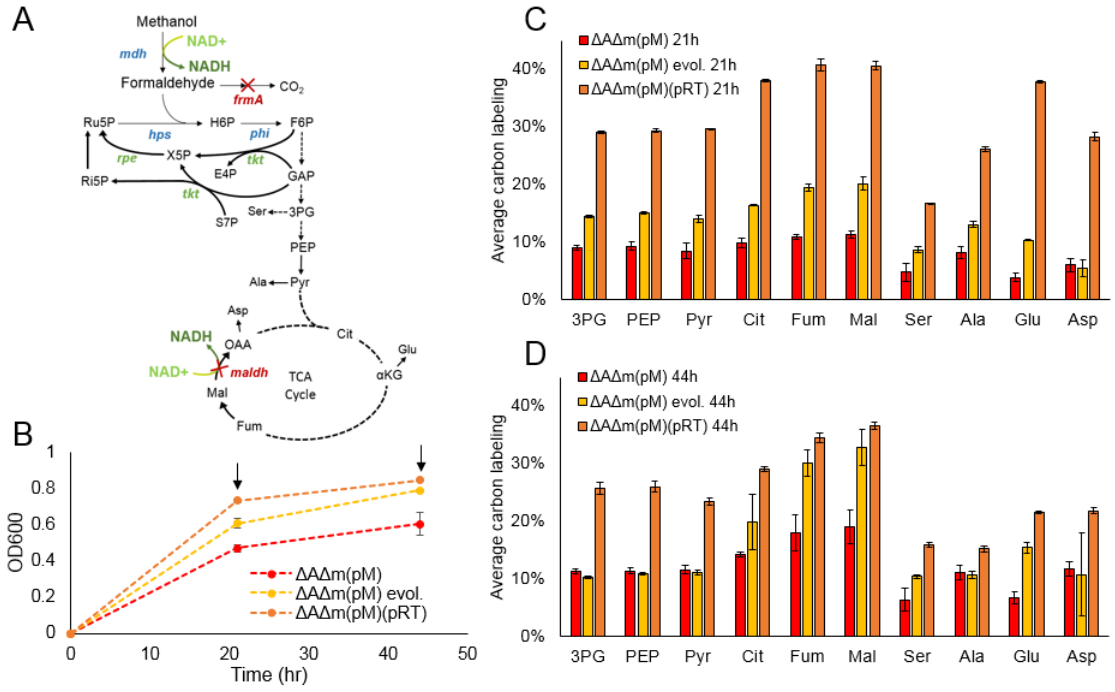

Figure 3.5 Episomal expression of *B. methanolicus rpe* and *tkt* in the ΔAΔm(pM) strain improves growth on 100 mM methanol and 1 g/L yeast extract, and assimilates higher levels of 500 mM ^{13}C-labeled methanol into intracellular metabolites over the course of an hour. (A) The activation of the non-oxidative PPP for the regeneration of Ru5P with *B. methanolicus rpe* and *tkt* is shown for the Δ*frmA* Δ*maldh* strain. Abbreviations are as previously described for Figure 3.3A. (B) Growth of the ΔAΔm(pM), ΔAΔm(pM) evol., and ΔAΔm(pM)(pRT) strains with 1 g/L yeast extract and 100 mM methanol. Arrows indicate sampling points for centrifugation and immediate resuspension in M9 minimal media and 500 mM ^{13}C-methanol. After one hour, pellets were collected and processed to determine ^{13}C-methanol assimilation into intracellular metabolites for the 21 h (C) and 44 h (D) samples. Abbreviations are as previously described for Figure 3.3A. Error bars represent standard deviation, n=2.

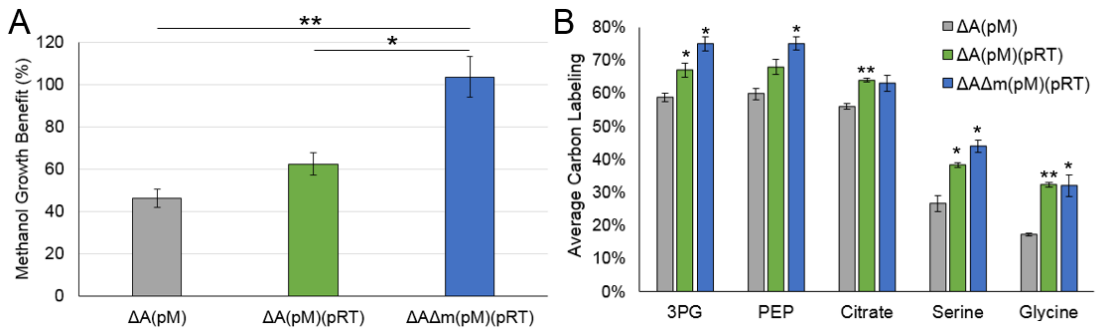

Figure 3.6 The ΔAΔm(pM)(pRT) strain achieves a higher methanol growth benefit (A) and similar levels of average carbon labeling from ^{13}C-methanol into intracellular metabolites (B) after two days of growth, compared to the ΔA(pM) and ΔA(pM)(pRT) strains during growth on 0.5 g/L yeast extract and 100 mM ^{13}C-methanol. Error bars indicate standard deviation, n=2. Two-tailed *t*-tests indicate significance compared to ΔA(pM). * P<0.05, ** P<0.01.

3.4 Conclusions

In this study we have applied multiple approaches to improve the utilization of both carbon and energy from methanol under yeast extract growth conditions in *E. coli*. These rational approaches are largely motivated by native methylotrophic bacteria such as *B. methanolicus*. While *E. coli* has no native mechanisms to recognize methanol as a substrate or to metabolize it, *B. methanolicus* upregulates genes involved in methanol metabolism during methanol growth.[41] We and others have attempted to replicate this ability in *E. coli* through the utilization of a formaldehyde-inducible promoter (P_{frm}). Here we demonstrated 8- to 30-fold upregulation of genes during growth on methanol and yeast extract. The use of similar dynamic regulatory approaches have been successful in the past. The production of fatty acids was improved in *E. coli* with the use of a synthetic malonyl-CoA switch using the

transcriptional regulator FapR,[102] and a semi-synthetic regulon was used to improve the growth of yeast on xylose compared to a constitutive approach.[103] Regulation approaches are particularly suited to engineering non-native substrate utilization, as they attempt to minimize the metabolic burden on the cell, and express heterologous protein only when necessary.[104]

B. methanolicus MGA3 requires a 19 kb plasmid which harbors the pentose phosphate pathway (PPP) genes *rpe*, *tkt*, *fba*, *glpX*, and *pfk*, to grow on methanol.[100] Previous work demonstrated that chromosomal integration of these five genes in *E. coli* resulted in improved methanol assimilation into intracellular metabolites, likely due to the increased regeneration of ribulose 5-phosphate required for formaldehyde fixation.[25] We expressed various combinations of the five PPP genes episomally under formaldehyde control, and achieved significantly improved ^{13}C-methanol assimilation during growth with 0.5 g/L yeast extract with the combination of *rpe* and *tkt* from *B. methanolicus*, including a 27% improvement in 3PG average carbon labeling.

Motivating our final approach, native obligate methanotrophs utilizing the RuMP pathway often have incomplete TCA cycles, such as lacking enzymes that include malate synthase, isocitrate lyase, and α–ketoglutarate dehydrogenase enzymes.[105] Through genome sequencing[48] and proteomic studies,[49] the obligate methylotroph *Methylobacillus flagellatus* was found to be missing α–ketoglutarate, malate, and succinate dehydrogenases. We deleted the *E. coli* NAD-dependent malate dehydrogenase gene (*maldh*) to perturb the cellular redox balance to favor methanol oxidation through the NAD-dependent methanol dehydrogenase. The methylotrophic

Δ*frmA* Δ*maldh* strain demonstrated a 3-fold higher methanol growth benefit when grown on 1 g/L yeast extract with 100 mM methanol compared to the Δ*frmA* control strain, suggesting the ability of the NADH produced through methanol oxidation to enable significant biomass formation. Adaptive laboratory evolution and the episomal expression of *B. methanolicus rpe* and *tkt* significantly improved methanol carbon utilization in the Δ*frmA* Δ*maldh* strain. A short-term ^{13}C-methanol assay was utilized to provide insight into the metabolic state of cells during growth with yeast extract and methanol, to examine the potential for cells to metabolize methanol based on enzyme levels and cofactor availability. The formaldehyde-controlled expression of *rpe* and *tkt* in the Δ*frmA* Δ*maldh* strain demonstrated dramatically higher ^{13}C-methanol incorporation into intracellular metabolites over the course of an hour compared to the parent and evolved ΔAΔm(pM) strains, suggesting ample availability of enzymes involved in continuous methanol assimilation. Similar assays can be used to assess whole cell pathway capability in the context of substrate utilization or production of metabolites. Ultimately the combination of formaldehyde-regulated PPP genes and the disruption of the NAD^+/NADH balance via the deletion of *maldh* resulted in high ^{13}C-methanol incorporation into intracellular metabolites as well as a 103% methanol growth benefit on 0.5 g/L yeast extract co-substrate.

Acknowledgements

This work was supported by the Advanced Research Projects Agency-Energy (ARPA-E) Reducing Emissions using Methanotrophic Organisms for Transportation Energy

(REMOTE) program (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 R01GM085232.

Chapter 4

ENGINEERING THE *Geobacillus stearothermophilus* METHANOL DEHYDROGENASE FOR HIGHER ACTIVITY AND SELECTIVITY TOWARDS METHANOL

4.1 Background

The desire to use of single-carbon feedstocks such as methane and methanol as substrates for microbial fermentations has increased in recent years due to an increased availability of cheap natural gas, composed primarily of methane. Microbial fermentations are used to produce a wide range of products, including alcohols, value-added chemicals, biofuels, amino acids, vitamins, and recombinant proteins. Engineering synthetic methylotrophy in a model organism such as *Escherichia coli* would allow for the replacement or supplementation of traditional sugar substrates with methanol, creating a drop-in pathway for producing a wide array of products from methanol carbon and reducing equivalents.

Implementing methanol assimilation in a non-native organism for synthetic methylotrophy relies first on an enzyme capable of converting methanol to formaldehyde. Native methylotrophs have developed three major classes of methanol oxidation enzymes, distinguished by their electron acceptor. Methylotrophic yeasts utilize alcohol oxidases (AOXs) which use oxygen as an electron acceptor, converting methanol and O_2 to formaldehyde and hydrogen peroxide.[19] Methylotrophic bacteria

utilize either NAD^+ or pyrroloquinoline quinone (PQQ) dependent methanol dehydrogenases (Mdhs). *E. coli* natively uses NAD^+ as a cofactor but is unable to produce PQQ without the heterologous expression of multi-gene clusters. In addition to the PQQ biosynthetic pathway, the expression of electron acceptor cytochrome c_L and cytochrome c oxidase are also required for the utilization of a PQQ-dependent Mdh in *E. coli*.[5] The NAD-dependent methanol oxidation reaction is less favorable at physiological conditions compared to AOX and PQQ-dependent Mdh catalyzed reactions.[5] However, with all factors considered, NAD-dependent Mdhs have still emerged as a clear choice for implementing synthetic methylotrophy in *E. coli*.

Müller et al. published the first attempt at identifying the most suitable NAD-dependent Mdh for use in *E. coli* by testing the *in vivo* activities of Mdhs sourced from native methylotrophs *Bacillus methanolicus*, *B. coagulans*, *Desulfitobacterium hafniense*, *Lysinibacillus fusiformis*, *L. sphaericus*, and *Desulfotomaculum kuznetsovii*.[20] Based on *in vitro* and *in vivo* studies, the Mdh2 from *B. methanolicus* MGA3 was determined to have the highest activity when the Mdhs were tested alone.[20] The first attempt at engineering an Mdh for synthetic methylotrophy targeted the NAD-dependent Mdh2 from *Cupriavidus necator* N-1.[21] *C. necator* is a mesophilic organism, which offers putative benefits for heterologous enzyme activity in mesophilic *E. coli*, however it is not a methylotroph. Through directed evolution, a variant denoted CT4-1, reportedly achieved a K_m for methanol of 22 mM, compared to 132 mM for the wild-type Mdh2.[21] The *B. methanolicus* Mdh2 was also engineered using a phage-assisted evolution approach, and achieved a 48% lower K_m for

methanol, decreasing from 636 to 329 mM in a variant with three amino acid substitutions.[22]

The Mdh from *Geobacillus stearothermophilus*, reported to have a K_m for methanol of 20 mM,[24] was also used for synthetic methylotrophic applications, after *in vivo* assays determined it was superior to the *B. methanolicus* Mdh2.[23] Here, we focus on engineering the *G. stearothermophilus* Mdh to further improve its *in vivo* activity and kinetic parameters towards methanol oxidation.

4.2 Materials and Methods

4.2.1 Chemicals and reagents

Methanol and glacial acetic acid were purchased from Fisher Scientific (Hampton, NH). Ammonium acetate and acetylacetone were purchased from Sigma-Aldrich (St. Louis, MO). ^{13}C-methanol (99% ^{13}C) was purchased from Cambridge Isotope Laboratories (Tewksbury, MA). *E. coli* NEB5α competent cells, Q5 High-Fidelity 2X Master Mix, Q5 Site-Directed Mutagenesis Kit, and NEBuilder HiFi DNA Assembly Master Mix were purchased from NEB (Ipswich, MA).

4.2.2 Strains and plasmids

Strains and plasmids used in this study are listed in Table C.1. Primers used for the construction of plasmids are listed in Table C.2. *E. coli* strain NEB5α (New England Biolabs (NEB), Ipswich, MA) was used for plasmid cloning and maintenance. The Δ*frmA* strain was used for growth, labeling, and flow cytometry experiments, and the

Δ*frmA* Δ*pgi* strain was used for butanol production experiments. Plasmids were constructed using Gibson cloning methods via NEBuilder HiFi DNA Assembly Master Mix, or the Q5 Site-Directed Mutagenesis Kit (NEB). Details of construction can be found in Table C.2.

4.2.3 Mdh library construction

Error-prone PCR targeting the coding region of the *G. stearothermophilus* Mdh was performed using GeneMorph II Random Mutagenesis Kit (Agilent, Santa Clara, CA). The resulting Mdh library was purified via the QIAquick Gel Extraction Kit (QIAGEN, Germantown, MD). The plasmid backbone was amplified and purified with a PCR purification kit (QIAGEN). The Mdh insert library with unmutated overhang sequences was cloned back into the plasmid backbone using the NEBuilder HiFi DNA Assembly Master Mix (NEB). A 20 µL reaction was transformed into a total of 13 aliquots of 30 µL Δ*frmA*(pAC_P_{frm}_GFP) electrocompetent cells, each transformation immediately recovered with 700 µL SOC media. Recovered aliquots were combined in a 250 mL baffled flask, brought up to 15 mL volume with LB and antibiotics, and grown overnight at 30°C. The combined library was stored frozen at -80°C in 15% v/v glycerol.

4.2.4 Flow cytometry and sorting

For flow cytometry analysis, the Δ*frmA* strain with pAC_P_{frm}_GFP and parent pETM6_P_{tac}_Mdh plasmids were grown overnight in LB media with appropriate

antibiotics. New cultures with and without 100 mM methanol in M9 minimal media supplemented with 1 g/L yeast extract and antibiotics were inoculated to an OD of 0.2 in 15 mL disposable culture tubes, and placed shaking (250 rpm) at 37°C. Samples were taken immediately and 1 hour after inoculation, and analyzed 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). Fluorescence histograms were generated for 10,000 events per sample.

For sorting, 1 mL of library frozen stock was thawed, centrifuged to remove excess glycerol, and used to inoculate 5 mL LB with appropriate antibiotics. Cultures were grown overnight at 30°C, then used to inoculate 3 mL LB cultures with antibiotics and with or without 100 mM methanol, which were placed at 37°C. Prior to sorting, the cytometer was calibrated using Accudrop Beads (BD) and SPHERO Rainbow Calibration Particles (Spherotech, Lake Forest, IL). The culture with methanol was sorted 1 hour after inoculation into five gates with approximately equivalent populations, and 500,000 events were collected from each gate directly into LB media. Populations were recovered at 37°C overnight and mini-prepped for sequencing.

Antibiotics were added at the following final concentrations when appropriate: carbenicillin at 100 µg/mL for maintenance of Amp^R plasmids, and chloramphenicol (dissolved in water) at 25 µg/mL for maintenance of Cm^R plasmids.

4.2.5 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 301 bases at the University of Delaware DNA Sequencing and Genotyping Center. Paired-end reads were stitched together by the Center for Bioinformatics and Computational Biology (CBCB) at the Delaware Biotechnology Institute (DBI). Reads were filtered based on length, and the number of amino acid mismatches between each read and the native sequence was calculated. We discarded reads with more than approximately 30 amino acid mutations. 400,000-500,000 reads for each bin met all quality standards and were used for further analysis. Within each bin in an experiment, we calculated the frequency of each base and amino acid at each position from the aligned reads, and divided it by the total number of reads in the experiment. Mutations of interest were identified through enrichment across the five expression bins.

4.2.6 Formaldehyde measurements and *in vivo* Mdh resting cell assay

Formaldehyde was assayed via the colorimetric Nash assay.[97] Samples were collected and centrifuged (2 min, 13000 rpm) followed by the transfer of 400 µL supernatant to a microcentrifuge tube. Formaldehyde concentration was determined through the addition of 800 µL Nash reagent (7.5 g ammonium acetate, 150 µL glacial acetic acid, and 100 µL acetylacetone brought to 50 mL with water) to each sample, incubation for 30 min at 37°C, and measurement using a DU730 spectrophotometer from Beckman Coulter at 412 nm wavelength.

For the *in vivo* Mdh resting cell assay, 5 mL overnight cultures were inoculated in LB and antibiotics for Δ*frmA* strains with parent or variant pETM6_P_{tac}_Mdh-his plasmids. Overnight cultures were used to inoculate new 10 mL LB and antibiotic cultures with a 2% inoculum. Cultures were grown for 3.5 hours (37°C, 250 rpm), centrifuged (10 min, 5000 rpm), and resuspended in 4 mL M9 minimal medium. Concentrated cells were added to pre-warmed M9 minimal medium with either 100 mM methanol or 500 mM methanol and 10 mM butanol to a final OD of 1-4. Samples for formaldehyde measurement were taken at 0, 5, 10, and 30 minutes, quickly centrifuged, and placed at 4°C until processing.

4.2.7 Mdh purification and *in vitro* assays

Isolation of purified his-tagged Mdh protein was performed using the QIAGEN Ni-NTA Spin Kit, largely following manufacturer's protocols. Briefly, 50 mL of LB and appropriate antibiotics in 250 mL baffled flasks were inoculated from colonies and grown overnight (37°C, 250 rpm). Cultures were centrifuged (3000 rpm, 15 min), washed with M9 minimal media, and pellets were stored at -20°C until needed. Pellets were resuspended in 900 µL of NPI-10 lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8) and 100 µL 10 mg/mL lysozyme, transferred to microcentrifuge tubes, and incubated on ice for 30 min. Sonication followed in a Fisher Scientific Model 505 Sonic Dismembrator, Qsonica Cup Horn (Fisher Scientific, Hampton, NH) (10 min ON, 30 s ON/30 s OFF, 50% amplitude). Lysate was centrifuged at 12,000 x *g* for 30 min (4°C) and supernatant was collected. Chilled Ni-NTA spin columns were equilibrated with 600 µL NPI-10 lysis buffer, then loaded twice with 600 µL lysate and centrifuged (10 min, 250 x *g*, 4°C). The column was washed twice with 600 µL NPI-20 (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM

imidazole, pH 8) for 2 min (880 x *g*, 4°C), then his-tagged native protein was eluted twice with 300 µL NPI-500 (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH 8) for 2 min (880 x *g*, 4°C) and collected.

Purified protein was further concentrated using Amicon Ultra-0.5 mL 10K Centrifugal Filters (MilliporeSigma, Burlington, MA) following manufacturers protocols. Briefly, eluate from Ni-NTA spin columns were loaded onto the device and centrifuged (14,000 x *g*, 10 min, 4°C). The filter was washed twice with 450 µL 0.1 M pH 7 potassium phosphate buffer (14,000 x *g*, 10 min, 4°C), and a third time with 400 µL 0.1 M pH 7 potassium phosphate buffer (14,000 x *g*, 15 min, 4°C). The device was flipped upside down into a new collection tube and concentrated purified protein was collected (2 min, 1000 x *g*, 4°C). Protein concentration was determined via the Pierce Coomassie (Bradford) Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA) using bovine serum albumin (BSA) as a standard following manufacturers protocols.

Mdh activities were measured spectrophotometrically via the generation of NADH at 340 nm in 1 mL quartz cuvettes as described previously.[106] The 1 mL reaction mixture contained 100 mM Glycine-KOH pH 9.5, 5 mM $MgSO_4$, 1 mM NAD^+, and approximately 50 µg/mL purified protein. Components were mixed in the cuvette and pre-warmed to 45°C, and the assay was initiated with the addition of methanol. The Beer-Lambert law was used to convert absorbance to NADH concentration (ε_{NADH}=6220 M^{-1} cm^{-1}). One unit (U) of Mdh activity was defined as the amount of enzyme needed to produce 1 µmol NADH per minute under tested conditions. The production of NADH was continuously measured at 340 nm using a DU730 spectrophotometer from Beckman Coulter (Brea, CA) with a peltier temperature control module maintaining the temperature at 45°C.

4.2.8 Analytical methods

Cell growth was determined by measuring absorbance at 600 nm in a DU730 spectrophotometer from Beckman Coulter. Butanol concentrations were measured with high performance liquid chromatography (HPLC; Agilent 1200 series) with an Aminex HPX-87H column (Bio-Rad, Hercules, CA).

Intracellular metabolites and amino acids were extracted and analyzed as previously described[23] with minor modifications. Briefly, cells were pelleted and stored in microcentrifuge tubes at -20°C until extraction. For extraction, 1 mL of 70% (v/v) ethanol was heated to 70°C, added to each frozen cell pellet and vortexed for 30 s. The suspension was then heated at 95°C for 5 min, and cooled on ice for 5 min. Thereafter, the suspension was separated by centrifugation at 14,000 rpm for 5 min. The supernatant was transferred to a clean microcentrifuge tube and dried under nitrogen at 37°C. To derivatize the samples for GC-MS analysis, 50 µL of 2 wt% methoxyamine in pyridine solution was first added to the sample, and the sample was incubated at 37°C for 1 h. Thereafter, 50 µL of MTBSTFA+1% tBDMS was added to the sample and incubated at 60°C for 30 min. The derivatized sample was then centrifuged for 5 min at 14,000 rpm, and 60 µL of the supernatant was transferred to a GC injection vial.

GC-MS analysis for intracellular metabolites was performed as described.[98] GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30m, 0.25 mm i.d., 0.25 µm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass spectrometer operating under ionization by electron impact (EI) at 70 eV. ^{13}C-labeling was determined from the measured mass isotopomer data. First, the mass isotopomer distributions were corrected for natural isotope abundance [99]. Next, average ^{13}C-labeling was determined

as follows: Average ^{13}C-labeling = sum (Mi*i)/n, where n is the number of carbon atoms for the measured MS fragment less carbon atoms from derivatization reagent, Mi is the corrected mass isotopomer abundance, and i represents the number of ^{13}C-atoms making up that fragment.

4.3 Results

4.3.1 High-throughput Mdh activity library screening utilizing flow cytometry

A library of mutated *G. stearothermophilus* Mdh enzymes was generated via error-prone PCR (see Methods), and the Mdh activity of individual variants was assessed utilizing a high-throughput flow cytometry assay. A two plasmid reporter system was used to efficiently convert *in vivo* Mdh activity into a detectable fluorescent output. The system consists of a plasmid with Mdh under the P_{tac} promoter (pETM6_P_{tac}_Mdh) and a plasmid with GFP under the control of the formaldehyde-inducible promoter P_{frm} (pAC_P_{frm}_GFP) (Figure 4.1A). One hour after inoculating cultures with and without 100 mM methanol, a clearly defined separation of populations can be detected using flow cytometry (Figure 4.1A). Cultures with the parent Mdh without methanol express negligible levels of GFP, while those with 100 mM methanol have 2.4-fold higher geometric mean fluorescence due to the conversion of methanol to formaldehyde by active Mdh and subsequent induction of GFP expression (Figure 4.1A). Previous experiments on the formaldehyde-inducible promoter P_{frm} indicate that when formaldehyde is added directly, GFP can be detected as early as 5-10 minutes after dosing depending on media and growth phase.[96]

Extensive testing of response time for the two plasmid reporter system determined one hour to be the optimal response time for Mdh activity, allowing time for the accumulation of formaldehyde and triggering the induction of GFP expression (Figure C.1). The formaldehyde produced through methanol conversion was measured exogenously one hour post inoculation for 0, 30, and 60 mM methanol treatments, and demonstrated a linear response (Figure 4.1B). Under the conditions tested, undosed cultures exhibited less than 1 μM formaldehyde, while 30 and 60 mM methanol treatments resulted in 13 and 27 μM formaldehyde, respectively (Figure 4.1B). The two plasmid reporter system successfully confirms Mdh activity via GFP expression based on the 2.4-fold higher fluorescence with 100 mM methanol, providing an easily measurable fluorescence output in response to low levels of formaldehyde production in a highly sensitive assay.

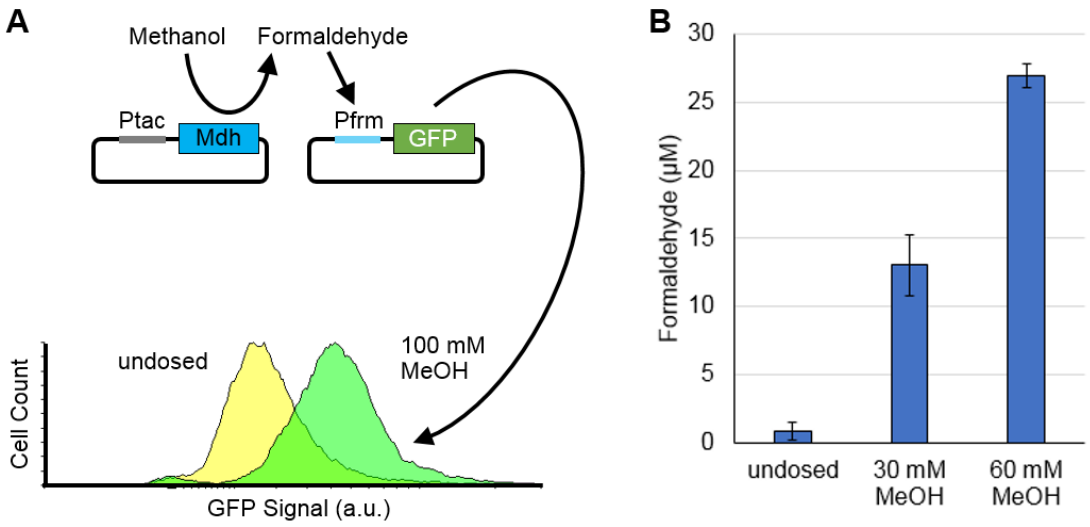

Figure 4.1 Flow cytometric Mdh activity reporter assay. (A) A two plasmid system serves as a reporter of Mdh activity through the production of Mdh (pETM6_P_{tac}_Mdh) for the conversion of methanol to formaldehyde, and the subsequent expression of GFP in response to formaldehyde (pAC_P_{frm}_GFP). Fluorescence histograms one hour after the addition of 0 or 100 mM methanol demonstrate the increased GFP expression in response to methanol. (B) Formaldehyde produced one hour after adding 0, 30, or 100 mM methanol, determined using the Nash assay. Error bars indicate standard deviation, n=2.

4.3.2 Sorting and sequencing approach enables rapid assessment of Mdh variants

Engineering the *G. stearothermophilus* Mdh first required the generation of thousands of Mdh variants, accomplished here utilizing error-prone PCR to quickly generate a high-diversity library of the Mdh coding region. The high-throughput flow cytometry assay enabled the simultaneous screening of the entire variant Mdh library. Relating GFP expression back to a specific mutation or variant, however, requires sequencing those variants. Next-generation sequencing allows for relatively inexpensive

sequencing of large libraries. Rather than isolating only highly fluorescent cells one hour after the addition of 100 mM methanol, the Mdh library was sorted into five fluorescent activity bins representing Mdh activity, each approximately equal in size (Figure 4.2A). The five populations were then sequenced, correlating sequence to function (Figure 4.2A). In this way, beneficial mutations were identified through their enrichment across activity bins, avoiding the false selection of highly abundant mutations which do not increase Mdh activity, and allowing for the identification of deleterious mutations. Analysis of the five sequenced populations led to the identification of nine Mdh variants of interest across the span of the Mdh coding region, each with a single amino acid substitution: V5I, L13M, R20H, P50S, V64M, V68I, V116C, Q227R, and V308L (Figure 4.2B).

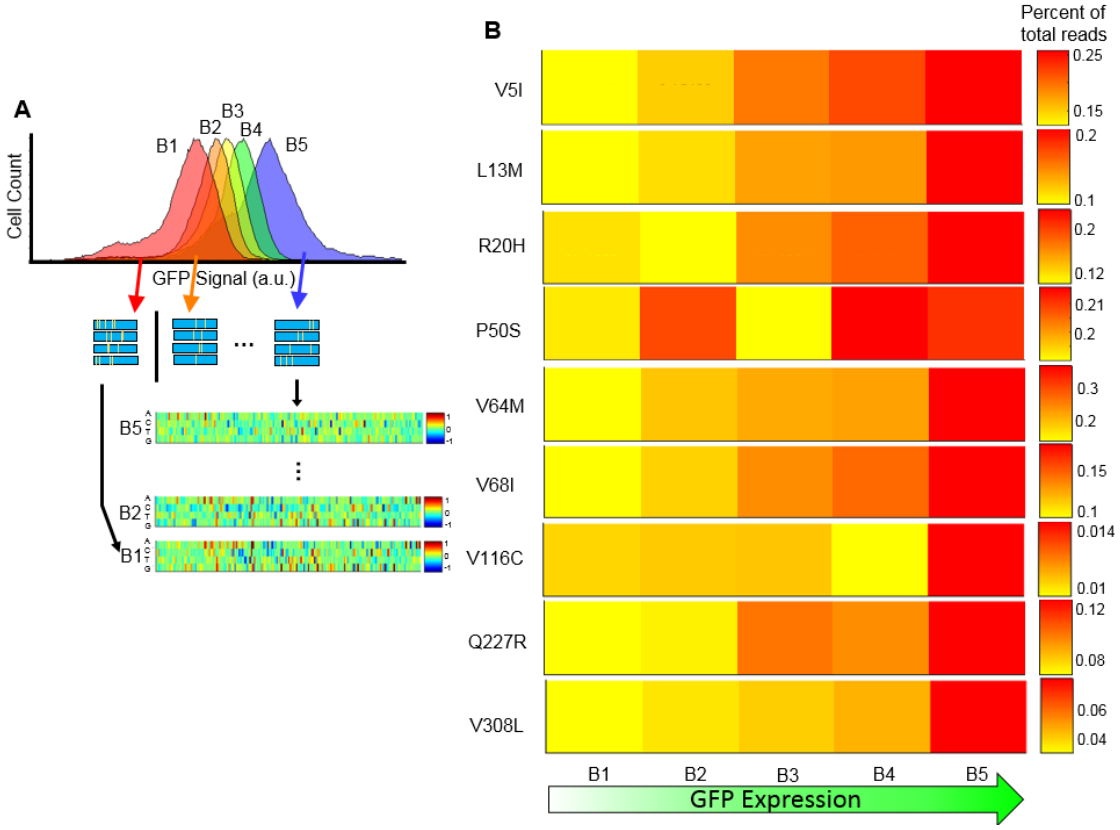

Figure 4.2 Sorting and sequencing method and the enrichment of variant Mdh sequences in high expression bins. (A) The Mdh library was sorted into five populations, labeled B1-B5, and each population was sequenced and analyzed for trends in nucleotide and amino acid mutations across the five bins. Heat maps show enrichment of each nucleotide (A, C, T, G) compared to the library for the first 200 nucleotides of *mdh* as an example of the sequencing data obtained. (B) Enrichment of nine Mdh variants with single amino acid substitutions in high GFP expression bins: V5I, L13M, R20H, P50S, V64M, V68I, V116C, Q227R, and V308L. The number of reads containing each mutation was divided by the total number of reads for each expression bin and multiplied by 100 to obtain percentage. Enrichment for each mutation across increasing GFP expression levels indicate a non-random pattern.

4.3.3 Improved *in vitro* and *in vivo* Mdh activity with variants identified via sequencing

Validation of the sorting and sequencing method was accomplished by further characterizing the identified Mdh variants through *in vitro* and *in vivo* assays. Kinetic parameters were determined from his-tag purified protein at 45°C and pH 9.5, as described previously for the characterization of *B. methanolicus* Mdhs.[106] In our hands, the parent *G. stearothermophilus* Mdh exhibited a K_m of 55 mM. The previously reported apparent K_m for methanol was 20 mM, at pH 7 and 37°C.[24] All nine variant Mdhs demonstrated lower K_m values compared to the parent when tested under identical conditions (Table 4.1). The V64M variant displayed the lowest methanol K_m value at 30.3 mM, 45% lower than the parent. P50S, V68I, and V5I displayed apparent methanol K_m parameters 35, 34, and 27% lower than the parent Mdh, respectively (Table 4.1). The K_m parameter is of particular interest for synthetic methylotrophy applications, as low methanol concentrations are desired for efficient substrate utilization. Attempts to ultimately develop an *E. coli* synthetic methanotroph through the expression of a soluble methane monooxygenase are also not likely to result in high methanol concentrations.[14] While all of the variants exhibited lower K_m values, most also showed slightly lower V_{max} values. Variants V308L and P50S had improved V_{max} values 7 and 13% higher than the parent Mdh, respectively (Table 4.1). *In vitro* kinetic parameters were determined from one biological replicate and therefore need to be repeated for significance testing.

In vitro enzyme assays provide reliable kinetic parameters of purified protein, however multiple other factors can impact Mdh activity in whole cells. A more accurate representation of Mdh activity *in vivo* and in the Δ*frmA* strain of interest was obtained via a resting cell assay, whereby overnight cultures were resuspended in M9 minimal media with 100 mM methanol or 500 mM methanol and 10 mM competitive butanol, and formaldehyde production was measured over the course of 30 minutes (Figure 4.3). Some variants, including L13M, P50S, and V116C, produced formaldehyde at a faster rate over the first 10 minutes compared to the parent Mdh after the addition of methanol (Figure 4.3A-D). The P50S Mdh also resulted in an improvement in selectivity towards methanol when the competing longer-chain alcohol 1-butanol was added, producing formaldehyde at a faster rate compared to the parent Mdh (Figure 4.3E).

Table 4.1 *In vitro* Mdh kinetic parameters of the parent and nine variant Mdhs, determined at 45°C and pH 9.5. Parameters were calculated via a Lineweaver-Burk plot from one replicate.

	MeOH K_m (mM)	MeOH V_{max} (mU/mg)
Parent	55.0	750.3
V5I	40.0	666.6
L13M	44.3	443.5
R20H	54.3	736.6
P50S	35.6	849.0
V64M	30.3	349.8
V68I	36.4	699.4
V116C	51.8	508.5
Q227R	41.5	717.5
V308L	41.8	805.3

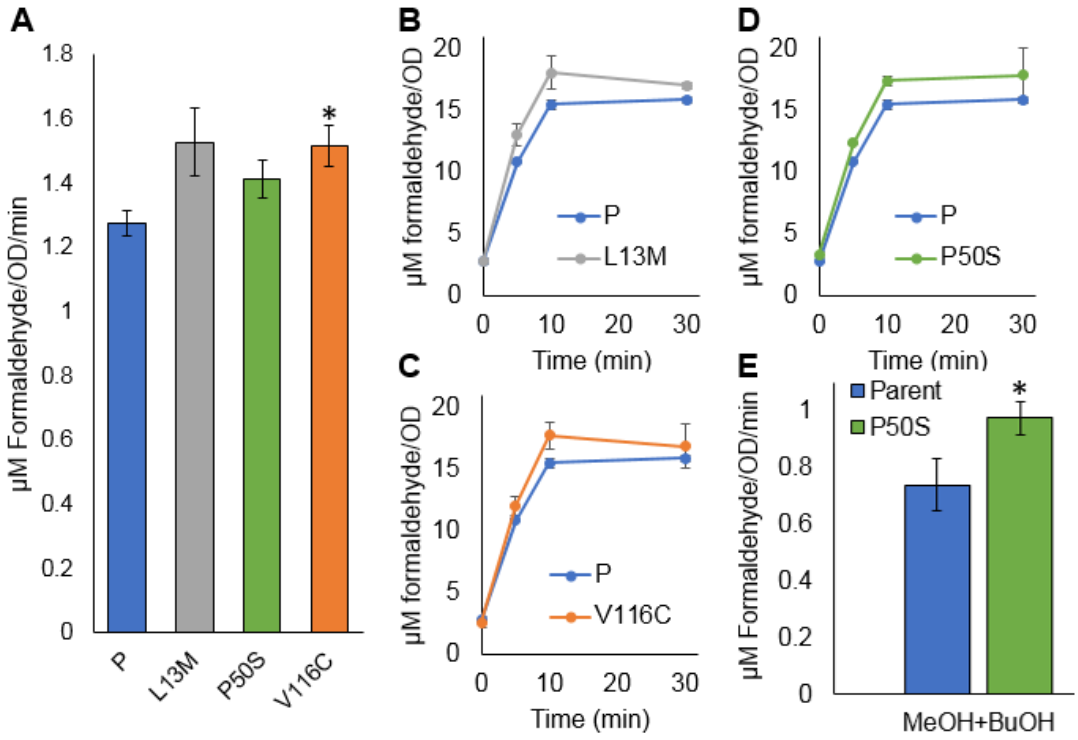

Figure 4.3 Mdh activity determined via an *in vivo* resting cell assay. (A) Formaldehyde production of the parent, P, and variant Mdhs L13M, P50S, and V116C, normalized to OD and time for the first 10 minutes after addition of 100 mM methanol to resting cells in M9 minimal medium. Individual plots for formaldehyde production normalized to OD for the parent, P, and the L13M (B), V116C (C), and P50S (D) variants 0, 5, 10, and 30 minutes after the addition of 100 mM methanol. (E) Formaldehyde production of the parent and P50S Mdh variant, normalized to OD and time for the first 10 minutes after addition of 500 mM methanol and 10 mM 1-butanol to resting cells in M9 minimal medium. Error bars indicate standard deviation, n=2-3. Two-tailed *t*-tests indicate significance compared to the parent strain. * $P<0.05$.

4.3.4 Higher methanol assimilation into intracellular metabolites when utilizing P50S Mdh variant

The P50S Mdh variant was chosen based on its favorable kinetic parameters and *in vivo* formaldehyde production for additional characterization. The Mdh parent and P50S variant were each placed in an operon with the *B. methanolicus hps* and *phi* to provide all the genes necessary for methanol oxidation and subsequent formaldehyde fixation to fructose 6-phosphate.[25] After 30 hours of growth with 100 mM ^{13}C-methanol and 0.5 g/L yeast extract, ^{13}C-methanol incorporation in intracellular metabolites was found to be significantly higher for the P50S strain compared to the parent (Figure 4.4). This included a 30% increase for upper glycolytic intermediate 3-phosphoglycerate (3PG), and 70 and 90% increases in TCA cycle intermediates citrate and fumarate, respectively. Methanol carbon incorporation into intracellular amino acids glycine, alanine, and aspartate also saw improvements of 85, 32, and 61%, respectively. In addition to improvements in average carbon labeling, there were improvements in the depth of labeling, particularly for TCA cycle intermediates citrate and fumarate (Figure 4.4B). Citrate molecules with either 5 or 6 carbons deriving from methanol increased from 2.8% in the parent strain to 10.5% in the P50S strain (Figure 4.4B). Under similar conditions, variants R20H, V68I, and Q227R also showed significant improvements in succinate labeling, and increases for other TCA cycle intermediates (Figure C.2). The improved methanol assimilation indicates that the increased formaldehyde produced from the P50S Mdh variant is successfully incorporated into intracellular metabolites. Interestingly, when the labeling experiment

was repeated under less aerobic conditions in sealed screw cap flasks rather than baffled flasks capped with tinfoil, the P50S Mdh variant strain exhibited similar ^{13}C-methanol carbon incorporation into intracellular metabolites compared to the parent strain (Figure C.3). We attribute this to the effect of oxygen availability on methanol oxidation.

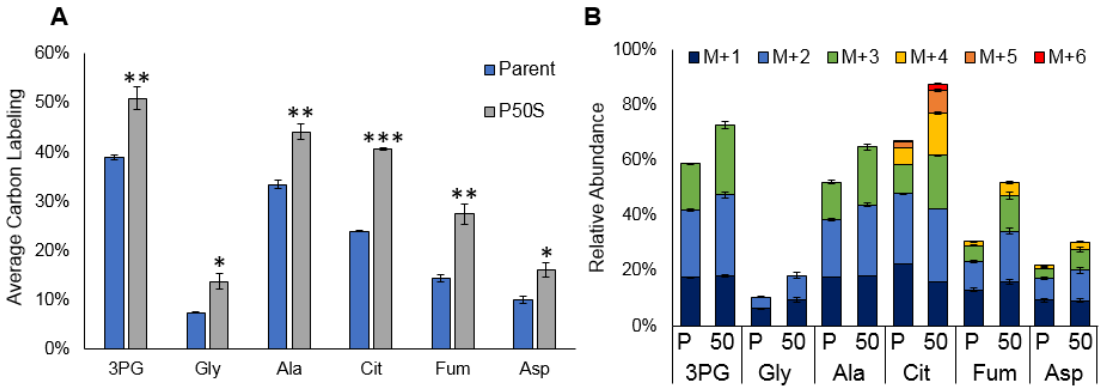

Figure 4.4 ^{13}C-methanol incorporation into intracellular metabolites for *E. coli* with the parent Mdh and P50S Mdh variant after 30 hours of growth with 100 mM ^{13}C-methanol and 0.5 g/L yeast extract. (A) and depth of labeling (B) in intracellular metabolites. 3PG, 3-phosphoglycerate; Gly, glycine; Ala, alanine; Cit, citrate; Fum, fumarate; Asp, aspartate. Error bars indicate standard deviation, n=2-3. Two-tailed *t*-tests indicate significance compared to the parent strain. * P<0.05, ** P<0.01, *** P<0.001.

4.3.5 Increased 1-butanol production when utilizing P50S Mdh variant

The P50S Mdh variant had previously shown higher formaldehyde production from methanol in the presence of competitive butanol. Higher Mdh selectivity *in vivo* has the potential to improve the titers of products, including longer chain alcohols such as butanol. Utilizing a butanol production pathway derived from *C. necator*, *Aeromonas caviae*, *Treponema denticola*, and *Clostridium saccharoperbutylacetonicum* developed[107] and used previously,[25] we tested butanol production during a short-term high-OD assay from glucose and methanol in a Δ*frmA* Δ*pgi* background strain (Figure 4.5A). The P50S strain displayed higher butanol titers at 19 and 44 hours, while measuring at the same OD as the parent strain (Figure 4.5B-C). This assay was performed with one biological replicate and therefore would need to be repeated for significance testing. If it reproduces, it would indicate that the higher selectivity towards methanol displayed in the P50S Mdh variant translates to improved butanol production compared to the parent Mdh strain.

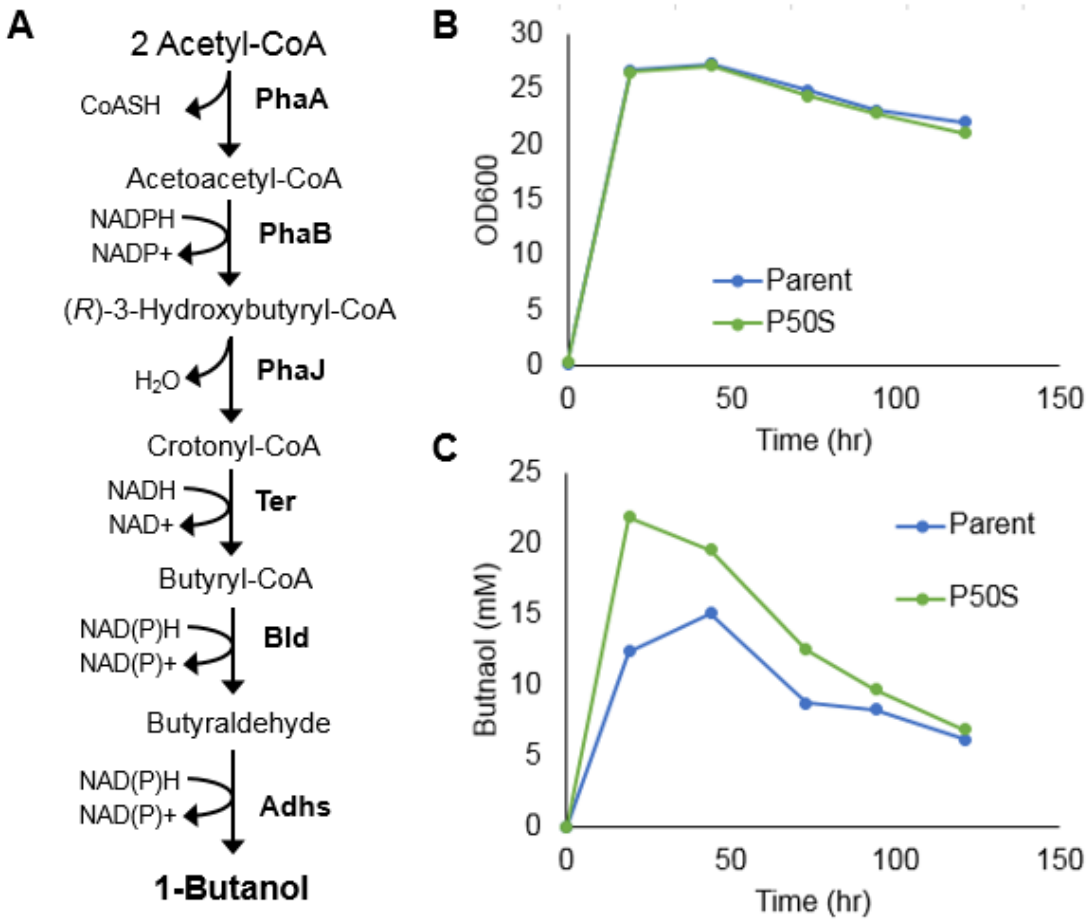

Figure 4.5 Butanol production with the parent Mdh and P50S Mdh variant in combination with *hps*, *phi*, and a heterologous butanol production pathway in a Δ*frmA* Δ*pgi* base strain, in TB media with 36 mM glucose and 500 mM methanol. (A) Heterologous butanol production pathway. PhaA, acetyltransferase; PhaJ, (*R*)-specific enoyl-CoA hydratase; Ter, *trans*-enoyl-CoA reductase; Bld, butyraldehyde dehydrogenase; Adhs, alcohol dehydrogenases. (B) Growth curves over the course of 120 hours. (C) Butanol production over the course of 120 hours of the parent and P50S strains for one biological replicate.

4.4 Discussion

Mdh engineering is a straightforward target for improving synthetic methylotrophy, especially considering the kinetic shortfalls of the NAD-dependent Mdhs frequently chosen for these applications. We engineered the *G. stearothermophilus* Mdh, aiming to increase activity and selectivity towards methanol through a high-throughput sorting and sequencing approach which allowed us to rapidly characterize thousands of Mdh mutations. The assessment of Mdh activity through GFP expression via a formaldehyde biosensor was accurate, however due to the two-step conversion necessary for GFP expression in response to methanol, the resolution of the enrichment data was lower than similar promoter elucidation experiments with the same biosensor.[96] Analysis of sequencing data therefore focused on increased enrichment across each of the five activity bins to minimize the effect of noise and identify promising mutations. Further characterization of chosen mutations revealed improvements *in vivo* for resting cell assays measuring formaldehyde production rates. The P50S Mdh variant also produced higher formaldehyde than the parent Mdh *in vivo* in the presence of competitive butanol, indicating a higher selectivity towards methanol than the parent Mdh. Higher selectivity in the P50S variant was further corroborated by preliminary data indicating the production of higher butanol titers in a Δ*frmA* Δ*pgi* strain of *E. coli* with *hps*, *phi*, and a heterologous butanol production pathway. Generally, extensive control over experimental conditions has been necessary to obtain reproducibility in these assays. While the P50S Mdh variant has shown improvements in multiple assays, the other identified variants often displayed

similar activities to the parent Mdh, exposing a weakness in the stringency of selection used to identify mutations for further characterization.

Acknowledgements

This work was supported by the Advanced Research Projects Agency-Energy (ARPA-E) Reducing Emissions using Methanotrophic Organisms for Transportation Energy (REMOTE) program (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 R01GM085232.

Chapter 5

CONCLUSIONS

5.1 Conclusions

We have addressed several metabolic bottlenecks and barriers here to the implementation of complete synthetic methylotrophy in *E. coli*. These approaches have been largely motivated by the regulation, gene expression, and metabolism of native methylotrophic bacteria. *B. methanolicus* in particular has a sizeable basis of research that has been invaluable for extracting the enzymes and regulatory framework which are likely to benefit *E. coli* synthetic methylotrophy.

First, we characterized and engineered the *E. coli* formaldehyde inducible promoter P_{frm}, achieving a significantly higher dynamic range of expression and developing a suite of formaldehyde-responsive promoters covering a wide range of uninduced and induced expression levels. This work also added a well-characterized biosensor to the *E. coli* genetic toolkit, expanding the repository of detectable metabolites with a defined, selective, and tunable system. Characterizing the formaldehyde-inducible promoter addresses two potential limitations of current synthetic methylotrophy: accumulation of cytotoxic formaldehyde leading to cell death, and the lack of *E. coli* regulatory networks driving expression of methylotrophic genes in response to methanol or formaldehyde.

Second, we expressed five *B. methanolicus* genes in various combinations under formaldehyde control, and utilized a knock-out of the NAD-dependent malate dehydrogenase (encoded by *maldh*) to increase ribulose 5-phosphate (Ru5P) regeneration and drive methanol oxidation, respectively. Formaldehyde is fixed to Ru5P, therefore maintaining adequate Ru5P pools is essential for avoiding formaldehyde accumulation, and ensuring continuous formaldehyde and methanol assimilation. While *E. coli* has all the enzymes necessary for regenerating Ru5P with the non-oxidative pentose phosphate pathway, they may be limited kinetically or transcriptionally downregulated. We show that with formaldehyde-controlled expression of just two *B. methanolicus* genes, *rpe* encoding ribulose 5-phosphate 3-epimerase and *tkt* encoding transketolase, methanol carbon assimilation to intracellular metabolites was significantly improved. Perturbation of the cellular redox balance, achieved with the *maldh* knockout, led to a dramatic improvement of the percent growth benefit with methanol when grown on yeast extract co-substrate. However low methanol carbon incorporation in this strain re-shaped our understanding of the relationship between the contribution of methanol carbon versus methanol energy, in the form of reducing equivalents, to biomass improvements. The provision of *rpe* and *tkt* in the Δ*maldh* strain fully compensated for the low labeling, achieving a strain with a powerful combination of high methanol carbon and methanol energy utilization, displayed via both high labeling from ^{13}C-methanol and an elevated growth benefit with methanol.

Third, we engineered the *G. stearothermophilus* Mdh for improved methanol activity and selectivity. Through a sorting and sequencing approach we were able to identify Mdh variants likely to possess improved Mdh activity. Many of these variants were verified to have improved Mdh activity via either *in vivo* resting cell assays, *in vitro* assays, or ^{13}C-methanol assimilation assays. Mdh activity can be made more thermodynamically favorable with lower formaldehyde concentrations as well, and therefore combining multiple approaches taken here should have synergistic effects on improving methanol assimilation.

Two of the included chapters utilized a sorting and sequencing approach to elucidate sequence-function relationships. These approaches have proven extremely powerful in identifying mutations of interest and determining their effects in a high-throughput way. The importance of careful design cannot be overstated, however. The size of the library generated, stringency of the assay design, number of expression bins, and sequencing coverage of each population can all heavily impact the quality of resulting data. The application of sort-seq can easily analyze promoter regions with sufficient coverage, however the analysis of proteins is far more difficult. The relationship between activity (Mdh activity here) and output (GFP expression here) is not as direct, and achieving coverage for 20 amino acid mutations at each position is far more difficult than 4 nucleotides. Full coverage of every amino acid mutation is not required for recovering meaningful data regarding protein activity and the impact of individual mutations, the cost-benefit analysis is just a necessary calculation to be considered prior to experimentation.

5.2 Perspective on the future of synthetic methylotrophy

In pursuit of a pure synthetic methylotroph, utilizing complex co-substrates such as yeast extract, or defined amino acid substrates, should continue to be pursued. Careful analysis of preferred methanol co-substrate and metabolite production from methanol carbon should elucidate steps forward for identifying and alleviating regulatory or enzymatic bottlenecks. Methanol carbon is not equally incorporated into all metabolites, and understanding why can inform an approach for upregulating specific biosynthetic pathways to ensure that methanol has available and functional metabolic pathways for generating every biomass component. Regulation will undoubtedly continue to be a major and complex hurdle to be reckoned with. While we have taken steps to institute dynamic formaldehyde regulation here, we have not addressed native *E. coli* global regulators which are likely creating a non-optimal expression profile for methanol utilization.

Methanol auxotrophy, or methanol-dependent strain engineering is also a promising avenue for the development of synthetic methylotrophy. It is likely the best approach for improving product titers and incorporating high levels of methanol carbon into products of interest in the short-term. It is also an option to scale-up for industrial production due to its sugar co-substrate and the recovered growth rates that have been achieved after relatively short periods of laboratory evolution. However the evolution or engineering of a pure synthetic methylotroph from a methanol-dependent strain seems highly unlikely.

REFERENCES

(1) (2019) Natural Gas Gross Withdrawals and Production. https://www.eia.gov/dnav/ng/ng_prod_sum_dc_NUS_mmcf_a.htm, U.S. Energy Information Administration, U.S. Department of Energy.

(2) Tenenbaum, D. J. (2008) Food vs. Fuel: diversion of crops could cause more hunger. *Environ. Health Perspect. 116*, 254-257.

(3) Hassan, S. S., Williams, G. A., and Jaiswal, A. K. (2018) Emerging technologies for the pretreatment of lignocellulosic biomass. *Bioresour. Technol. 262*, 310-318.

(4) Schrader, J., Schilling, M., Holtmann, D., Sell, D., Filho, M. V., Marx, A., and Vorholt, J. A. (2009) Methanol-based industrial biotechnology: current status and future perspectives of methylotrophic bacteria. *Trends Biotechnol. 27*, 107-115.

(5) 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.

(6) Olah, G. A. (2013) Towards Oil Independence Through Renewable Methanol Chemistry. *Angew. Chem. Int. Ed. 52*, 104-107.

(7) (2019) The George Olah Renewable Methanol Plant. https://www.carbonrecycling.is/george-olah, Carbon Recycling International.

(8) Leonzio, G. (2018) State of art and perspectives about the production of methanol, dimethyl ether and syngas by carbon dioxide hydrogenation. *Journal of CO2 Utilization 27*, 326-354.

(9) Al-Kalbani, H., Xuan, J., García, S., and Wang, H. (2016) Comparative energetic assessment of methanol production from CO2: Chemical versus electrochemical process. *Applied Energy 165*, 1-13.

(10) Matzen, M., Alhajji, M., and Demirel, Y. (2015) Chemical storage of wind energy by renewable methanol production: Feasibility analysis using a multi-criteria decision matrix. *Energy 93*, 343-353.

(11) Matzen, M., and Demirel, Y. (2016) Methanol and dimethyl ether from renewable hydrogen and carbon dioxide: Alternative fuels production and life-cycle assessment. *Journal of Cleaner Production 139*, 1068-1077.

(12) Chen, Q., Gu, Y., Tang, Z., and Sun, Y. (2019) Comparative environmental and economic performance of solar energy integrated methanol production systems in China. *Energy Conversion and Management 187*, 63-75.

(13) Hwang, I. Y., Nguyen, A. D., Nguyen, T. T., Nguyen, L. T., Lee, O. K., and Lee, E. Y. (2018) Biological conversion of methane to chemicals and fuels: technical challenges and issues. *Appl. Microbiol. Biotechnol. 102*, 3071-3080.
(14) Kim, H. J., Huh, J., Kwon, Y. W., Park, D., Yu, Y., Jang, Y. E., Lee, B.-R., Jo, E., Lee, E. J., Heo, Y., Lee, W., and Lee, J. (2019) Biological conversion of methane to methanol through genetic reassembly of native catalytic domains. *Nature Catalysis 2*, 342-353.
(15) Pfeifenschneider, J., Brautaset, T., and Wendisch, V. F. (2017) Methanol as carbon substrate in the bio-economy: Metabolic engineering of aerobic methylotrophic bacteria for production of value-added chemicals. *Biofuels, Bioproducts and Biorefining 11*, 719-731.
(16) Irla, M., Heggeset, T. M. B., Nærdal, I., Paul, L., Haugen, T., Le, S. B., Brautaset, T., and Wendisch, V. F. (2016) Genome-based genetic tool development for Bacillus methanolicus: Theta- and rolling circle-replicating plasmids for inducible gene expression and application to methanol-based cadaverine production. *Front. Microbiol. 7*, 1481.
(17) Nærdal, I., Pfeifenschneider, J., Brautaset, T., and Wendisch, V. F. (2015) Methanol-based cadaverine production by genetically engineered Bacillus methanolicus strains. *Microbial Biotechnology 8*, 342-350.
(18) Brautaset, T., Jakobsen, Ø. M., Josefsen, K. D., Flickinger, M. C., and Ellingsen, T. E. (2007) Bacillus methanolicus: A candidate for industrial production of amino acids from methanol at 50°C. *Appl. Microbiol. Biotechnol. 74*, 22-34.
(19) Goswami, P., Chinnadayyala, S. S. R., Chakraborty, M., Kumar, A. K., and Kakoti, A. (2013) An overview on alcohol oxidases and their potential applications. *Appl. Microbiol. Biotechnol. 97*, 4259-4275.
(20) 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.
(21) Wu, T.-Y., Chen, C.-T., Liu, J. T.-J., Bogorad, I. W., Damoiseaux, R., and Liao, J. C. (2016) Characterization and evolution of an activator-independent methanol dehydrogenase from Cupriavidus necator N-1. *Appl. Microbiol. Biotechnol. 100*, 4969-4983.
(22) Roth, T. B., Woolston, B. M., Stephanopoulos, G., and Liu, D. R. (2019) Phage-Assisted Evolution of Bacillus methanolicus Methanol Dehydrogenase 2. *ACS Synth. Biol. 8*, 796-806.
(23) 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.

(24) 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.
(25) Bennett, R. K., Gonzalez, J. E., Whitaker, W. B., Antoniewicz, M. R., and Papoutsakis, E. T. (2018) Expression of heterologous non-oxidative pentose phosphate pathway from Bacillus methanolicus and phosphoglucose isomerase deletion improves methanol assimilation and metabolite production by a synthetic Escherichia coli methylotroph. *Metab. Eng. 45*, 75-85.
(26) Woolston, B. M., King, J. R., Reiter, M., Van Hove, B., and Stephanopoulos, G. (2018) Improving formaldehyde consumption drives methanol assimilation in engineered E. coli. *Nat. Commun. 9*.
(27) Chen, C.-T., Chen, F. Y.-H., Bogorad, I. W., Wu, T.-Y., Zhang, R., Lee, A. S., and Liao, J. C. (2018) Synthetic methanol auxotrophy of Escherichia coli for methanol-dependent growth and production. *Metab. Eng. 49*, 257-266.
(28) Orita, I., Sakamoto, N., Kato, N., Yurimoto, H., and Sakai, Y. (2007) Bifunctional enzyme fusion of 3-hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase. *Appl. Microbiol. Biotechnol. 76*, 439-445.
(29) 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.
(30) Meyer, F., Keller, P., Hartl, J., Gröninger, O. G., Kiefer, P., and Vorholt, J. A. (2018) Methanol-essential growth of Escherichia coli. *Nat. Commun. 9*.
(31) Gonzalez, J. E., Bennett, R. K., Papoutsakis, E. T., and Antoniewicz, M. R. (2018) Methanol assimilation in Escherichia coli is improved by co-utilization of threonine and deletion of leucine-responsive regulatory protein. *Metab. Eng. 45*, 67-74.
(32) Görke, B., and Stülke, J. (2008) Carbon catabolite repression in bacteria: Many ways to make the most out of nutrients. *Nature Reviews Microbiology 6*, 613-624.
(33) Shimada, T., Fujita, N., Yamamoto, K., and Ishihama, A. (2011) Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. *PLoS ONE 6*, e20081.
(34) Peña, D. A., Gasser, B., Zanghellini, J., Steiger, M. G., and Mattanovich, D. (2018) Metabolic engineering of Pichia pastoris. *Metab. Eng. 50*, 2-15.
(35) Vogl, T., Sturmberger, L., Kickenweiz, T., Wasmayer, R., Schmid, C., Hatzl, A. M., Gerstmann, M. A., Pitzer, J., Wagner, M., Thallinger, G. G., Geier, M., and Glieder, A. (2016) A Toolbox of Diverse Promoters Related to Methanol Utilization: Functionally Verified Parts for Heterologous Pathway Expression in Pichia pastoris. *ACS Synth. Biol. 5*, 172-186.
(36) Dubey, A. A., and Jain, V. (2019) MnoSR is a bona fide two-component system involved in methylotrophic metabolism in Mycobacterium smegmatis. *Appl. Environ. Microbiol. 85*, e00535-00519.

(37) Selvamani, V., Maruthamuthu, M. K., Arulsamy, K., Eom, G. T., and Hong, S. H. (2017) Construction of methanol sensing Escherichia coli by the introduction of novel chimeric MxcQZ/OmpR two-component system from Methylobacterium organophilum XX. *Korean Journal of Chemical Engineering 34*, 1734-1739.
(38) Harms, N., Reijnders, W. N. M., Koning, S., and Spanning, R. J. M. V. (2001) Two-component system that regulates methanol and formaldehyde oxidation in Paracoccus denitrificans. *J. Bacteriol. 183*, 664-670.
(39) Skovran, E., Palmer, A. D., Rountree, A. M., Good, N. M., and Lidstrom, M. E. (2011) XoxF is required for expression of methanol dehydrogenase in Methylobacterium extorquens AM1. *J. Bacteriol. 193*, 6032-6038.
(40) Springer, A. L., Auman, A. J., and Lidstrom, M. E. (1998) Sequence and characterization of mxaB, a response regulator involved in regulation of methanol oxidation, and of mxaW, a methanol-regulated gene in Methylobacterium extorquens AM1. *FEMS Microbiol. Lett. 160*, 119-124.
(41) 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.
(42) Yurimoto, H., Hirai, R., Matsuno, N., Yasueda, H., Kato, N., and Sakai, Y. (2005) HxlR, a member of the DUF24 protein family, is a DNA-binding protein that acts as a positive regulator of the formaldehyde-inducible hxlAB operon in Bacillus subtilis. *Mol. Microbiol. 57*, 511-519.
(43) Mitsui, R., Kusano, Y., Yurimoto, H., Sakai, Y., Kato, N., and Tanaka, M. (2003) Formaldehyde Fixation Contributes to Detoxification for Growth of a Nonmethylotroph, Burkholderia cepacia TM1, on Vanillic Acid. *Appl. Environ. Microbiol. 69*, 6128-6132.
(44) Irla, M., Neshat, A., Brautaset, T., Rückert, C., Kalinowski, J., and Wendisch, V. F. (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.
(45) Selvamani, V., Ganesh, I., Maruthamuthu, M. k., Eom, G. T., and Hong, S. H. (2017) Engineering chimeric two-component system into Escherichia coli from Paracoccus denitrificans to sense methanol. *Biotechnology and Bioprocess Engineering 22*, 225-230.
(46) Ganesh, I., Vidhya, S., Eom, G. T., and Hong, S. H. (2017) Construction of methanol-sensing Escherichia coli by the introduction of a Paracoccus denitrificans MxaY-based chimeric two-component system. *Journal of Microbiology and Biotechnology 27*, 1106-1111.
(47) 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.

(48) Chistoserdova, L., Lapidus, A., Han, C., Goodwin, L., Saunders, L., Brettin, T., Tapia, R., Gilna, P., Lucas, S., Richardson, P. M., and Lidstrom, M. E. (2007) Genome of Methylobacillus flagellatus, molecular basis for obligate methylotrophy, and polyphyletic origin of methylotrophy. *J. Bacteriol. 189*, 4020-4027.
(49) Hendrickson, E. L., Beck, D. A. C., Wang, T., Lidstrom, M. E., Hackett, M., and Chistoserdova, L. (2010) Expressed genome of Methylobacillus flagellatus as defined through comprehensive proteomics and new insights into methylotrophy. *J. Bacteriol. 192*, 4859-4867.
(50) Müller, J. E. N., Meyer, F., Litsanov, B., Kiefer, P., and Vorholt, J. A. (2015) Core pathways operating during methylotrophy of Bacillus methanolicus MGA3 and induction of a bacillithiol-dependent detoxification pathway upon formaldehyde stress. *Mol. Microbiol. 98*, 1089-1100.
(51) Müller, J. E. N., Litsanov, B., Bortfeld-Miller, M., Trachsel, C., Grossmann, J., Brautaset, T., and Vorholt, J. a. (2014) Proteomic analysis of the thermophilic methylotroph Bacillus methanolicus MGA3. *Proteomics 14*, 725-737.
(52) Peyraud, R., Schneider, K., Kiefer, P., Massou, S., Vorholt, J. A., and Portais, J.-C. (2011) Genome-scale reconstruction and system level investigation of the metabolic network of Methylobacterium extorquens AM1. *BMC Syst. Biol. 5*, 189.
(53) Wang, X., Wang, X., Lu, X., Ma, C., Chen, K., and Ouyang, P. (2019) Methanol fermentation increases the production of NAD(P)H-dependent chemicals in synthetic methylotrophic Escherichia coli. *Biotechnology for Biofuels 12*, 17.
(54) 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.
(55) Glick, B. R. (1995) Metabolic load and heterogenous gene expression. *Biotechnol. Adv. 13*, 247-261.
(56) Rogers, J. K., Taylor, N. D., and Church, G. M. (2016) Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol. 42*, 84-91.
(57) 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.
(58) 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.
(59) 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.

(60) Mahr, R., and Frunzke, J. (2016) Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl. Microbiol. Biotechnol. 100*, 79-90.
(61) Libis, V., Delépine, B., and Faulon, J.-L. (2016) Sensing new chemicals with bacterial transcription factors. *Curr. Opin. Microbiol. 33*, 105-112.
(62) 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.
(63) Atwal, G. S., and Kinney, J. B. (2016) Learning Quantitative Sequence–Function Relationships from Massively Parallel Experiments. *J. Stat. Phys. 162*, 1203-1243.
(64) 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.
(65) 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.
(66) Treves, A., and Panzeri, S. (1995) The Upward Bias in Measures of Information Derived from Limited Data Samples. *Neural Comput. 7*, 399-407.
(67) Gasperini, M., Starita, L., and Shendure, J. (2016) The power of multiplexed functional analysis of genetic variants. *Nat. Protoc. 11*, 1782-1787.
(68) 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.
(69) 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.
(70) 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.
(71) 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.
(72) 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.

(73) 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.
(74) 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.
(75) 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.
(76) 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.
(77) 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.
(78) 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.
(79) 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.
(80) Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974) Culture medium for enterobacteria. *J. Bacteriol. 119*, 736-747.
(81) Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: A laboratory manual*. 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
(82) 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.
(83) 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.
(84) 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.
(85) 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.

(86) 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.
(87) 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.
(88) Huber, P. J. (1981) *Robust Statistics*, John Wiley & Sons.
(89) 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.
(90) Bennett, R. K., Steinberg, L. M., Chen, W., and Papoutsakis, E. T. (2018) Engineering the bioconversion of methane and methanol to fuels and chemicals in native and synthetic methylotrophs. *Curr. Opin. Biotechnol. 50*, 81-93.
(91) 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.
(92) Yu, R., Lai, Y., Hartwell, H. J., Moeller, B. C., Doyle-Eisele, M., Kracko, D., Bodnar, W. M., Starr, T. B., and Swenberg, J. A. (2015) Formation, accumulation, and hydrolysis of endogenous and exogenous formaldehyde-induced DNA damage. *Toxicol. Sci. 146*, 170-182.
(93) Long, C. P. (2018) Interrogating Central Carbon Metabolism in Escherichia coli via the mapping of flux responses to gene knockouts and adaptive evolution by Christopher. pp 188-209.
(94) 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.*, 2006.0008.
(95) 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.
(96) Rohlhill, J., Sandoval, N. R., and Papoutsakis, E. T. (2017) Sort-Seq Approach to Engineering a Formaldehyde-Inducible Promoter for Dynamically Regulated Escherichia coli Growth on Methanol. *ACS Synth. Biol. 6*, 1584-1595.
(97) Nash, T. (1953) The colorimetric estimation of formaldehyde by means of the Hantzsch reaction. *Biochem. J. 55*, 416-421.
(98) Long, C. P., and Antoniewicz, M. R. (2014) Quantifying biomass composition by gas chromatography/mass spectrometry. *Analytical Chemistry 86*, 9423-9427.
(99) Fernandez, C. A., Rosiers, C. D., Previs, S. F., David, F., and Brunengraber, H. (1996) Correction of 13C Mass Isotopomer Distributions for Natural Stable Isotope Abundance. *J. Mass Spectrom. 31*, 255-262.

(100) Brautaset, T., Jakobsen, Ø. M., Flickinger, M. C., Valla, S., and Ellingsen, T. E. (2004) Plasmid-Dependent Methylotrophy in Thermotolerant Bacillus methanolicus. *J. Bacteriol. 186*, 1229-1238.
(101) Van der Rest, M. E., Frank, C., and Molenaar, D. (2000) Functions of the membrane-associated and cytoplasmic malate dehydrogenases in the citric acid cycle of Escherichia coli. *J. Bacteriol. 182*, 6892-6899.
(102) Xu, P., Li, L., Zhang, F., Stephanopoulos, G., and Koffas, M. (2014) Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc. Natl. Acad. Sci. U.S.A. 111*, 11299-11304.
(103) Endalur Gopinarayanan, V., and Nair, N. U. (2018) A semi-synthetic regulon enables rapid growth of yeast on xylose. *Nat. Commun. 9*, 1233.
(104) 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.
(105) Trotsenko, Y. A., and Murrell, J. C. (2008) Metabolic Aspects of Aerobic Obligate Methanotrophy. *Adv. Appl. Microbiol. 63*, 183-229.
(106) Krog, A., Heggeset, T. M. B., Muller, J. E. N., Kupper, C. E., Schneider, O., Vorholt, J. A., Ellingsen, T. E., and Brautaset, T. (2013) Methylotrophic Bacillus methanolicus Encodes Two Chromosomal and One Plasmid Born NAD+ Dependent Methanol Dehydrogenase Paralogs with Different Catalytic and Biochemical Properties. *8*, e59188.
(107) Kataoka, N., Vangnai, A. S., Pongtharangkul, T., Tajima, T., Yakushi, T., Matsushita, K., and Kato, J. (2015) Construction of CoA-dependent 1-butanol synthetic pathway functions under aerobic conditions in Escherichia coli. *J. Biotechnol. 204*, 25-32.

CHAPTER 2 SUPPLEMENTARY INFORMATION

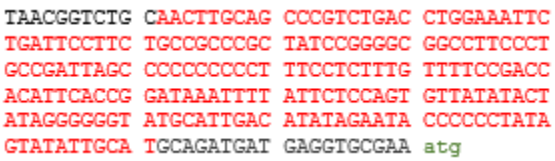

Figure A.1 Region upstream of the *frmRAB* operon. Red letters were isolated as the 200 bp P_{frm} sequence and targeted for P_{frm} library generation. Green 'atg' is the start codon for *frmR*. Adapted from EcoCyc database (Accession ID: G6209).

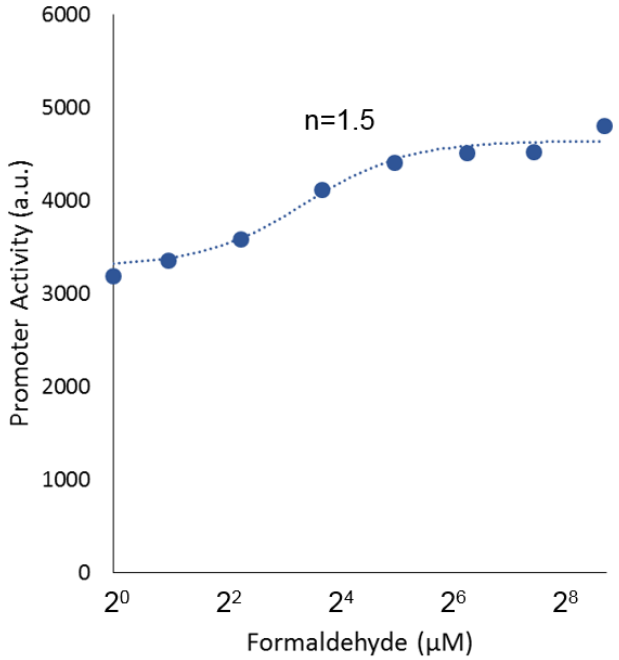

Figure A.2 Response curve for P_{frm}-GFP without plasmid-expressed FrmR in the NEB5α strain fit to the Hill equation. The Hill coefficient is printed above the curve.

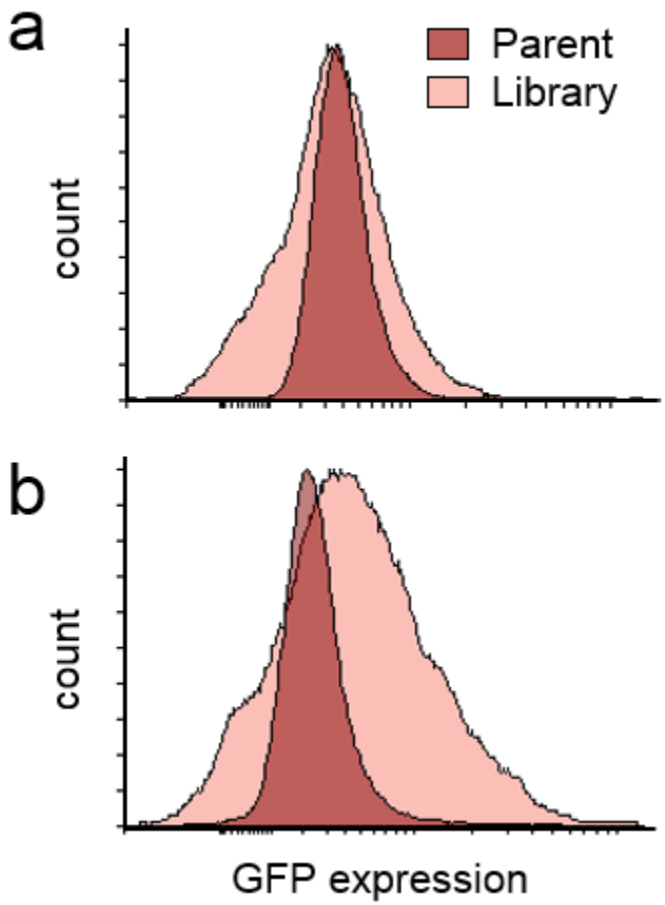

Figure A.3 Fluorescence distributions of 10,000 cells for the parent P_{frm}-GFP-P_{lac}-FrmR plasmid and mutated library in the (a) NEB5α and (b) Δ*frmR* strains.

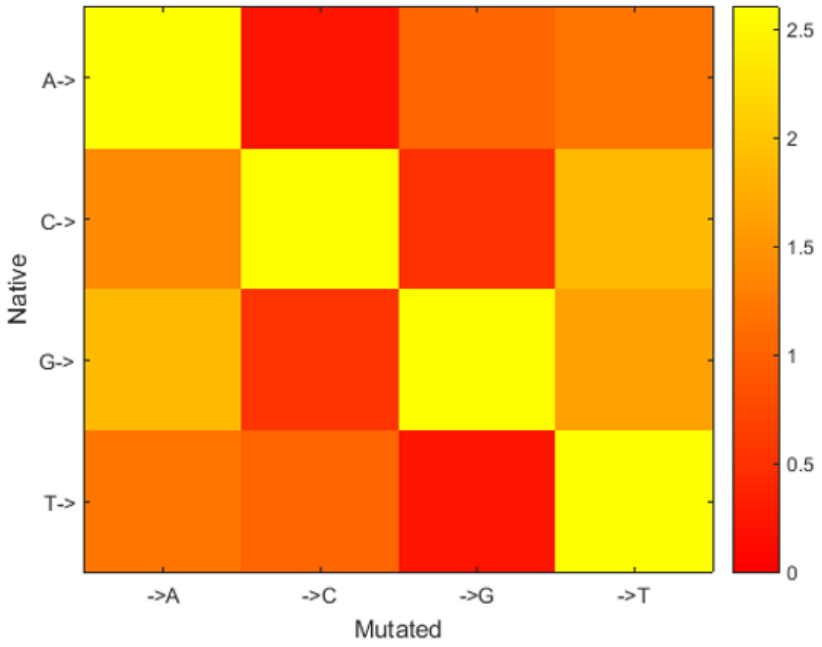

Figure A.4 Mutation bias of the unsorted P_{frm} library, shown as percent mutated.

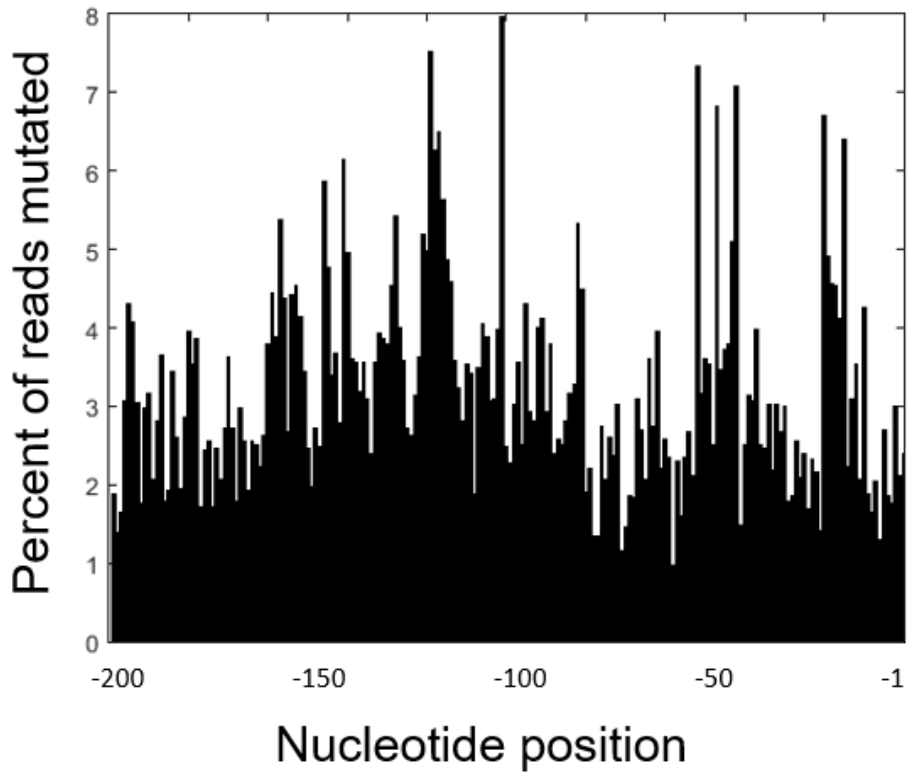

Figure A.5 The percent of reads mutated at each nucleotide position within the 200 bp P_{frm} promoter. The results from 327693 reads of the unsorted P_{frm} library are shown.

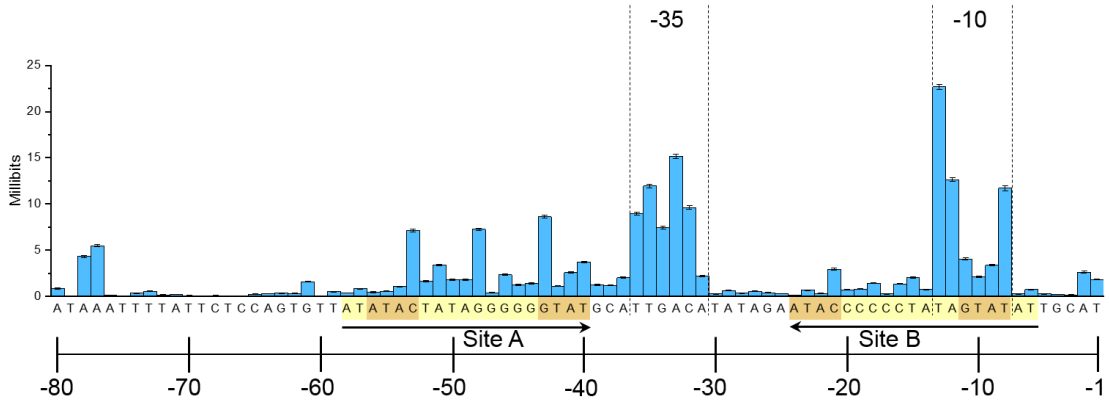

Figure A.6 Information footprint of the *E. coli* P_{frm} in the Δ*frmR* strain with plasmid-expressed FrmR under uninduced (0 μM formaldehyde) conditions. Yellow highlight shows a large inverted repeat while the two left red sites and two right red sites indicate smaller inverted repeats relevant to FrmR binding. Error bars indicate uncertainties inferred from subsampling.

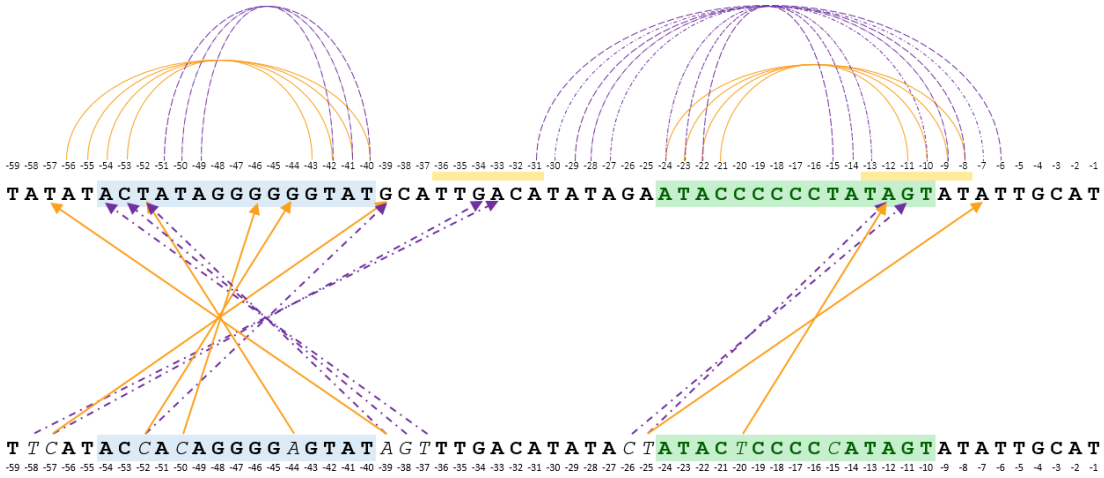

Figure A.7 Multiple hairpin formations affecting FrmR binding strength. For each identified FrmR binding site (site A and site B), there are two different hairpin structures which may form to enable FrmR binding (shown here in purple dashed lines or orange solid lines). From sort-seq data, mutations likely to cause stronger FrmR binding were identified as shown (bottom sequence) compared to the native sequence (top sequence). The bold letters represent the unmutated native P_{frm} sequence while the italicized letters represent mutations which were found to decrease GFP expression the most, and thus likely to be indicative of a stronger FrmR binding event. The mutations shown were found to be in common in at least three of the four selection criteria: the most enriched base found at that site in the lowest two expression bins (bins 1 and 2) and the most depleted base in the highest two expression bins (bins 7 and 8), all with respect to the unsorted library. Mutations which indicate a larger hairpin structure are shown by arrows indicating the corresponding base with which they would interact in a hairpin structure in each of the two versions (solid orange or dashed purple arrows). It is of note that four such mutations would work for both versions of the hairpin structures. These mutations indicate FrmR binds DNA in either form of the hairpin.

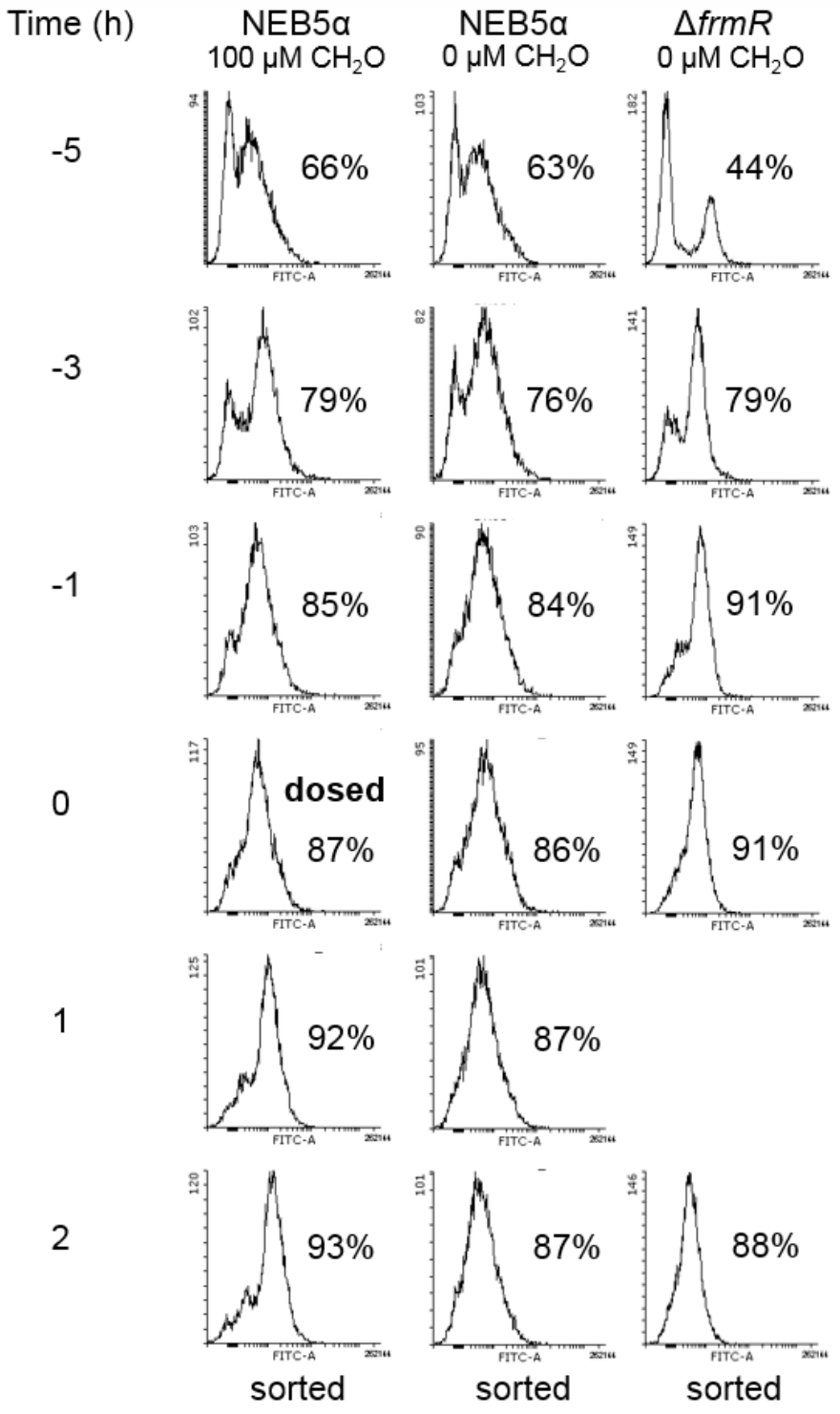
Time (h)
NEB5α
100 μM CH_2O
NEB5α
0 μM CH_2O
Δ*frmR*
0 μM CH_2O
-5
66%
63%
44%
-3
79%
76%
79%
-1
85%
84%
91%
0
dosed
87%
86%
91%
1
92%
87%
2
93%
87%
88%
FITC-A
sorted
sorted
sorted

Figure A.8 Fluorescence histograms tracking the recovery of the mutant P_{frm} library before sorting for the NEB5α and Δ*frmR* strains. Percent GFP positive is annotated on each histogram, where the GFP positive gate was drawn to include 1% of the promoterless negative control. Time zero was determined to be when the distribution was 85-90% GFP positive and largely unimodal.

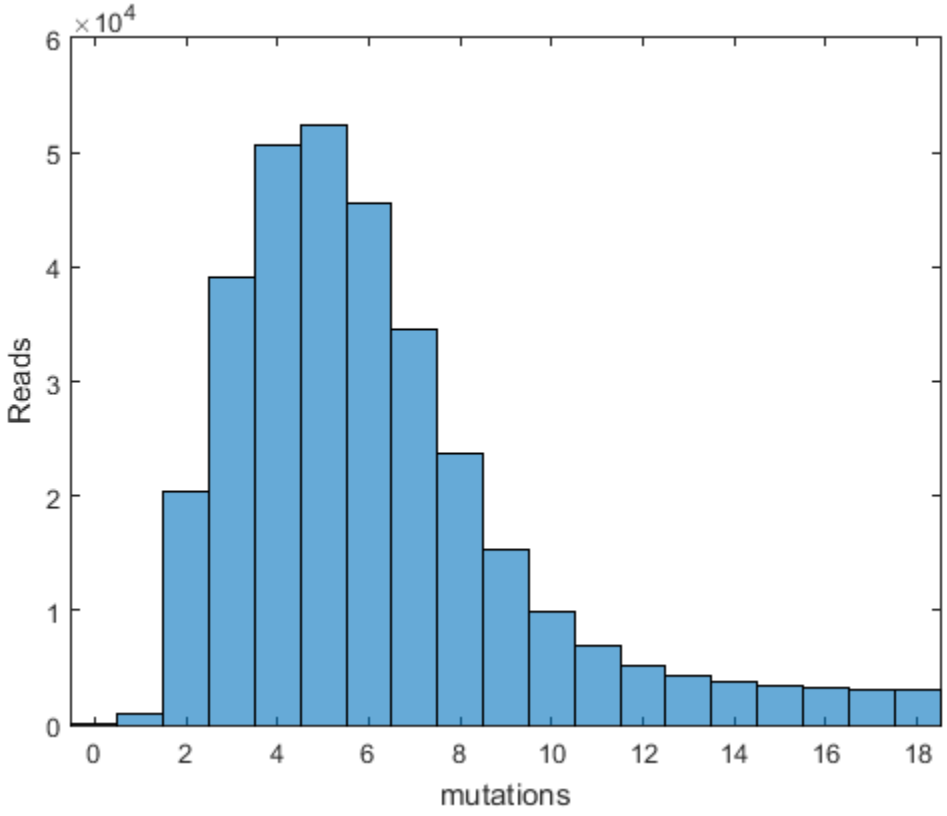

Figure A.9 Histogram of the number of reads with different numbers of mutations for the unsorted P_{frm} library. The results for 327693 reads are shown with a median of 5 mutations per 200 bp promoter.

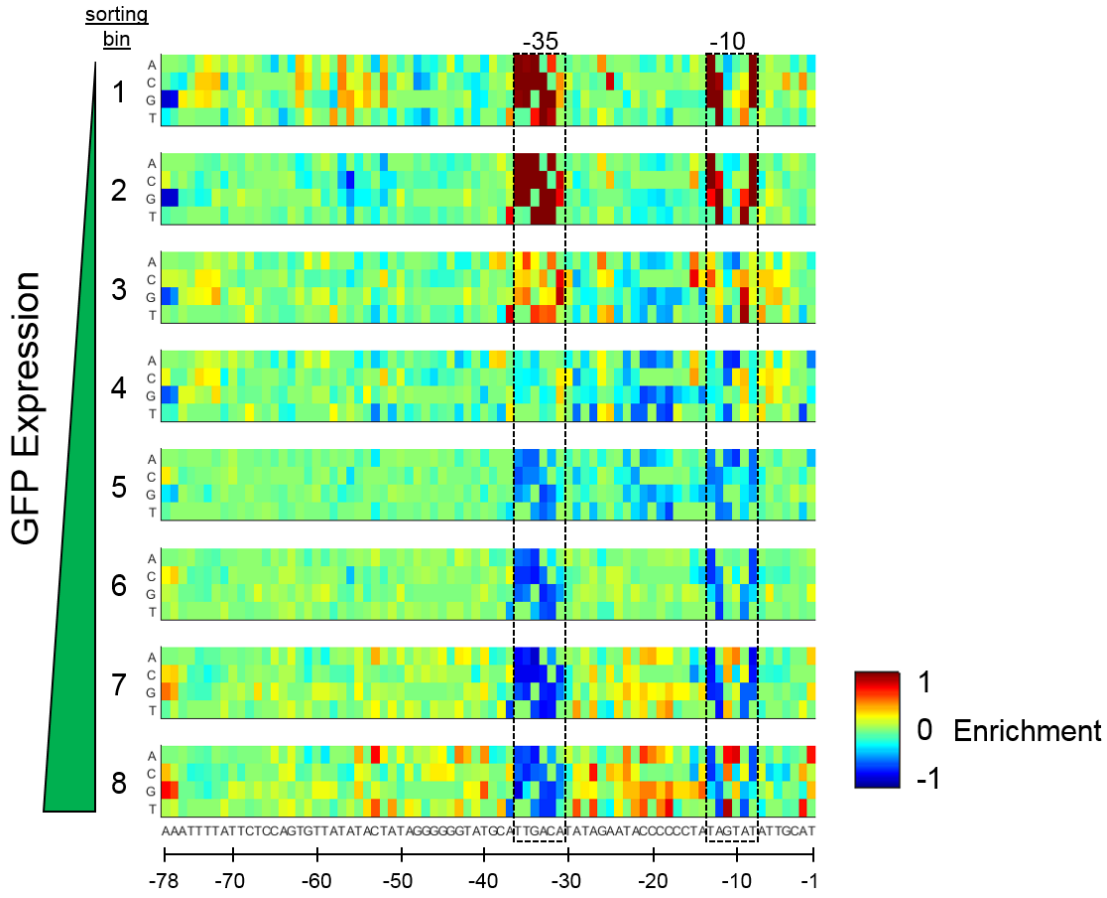

Figure A.10 Identification of up and down mutations through sequence analysis. Heat maps of the P_{frm} sequence in each of eight sorting bins in the NEB5α strain under induced (100 µM formaldehyde) conditions. Mutated promoters isolated from low GFP expression bins are represented in the top heatmaps 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.

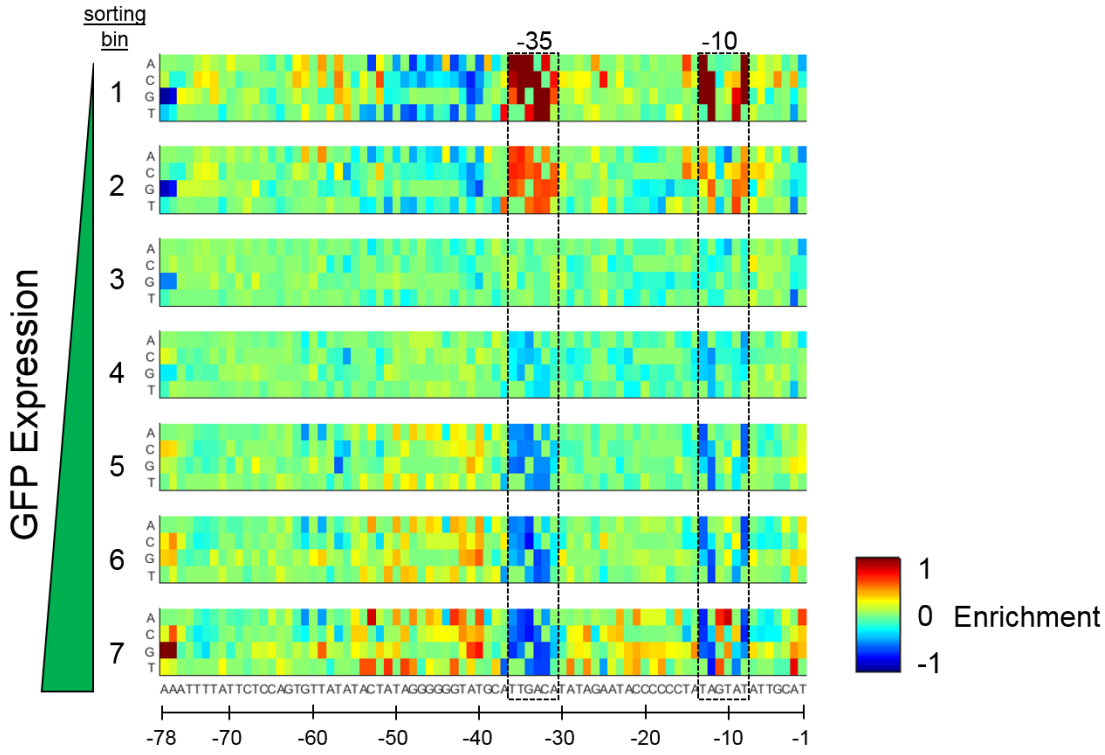

Figure A.11 Identification of up and down mutations through sequence analysis. Heat maps of the P_{frm} sequence in each of seven sorting bins in the Δ*frmR* strain with plasmid expressed FrmR under uninduced conditions. Mutated promoters isolated from low GFP expression bins are represented in the top 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.

Table A.1 Strains and plasmids used in this study, their relevant characteristics, and sources.

Strain or plasmid	Relevant characteristics	Source
E. coli strains		
BW25113	F-, *Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻, rph-1, Δ(rhaD-rhaB)568, hsdR514*	Datsenko and Wanner, 2000
Δ*frmR*	BW25113; ΔfrmR (b0357), KanR	Baba et al, 2006
Δ*frmA*	BW25113; ΔfrmA (b0356), KanR	Baba et al., 2006
Δ*frmA* Δ*pgi*	BW25113; ΔfrmA (b0356), Δpgi (b4025), KanR	This study
NEB5α	F-, *fhuA2, Δ(argF-lacZ)U169, phoA, glnV44, Φ80, Δ(lacZ)M15, gyrA96, recA1, relA1, endA1, thi-1, hsdR17*	NEB
Plasmids		
pETM6	AmpR, PT7, ColEI ori	Xu et al., 2012
P_{frm}-GFP	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study (Part ID: ACS_000558)
P_{frm}1-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}2-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}3-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}4-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}5-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}6-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}7-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}8-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}14-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}15-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}16-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}17-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}18-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}19-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study

P_{frm}20-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}21-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}21-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}22-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}23-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}24-GFP-P_{lac}-FrmR	pETM6 backbone (AmpR, ColEI ori)	This study
P_{frm}-Mdh-Hps-Phi	pETM6 backbone (AmpR, ColEI ori), *B. stearothermophilus mdh*, *B. methanolicus hps* and *phi*	This study (Part ID: ACS_000560)
P_{frm}20-Mdh-Hps-Phi	pETM6 backbone (AmpR, ColEI ori), *B. stearothermophilus mdh*, *B. methanolicus hps* and *phi*	This study

Table A.2 Primers used in this study and their functions.

Primer		Sequence (5'->3')	Function
1	pfrm_EPP_F	GAA GCA TAA AGT GTA AAG CCC TAG G	Amplify P_{frm} during error-prone PCR from P_{frm}-GFP-P_{lac}-FrmR plasmid backbone
2	pfrm_EPP_R	CCA TGC TAG GTT CTC CTT CTT AAT CTA GA	
3	vec_F_EPP	TCT AGA TTA AGA AGG AGA ACC TAG CAT GG	Amplify backbone of P_{frm}-GFP-P_{lac}-FrmR plasmid without pfrm
4	vec_R_EPP	CCT AGG GCT TTA CAC TTT ATG CTT C	
5	WT_F	ATA GAA TAC CCC CCT ATA GTA TAT TGC	Primers to generate P_{frm} constructs 1-8 (see Table A.3)
6	WT_R	ATG TCA ATG CAT ACC CCC CT	
7	Scrambled_R	ATG TCA ATG CCT AAT AGT GCA TAG CAT ATA ACA CTG GAG AAT AAA ATT TAT CC	
8	Scrambled_F	ATA GAG CTA TGC ACT ATA GTA TAT TGC ATT CTA GAT TAA GAA GG	
9	Shift_opt10_F	ATA CCC CCC TAT AGT ATA TAA TAT TGC ATT CTA GAT TAA GAA GGA G	
10	Flip_R	ATG TCA ATG CAC TAT AGG GGG GTA TAT ATA ACA CTG GAG AAT AAA ATT TAT CC	
11	Flip_F	ATA GAC TAT AGG GGG GTA TAA TAT TGC ATT CTA GAT TAA GAA GGA G	
12	down_F	ATA CTA TAC TCC CCT ATA GTA TAT TGC ATT C	

13	down_v1_R	ATG TCA AAC TAT ACT CCC CTA TAG TAT GAA ACA CTG GAG	Primers to generate P_{frm} constructs 14-25 (see Table A.3)
14	down_v2_R	ATG TCA AAC CAT ACT CCC CTA TGG TAT GAA ACA CTG GAG	
15	up_F	ATA GAA TAC CCC CCT ATA TAC TAT TGC TAT CTA GAT TAA GAA GGA G	
16	up_R	ATG TCA ATG CAT ACC CCC CTA TAG TAT ATA ACA CTG GAG AAT AAA ATC CAT CCG GTG AAT	
17	Range_F	ATA GTA TAC TCC CCT ATA TAC TAT TGC TAT CTA GAT TAA GAA GGA G	
18	Range_R	ATG TCA AAC TAT ACT CCC CTA TAG TAT GAA ACA CTG GAG AAT AAA ATC CAT CCG GTG AAT	
19	Pfrm20ins_F	tgc gtc cgg cgt agc cta ggA ACT TGC AGC CCG TCT GAC	Amplify P_{frm}20 with overhangs compatible with P_{frm}-Mdh-Hps-Phi plasmid
20	Pfrm20ins_R	ctc ctg cat ctc ttc gga gaT AGC AAT AGT ATA TAG GGG GGT ATT CTA TAT G	
21	pgi KO F	TCT CCG AAG AGA TGC AGG AGG	Used to knockout *pgi* in Δ*frmA* strain
22	pgi KO R	CCT AGG CTA CGC CGG ACG	
23	MeOHvec_F	TCT CCG AAG AGA TGC AGG AGG	Amplify backbone of P_{frm}-Mdh-Hps-Phi without promoter
24	MeOHvec_R	CCT AGG CTA CGC CGG ACG	

Table A.3 Primers used to generate P_{frm} constructs 1-8 and 14-25 through inverse PCR using the P_{frm}-GFP-P_{lac}-FrmR plasmid as template. The number for each primer refers to Table A.2.

Construct	F Primer	R Primer
1	WT_F (5)	Scrambled_R (7)
2	WT_F (5)	Flip_R (10)
3	Scrambled_F (8)	WT_R (6)
4	Scrambled_F (8)	Scrambled_R (7)
5	Scrambled_F (8)	Flip_R (10)
6	Shift_opt10_F (9)	WT_R (6)
7	Shift_opt10_F (9)	Flip_R (10)
8	Shift_opt10_F (9)	Flip_R (10)
14	down_F (12)	down_v1_R (13)
15	down_F (12)	down_v2_R (14)

16	down_F (12)	up_R (16)
17	down_F (12)	Range_R (18)
18	up_F (15)	down_v1_R (13)
19	up_F (15)	down_v2_R (14)
20	up_F (15)	up_R (16)
21	up_F (15)	Range_R (18)
22	Range_F (17)	down_v1_R (13)
23	Range_F (17)	down_v2_R (14)
24	Range_F (17)	up_R (16)
25	Range_F (17)	Range_R (18)

Table A.4 Fitted parameters for the Hill equation corresponding to Figure 2.2C.

Strain	Plasmid FrmR	*n*	*K* (μM)	*Pmin (a.u.)*	*Pmax (a.u.)*
NEB5α	+	1.18 ± 0.13	26.3 ± 15.6	212.6 ± 157.7	1211.0 ± 290.5
ΔfrmR	+	0.46 ± 0.02	189392 ± 322961	629.1 ± 273.6	38366 ± 53481

Table A.5 Spread of GFP fluorescence as determined by percent robust coefficient of variation.

Strain	Sample	%rCV	%rCV fold increase
NEB5α	Library	96.8	2.2
	Parent	44.4	
ΔfrmR	Library	75.7	2.0
	Parent	37.6	

Table A.6 Mutation bias of the unsorted P_{frm} library.

Mutation	Rate (%)
A->T, T->A	0.987
A->C, T->G	0.152
A->G, T->C	0.916
G->A, C->T	1.711
G->C, C->G	0.490
G->T, C->A	1.266

CHAPTER 3 SUPPLEMENTARY INFORMATION

Table B.1 Primers used in this study. For Gibson assembly reactions and knock-out primers, binding regions are black and homology regions are red. The Site-Directed Mutagenesis Kit (NEB) is referred to as SDMK.

Primer	Sequence (5'-3')	Purpose
Bm rpe F	CCG CCA ATT TCG CAC GCT TGG	qPCR
Bm rpe R	CGG AAG ATT GGT AAC GGG GCG	
Bm tkt F	GGG CTC TGG CCA CCC TGG TAT GCC	qPCR
Bm tkt R	GCA CTT AAT ACG AAG CGA TCA CGG	
Bm fba F	CCG GTA ATC ATC GGC GTC AGC	qPCR
Bm fba R	CCC ACG GAA TGT GCA ATT TCG	
Bm glpX F	GGA GAG GGG GAA ATG GAC GAA GCG	qPCR
Bm glpX R	GGT GCA TGA AGC AAA GAG CCG CG	
Bm pfk F	GGA TTA GAC GTT TAT GGG	qPCR
Bm pfk R	CCC TTC AAA TGT CCC GTC ACC ACC	
Bm ihfB F	CCC AGC AAT CGC ACA TTC CCG	qPCR
Bm ihfB R	CGT CCG GTA CGT GGT GCG	
pAC fgp F	GTC GAC CGG GTC GAA TTT GC	Amplify pAC backbone, pfrm insert, and *fba-glpX-pfk* insert for Gibson assembly to form plasmid pFGP with 3-fragment reaction
pAC fgp R	GCT ACG CCG GAC GCA TCG TG	
pfrm fgp F	CAC GAT GCG TCC GGC GTA GC AAC TTG CAG CCC GTC TGA C	
pfrm fgp R	TAA GAT GTC A ATG CAA TAT ACT ATA GGG GGG TAT TCT ATA TG	
fba glpX pfk F	TAT ATT GCA T TGA CAT CTT ATA AGG AGG AAA TAA AAT G	
fba glpX pfk R	CGA AAG CAA ATT CGA CCC GG GAA ACA AAA AAA CAC CCT TTC	
pAC pfrm F	CCG GGT CGA ATT TGC TTT C	

pAC pfrm R	ATG CAA TAT ACT ATA GGG GGG	Amplify pAC-pfrm backbone and *rpe-tkt* insert for Gibson assembly to form plasmid pRT with 2-fragment reaction
rpe tkt F	TAC CCC CCT ATA GTA TAT TGC AT AAC AAC TAA GGA GGT CCA ATA C	
rpe tkt R	CGA AAG CAA ATT CGA CCC GG GAA ACA AAA AAA CAC CCT TTC	
rpe vec F	TTA TTG AAC ACC CGA AAG GGT GTT TTT TTG	Amplify pRT plasmid and use SDMK to generate pR
rpe vec R	TCA ATT ACG AAT CTT AGC AAT TGC TTG AGA GC	
tkt vec F	CGT TGA GAA ATC TAA GGA CTC C	Amplify pRT plasmid and use SDMK to generate pT
tkt vec R	ATG CAA TAT ACT ATA GGG GGG TAT TCT ATA TG	
pFrt F	TTA TTG AAC ACC CGA AAG GGT G	Amplify pRT backbone and *fba* insert (from pFGP) for Gibson assembly to form plasmid pRTF
pFrt R	TTA CAA AAG TTT TTT GAA ATG AGA GAC TAC ATT CTC G	
fba ins F	CAT TTC AAA AAA CTT TTG TAA TGA CAT CTT ATA AGG AGG AAA TAA AAT GC	
fba ins R	CCC TTT CGG GTG TTC AAT AAT CAG GCT TTC CCC GAA GAT C	
glpX ins F	CAT TTC AAA AAA CTT TTG TAA AAA ATA ACA TTG AGG GAG AAA C	Amplify *glpX* insert (from pFGP) for Gibson assembly (w/ above pfrm rpe tkt vector) to form plasmid pRTG
glpX ins R	CCC TTT CGG GTG TTC AAT AAT TAG TGA TAA GCC TCA ATA AAT TG	
pfk ins F	CAT TTC AAA AAA CTT TTG TAA ATA ACA ACC GTT GGG GAG	Amplify *pfk* insert (from pFGP) for Gibson assembly (w/ above pfrm rpe tkt vector) to form plasmid pRTP
pfk ins R	CCC TTT CGG GTG TTC AAT AAT CAA AGG GAC AAA GAA CG	
maldh KO F	CAG CGG AGC AAC ATA TCT TAG TTT ATC AAT ATA ATA AGG AGT TTA GGA TGA TTC CGG GGA TCC GTC GAC C	*maldh* knock-out primers
maldh KO R	GCT CCG GTT TTT TAT TAT CCG CTA ATC AAT TAC TTA TTA ACG AAC TCT TCT GTA GGC TGG AGC TGC TTC G	
maldh upstream F	TTC TTG CTT AGC CGA GCT TC	Confirmation primers for *maldh* knock-out
maldh downstream R	GGG CAT TAA TAC GCT GTC GT	

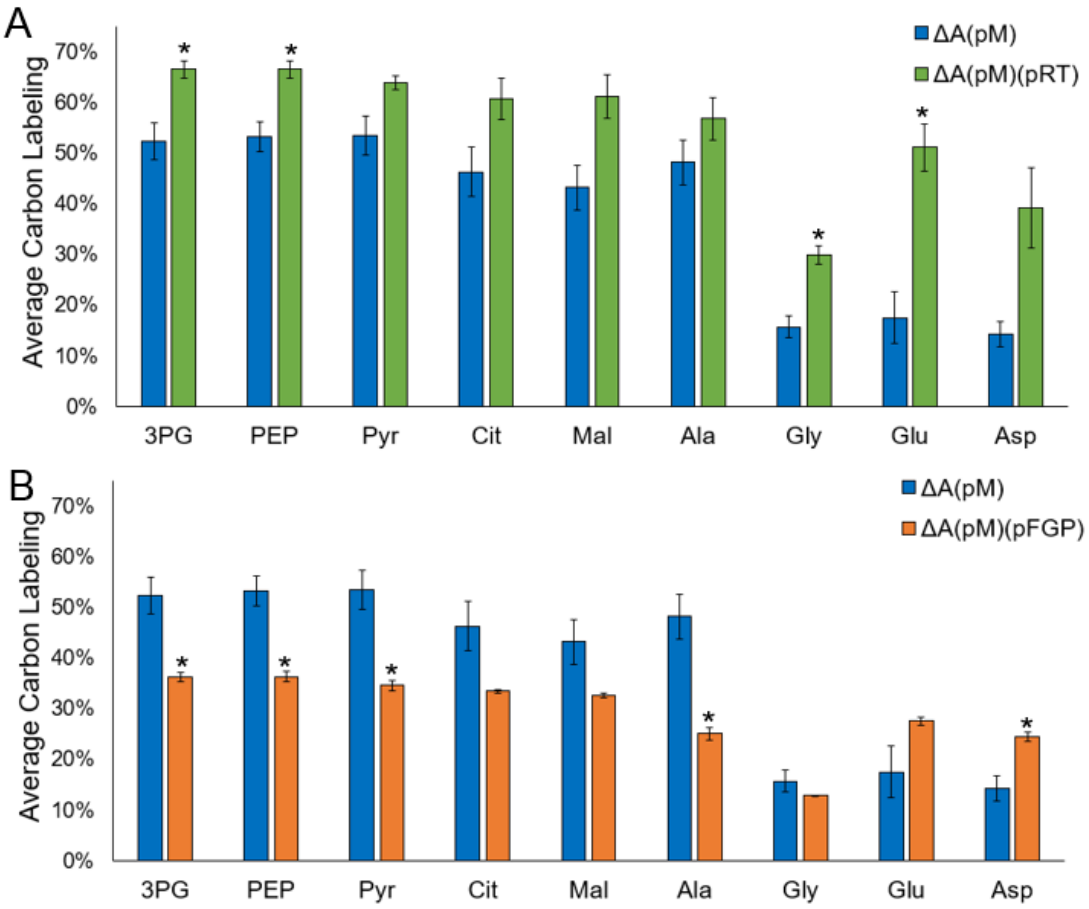

Figure B.1 Average carbon labeling in intracellular metabolites from ^{13}C-methanol after two days of growth on 0.5 g/L yeast extract and 100 mM ^{13}C-methanol. The ΔA(pM) strain is compared to the ΔA(pM)(pRT) (A) and the ΔA(pM)(pFGP) (B) strains. 3PG, 3-phosphoglycerate; PEP, phosphoenolpyruvate; Pyr, pyruvate; Cit, citrate; Mal, malate; Ala, alanine; Gly, glycine; Glu, glutamate; Asp, aspartate. Error bars represent standard deviation, n=2. Two-tailed *t*-tests indicate significance compared to ΔA(pM). * $P<0.05$.

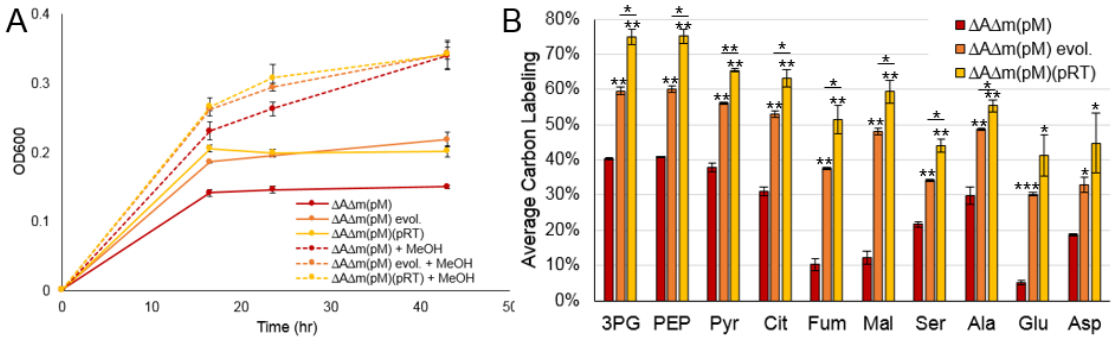

Figure B.2 (A) Growth curves in 0.5 g/L yeast extract with (dashed lines) and without (solid lines) 100 mM ^{13}C-methanol for the ΔAΔm(pM), ΔAΔm(pM) evol., and ΔAΔm(pM)(pRT) strains. (B) Average carbon labeling of intracellular metabolites from ^{13}C-methanol after two days of growth in 0.5 g/L yeast extract and 100 mM ^{13}C-methanol for the ΔAΔm(pM), ΔAΔm(pM) evol., and ΔAΔm(pM)(pRT) strains. 3PG, 3-phosphoglycerate; PEP, phosphoenolpyruvate; Pyr, pyruvate; Cit, citrate; Fum, fumarate; Mal, malate; Ser, serine; Ala, alanine; Glu, glutamate; Asp, aspartate. Error bars represent standard deviation, n=2. Two-tailed *t*-tests indicate significance compared to ΔAΔm(pM) unless indicated otherwise. * P<0.05, ** P<0.01, *** P<0.001.

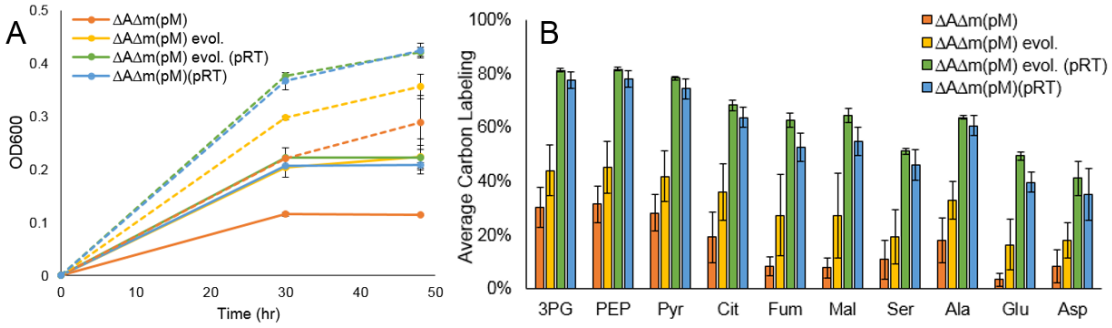

Figure B.3 (A) Growth curves in 0.5 g/L yeast extract with (dashed lines) and without (solid lines) 100 mM ^{13}C-methanol for the ΔAΔm(pM), ΔAΔm(pM) evol., ΔAΔm(pM) evol. (pRT), and ΔAΔm(pM)(pRT) strains. (B) Average carbon labeling of intracellular metabolites from ^{13}C-methanol after two days of growth in 0.5 g/L yeast extract and 100 mM ^{13}C-methanol for the ΔAΔm(pM), ΔAΔm(pM) evol., ΔAΔm(pM) evol. (pRT), and ΔAΔm(pM)(pRT) strains. 3PG, 3-phosphoglycerate; PEP, phosphoenolpyruvate; Pyr, pyruvate; Cit, citrate; Fum, fumarate; Mal, malate; Ser, serine; Ala, alanine; Glu, glutamate; Asp, aspartate. Error bars represent standard deviation, n≥2.

CHAPTER 4 SUPPLEMENTARY INFORMATION

Table C.1 Strains and plasmids used in this study.

Strain or plasmid	Relevant characteristics	Source or reference
	E. coli strains	
NEB5α	F-*, fhuA2, Δ(argF-lacZ)U169, phoA, glnV44, Φ80, Δ(lacZ)M15, gyrA96, recA1, relA1, endA1, thi-1, hsdR17*	NEB
BW25113	*Δ(araD-araB)567 ΔlacZ4787(::rrnB-3) λ- rph-1 Δ(rhaD-rhaB)568 hsdR514*	Datsenko and Wanner, 2000
Δ*frmA*	BW25113; ΔfrmA (b0356), KanR	Baba et al., 2006
Δ*frmA* Δ*pgi*	BW25113; ΔfrmA (b0356), Δpgi (b4025), KanR	Bennett et al., 2018
	Plasmids	
pAC_P_{frm}_GFP	Cm^R	This study
pETM6_P_{tac}_Mdh	pETM6-P_{tac}-*Gs mdh*, Amp^R	Whitaker et al., 2017
pETM6_P_{tac}_Mdh-his	pETM6-P_{tac}-*Gs mdh* his-tagged, Amp^R	This study
pETM6_P_{tac}_V5I Mdh-his	pETM6-P_{tac}-*Gs mdh* V5I his-tagged, Amp^R	This study
pETM6_P_{tac}_L13M Mdh-his	pETM6-P_{tac}-*Gs mdh* L13M his-tagged, Amp^R	This study
pETM6_P_{tac}_R20H Mdh-his	pETM6-P_{tac}-*Gs mdh* R20H his-tagged, Amp^R	This study
pETM6_P_{tac}_P50S Mdh-his	pETM6-P_{tac}-*Gs mdh* P50S his-tagged, Amp^R	This study
pETM6_P_{tac}_V64M Mdh-his	pETM6-P_{tac}-*Gs mdh* V64M his-tagged, Amp^R	This study
pETM6_P_{tac}_V68I Mdh-his	pETM6-P_{tac}-*Gs mdh* V68I his-tagged, Amp^R	This study
pETM6_P_{tac}_V116C Mdh-his	pETM6-P_{tac}-*Gs mdh* V116C his-tagged, Amp^R	This study
pETM6_P_{tac}_Q227R Mdh-his	pETM6-P_{tac}-*Gs mdh* Q227R his-tagged, Amp^R	This study
pETM6_P_{tac}_V308L Mdh-his	pETM6-P_{tac}-*Gs mdh* V308L his-tagged, Amp^R	This study
pETM6_P_{trc}_Mdh-Hps-Phi	pETM6-P_{trc}-*Gs mdh -Bm hps -Bm phi*, Amp^R	Bennett et al., 2018

pETM6_P_{trc}_V5I Mdh-Hps-Phi	pETM6-P_{trc}-*Gs mdh* V5I -*Bm hps* -*Bm phi*, Amp^R	This study
pETM6_P_{trc}_L13M Mdh-Hps-Phi	pETM6-P_{trc}-*Gs mdh* L13M -*Bm hps* -*Bm phi*, Amp^R	This study
pETM6_P_{trc}_R20H Mdh-Hps-Phi	pETM6-P_{trc}-*Gs mdh* R20H -*Bm hps* -*Bm phi*, Amp^R	This study
pETM6_P_{trc}_P50S Mdh-Hps-Phi	pETM6-P_{trc}-*Gs mdh* P50S -*Bm hps* -*Bm phi*, Amp^R	This study
pETM6_P_{trc}_V64M Mdh-Hps-Phi	pETM6-P_{trc}-*Gs mdh* V64M -*Bm hps* -*Bm phi*, Amp^R	This study
pETM6_P_{trc}_V68I Mdh-Hps-Phi	pETM6-P_{trc}-*Gs mdh* V68I -*Bm hps* -*Bm phi*, Amp^R	This study
pETM6_P_{trc}_V116C Mdh-Hps-Phi	pETM6-P_{trc}-*Gs mdh* V116C -*Bm hps* -*Bm phi*, Amp^R	This study
pETM6_P_{trc}_Q227R Mdh-Hps-Phi	pETM6-P_{trc}-*Gs mdh* Q227R -*Bm hps* -*Bm phi*, Amp^R	This study
pETM6_P_{trc}_V308L Mdh-Hps-Phi	pETM6-P_{trc}-*Gs mdh* V308L -*Bm hps* -*Bm phi*, Amp^R	This study
pAC_P_{trc}_phaA_phaB_bld_P_{trc}_phaJ_ter	pAC-P_{trc}-*Cn phaA* -*Cn phaB* -*Cs bld* -P_{trc}-*Ac phaJ* -*Td ter*, Cm^R	Bennett et al., 2018

Table C.2 Primers used in this study. For Gibson assembly reactions binding regions are black and inserted, mutated, or homology nucleotides are red. The Site-Directed Mutagenesis Kit (NEB) is referred to as SDMK.

Primer	Sequence (5'-3')	Purpose
his_BsMdh_F	CAT CAC CAT CAT CAC CAC TAA CTC GAG TCT GGT AAA GAA ACC G	Create pAC_P_{tac}_Mdh-his with SDMK
his_BsMdh_R	C TTC TTT CAG TTT CAG CAC GAT ACG ACC	
Mdh_EPP_F	CCC TCT AGA AAT AAT TTT GTT TAA CTT TAA GAA GGA GAT ATA CAT ATG	Amplify Mdh during error-prone PCR from pAC_P_{tac}_Mdh
Mdh_EPP_R	CGC AGC AGC GGT TTC TTT ACC AGA CTC GAG TTA	
vec_F_EPP	TAA CTC GAG TCT GGT AAA GAA ACC GCT GCT GCG	Amplify backbone of pAC_P_{tac}_Mdh plasmid without Mdh to clone EPP library with 2 fragment Gibson assembly
vec_R_EPP	CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG	
ORI_F	ACC TCG CTC TGC TAA TCC TGT TAC C	use ORI_F with variant R primers and ORI_R with variant F primers to create Mdh variants with a 2-fragment Gibson assembly
ORI_R	GGT AAC AGG ATT AGC AGA GCG AGG	
V5I_F	ATT GTC AAT GAA TTT AAG AAA GCC CTG GAA ATC	Make V5I mutation with ORI_F and ORI_R primers
V5I_R	TTC TTA AAT TCA TTG ACA ATC GCT GCT TTC ATA TGT ATA TCT CC	
R20H_F	CAC CCG AAA CTG GAA GAA GGT GAA G	Make R20H mutation with ORI_F and ORI_R primers
R20H_R	CTT CCA GTT TCG GGT GTT CCA CTT CTT TGA TTT CC	
P50S_F	TCG ATT AAA CCG AAA CTG CCG CTG ATC C	Make P50S mutation with ORI_F and ORI_R primers
P50S_R	GTT TCG GTT TAA TCG ACC AGT CAC CGT GGG C	
V64M_F	ATG GGT ATC GTG GTT GAA GTG GCC	Make V64M mutation with ORI_F and ORI_R primers
V64M_R	AAC CAC GAT ACC CAT GCC TTC GTG ACC	
V68I_F	ATT GAA GTG GCC AAA GGT GTT AAA TCA ATT AAA GTC GG	Make V68I mutation with ORI_F and ORI_R primers
V68I_R	TTA ACA CCT TTG GCC ACT TCA ATC ACG ATA CCC ACG	
V116C_F	TGT GAT GGC GGT TAT GCA GAA TAC TGC	

V116C_R	TAT TCT GCA TAA CCG CCA TCA CAG GAA TAA CCG CCG TTC	Make V116C mutation with ORI_F and ORI_R primers
Q227R_F Q227R_R	CGT GTT GGC GGT GTC CAC GC CGT GGA CAC CGC CAA CAC GGT CAT GAA TCG CTT TAA C	Make Q227R mutation with ORI_F and ORI_R primers
V308L_F V308L_R	CTC GAA ACG GCG GAA CTG GAA GAA ATC AGT TCC GCC GTT TCG AGA ATC GGG CGA AC	Make V308L mutation with ORI_F and ORI_R primers
Mdh_F Mdh_R	TCC GAA GAG ATG CAG GAG GAG CTA CAT GAA AGC AGC GGT TGT C GTA CGT CCT CCT TCT AGT CAT CTC ATT AAT CTT CTT TCA GTT TCA GCA C	Amplify parent or variant Mdh w/ overhangs to make pETM6_P_{trc}_Mdh_Hps_Phi plasmids with 2-fragment Gibson assembly
pM6_hps phi_F pM6_hps phi_R	TGA GAT GAC TAG AAG GAG GAC GTA GCT CCT CCT GCA TCT C	

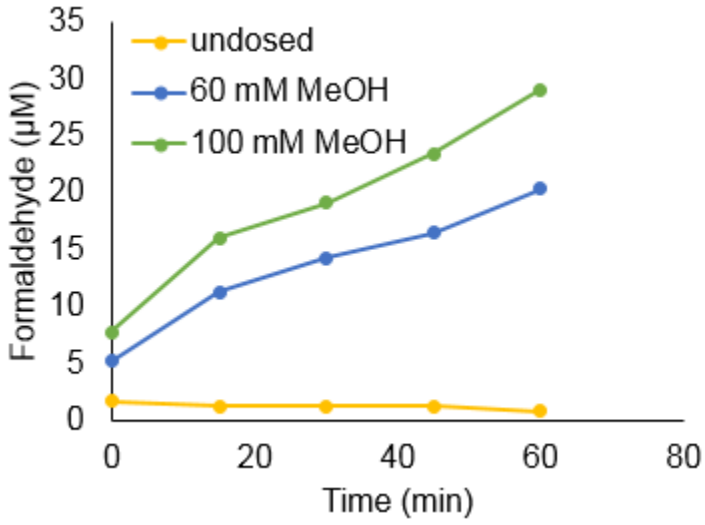

Figure C.1 Formaldehyde production after inoculation into M9 minimal medium supplemented with 1 g/L yeast extract with 0, 60, or 100 mM methanol, for Δ*frmA*(pETM6_P_{tac}_Mdh)(pAC_P_{frm}_GFP) parent strain, determined using Nash assay.

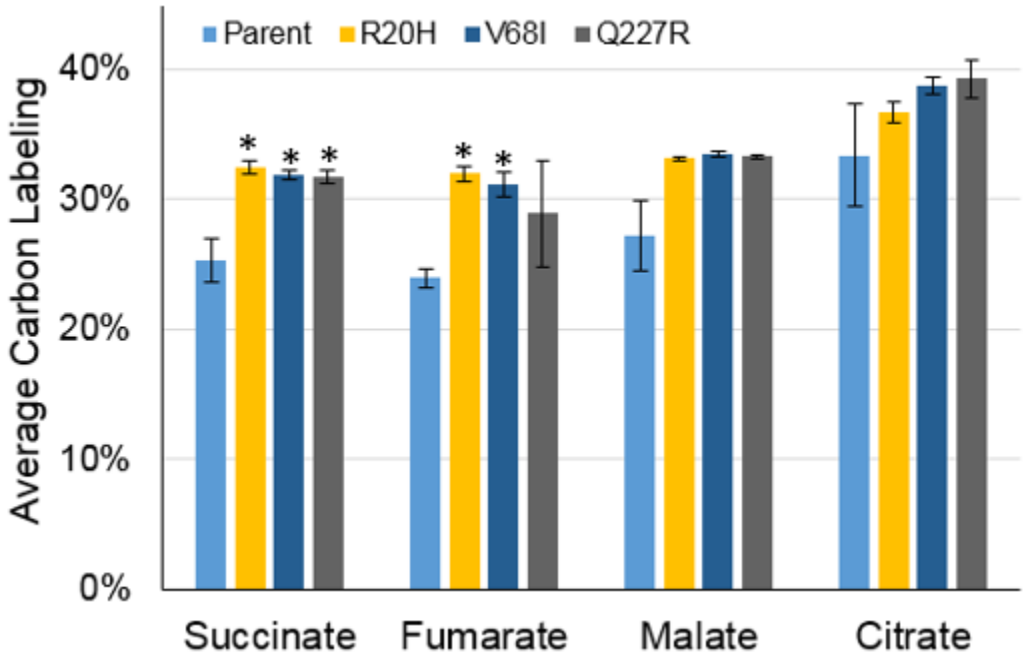

Figure C.2 ^{13}C-methanol incorporation into intracellular metabolites for *E. coli* with the parent Mdh or Mdh variants R20H, V68I, or Q227R, after 30 hours of growth with 100 mM ^{13}C-methanol and 0.5 g/L yeast extract. Error bars indicate standard deviation, n=2. Two-tailed *t*-tests indicate significance compared to the parent strain. * P<0.05.

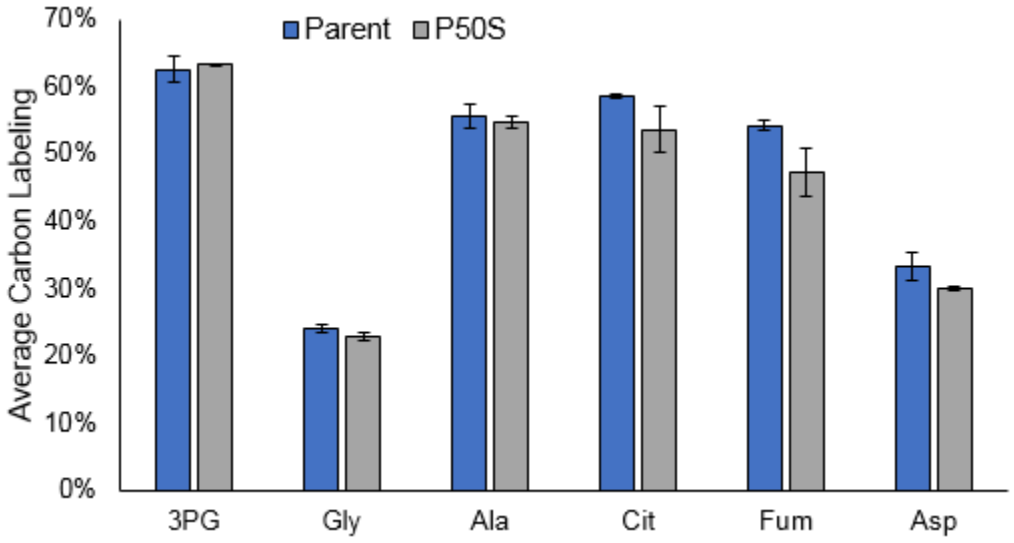

Figure C.3 ^{13}C-methanol incorporation into intracellular metabolites for *E. coli* with the parent Mdh and P50S Mdh variant after 44 hours of growth with 100 mM ^{13}C-methanol and 0.5 g/L yeast extract. 3PG, 3-phosphoglycerate; Gly, glycine; Ala, alanine; Cit, citrate; Fum, fumarate; Asp, aspartate. Error bars indicate standard deviation, n=2.

PUBLICATION PERMISSIONS

Dear Julia Rohlhill,

Your permission requested is granted and there is no fee for this reuse. In your planned reuse, you must cite the ACS article as the source, add this direct link https://pubs.acs.org/doi/abs/10.1021/acssynbio.7b00114, and include a notice to readers that further permissions related to the material excerpted should be directed to the ACS.

If you need further assistance, please let me know.

Sincerely,

Raquel Picar-Simpson
ACS Publications Support
Customer Services & Information
Website: https://help.acs.org/

Incident Information:

Incident #: 2940297
Date Created: 2019-08-11T11:48:26
Priority: 3
Customer: Julia Rohlhill
Title: Copyright permission request -- DOI: 10.1021/acssynbio.7b00114
Description: To whom it may concern:

I am seeking permission to reuse content from the following article:
https://pubs.acs.org/doi/abs/10.1021/acssynbio.7b00114

I wish to reuse the entire article for my dissertation at the University of Delaware to satisfy requirements for my degree.

Thanks,
Julia Rohlhill
PhD Candidate
Papoutsakis Lab
University of Delaware