Enabling C1-Based Bioconversion with Metabolic Engineering

by

Benjamin Michael Woolston

B.S., The Pennsylvania State University, 2011

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

Author . . .
> **Signature redacted**
> . . . .
Department of Chemical Engineering
May 19, 2017

Certified by . . . .
> **Signature redacted**
> . . . .
Gregory Stephanopoulos
Willard Henry Dow Professor in Chemical Engineering
Thesis Supervisor

Accepted by . . . .
> **Signature redacted**
Richard D. Braatz
Edward R. Gilliland Professor of Chemical Engineering
Chairman, Committee for Graduate Students
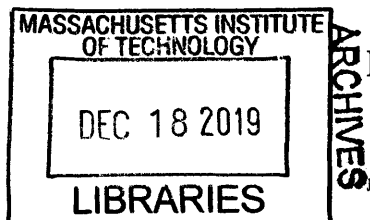
Enabling C1-Based Bioconversion with Metabolic Engineering

by

Benjamin Michael Woolston

Submitted to the Department of Chemical Engineering
on May 22, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Chemical Engineering

Abstract

Single-carbon (C1) substrates, such as synthesis gas and methanol, are attractive feedstocks for biochemical processes, as they are widely available, can be produced renewably, and do not compete with food supply. However, their use in industrial bioprocessing remains limited, primarily because microbes that utilize these substrates are poorly characterized biochemically, and limited tools exist for their genetic modification. This leaves the metabolic engineer with a choice: to develop genetic tools to enable engineering in the desired host, or to import the relevant catabolic pathway into a more tractable organism, such as *Escherichia coli*. This thesis explores both options within the context of developing strains for the conversion of C1 substrates into value-added chemicals and fuels.

Clostridium ljungdahlii is an acetogen that grows autotrophically on synthesis gas (CO, H_2, and CO_2) using the Wood-Ljungdahl pathway, and is a promising candidate for non-photosynthetic CO_2 fixation. In the first section, we extended its primitive genetic tools by developing a CRISPRi system for the targeted knockdown of specific genes. Constitutive downregulation of several genes with putative roles in energy conservation and carbon flux by up to 30-fold was demonstrated, and the associated phenotypes analyzed. Optimization of the promoter controlling dCas9 expression allowed for inducible knockdown, paving the way for dynamic metabolic control strategies to redirect carbon flux in engineered strains. To demonstrate this concept, several variants of a heterologous pathway for the biosynthesis of 3-hydroxybutyrate (3HB) were constructed, to probe 3HB production in the wild-type background and with various CRISPRi plasmids. The CRISPRi system represents a valuable contribution to the metabolic engineering field for its ability to redirect carbon flux, and is also useful to the microbiology community to probe gene function to answer open questions in the biochemistry underlying the Wood-Ljungdahl pathway. Efforts to develop genetic tools for the related acetogen *Moorella thermoacetica* are also described.

To explore the alternative approach of importing a single-carbon catabolic pathway into a tractable host, in the second section *E. coli* was engineered to metabolize methanol. Screening various candidates of the three heterologous pathway enzymes enabled robust incorporation of ^{13}C-labeled methanol into central carbon metabolism.

To further improve methanol assimilation, a kinetic-thermodynamic modeling framework was developed and combined with novel isotopic tracing experiments to probe potential pathway limitations. Flux leakage from the cyclical ribulose monophosphate (RuMP) pathway was identified as the primary bottleneck, as this led to the build-up of the toxic intermediate formaldehyde and ablation of the thermodynamic driving force for methanol oxidation. Strategies were developed to re-wire central metabolism accordingly, which restored the driving force and led to the identification of the kinetics of the first enzyme of the pathway - methanol dehydrogenase (MDH) - as the next limitation. These results represent the first systematic analysis of flux limitations in *E. coli* engineered for methanol metabolism, and provide clear targets for further metabolic engineering to enable synthetic methylotrophy. Finally, the development of a formaldehyde biosensor in support of evolutionary approaches to enhance MDH activity and partition carbon flux between formaldehyde assimilation and growth is described.

Thesis Supervisor: Gregory Stephanopoulos
Title: Willard Henry Dow Professor in Chemical Engineering

Acknowledgments

Graduate school has been an incredible journey, primarily because of the many people who have helped along the way. First and foremost, my research advisor, Professor Stephanopoulos, has provided me the resources and encouragement to pursue several diverse projects during my time here, allowing my research to freely develop along lines I find fruitful and interesting. He has provided me ample opportunities to teach, to mentor students, and to gain experience in grant writing. His support, enthusiasm, and wisdom have been invaluable.

I am also grateful to the members of my thesis committee. Professor Sinskey has challenged me from the very beginning to work on projects with a strong fundamental component, and I hope I have lived up to his expectations in that regard. Professor Prather has always asked insightful technical questions and had excellent advice. Professor Drennan has provided a welcome fresh perspective from her chemistry background, and sound career advice. I would like to additionally thank Professors Sinskey and Prather for allowing me to develop my teaching practice in their classrooms, and their valuable feedback on these lectures.

I owe a great debt of gratitude to the members of the Stephanopoulos lab, both for their assistance, and for putting up with me for the last six years. Dr. Hamid Rismani took me under his wing when I joined the lab, and taught me the techniques required for anaerobic microbiology. Working with Hamid, an experienced and capable researcher, helped me learn how to conduct good research early in my graduate school career. Dr. Tom Wasylenko got me started on the analysis of intracellular metabolites by LC-MS/MS, and without his patient advice I would not have been able to conduct many of the experiments in this thesis. I also respected and learned a lot from Tom's methodical approach to understanding metabolism. Over the last two years, I have come to rely on the insight of Dr. Jun Park, whose extensive experience with isotopic tracers and pathway thermodynamics has guided a large portion of my work. Dr. Steven Edgar joined the lab at the same time I did, and we shared many good times discussing research and troubleshooting each other's experiments

throughout graduate school. I am particularly thankful to the members of the RE-MOTE subgroup: Dr. Devin Currie, Dr. Jason King, and David Emerson. All three of them have contributed directly to the work in this thesis, and it has been fantastic to work with each of them for the last few years, and to call them friends. Devin joined the lab as a microbiologist interested in the applied side of industrial microbiology, and at the time I was a chemical engineer trying to develop more basic microbiology skills. It has been fun to watch these opposite trajectories meet somewhere in the middle. Jason's chemistry knowledge has been an asset not just to me, but to the entire lab, and I have enjoyed our many wide-ranging discussions, from optimal lab data management strategies, to the general role of water in biology. David's attention to detail has been immensely helpful in many of the projects in this thesis, and to the writing of the grant proposal that funded them. His willingness to perform qRT-PCR is something for which I am extremely grateful! I would also like to thank the many people in the lab who contributed to the enjoyable atmosphere, from the board game nights, to the karaoke trips, and the (many) Mead Hall lunches and dinners. Dr. Brian Pereira deserves special mention for more-or-less single handedly keeping the lab running during my time here.

I also wish to thank the various students I have had the privilege of mentoring. Ishwar Kohale's efforts on the early formaldehyde biosensor went above and beyond what I could have expected for a UROP, and I have no doubt he will succeed in his own graduate school experience. Michael Reiter made great strides on *E. coli* methanol metabolism, and to this day I still can't match his speed in performing metabolite extractions. Bob van Hove has only been at MIT for a short time, but has already contributed significantly to this work, and I look forward to our continued collaboration in my Post-Doc.

I would also like to thank my collaborators, Tim Roth and Professor David Liu. Working with them on the formaldehyde biosensor has given me an insight into the thought process of the synthetic biologist, and their efforts have dramatically improved the basic system that we started with.

Graduate school can be tough at times, and having an escape mechanism is impor-

tant. I have been very fortunate for the life-long friendships I've developed through the MIT Cycling Team and the Triathlon Team, and the fun times we've had racing up and down the East coast, at training camps in California, and during the hours spent riding around Massachusetts (often looking for ice cream).

To my friends and classmates: thank you for the many fun times we've shared, for commiserating with me in the bad times, and for helping me through the first semester classes. Ray Smith, Sakul Ratanalert, and Katie Maass in particular - thank you for always being there. To John Costello, the last three years have been amazing, and I'm lucky to have you in my life. Thank you for putting up with the many late nights in labs, the unforeseen LC-MS emergencies, the uncooperative cells, and my inability to predict how long lab-related tasks will take.

Finally, thank you to my parents - Sally and Crispin Woolston - who have always loved, encouraged and supported me. Your prediction that I 'might enjoy molecular biology' turned out to be completely accurate, despite my emphatic objections during high school. I wouldn't be where I am today without your guidance.

Dedication

This thesis is dedicated to the memory of my grandfather, Michael Richard Curson, whose patience, work ethic and moral fiber are a continual source of inspiration. I know exactly what he would say if I told him this degree made me an expert...

Contents

List of Figures

15

17

List of Tables

Chapter 1

Introduction

1.1 Summary

Concerns over the depletion of readily available fossil fuels and their effects on environmental pollution and climate change have led to heightened interest in alternative sources of liquid fuels, such as biofuels [1]. The first generation of biofuels utilized easily accessible sugars as feedstocks for the production of compounds such as ethanol and biodiesel. However, the sources of these sugars were generally food crops (e.g. corn, sugarcane), thus advanced biofuels are now being sought that do not compete with food production [2]. Various alternative options are under investigation, including lignocellulosic biomass [3], and photosynthetic microalgae [4], but commercial success has so far been limited. In this thesis, alternative bioprocesses based on single carbon compound (C1) feedstocks are investigated. These substrates can be supplied in aggregate large quantities and can be produced renewably, but their use in commercial bioprocesses is limited by the poor understanding of the organisms that consume them, and the lack of metabolic engineering tools for their genetic manipulation. This work addresses this limitation by developing improved genetic tools for acetogen metabolic engineering, and by developing strains of the biotechnology workhorse *E. coli* for the consumption of methanol.

In this chapter, the use of C1 compounds is motivated by analysis of their production and availability. A thorough overview of what is known about the metabolic

pathways involved in C1 metabolism follows, with specific attention paid to the Wood-Ljungdahl pathway of acetogenesis, and the pathways of aerobic methylotrophy. The chapter concludes with an assessment of the state-of-the-art in acetogen and methylotroph metabolic engineering, identifying the critical gaps that are adressed in the rest of the thesis.

1.2 C1 Substrates as Promising Feedstocks for Renewable Fuel and Chemical Production

C1 compounds vary in oxidation state all the way from +4 (CO_2) to -4 (CH_4), and thus represent a highly diverse class of energy-containing compounds with a wide range of reactivity. Various microbes have evolved exquisite enzyme systems to enable growth on C1 compounds of every oxidation state. As our understanding of these metabolic processes grows, so too does the ability to leverage the ability of these microbes to catalyze the conversion of these substrates into value-added chemicals and fuels. The potential benefits of the most common C1 substrates are described below.

1.2.1 Methane (CH_4, -4)

Over the last decade, widespread implementation of hydraulic fracturing (fracking) has increased the availability of unconventional natural gas in the United States. On an energy basis, this resource could meet the demands of the transportation industry for the next 50 years [5]. However, conversion of natural gas to a fuel of sufficient energy density remains problematic. Chemical conversion through the capitally intensive Fischer-Tropsch synthesis is economically viable only at large scale [6], incompatible with the distributed nature of the resource [7]. Compression to LNG and CNG is an expensive endeavor, further mired by a lack of refueling infrastructure [8]. In the absence of an economically viable alternative, up to 32% of the natural gas produced in some areas of the country is currently flared as a waste product [9]. A low-capital process for converting methane to a liquid fuel, compatible with existing

engine design, can therefore be a disruptive technology, with the potential to tap this underutilized resource and reduce the environmental impact of fracking [10]. A biotechnological process using an engineered microbe to produce fuel from methane could meet this need [11, 12].

1.2.2   Methanol (CH$_3$OH, -2)

With a worldwide production of 53 million tons per year, methanol represents a promising candidate for biochemical production [13]. Its current price is comparable to that of glucose, despite being more energy-dense, with 50% more electrons per carbon. Methanol can be produced renewably from municipal solid wastes and biogas through synthesis gas (syngas) as an intermediate, as exemplified by Enerkem. It can also be produced using renewable electricity from CO$_2$, for instance in the process of CRI in Iceland. Conversion of methane to methanol as a starting material for value-added processes has also been proposed as a way to monetize stranded natural gas. For these reasons, the concept of a 'Methanol Economy', advocated by George Olah [14], has received considerable attention. Until recently, limited understanding of methanol metabolism limited the industrial use of methylotrophs to the production of Single Cell Protein (SCP) in the late twentieth century [15]. Methanol bioprocessing is also associated with technical challenges, including high oxygen requirements and concomitant heat production, which leads to high cooling requirements [16].

1.2.3   Formate (HCOOH, +2)

Formate is an attractive substrate for bioprocessing. It can be produced from highly available resources, including electro- or photochemical reduction of CO$_2$, hydrogenation of CO$_2$, selective oxidation of biomass, or the partial oxidation of natural gas [17]. The majority of these processes are renewable and under active development. It is metabolized by a wide range of organisms, as discussed below, presenting many options for its conversion to value-added products.

1.2.4   Carbon monoxide (CO, +2)

A variety of organisms, including the acetogens and the so-called 'Knallgas bacteria', grow robustly using carbon monoxide as their sole carbon source, despite its general toxicity. Carbon monoxide is the main component of syngas, thus biological conversion of this substrate is often referred to as 'Biological Fischer-Tropsch', or 'Bio-GTL'. An advantage of the biological version of the process is the increased product specificity: Whereas the industrial chemical process produces hydrocarbons of a variety of chain lengths, which requires costly separation and markets for each species, acetogeneis can theoretically produce a single targeted compound. The organisms are also generally less sensitive to low-level contamintants in the gas stream that poison the catalytic process [18]. Syngas is produced primarily from methane, and therefore not considered renewable. However, biomass gasification for renewable syngas production, though not currently economically viable, is under intensive development [19]. In addition, steel-mill offgas has composition similar to syngas, and is therefore a negative cost substrate, as exemplified by the demonstration plant of LanzaTech which uses waste gas from Bao Steel as a substrate for fermentation [20]. The key challenges associated with syngas fermentation are the poor solubility of CO in water, which necessitates very high mass transfer rates ($k_L a$) to sustain reasonable productivities, and the limited genetic tools for engineering acetogens.

1.2.5   Other C1 substrates

This thesis is restricted to discussion of the C1 substrates described above, but there are others that can also support growth that are worth mentioning here for completeness. Formaldehyde (CH_2O) is highly toxic and not generally considered a growth substrate, but various microbes can grow using formaldehyde as the sole carbon and energy source when the concentration is kept low, as in a chemostat [21]. Aminated methyl compounds (methylamine, dimethylamine, and trimethylamine) serve as growth substrates in some aerobic methylotrophs [22] and anaerobic methanogens [23]. Some halogenated methyl compounds (e.g. dichloromethane) can also support

24

growth in organisms such as *Dehalicoccoides mccartyi* [24].

In the next section, the biochemical basis for microbial C1 metabolism is discussed, with emphasis on the substrates described above.

1.3 Pathways for C1 Assimilation

1.3.1 The Wood-Ljungdahl Pathway

The process of CO_2 fixation

The reductive acetyl-CoA pathway, also known as the Wood-Ljungdahl pathway after the investigators who unraveled the principal enzymatic pathway steps, catalyzes the fixation of two molecules CO_2 into the acetyl moiety of acetyl-CoA (AcCoA), using electrons derived from a variety of organic and inorganic sources (Figure 1-1). The pathway is divided into two branches: the Western (Carbonyl) branch, and the Eastern (Methyl) branch. In the Western branch, CO_2 is reduced to CO at the expense of reduced ferredoxin (Fd_{red}) in a reaction catalyzed by the bifunctional enzyme complex of acetyl-CoA synthetase (ACS) and carbon monoxide dehydrogenase (CODH) [25]. In the Eastern branch, CO_2 is reduced to formate, which is then activated by ATP and transferred to the C1 carrier tetrahydrofolate (THF). A series of reduction steps lead to the formation of methyl-THF, which is then delivered to the ACS/CODH by a corrinoid methyltransferase protein, where it is condensed with CO produced in the Western branch and free coenzyme A to yield AcCoA. AcCoA can be used either for the production of biomass, or converted to acetate to generate ATP. Overall, the stoichiometry of acetogenesis from H_2 and CO_2 can be represented as:

$$2CO_2 + 4H_2 \rightarrow CH_3COOH + 2H_2O \quad (\Delta G^{0'} = -95 \ kJ \ mol^{-1}) \qquad (1.1)$$
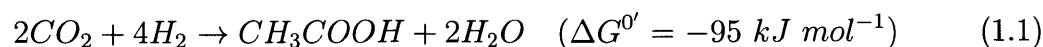
The negative free energy change of the process allows acetogenesis to be coupled to energy conservation. The Wood-Ljungdahl pathway is therefore unique among carbon fixation pathways in that it can support cell growth. It also represents the most
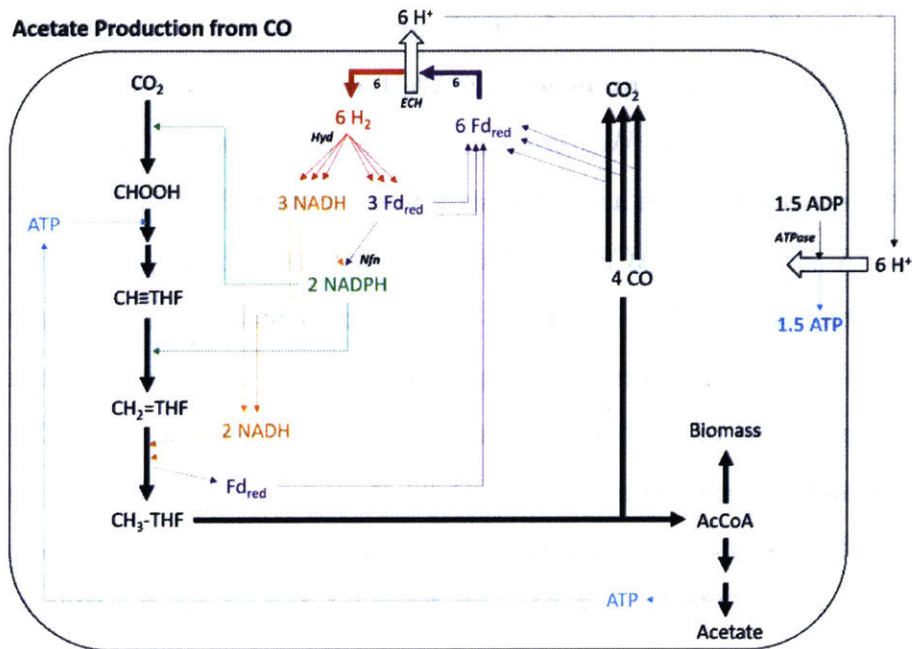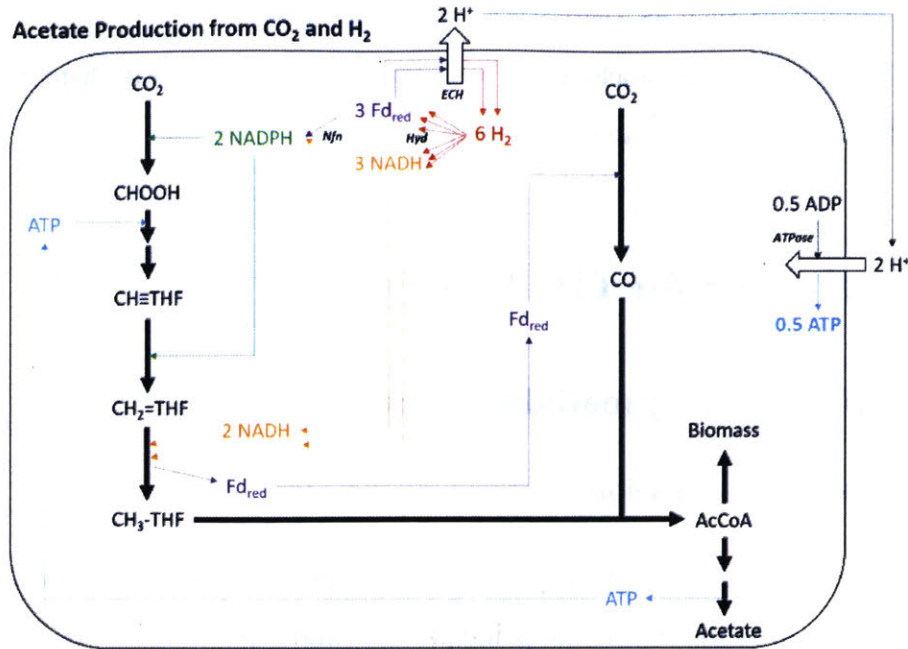
Figure 1-1: **The Wood-Ljungdahl Pathway in *Moorella thermoacetica*.** Biochemical pathways showing the energy-conserving production of acetate from CO_2 and H_2 (top) and CO (bottom). Black lines represent carbon flux, and colored lines represent interconversion of electron carriers.

energy-efficient pathway for CO_2 fixation [26], and is therefore of practical interest for the renewable production of chemicals and fuels.

26

Electron bifurcation and energy conversation in acetogenesis

How exactly energy is conserved through acetogenesis has been a long-standing question in the field, and has implications for the industrial potential of the pathway [27]. The ATP produced in the conversion of acetyl-CoA to acetate is used earlier in the pathway for the activation of formate, ruling out substrate-level phosphorylation as the mechanism. In *Moorella thermoacetica* (f. *Clostridium thermoaceticum*), the organism that was used by Wood and Ljungdahl in the initial characterization of the pathway, the presence of cytochromes and quinones originally suggested a membrane-dependent electron transport chain [28]. Determination of the redox potentials of the various reductive reactions suggested that only the reduction of methylene-THF to methyl-THF could be associated with energy conservation [29], though no mechanism was put forth. Analysis of the genomes of two other model acetogens, *Acetobacterium woodii*, and *Clostridium ljungdahlii*, however, showed they do not contain cytochromes, ruling out a membrane-dependent electron transport chain in these organisms [30, 31].

We now know that energy conservation in all three organisms is mediated by the recently described phenomenon of electron bifurcation. In this mechanism, which is being hailed as a third mode of energy conservation in addition to substrate-level phosphorylation and oxidative phosphorylation [32], an endergonic redox reaction is coupled with an exergonic one in a single flavin-dependent soluble enzyme complex, allowing energy conservation without a membrane. Electrons from a mid-potential donor (e.g. H_2) are bifurcated between a high-potential acceptor (e.g. NADH) and a low-potential acceptor (ferredoxin):

$$H_2 + Fd \rightarrow Fd^{2-} + 2H^+ \quad (\Delta G^{0'} = +7 \ kJ \ mol^{-1}) \tag{1.2}$$

$$H_2 + NAD^+ \rightarrow NADH + H^+ \quad (\Delta G^{0'} = -18 \ kJ \ mol^{-1}) \tag{1.3}$$

$$2H_2 + Fd + NAD^+ \rightarrow Fd^{2-} + NADH + 3H^+ \quad (\Delta G^{0'} = -11 \ kJ \ mol^{-1}) \tag{1.4}$$
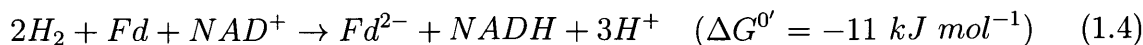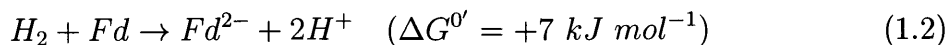
Re-oxidation of the ferredoxin at a membrane-bound proton translocase is coupled

to electrogenic proton extrusion, generating a proton-motive force (PMF) that can be coupled to ATP generation with the proton-dependent ATPase. In *A. woodii* and *C. ljungdahlii*, the membrane-bound complex is an RNF complex (for $\underline{R}$hodobacter $\underline{N}$itrogen $\underline{F}$ixation) [33], and in *M. thermoacetica* it is an ECH (for $\underline{E}$nergy-$\underline{C}$onserving $\underline{H}$ydrogenase). *A. woodii* differs from the other organisms in that both the RNF complex and ATPase are sodium-dependent [34, 31]. The important contribution of electron bifurcation is therefore the ability to conserve free energy changes associated with metabolic reactions occurring *in the cytosol* in the low-potential acceptor ferredoxin, whose oxidation at a membrane-bound complex can then be coupled to oxidative phosphorylation. Since the first demonstration of electron bifurcation in 2008 [35], many more enzymes catalyzing similar processes have been discovered. Though the mechanism still remains enigmatic [36, 37, 38], the importance of this phenomenon in energy conservation in anaerobic microbes is undisputed.

Though the main principles underlying energy conservation in acetogens are now established, *A. woodii* is the only acetogen for which we have a complete model of acetogenesis. In *C. ljungdahlii* and *M. thermoacetica*, there are still open questions regarding some of the enzymatic reactions involved. In *M. thermoacetica*, there is evidence that the reduction of methylene-THF to methyl-THF is a bifurcating reaction that uses NADH as the electron donor. However, the physiological electron acceptor has not yet been identified [39]. In *C. ljungdahlii*, there is evidence for both a formate dehydrogenase and a bifurcating hydrogen-formate lyase [40]. The specific cofactors used in CO_2 reduction in this organism are therefore unclear, and various scenarios predict the formation of between 0.13 and 0.63 molecules of ATP for each acetate produced [41]. The wide variation in these estimates has important consequences when thinking about the production of value-added chemicals from this pathway. Another outstanding question is how sufficient driving force for the endergonic reduction of CO_2 to CO with reduced ferredoxin is generated. One possibility is that the ferredoxin pool is highly reduced, with a reduction potential of approximately -500 mV. Others argue that by coupling the activity of CODH and ACS, the exergonic acetyl-CoA synthesis reaction provides driving force for the reduction at more moderate potential

[42]. The point here is that, compared to more typical metabolic engineering hosts such as *E. coli* and *S. cereveisae*, there are many open fundamental questions about acetogenic metabolism, whose answers will be important in developing engineering strategies to allow them to produce high levels of value-added products.

Substrate flexibility in the WL pathway

A variety of different reduced C1 compounds serve as substrates for the WL pathway. Carbon monoxide (CO), toxic to many organisms, provides increased growth yields over hydrogenotrophic acetogenesis because of the low potential of the CO/CO_2 redox couple. With this substrate, CODH works in the oxidative direction, generating reduced ferredoxin to power reduction in the methyl branch of the pathway. The overall stoichiometry is

$$4CO + 2H_2O \rightarrow 2CO_2 + CH_3COOH \quad (\Delta G^{0'} = -206 \; kJ \; mol^{-1}) \qquad (1.5)$$

The ability to grow on H_2, CO, and CO_2 has led to signifcant commercial interest in using acetogens to convert syngas into value-added fuels and chemicals. An open question here is the potential for diauxie. It was recently shown that addition of CO to H_2-grown cells of *A. woodii* caused an initial growth inhibition, and a second growth phase after CO was depleted, presumably due to CO inhibition of the H_2-dependent CO_2 reductase [43]. The same phenomenon was observed for batch cultures of *C. ljungdahlii* [44], highlighting the need for careful control of syngas composition in a commercial process, or the use of a CO-insensitive hydrogenase (e.g. from *Carboxydothermus hydrogenoformans* [45]).

As an intermediate in the WL pathway, formate can also support acetogenesis, and formate-dependent growth has been reported for *M. thermoacetica* [46], *A. woodii* [43], and *C. ljungahlii* [47]. The use of formate avoids the mass-transfer limitations associated with H_2 and CO, both of which are poorly soluble in water. As the sole

carbon and energy source, formate is disproportionated according to the stoichiometry

$$4HCOOH \rightarrow 2CO_2 + CH_3COOH + 2H_2O \quad (\Delta G^{0'} = -95.8 \ kJ \ mol^{-1}) \quad (1.6)$$
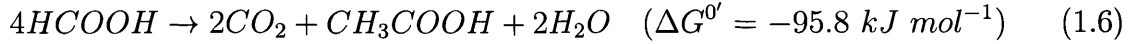
In *A. woodii*, it was shown that formate and CO can be co-utilized [43]. This represents an interesting way to avoid the complication of CO inhibition of hydrogenase in syngas-grown cells.

Some acetogens are also capable of using methanol as the sole energy source, although reports of this are somewhat contradictory, and the biochemistry is not entirely clear. In *M. thermoacetica*, methanol-dependent growth is well established both in the literature [46, 48] and in our hands, and thought be to be mediated by the transfer of the methyl group from methanol to THF via methyltransferases similar to those used in methylotrophic methanogenesis [49, 50]. Oxidation of methyl-THF by reversal of the methyl branch provides reducing equivalents for cell growth and ferredoxin CO_2 fixation in the carbonyl branch. Interestingly, reversing the methyl branch in *M. thermoacetica* is predicted to require reverse electron transport to drive the H_2-dependent reduction of ferredoxin at the ECH complex, which is in turn required for production of CO by the CODH (Figure 1-2). Thus, under these conditions, ATP conservation is mediated entirely by substrate-level phosphorylation. Methanol assimilation has also been demonstrated in *A. woodii* [51, 52, 43], but not in *C. ljungdahlii*. Methanol is the most reduced C1 substrate that can feed into the Wood-Ljungdahl pathway. The efficiency of the WL pathway leads to the highest carbon yields for conversion of methanol to acetate among known organisms, according to the stoichiometry

$$4CH_3OH + 2CO_2 \rightarrow 3CH_3COOH + 4H_2O \quad (\Delta G^{0'} = -228 \ kJ \ mol^{-1}) \quad (1.7)$$
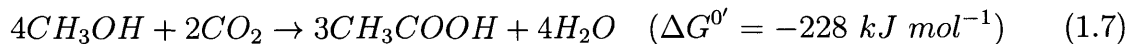
As with formate, methanol is soluble, mitigating mass transfer-related issues. This motivates further investigation into the mechanisms and limitations of methylotrophic acetogenesis to evaluate the potential for anaerobic methanol bioconversion via the WL pathway.
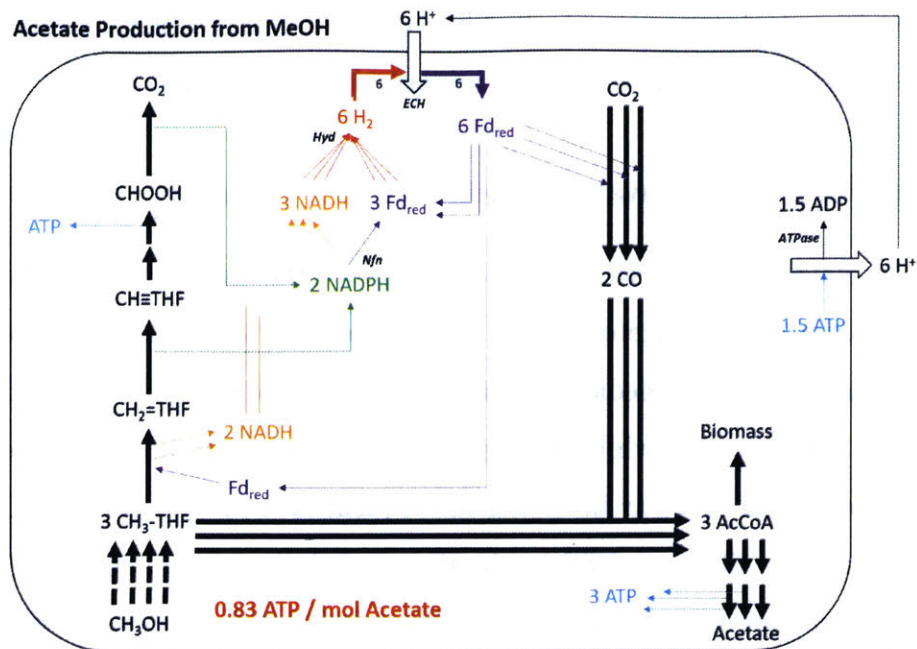
Figure 1-2: **Putative Methanol Metabolism Pathway in *Moorella thermoacetica*.** Putative biochemical pathways showing the energy-conserving production of acetate from methanol. Black lines represent carbon flux, and colored lines represent interconversion of electron carriers.

Beyond the C1 substrates listed above, acetogens can make use of a variety of other electron donors [46, 53]. *M. thermoacetica* specializes in the demethylation of aromatic O-methyl esters, such as vanillate [46]. *A. woodii* also respires caffeate, once again relying on electron bifurcation for energy conservation [54]. A variety of organic acids can also serve as growth substrates independent of the WL pathway. Most acetogens can also ferment the common hexoses fructose and glucose, and the pentose xylose. The WL pathway is active during this heterotrophic growth regime, and serves as a sink for the electrons removed during glycolysis. Thus, unlike typical heterotrophs, which can produce at most two molecules of acetate per gluocse, acetogens can achieve almost complete carbon conservation with three acetate. This makes engineered acetogens attractive candidates for the high-yield production of various acetyl-CoA derived products from heterotrophic substrates [55]. It has also been shown that a variety of acetogens can simultaneously ferment glucose and H_2. This ability allows the high carbon conversion efficiency of the WL pathway to be coupled with reducing

31

equivalents from hydrogen to make highly reduced molecules (e.g. butanol) at high yields, in a concept dubbed 'acetogenic mixotrophy' [56, 57, 58].

Fermentation products available from the WL pathway

Thus far we have discussed acetate as the primary fermentative end-product of acetogenesis. However, some acetogens naturally produce a variety of other compounds, including ethanol, 2,3-butanediol, butyrate, and butanol [59, 60]. How ethanol can be produced from H_2/CO_2 while allowing energy conservation has only recently been resolved. The typical pathway involving the sequential reduction of acetyl-CoA to acetaldehdye does not allow net ATP production, and thus could not sustain growth. Instead, in *Clostridium autoethanogenum*, a close relative of *C. ljungdahlii*, it is proposed that ethanol is produced from acetate. By this scheme, acetate is reduced directly at the expense of reduced ferredoxin, in a reaction catalyzed by an aldehyde-ferredoxin oxidoreductase (AFOR or AOR), then further reduced from acetaldehyde to ethanol, presumably by an alcohol dehydrogenase [61]. The thermodynamics of the reduction of non-activated acetate with ferredoxin are unfavorable ($\Delta G^0 = +30.4 \ kJ \ mol^{-1}$), such that the reaction is only favorable at high acetate concenteration and low pH. By converting acetate to ethanol under these conditions, this pathway allows for the continued production of ATP while mitigating acetate toxicity. The ability of several acetogens to produce high concentrations of ethanol was what motivated the initial interest in syngas fermentation for biofuel production.

In addition to acetate and ethanol, *C. ljungdahlii* produces small quantities of 2,3-butanediol (2,3 BDO). With steel mill waste gas a substrate, 2,3 BDO was produced in a ratio of 1:10:15 with acetate and ethanol [60]. A combination of genetic and enzymatic analysis later uncovered the synthesis pathway, in which two molecules of pyruvate are condensed to produce acetolactate, which is subsequently decarboxylated to acetoin and further reduced to 2,3 BDO [60].

Butyribacterium methylotrophicum ferments CO through the WL pathway to produce various ratios of acetate, ethanol, butyrate, and butanol. While the exact process is not clearly understood, the pathway likely shares similarity with the *Clostridia* bu-

tanol pathway typified by *Clostridium acetobutylicum* [62]. With the recent announcement of the genome for *B. methylotrophicum*, the genes involved can be determined [63]. Overall, with the range of both substrates and products available to them, acetogens display a highly diverse metabolic repertoire that makes them attractive candidates for industrial bioprocessing.

1.3.2   Aerobic methanotrophy and methylotrophy

Carbon flow in methanotrophy

Methanotrophs are a diverse group of microbes who specialize in respiring methane. They are a subset of methylotrophs, which refers more generally to organisms that can metabolize C1 compounds. The critical enzyme in aerobic methanotrophs is Methane Monooxygenase (MMO). This enzyme catalyzes the oxygen-dependent activation of the C-H bond of methane to form methanol, with reduction of the second oxygen atom to water:

$$CH_4 + O_2 + AH_2 \rightarrow CH_3OH + H_2O + A \tag{1.8}$$
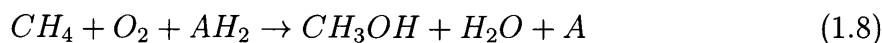
There are two variants of this enzyme: the copper-dependent particulate version (pMMO), and the iron-dependent soluble form sMMO. In Gram negative methylotrophs, methanol is oxidized to formaldehyde by a PQQ-dependent methanol dehydrogenase (MDH) in the periplasm. The involvement of methanol as an intermediate in methanotrophy enables many methanotrophs to also grow on methanol. Formaldehyde represents a node between carbon assimilation, and further oxidation to formate and CO_2 for energy conservation. In Type I methanotrophs, formaldehyde is assimilated through the Ribulose Monophosphate (RuMP) pathway. Formaldehyde is first condensed with ribulose 5-phosphate (Ru5P) by an aldolase, and then isomerized to fructose 6-phosphate (F6P). A variety of different variants of the RuMP pathway exist for the regeneration of Ru5P, that differ in their requirements for ATP or NADH [64]. In each case, three turns of the cycle lead to the net conversion of three molecules of formaldehyde to a three carbon sugar phosphate of central carbon metabolism, dihydroxyacetone phosphate (DHAP) or pyruvate (PYR). In Type II methanotrophs,

carbon is assimilated by the serine cycle, and formate serves as the node between biosynthesis and energy conservation [65]. Formate is reduced in an ATP-dependent manner to the formaldehyde oxidation level with NADPH and THF. In the serine cycle, carbon enters from formaldehyde and CO_2. The pathways for regenerating the serine vary based on presence or absence of isocitrate lyase, and have only been recently fully elucidated and demonstrated [66]. In all cases, the fixation of CO_2 is energetically costly, and growth yields in Type II methanotrophs are significantly lower than in Type I [67]. However, because formate is a branch point in their metabolism, the prototypical strain *Methylobacterium extorquens* AM1 can grow on formate as a sole carbon and energy source [68].

Variations on methylotrophy

Aerobic growth on methanol is not limited to the Gram negative methylotrophs above. In the late 1980s, a Gram positive thermophilic Bacillus species, *Bacillus methanolicus*, was discovered that could grow with methanol at temperatures up to 55 °C [69]. It was soon discovered that this species oxidized methanol via an NAD-dependent methanol dehydrogenase whose expression accounted for up to 30% of the total protein. Ongoing biochemical and genetic investigations into this organism have revealed many interesting details about its metabolism: in *in vitro* assays of MDH, the presence of a Nudix hydrolase encoded by the *act* gene increases k_{cat} and decreases the K_m for methanol by hydrolyzing the bound NAD cofactor and changing the catalytic mechanism [70, 71]. The role of this modification *in vivo* is unclear, but the *in vitro* activation of alcohol dehydrogenases by Nudix hydrolases appears to be a widespread phenomenon [72]. In addition, methylotrophy in this strain is plasmid-dependent, and lost when the plasmid is cured [73]. Sedoheptulose bisphosphatase activity of an enzyme encoded by the *glpX* gene on the plasmid suggests that *B. methanolicus* uses the SBPase variant of the RuMP pathway, in which sedoheptulose phosphate is generated by the successive action of the fructose bisphosphate aldolase – which produces SBP from DHAP and erythrose 4-phosphate (E4P) – and SBPase, which hydrolyzes the phosphate to form the monophosphorylated sugar [74]. Transcriptomic,

metabolomic and proteomic characterization has also been carried out, and provides further insight into the metabolism of this microbe [75, 76, 77].

Some yeast species, notably *Pichia pastoris*, are capable of growth on methanol. In these organisms, methanol is oxidized to formaldehyde by an alcohol oxidase, which produces peroxide as a byproduct. Due to the toxicity of peroxide, this process occurs in an intracellular organelle called the peroxisome. Formaldehyde assimilation is accomplished in a process similar to the RuMP pathway, in which xylulose 5-phosphate serves as the sugar phopshate acceptor for formaldehyde [78]. Since the electrons from the oxidation of methanol are wasted in the generation of peroxide, the product yield on methanol is predicted to be low in these organisms, and they will not be discussed further here. It should be pointed out, however, that because of its high rate of methanol uptake, *Pichia* is an excellent microbe for the production of heterologous proteins from methanol [79].

Energy conservation in aerobic methanotrophs and methylotrophs

As with acetogenesis, there are aspects of energy conservation in aerobic methylotrophy that are still not fully understood. The primary mode of energy conservation in these organisms is oxidative phosphorylation, where the reoxidation of reduced cofactors – generated from the various steps of methane oxidation – at terminal oxidases generates a PMF for ATP production. As such, understanding the flow of electrons between the various co-factors involved is critical. The electron donor for the sMMO is NADH, however, despite intensive effort, the physiological electron donor for the pMMO is still unknown [12]. Experimental growth yield data and CH_4/O_2 values support the existence of an electron transport chain between MDH and the pMMO [80, 67]. In support of this, evidence has been found for the association of MDH with pMMO, suggesting a possible direct electron transport between these enzymes [81].

35

1.3.3   Other native pathways for C1 assimilation

This thesis focuses on C1 assimilation pathways whose manipulation could lead to the formation of organic compounds. As such, we will not discuss methanogenesis, by which a variety of carbon sources, including methanol, are converted to CO_2 and CH_4. Also ignored are autorophic aerobic CO oxidizers, who use the electrons generated from CO oxidiation to drive carbon fixation through the CBB pathway. For excellent reviews on these important subjects, the interested reader is directed to [82] and [83], respectively.

1.3.4   Synthetic pathways for C1 Assimilation

Motivated by the availability of over 5,000 known natural metabolic reactions and computational tools to automate pathway design, an emerging interest in the field is the design of non-native pathways for C1 assimilation that may be more efficient than natural ones [84]. This approach aims to combine enzymes from a variety of different organisms in synthetic pathways to overcome thermodynamic or kinetic limitations of known pathways [85]. The effectiveness of this approach was recently shown *in vitro* with the design and construction of a highly efficient CO_2 fixation pathway [86]. Several hypothetical pathways for formate assimilation have been put forward [17], though these have not yet been realized experimentally. In an impressive feat of enzyme engineering and pathway design, a novel linear pathway for formate assimilation via formaldehyde was conceived and shown by ^{13}C labeling analysis to enable the conversion of formate to DHAP *in vitro*. The cornerstone of this pathway was the re-engineering of benzaldehyde lyase to catalyze the coupling of three molecules of formaldehyde into dihydroxyacetone in a reaction known as the formolase reaction [87]. Though the catalytic ability of the enzyme was insufficient to support growth, these results highlight the potential for the improving tools in functional protein design to allow access to previously non-existent metabolic pathways.

1.4 Metabolic Engineering in C1 Organisms

1.4.1 Acetogen Metabolic Engineering

Genetic tools for acetogens

The ability of acetogens to generate a variety of reduced products from syngas has led to recent interest in metabolic engineering of these hosts to allow access to higher-value products [88]. In general, the lack of well developed and convenient genetic tools is the major bottleneck in this area [89]. The first genetic engineering tools for an acetogen were repoted in 2010, for *Clostridium ljungdahlii* [90]. The authors generated a plasmid encoding the butanol biosynthesis pathway from *C. acetobutylicum*, and upon introduction of the plasmid into *C. ljungdahlii*, demonstrated the production of about 2 mM butanol from syngas. Since then, the preliminary tools have been improved [91, 92], and now include an inducible promoter [93], a modestly functional fluorescent protein and temperature-sensitive replicon[92], and CRISPR-cas9 tools for genome editing [94]. Efforts to engineer *C. ljungdahlii* to produce additional products such as butyrate [95] and acetone [93] have been reported, though the titers are still low compared to native products. In all cases, efforts to eliminate formation of ethanol and acetate have met with limited success, highlighting the limitations both of the tools used to generate knockouts, which are currently slow, and of our understanding of the genes and enzymes responsible for mediating these conversions in these hosts. Ultimately the ability to eliminate acetate production may be limited by the ATP requirements of the cell. Genetic tools have also been reported for *A. woodii*, and used to overexpress components of the WL pathway in an effort to improve the rate of carbon assimilation [96]. This organism was also recently engineered for acetone production [97]. A protocol for the transformation of *M. thermoacetica* has also been published, in which the authors overexpressed a lactate dehydrogenase gene to redirect carbon flux to lactate under heterotrophic conditions [98].

Expected yields in syngas fermentation

From Equations 1.1 and 1.5 it is clear that acetogenesis operates close to the thermodynamic limit of life [41]. As such, even if strains can be engineered to produce products more reduced than acetate, the yield will likely be low [27]. Furthermore, since acetate production is critical to the ATP balance of the cell, the production of significant byproduct may be necessary in order to maintain cell viability, which will lead to increased separation costs. Given the mostly understood biochemistry of acetogenesis, and the known pathways for the synthesis of the potential products, the maximal theoretical yield can be estimated, along with the molar ratio of acetate that will have to be produced to balance ATP. The results of these calculations are are shown in Figure 1-3, in the form of the maximum hydrogen cost that will make acetogenic production equivalent in raw material costs to heterotrophic production from glucose. Products derived from acetyl-CoA are economically viable with hydrogen prices between \$1.75 and \$4 per kilogram, depending on the ATP cost of the biosynthetic pathway. Products that arise from pyruvate require much lower hydrogen prices, owing to the energetic penalty of fixing additional CO_2 to produce the 3-carbon precursor. The situation is improved with CO as a feedstock, because of its higher energy content, but still, the message is plain: the low ATP generated from the WL pathway is a limiting factor in the expected yield of most products of interest. Additional methods to bolster ATP generation, perhaps in the form of mixotrophy [58], or with additional electron acceptors that allow respirative ATP production [99], will therefore be of high importance for the industrialization of engineered acetogenesis. Another method for avoiding the ATP penalty is to use a two-stage system, where an acetogen produces acetate, which is then separated and fed to a second organism for upgrading to higher value product, for example lipids by the oleaginous yeast *Yarrowia lypolytica* [100]. Here the extra energy associated with *aerobic* consumption of acetate by the second organism allows high-yield production of an energy-intensive biodiesel-like product, while taking advantage of the ability of *M. thermoacetica* to ferment syngas to acetate at exceptionally high yield.
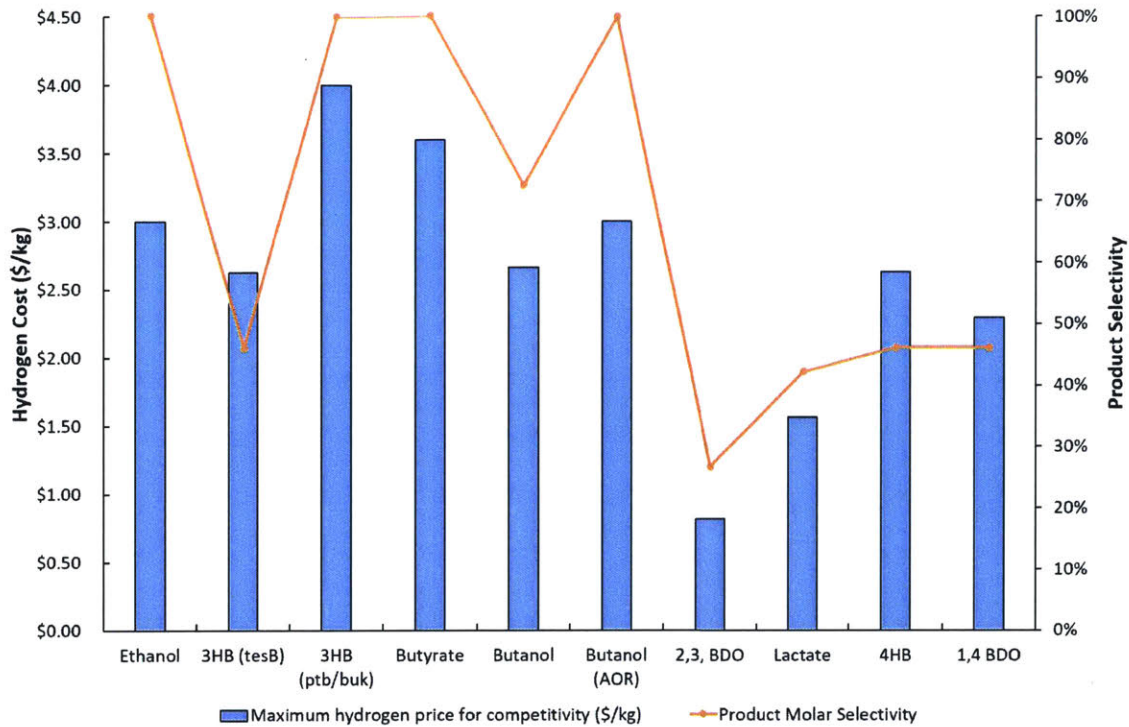
Figure 1-3: **Raw Material Cost Analysis of Syngas Fermentation.** Hydrogen yields were calculated for *C. ljungdahlii* assuming the most favorable ATP conservation conditions in [41], and used to calculate the maximum cost of H_2 that would make acetogenic production equal in raw material cost to glucose-based production (blue bars). The product molar selectivity (orange line) was calculated as the mol fraction of total fermentation product that was the desired product. All other product is acetate to balance ATP consumption. Details of the calculations are shown in the Appendix.

1.4.2 Aerobic Methanotroph and Methylotroph Metabolic Engineering

Due to the availability of natural gas in the United States, there has recently been renewed interest in engineering methanotrophs to produce value-added chemicals [12, 101]. As with acetogens, these efforts have been hampered until recently by the limited availability of robust genetic tools and fundamental understanding of carbon metabolism [102]. In particular, metabolic models of methane utilization pathways have largely failed to predict experimentally measured fermentation parameters [12], which makes the rational prediction of gene targets for improved product flux challenging. The application of -omics techniques in recent years has greatly improved

our understanding of these pathways [103, 104], and is expected to facilitiate the engineering of novel strains. The domestication and development of genetic tools of *Methylomicrobium buryatense*, which benefits from a short doubling time of only three hours, will also aid this effort [105].

To-date, more success has been achieved in engineering methanol-consuming strains for heterologous production. Recently, introduction of the butanol pathway into *M. extorquens* led to the production of small quantities (15 mg L^{-1}) of butanol from ethylamine, highlighting the drastic improvements needed to reach commercially relevant metrics [106]. More success was achieved with the engineering of α-humulene production in the same organism, which after ribosome binding site and fermentation optimization led to titers of 1.65 g L^{-1} [107]. Other efforts to engineering methanol-based production in *M. extorquens* have generated strains that can produce polyhydroxyalkanoate co-polymers [108] and dicarboxylic acids [109, 110]

Primitive genetic engineering tools have also been developed for *Bacillus methanolicus* [16]. A challenge associated with this organism is its high optimum growth temperature, which renders many of the traditional reporter genes, such as GFP, non-functional. Nevertheless, introduction of *cadA* and *ldcC* genes from *E. coli* led to the production of cadaverine at titers of 500 mg L^{-1} from methanol in shake flasks [111].

1.5 Thesis Overview

Motivated by the promise of C1 compounds as feedstocks, this thesis explores metabolic engineering strategies to overcome some of the barriers described above. The first chapter describes efforts to develop genetic engineering tools for *Moorella thermoacetica*, which would enable metabolic engineering to re-direct carbon flux from acetate to value-added chemicals, and facilitate investigation into the open questions surrounding aspects of its metabolism. From the metabolic engineering work conducted with *C. ljungdahlii* so far, it was concluded that the major bottlenecks are the absence of genetic tools to reduce the unwanted formation of the byproducts acetate and ethanol,

and insufficient understanding of the genes to target to achieve the desired product profile. In the second chapter, therefore, a CRISPRi system is developed to allow the simultaneous knockdown of multiple genes with putative roles in carbon flux and energy conservation. This platform can be used for both pathway engineering and experimental annotation of gene function. In the third chapter, various pathways for the conversion of methane to a biofuel (here, n-butanol), are assessed, to provide direction for the nascent field of methanotroph metabolic engineering. Methanol is an attractive substrate, but, as shown above, metabolic engineering efforts in both *Methylobacterium extorquens* and *Bacillus methanolicus* have not resulted in industrially relevant process metrics. Taking advantage of the relative simplicity of the methanol consumption pathway in *B. methanolicus* – which requires only three heterologous enzymes, all of which have previously been functionally expressed in *E. coli* [73, 112] – the development of an *E. coli* strain capable of methanol assimilation is described in the next chapter. A combination of theoretical and experimental approaches were used to analyze bottlenecks in the engineered strain in support of further metabolic engineering efforts to enable growth on methanol as a sole carbon and energy source. In the final chapter, the development of a highly sensitive formaldehyde biosensor to aid in methanol metabolic engineering is described.

The two main projects in this thesis – the development of CRISPRi for *C. ljungdahlii* and the importation of the methanol assimilation pathway into *E. coli* – represent complementary metabolic engineering approaches to enabling the use of alternative substrates. In the first, we take advantage of the metabolic capabilities of the organism in question, and develop tools for the engineering of downstream production. In the second, we take advantage of the wealth of biochemical and genetic understanding of *E. coli*, and the large number of successes of engineering it for the biosynthesis of a variety of products, and attempt to confer it with additional metabolic capabilities. Deciding on which approach to pursue is dependent on a number of factors [113], not least of which is the pathway complexity: the Wood-Lungdahl pathway consists of at least 24 genes, encoding enzymes with highly complex 3-dimensional architecture, multiple labile iron-sulfur clusters, and exotic trace metal requirements

[18]. The chances of successfully importing this scale of heterologous machinery into a heterotrophic host seem slim[1]. By contrast, the enzymes of the methanol pathway use cofactors found naturally in *E. coli*, and are encoded by a single gene each, and is therefore conceptually straightforward. The thesis concludes with reflections about the merits of these approaches, based on the insight gained from pursuing these projects.

---

[1]Although the complete refactoring of the nitrogenase gene cluster from *K. oxytoca* suggests this possibility is growing nearer [114].

Chapter 2

Development of Genetic Tools for *Moorella thermoacetica*

2.1 Summary

M. thermoacetica is a model homoacetogen and attractive candidate for syngas bio-processing, but both fundamental and applied work in this organism is hampered by the absence of robust genetic tools. In this chapter, I describe the approaches taken during the early part of my PhD to develop a genetic system, and to devise experimental procedures to analyze the steps of transformation in order to troubleshoot problems. First, a generalizable flow-cytometry based assay for assessing the delivery of fluorescently labeled plasmids or oligonucleotides to bacteria was developed, and used to optimize electroporation conditions in *Moorella*. The potential to modify the basic assay to detect intracellular DNA degradation by FRET was also examined. Second, expression of the methyltransferases from *Moorella's* restriction-modification system in the *E. coli* strain used for plasmid propagation was shown to partially protect plasmid DNA from digestion by ApaI, an isoschizomer of one of *Moorella's* restriction enzymes. Finally, a series of replicative vectors (carrying a thermostable kanamycin resistance marker) and integrative suicide plasmids (targeting the *pyrF* locus to impart FOA resistance) were constructed and transformed into *Moorella*. While the integrative plasmid led to FOA resistance, sequencing of the resulting iso-

lates revealed only partial deletion of *pyrF*, indicating low efficiency of homologous recombination. By contrast, one of the replicating plasmids, pQexp-kanR, conferred kanamycin resistance. Overlapping PCR of the plasmid backbone suggested the vector was maintained as an autonomusly replicating unit. Difficulties in reproducing this result, coupled with the development of robust transformation protocols for the related organism *Clostridium ljungdahlii* led to the decision to suspend these efforts and re-focus on that organism instead.

2.2 Motivation

Previous work in our lab has established *M. thermoacetica* as a promising platform organism for the production of acetate from syngas [115, 116]. *M. thermoacetica* is also the model homoacetogen used in the seminal studies in the laboratories of Harland Wood and Lars Ljungdahl to establish the biochemistry of the Wood-Ljungdahl pathway [46, 117]. As such, the development of genetic engineering tools for this bacterium would enable both fundamental genetic studies to answer outstanding questions about acetogenic metabolism [39], and more applied efforts to enhance acetogenesis or enable the production of higher value chemicals. The process of developing these sorts of tools has typically been a high-risk trial-and-error approach, thus an additional motivation for this work was to simultaneously devise general strategies to systematize the development of genetic tools for intractable organisms.

2.3 Development of a Flow Cytometry Assay for Optimizing DNA Delivery to Bacteria

2.3.1 Introduction

Metabolic engineering in non-traditional host organisms is limited by the lack of transformation protocols for these organisms [118, 53, 119]. Developing such protocols is a difficult undertaking, primarily because of the many variables in the transformation

process that must be optimized individually for the species in question and whose effects are impossible to predict *ab initio* or assess in the absence of a protocol that already produces at least some transformants [120]. Typically, this optimization is carried out by varying a single parameter, and assessing its effect on the efficacy of the overall process by counting the relative number of colonies that appear after plating on selective medium. The process is iterated with each variable until a sufficiently robust protocol is developed. As this black box methodology relies on transformation efficiency as an objective function, it can only be applied to a protocol that already produces at least some transformants. But developing a protocol even to this point can be a protracted process, requiring substantial guesswork and time-intensive full factorial experiments, which in the early stages are generally unsuccessful, and more importantly yield no insight into which of the many parameters is causing the failure. Thus, transformation protocol development can be facilitated by methodologies that can directly assess the impact of various parameters on specific steps within the transformation process, regardless of the overall outcome.

Here we describe a simple flow cytometry (FC) assay that can be used to directly assess the delivery of fluorescently labeled DNA to a bacterial cell, regardless of whether transformants eventually arise. The concept of using fluorescent molecules as analogs to DNA to gauge transformation efficiency is not new. In the development of a transformation system for *Thermoanaerobacter sp X514*, Lin et al used Texas red-conjugated dextran for electroporation and sonoporation experiments, observing the cells post-transformation via fluorescent microscopy [121]. Azencott et al used BSA-FITC and Calcein, along with propidium iodide (PI), to quantify uptake into the algae *Chlamydomonas reinhardtii* using flow cytometry [122]. Numerous similar reports are available in the literature [123, 124, 125, 126], However, in each of these cases the chemical and physical properties of the fluorescent molecules used for transformation are quite different from DNA, and their delivery to the cell may not necessarily correlate with successful delivery of DNA by the same method. Indeed, Calcein and BSA-FITC have molecular weights of 662 Da and 66 kDa, respectively, and even the Texas Red conjugate has a weight only of 70 kDa. By contrast, double-stranded DNA

has a molecular weight of 650 Da bp^{-1}, such that even a modest 5 kB plasmid has a molecular weight of >3,000 kDa. This translates to dramatic differences in the diffusion coefficient – 5.9x10^{-7}cm^2s^{-1} and 4.11x10^{-8}cm^2s^{-1} for BSA[1] and a 3.7 kb plasmid [127], respectively, which may have important consequences in the transport of the molecule across the cell membrane.

In the method described below, the actual DNA used in the transformation is labeled directly with Cy5. After the cells are processed, they are treated with a membrane-impermeable DNA binding dye, propidium iodide (PI), to assess viability and analyzed via flow cytometry. Cells that are PI-negative with Cy5 fluorescence above that of a non-fluorescent control are considered intact and to have taken up the DNA. This approach has the advantage of using DNA, not an analog, and not relying on *in vivo* expression. To our knowledge, no such assay has been reported for bacterial systems. James and Giorgio [128] used a similar labeled plasmid protocol for analyzing transfection of HeLa cells, but did not use PI to analyze viability. In addition, our assay functions with both labeled plasmid DNA and labeled oligonucleotides. While transformation with plasmid DNA is the standard method for introducing heterologous DNA in bacteria, in recent years oligonucleotide-mediated recombineering has emerged as a powerful technology for metabolic engineering, generating knock-outs, insertions and point mutations. This process occurs via a highly conserved mechanism that makes its application feasible in a wide variety of species [129, 130, 131, 132]. Thus, the ability to detect delivery of both plasmid and oligonucleotide DNA will help enable metabolic engineering in previously intractable hosts.

2.3.2 Methods

Cultivation Conditions

E. coli DH5α was cultivated in LB medium according to standard procedures. *M. thermoacetica* (ATCC 39073) was cultivated at 55°C in a shaking incubator (250 rpm) in N-SCM, an anaerobic rich medium (Per liter: Glucose 8g, NaHCO$_3$ 7.5g, KH$_2$PO$_4$

---

[1]From the Sigma product spec sheet, see Appendix

7g, K_2HPO_4 5.5g, BACTO Yeast Extract 5g, BACTO Tryptone 5g, $(NH_4)_2SO_4 \cdot 7H_2O$ 2g, $MgSO_4 \cdot 7H_2O$ 0.5g, $CaCl_2 \cdot H_2O$ 0.02g, Wolfe's Trace Minerals 10 mL, Resazurin 1 mg, Cysteine 300mg, pH 6.8) following standard techniques for working with obligate anaerobes.

Preparation of Labeled DNA

Cy5-labeled plasmid was prepared by the LabelIT Tracker kit (MIR 7021, MirusBio, Madison, WI), according to the manufacturer's instructions, using column-purified pUCG18 [133] as plasmid template. The reaction mixture contained: 1x Labeling Buffer A, a ratio of 2:1 ($\mu L/\mu g$) LabelIT Tracker reagent to plasmid, and the balance water. The reaction volume was determined such that the LabelIT reagent would account for 12% of the total. After incubation at 37°C for 3 hours, labeled plasmid was purified via Qiagen mini-prep spin column and eluted in 100 μL pre-warmed elution buffer. Labeling intensity was calculated by the manufacturer's protocol, using a Nano-Drop (ThermoScientific, Wilmington, DE) to determine absorption at 260nm and 649nm. Fluorescein-labeled plasmid was prepared following the procedure described above, but using the fluorescein LabelIT Tracker kit (MIR 7025, MirusBio, Madison, WI) and assessing absorption at 494nm. Generally, plasmid was labeled at a density of one fluorophore every approximately 40 base pairs, regardless of fluorophore. Cy5-labeled oligonucleotide was purchased from IDT as an internal modification to a 90 bp oligonucleotide complementary to the putative pyrF gene in M. thermoacetica (*Moth0883*) with base-pair mismatches at the active site to inactivate the cognate protein and allow selection with 5-FOA. The full sequence of the primer is: 5'-GATGGTAAAGGAGCGTGGTGGCCGGGTATTTGCGCTGG/iCy5/CCG CCTTCCACGACATCCCCAACACCGTGGCCGGAGCGGCGCG-3'.

Sonoporation

100 mL N-SCM medium in sealed serum bottles was inoculated with 250 μL of overnight culture of *M. thermoacetica* and cells were allowed to grow to O.D. 0.6. Culture bottles were chilled on ice for 20 minutes, and cells were harvested by cen-

trifugation (3,700 g, 4°C, 15 min) under anaerobic conditions, washed once with 20 ml anaerobic ice-cold PBS and resuspended in 2.25 ml of PBS-M buffer (Phosphate Buffered Saline with 1 mM $MgCl_2$, 1 mM $CaCl_2$). Resuspended cells were mixed with labeled or unlabeled plasmids (10 μg/mL), and transferred to 7-mL screw-cap flat bottom vials. Vials were placed in the ice-water bath of a Branson B300 sonicator [121], located inside the anaerobic glove box (Coy Labs, Grass Lake, MI), and sonicated for 10, 20, 30, 40 and 50 seconds. After sonication, 1 mL of pre-warmed N-SCM medium was immediately added to the vials, and the resuspended cells were transferred to culture tubes containing 5 mL pre-warmed N-SCM medium.

Electrocompetent Cell Preparation

E. coli DH5α was made electrocompetent by standard procedures. To make electrocompetent *Moorella thermoacetica* for plasmid transformation, 100 mL anaerobic N-SCM medium in an anaerobic serum bottle was inoculated with 1 mL actively growing culture, then allowed to grow overnight at 55°C. When the OD_{600} reached approximately 0.2, the culture was chilled on ice for 30 minutes, then centrifuged and washed three times with 50 mL ice cold anaerobic buffered SPG solution (272 mM sucrose, 7 mM phosphate, 15% glycerol, pH 6.8). Cells were resuspended in a final volume of 1mL ice cold SPG and stored at -80°C until needed. To make electrocompetent cells for oligonucleotide transformation, cells were grown as above but in a minimal version of N-SCM (containing no yeast extract or tryptone), and then made competent according to one of two protocols: 1) The same as above for plasmid transformation, or 2) Three washes with 50 mL ice cold SMP solution (272mM sucrose, 1mM MgCl2, 7 mM phosphate, pH 6.0), followed by resuspension in 1 mL ice cold SMP with 10% DMSO, as in Leang et al [91]. All manipulations of *M. thermoacetica* cells except centrifugation were performed in an anaerobic glove box and following standard anaerobic microbiology techniques.

Electrotransformation

E. coli DH5α competent cells were thawed on ice, and 25 μL combined with 250 ng labeled or unlabeled plasmid, or 0.2-10 μg labeled or unlabeled oligonucleotide. The mixture was transferred to a chilled 1 mm electro-cuvette and electroporated at 1.7 kV, 200 Ω, 25 μF, then immediately resuspended in 1 mL SOC medium. For assessing transformation success, cells were transferred to a 37°C shaking incubator and recovered for 1 hour, before plating an appropriate dilution on LB plates containing ampicillin (100 μg/mL) and incubating overnight. *M. thermoacetica* competent cells were thawed on ice, and 100 μL combined with 1 μg labeled or unlabeled plasmid, or 20 μg labeled or unlabeled olionucleotide. The mixture was transferred to a chilled 1mm electro-cuvette and electroporated at 0-1.25kV, 500 Ω, 50 μF. Immediately following electroporation, 1 mL pre-warmed N-SCM medium was added to the cuvette, and the resuspended cells were transferred to a Hungate tube containing 5 mL of pre-warmed N-SCM medium. All of the above steps for *M. thermoacetica* were carried out in an anaerobic glove box.

Flow Cytometry Preparation and Analysis

Unless stated, all cell handling steps were carried out aerobically at room temperature. An aliquot of 1 mL cells in recovery media (SOC or N-SCM) were centrifuged at maximum speed in microcentrifuge tubes for 1 minute. Supernatant was removed and the cells were washed once in PBS, and resuspended to a final concentration of approximately 1×10^6 CFU mL^{-1}. 2 μL propidium iodide (PI) (1 mg mL^{-1}, Sigma-Aldrich) was added, and cells were incubated at 4°C in the dark for 5 minutes before being strained into 5 mL polystyrene FC tubes (BD Falcon 352235). Cells were stored in the dark at 4°C until analysis, which was always within 3 hours of electroporation.

Flow cytometry analysis was performed using a FACS Calibur (BD Biosciences, city, state) equipped with 488nm and 635nm lasers, and the following filters: 530/30 nm (FL1), 585/42 nm (FL2), 650nm LP (FL3), and 661/16 nm (FL4). The gating strategy is shown in Figure 2-1, and is described here: The cell population was gated

based on forward and side scatter, and acquisition continued until 20,000 events had fallen in this range. PI fluorescence was measured on the FL3 channel and gates established using an unstained control sample. Cy5 fluorescence was detected on the FL4 channel with a gate established using an unstained sample. Cells with FL3 signal below threshold and FL4 signal above threshold were considered to be intact and transformed. Data analysis was carried out using Cyflogic.

Quantification of Intracellular Labeled Oligonucleotide

The mean fluorescence of intact, transformed cells was converted to Cy5 MESF using calibrated microbeads from Bangs Labs (Quantum Cy5 MESF, 822). Because the Cy5 molecule conjugated to an oligonucleotide may not have the same fluorescent properties as the free fluorophore in solution, an additional conversion factor comparing the fluorescence of equimolar amounts of Cy5-labeled oligonucleotide and Sulfo-Cy5 NHS Ester (Lumiprobe, Hallandale Beach, FL) was generated using a SpectraMax M2e fluorescent plate reader (Molecular Devices, Sunnydale, CA) with excitation and emission wavelengths set the same as those available on the Flow Cytometer, as described in [123]. This correlation was determined in PBS buffer to mimic intracellular conditions as closely as possible. It was found that the ratio of DNA-bound Cy5 fluorescence to free Cy5 fluorescence under these conditions was 0.618, indicating that the linkage of Cy5 to the oligonucleotide reduced its effective fluorescent signal. The Cy5 MESF from the transformed cells was divided by this ratio to convert it to a number of Cy5-labeled oligonucleotide molecules.

Fluorescence Microscopy Preparation and Analysis

To visualize labeled plasmid using microscopy, cells of *M. thermoacetica* were electroporated as described above, in the presence of fluorescein-labeled plasmid, and transferred immediately to 1 mL pre-warmed N-SCM medium. The cell wall was labeled using wheat germ agglutinin (WGA) conjugated to AlexaFluor 594 (Life Technologies, Carlsbad, CA). For this, cells were centrifuged (13,000 g, 2 minutes, room temperature) and washed with 1 mL PBS buffer three times, then resuspended

in PBS solution containing 15.0 μg mL^{-1} of WGA conjugate and incubated for 10 minutes at 37°C. Labeled cells were then washed twice in 1 mL PBS buffer and fixed with 4% formaldehyde for 15 minutes at 37°C, followed by two more washes in 1 mL PBS buffer. An aliquot of fixed labeled cells was spotted onto a poly-l-lysine coated glass coverslip (BD Biosiences, Franklin Lakes, NJ), and then mounted on a glass slide using Fluormount, a non-fluorescing mounting medium (BDH Laboratory Supplies, Poole, UK). Glass slides were stored in the dark until they were imaged.

Epifluorescence imaging of the cells was performed using either an Applied Precision DeltaVision deconvolution microscope equipped with a Hg lamp, excitation and emission filters wheels, 100 X, NA 1.4 oil objective, a CCD Photometrics camera and Softworx deconvolution software, or a DeltaVision-OMX super-resolution microscope equipped with 405 nm, 488, nm and 598 nm lasers, a 100 X, NA 1.4 oil objective, 3 EMCCD cameras and Applied Precision Softworx OMX software. Three μm Z-stacks were collected at 0.125 μm spacing and then either, deconvolved (when using DeltaVision microscope), or 3D-SIM reconstructed (when using 3D SIM Delta OMX) with Softworx software. Z-stack projections were generated using the maximum intensity method, and saved as TIF files.

2.3.3 Results and Discussion

FC assay validation with *E. coli*

To validate that our FC protocol could be used to detect delivery of DNA to cells, we used electroporation to deliver 250 ng of labeled plasmid (at approximately 1 fluorophore per 40 bp) to electrocompetent *E. coli* cells.

Figure 2-1 shows the gating strategy and results of a triplicate experiment in which labeled plasmid was added to electrocompetent cells, and then cells were either electroporated or not. A clear difference in the percentage of intact cells with Cy5 fluorescence above the threshold was observed between the untransformed and the electroporated cells, 0.02% vs. 16% respectively (Fig. 2-1c,f and Table 2.1). Successful stable transformation with the labeled plasmid was verified by plating recovered
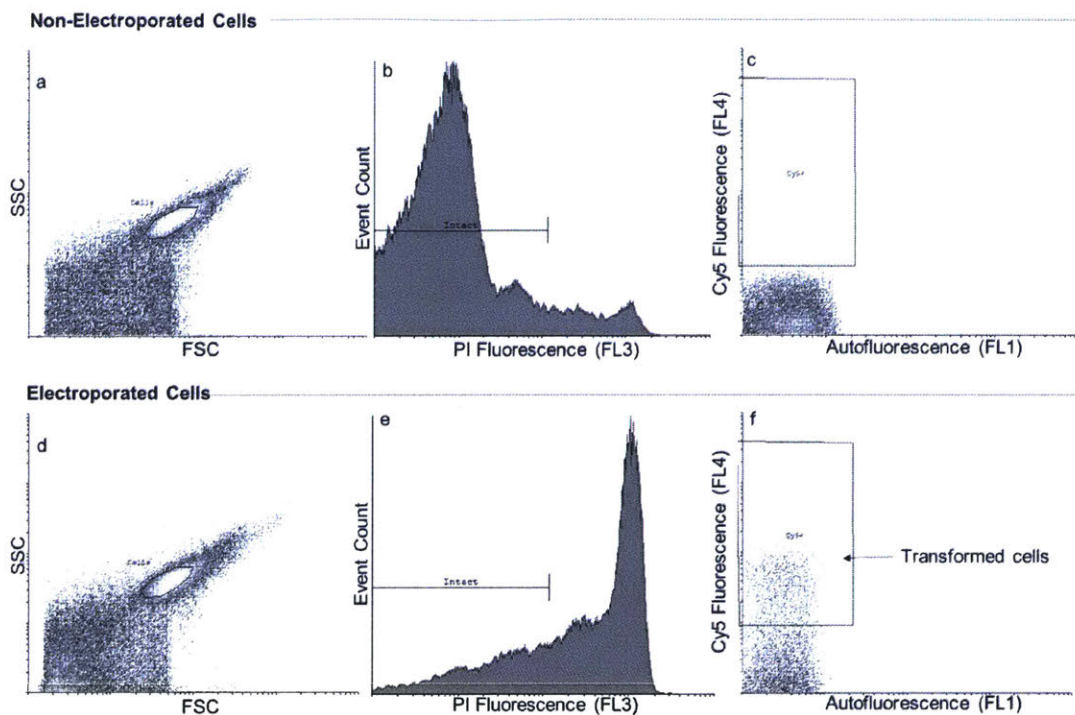
Figure 2-1: **Gating strategy and sample data from *E. coli* transformation.**
The top row shows FC analysis of cells that were mixed with labeled plasmid, but
not electroporated, as a negative control. The bottom row shows cells that were
electroporated with labeled plasmid. In both cases, cells were gated based on forward
and side scatter (a, d), and intact cells were selected by low PI fluorescence measured
on the FL3 channel (b, e). The Cy5$^+$ gate was set up such that 0.05% of the intact
cells from a non-electroporated control sample with unlabeled plasmid (not shown)
were above the FL4 threshold, using FL1 as an autofluorescence check (c, f). Intact
cells falling above the Cy5$^+$ threshold were considered to be transformed (f).

cells from both treatments on LB medium containing ampicillin, on which only the
electroporated cells developed colonies (Supplementary Figure A-1). The Cy5 fluo-
rescence from the untransformed cells (Fig. 2-1e) was virtually identical to a control
sample where cells were transformed with unlabeled plasmid, confirming complete
removal of unincorporated DNA by the washing process (Data not shown). Electro-
poration reduced the fraction of intact cells from 85% to 19%, as determined by PI
staining, which is consistent with previous reports [134]. Taken together, these results
demonstrate that labeled plasmid could be detected in intact, transformed cells.

52

Treatment	Intact Cells	Transformed Cells
Non-electroporated	85% ± 2%	0.02% ± 0.01%
Electroporated	19% ± 2%	16% ± 0.5%

Table 2.1: **Assay Validation: Detection of Cy5-Labeled Plasmid and cell membrane integrity in electroporated _E. coli_.** Stated errors represent the standard deviation of triplicate experiments. Transformed cells are taken as the percentage of the intact cells (as determined by PI staining) that showed Cy5 fluorescence above the threshold.

Sonoporation is ineffective for plasmid delivery to _M. thermoacetica_

With a robust assay in hand, we set about establishing a set of transformation conditions for _M. thermoacetica_. There have been reports of transformation of other gram-positive thermophiles using sonication to deliver plasmid DNA [121], therefore we were interested to see if this procedure could be applied to _M. thermoacetica_. This method is potentially advantageous over electroporation as it requires fewer steps to prepare cells for transformation and is less detrimental to cell viability. Cells were mixed with labeled plasmid and sonicated for durations up to 60 seconds, then processed and analyzed via flow cytometry using the same gating process as with _E. coli_. The results are shown in Figure 2-2. There was no fluorescence difference between cells that were sonicated and those that were not, indicating that the chosen sonoporation conditions were not successful in enabling plasmid delivery to the cell. The percentage of the cell population remaining intact after sonication was consistent among all treatments, suggesting that the sonication conditions were not powerful enough to produce pores in the cell wall and membrane.

Electroporation effectively delivers plasmid to _M. thermoacetica_

We next tried electroporation as a method for DNA delivery to _M. thermoacetica_. A range of voltages from 0 to 12.5 kV cm^{-1} were tested, with 1 μg of labeled plasmid. The results are shown in Figure 2-2. As the voltage increased, the number of cells remaining intact after electroporation diminished in a sigmoidal fashion. At applied voltages between 0.5 and 1.25 kV cm^{-1}, a significant fraction of the intact cells had taken up plasmid, with a maximum of 3% occurring at 1 kV cm^{-1}. In the negative
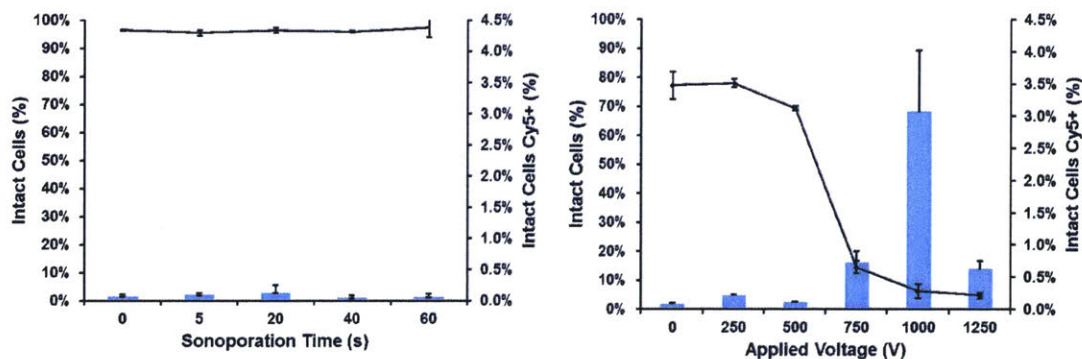
Figure 2-2: **Delivery of Cy5-labeled plasmid to *M. thermoacetica* via sonporation (left) and electroporation (right).** Black line: Percentage of all cells intact (PI-) after sonication or electroporation (left axis). Blue bars: Percentage of intact (PI-) cells with Cy5 fluorescence above threshold after sonication for certain durations, or electroporation at certain voltages (right axis). While sonication with the chosen conditions resulted in no detectable plasmid uptake, electroporation could deliver plasmid to 3% of the surviving cells. Error bars represent standard deviation of two biological replicates.

control, the percentage of intact cells is lower than in the corresponding negative control during sonication, presumably because of the increased amount of cell handling necessary for preparation of electrocompetent cells that reduced cell viability.

Verification of plasmid delivery with fluorescence microscopy

In order to verify that the plasmid had indeed been delivered to the cytoplasm, and was not simply attached non-specifically to the cell wall, electroporated cells and untreated cells that had been incubated with plasmid were analyzed via fluorescence microscopy. The cell wall was labeled with wheatgerm agglutinin conjugated to AlexaFluor 594, which binds to the N-acetylglucosamine of the Gram-positive cell wall [135]. For compatibility with the available filters, the plasmid used in this experiment

was labeled with fluorescein at the same density as the Cy5 experiments. Previous experiments showed that similar quantitative FC results were obtained with fluorescein as with Cy5 (Supplementary Figure A-2). The microscopy results are shown in Figure 2-3. A high level of signal can clearly be seen in the central area of the transformed cells (Figure 2-3a), demonstrating the correct localization of the plasmid. No fluorescein signal was detected in untransformed cells (Figure 2-3b).
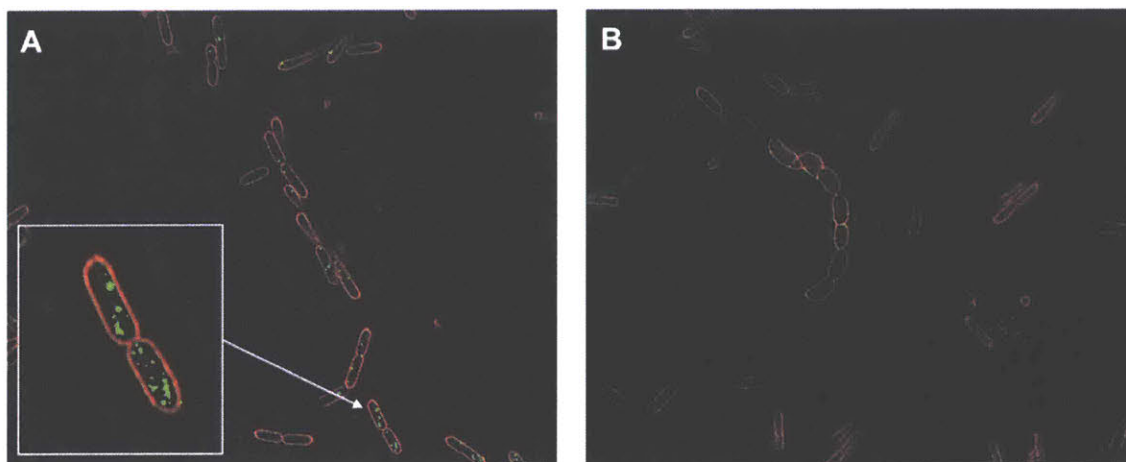
Figure 2-3: **Fluorescence microscopy confirmation of plasmid delivery to *Moorella thermoacetica*.** Fluorescence microscopy images of *M. thermoacetica* mixed with fluorescein-labeled plasmid and A) electroporated, and B) not electroporated. Cells were fixed, washed, and stained with WGA-AlexaFluor 594 before visualization. Green: Fluorescein channel. Red: AlexaFluor 594 Channel. Labeled plasmid is clearly visible in the cytoplasm in several electroporated cells, and absent from the non-transformed control.

Quantification of Cy5-labeled oligonucleotide delivery to *E. coli*

The FC assay was next used to quantify the delivery of oligonucleotides to bacterial cells, in support of recombineering approaches to genetic engineering. A variety of concentrations (0.2-10 μg) of Cy5-labeled oligonucleotides were electroporated into *E. coli* at the optimal conditions (1.7 kV, 200 Ω, 25 μF). The downstream processing steps were the same as described above for the labeled plasmid studies. The results are shown in Figure 2-4. Transformed cells were detected for all concentrations. At the highest three concentrations, the cells had taken up sufficient labeled DNA that

the transformed and untransformed populations were easily distinguishable on a Cy5 histogram (Figure 2-4), allowing calculation of the average number of oligonucleotides delivered to the transformed cells by using calibrated Cy5 beads. The optimal transformation protocol could deliver up to 11,000 copies of the oligonucleotide to *E. coli*, in a dose-dependent manner (Figure 2-4).

The concentration dependence of the mean fluorescence of the transformed population suggests that the inability to completely separate transformed cells from nontransformed ones at low DNA concentrations could be due to the low fluorescent signal of the small number of DNA molecules that are taken up in transformation that is indistinguishable from background autofluorescence, rather than from the absence of transformation under these conditions. This in turn suggests that categorizing cells as non-transformed based on sub-threshold Cy5 fluorescence may lead to a significant underestimation of the number of transformed cells. These results have implications in interpretation of the plasmid experiments: At a labeling density of 1 Cy5 per 40 bp, 250 ng of labeled pUCG18 plasmid corresponds to $6\text{x}10^{12}$ fluorophores, similar to the $4.7\text{x}10^{12}$ fluorophores for the 200 ng of the singly labeled oligonucleotide. Assuming a best case scenario of equal transformation efficiency for plasmids and oligonucleotides, which seems unlikely given the size difference, the number of fluorophores per cell would be at most 1.5x higher in the plasmid experiment, which would also be indistuinguishable from autofluorescence. Again, this suggests that the gating strategy used to assign transformed cells likely significantly *underestimates* the true percentage of cells that have been transformed. This is in agreement with the fluorescence microscopy analysis, which shows plasmid fluorescent signal in the majority of cells. In other words, the propensity for false negatives in this assay is high, and the presence of $Cy5^{+}$ cells in plasmid experiments should be interpreted qualtitatively as confirmation of transformation.
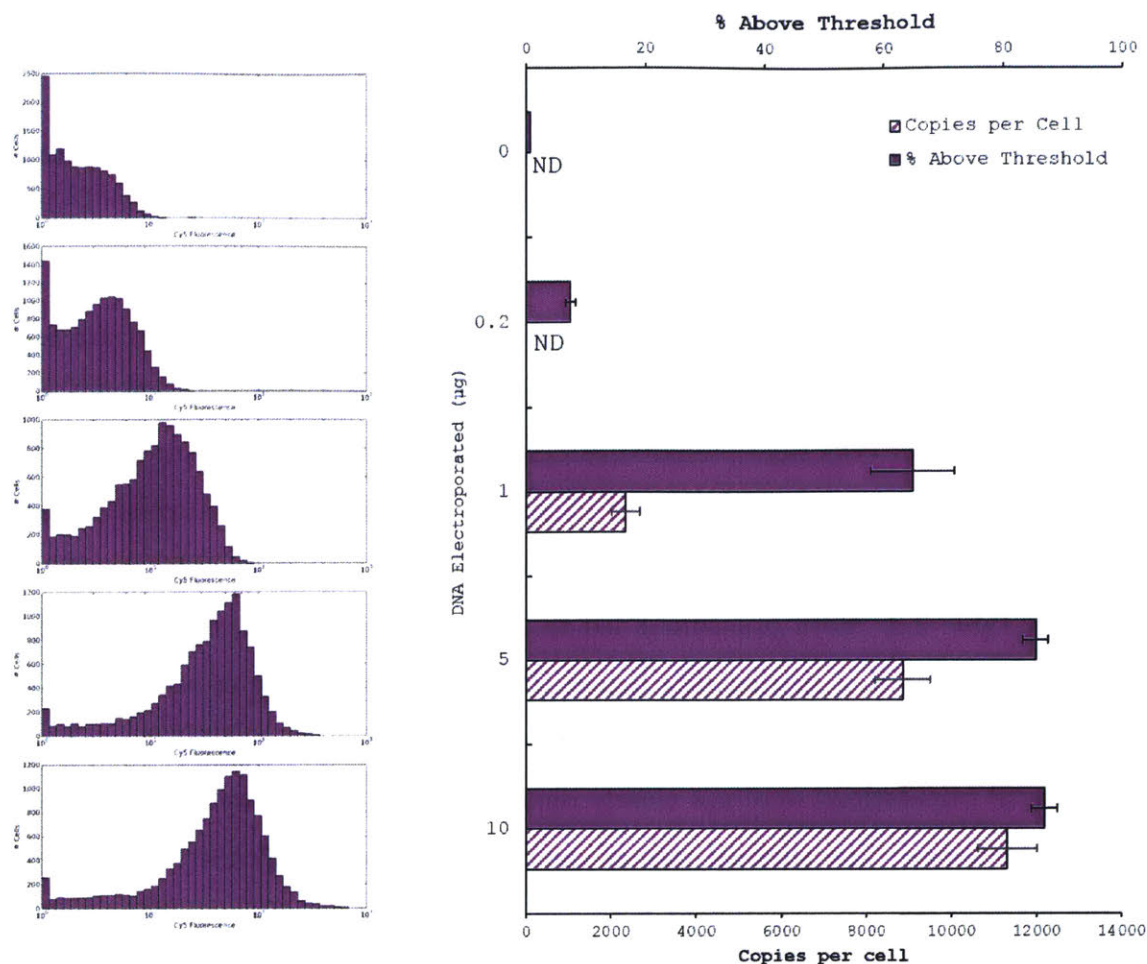
Figure 2-4: **Quantification of Cy5-Labeled Oligonucleotide Delivery to *E. col*.** Varying amounts of labeled oligonucleotide were electroproated into *E. coli*, and quantified using calibrated Cy5 beads, converting the mean fluorescence of the intact transformed cell population to the number of fluorescent oligonucleotides per cell (right). At low DNA concentrations, the transformed and untransformed histograms could not be fully resolved (left), preventing this analysis. Error bars represent the standard deviation of two biological replicates.

Comparison of *M. thermoacetica* competent cell preparations using Cy5-labeled oligonucleotides

To demonstrate the utility of the oligonucleotide-based assay, we used it to compare two different methodologies for the preparation of electrocompetent cells of *M. thermoacetica*: Our original procedure, and the procedure optimized for transformation of a related organism, *Clostridium ljungdahlii* [91]. The two protocols are very

Preparation Protocol	Intact Cells	Transformed Cells
Original (SPG Buffer)	32.4% ± 3.7%	5.8% ± 2.0%
Leang et al (SMP Buffer, DMSO)	37.7% ± 0.5%	2.1% ± 0.5%

Table 2.2: **Comparison of methodologies for preparing electrocompetent** *M. thermoacetica.* Stated errors represent the standard deviation of triplicate experiments.

similar, except that our protocol uses a wash buffer with pH 6.8 and glycerol as a cryopreservant, whereas the *C. ljungdahlii* protocol uses a buffer with pH 6.0, and DMSO. Competent cells were prepared as described in the methods section, electroporated in the presence of 20 μg labeled oligonucleotide at the optimal conditions as determined above (1000V, 500 Ω, 50 μF), and processed as before. The results are shown in Table 2.2. Both procedures led to similar survival rates, but our original procedure resulted in a significantly higher (P < 0.05) percentage of cells taking up the oligonucleotide. Interestingly, despite the high concentration of DNA used, the transformed populations were not clearly distinguishable as they were in the *E. coli* experiment above. The survival rate of the cells was also significantly higher than in the plasmid delivery experiment with the same electroporation conditions. The reason for this is unclear, but one possible explanation could be a high sensitivity of the optimal electroporation conditions to the cell preparation procedures, such that small variations between batches of cells could lead to non-transformable cells. This would certainly help to explain the low frequency of success of electroporation in Gram positive cells.

Beyond Transformation Efficiency: Using Dual-Labeled Oligonucleotides to Assess *in vivo* Nuclease Activity

An exciting benefit of using labeled oligonucleotides in assessing the delivery of DNA is the ability to precisely control the position of the fluorescent label, and even to encode multiple labels on a single oligonucleotide. This opens the door to a wide variety of additional experiments that can provide information about the intracellular fate of the DNA. As an example, by encoding a FRET pair (such as Cy3 and Cy5)

on either side of an endonuclease recognition site on an annealed double stranded oligonucleotide, the *in vivo* digestion of the oligonucleotide could be followed in real time by fluorescence microscopy or flow cytometry. One could then imagine employing this technique with a library of oligonucleotides containing a myriad of different recognition sites and using FACS and next-generation sequencing (NGS) to select and identify sequences that are degraded, in order to identify endogenous endonucleases in a high-throughput manner. As a proof of principle, we designed a 12bp nucleotide covalently modified with a 5' Cy3 and a 3' Cy5 fluorophore. The digestion of 1 μg of the oligonucleotide by DNAseI was followed spectrophometrically over the course of one hour, as shown in Figure 2-5. Compared to a control sample with no nuclease, the FRET signal diminished, while the Cy3 emission signal increased, indicating the increased spatial separation between the FRET pair as the connecting oligonucleotide backbone was degraded.
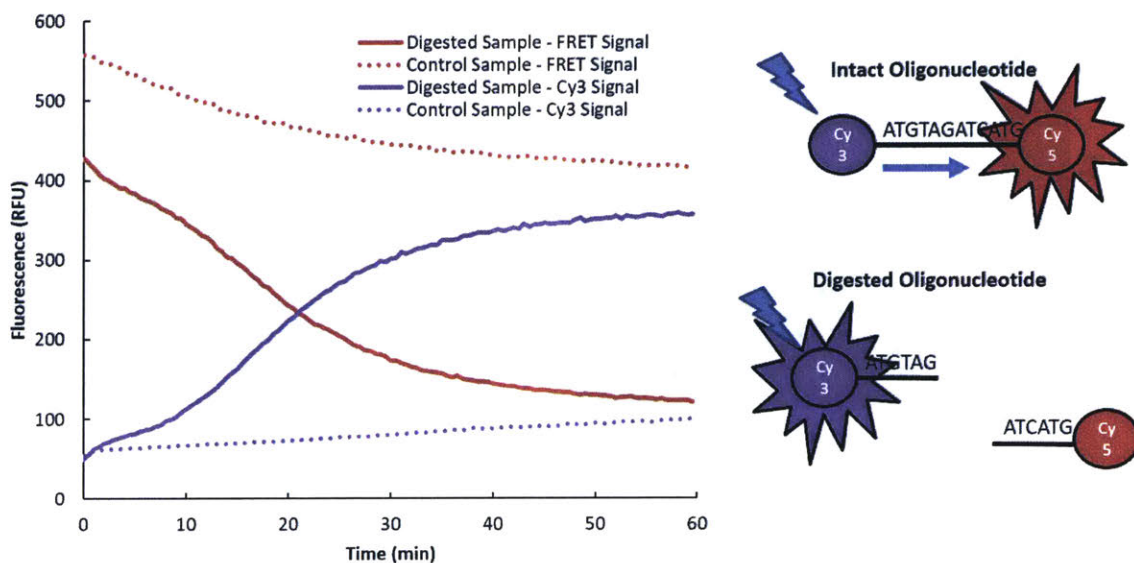
Figure 2-5: *In vitro* **DNAseI digestion of a dual-labeled oligonucleotide monitored by FRET**. 1 μg dual-labeled oligonucleotide incubated for one hour at 37°C with or without DNAseI. Decrease in Cy5 fluorescence and increase in Cy3 fluorescence indicates increased distance between the two fluorophores, due to digestion of the DNA backbone.

In a preliminary experiment to assess the feasibility of this approach *in vivo*, we

electroporated 10 μg of the dual-labeled oligonucleotide into *E. coli*, and used FC to quantify both the proportion of the population that had taken up the DNA (using the Cy5 signal), and what fraction of those cells contained intact DNA (FRET signal above the background obtained with a singly labeled Cy5 oligonucleotide). The results are shown in Figure 2-6. Approximately 16% of the cells showed Cy5 signal above the threshold, indicating DNA uptake. Of those cells, 5.2% showed FRET signal, indicating significant degradation of the nucleotide *in vivo*. This is not surprising given that WT *E. coli* is known to rapidly degrade single stranded DNA [136]. As discussed above, the quantitative results should be treated with caution. It is possible that the FRET signal is only strong enough to be seen above threshold in cells with high concentrations of oligonucleotides, such that in actuality a higher percentage of cells still have intact DNA.
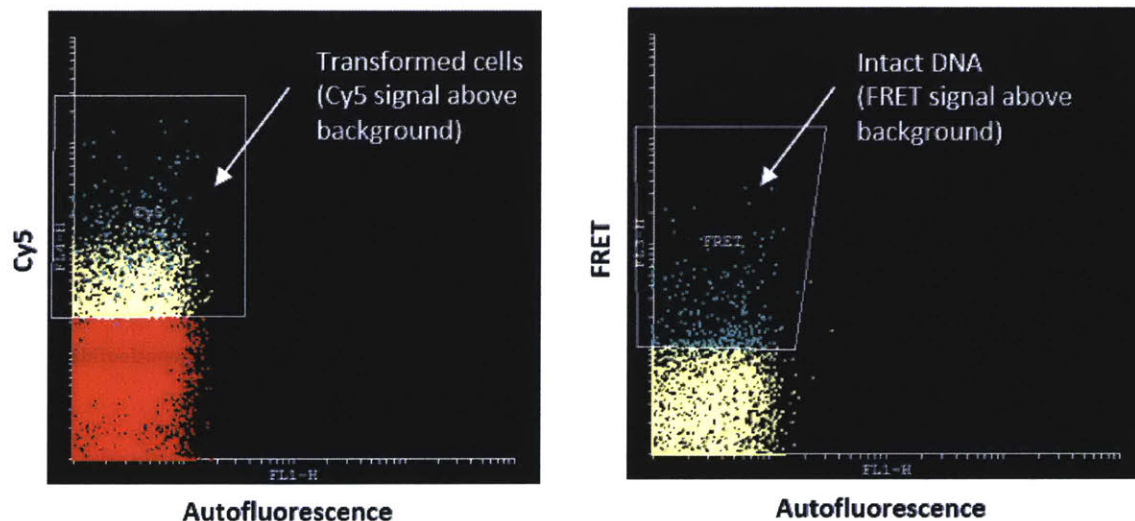
Figure 2-6: *In vivo* **digestion of a dual-labeled oligonucleotide monitored by FC.** Left: Cy5 signal used to establish percentage of cells taking up plasmid. Right: FRET signal used to identify fraction of transformed cells with intact DNA. Cy3 signal could not be monitored due to incompatibilty of the laser and filter set-up on the FC.

2.3.4 Conclusions

In summary, we have developed and a validated a simple flow cytometry method for quickly assessing different methods of DNA delivery to bacterial cells. By ap-

plying this technique to *M. thermoacetica*, we were able to determine an optimal set of electroporation conditions and confirm that plasmid was successfully delivered to cells. The assay described here is widely applicable to a range of bacterial species and should facilitate the development of genetic engineering tools for hitherto genetically intractable organisms. The ability to precisely locate multiple fluorophores on a single oligonucleotide provides additional opportunities to study the fate of DNA post-electroporation, for example by using FRET to assess DNA degradation by nucleases. While quantitative results from these assays should be treated with caution, owing to the incomplete separation of positive and negative cells, the presence of cells with fluorescence above the threshold can still be used as a binary indication of the success of DNA delivery. Future work in this area should therefore focus on the improvement of the specific fluorescence of the DNA molecule, either by finding conditions for higher fluorophore loading, or by exploring brighter fluorophores. In this work, we chose Cy5 because, of the dyes available from MirusBio for covalent modification, it had the highest quantum yield and spectral properties that put it outside of the main region of *E. coli* and *M. thermoacetica* autofluorescence. One possible alternative would be the use of Quantum Dots, which generally have superior fluorescence properties compared to organic dyes, and are available with a wide variety of emission wavelengths. These can be purchased as streptavidin conjugates, and bound to the DNA through biotinylation of the plasmid using commercially available reagents.

2.4   Evasion of *Moorella's* Restriction-Modification System

2.4.1   Introduction

After confirmation of DNA delivery, the next obstacle in transformation is the degradation of plasmid DNA by the endogenous restriction-modification system. Restriction endonucleases function as a bacterial innate immune response, and evolved as a defense mechanism against viral DNA. Separated into four major types by their

structure, their preference for methylated or unmethylated DNA, and where they cut in relation to their recognition site, these enzymes selectively degrade non-native DNA by recognizing differences in the methylation pattern of base pairs within the recognition site. Transformation of bacteria with plasmids containing unmodified recognition sites leads to low (or non-existent) transformation efficiencies, as linearized plasmid becomes a target for rapid degradation by ubiquitous exonucleases [120]. Three general strategies can be conceived for avoiding digestion: 1) Inactivation of the endonucleases, either by targeted knock-out of the endonuclease from the genome or other means; 2) Modifying vectors to avoid DNA sequences containing the recognition sites of identified endonucleases; and, 3) Circumventing endonuclease activity by pre-methylation of the plasmid DNA. The first strategy is of limited use before successful transformation of an organism has been demonstrated. The second requires identification of the specificity of the endonucleases. Until recently, this was a slow procedure that relied on using crude extracts from the target organism to digest large DNA fragments of known sequence, followed by identification of the cut sites based on the lengths of the fragments obtained [137]. With the advent of Single Molecule Real Time (SMRT) sequencing [138], which can detect methylation sites directly in genomic DNA [139], cut sites can now be identified in a high-throughput manner by recognizing the cognate methylation patterns that protect the host organism from self-restriction [140]. In the absence of knowledge of nuclease recognition sites, the third strategy is the preferred option. Plasmid DNA for transformation can be methylated *in vitro* by incubation with crude extract from the target organism supplemented with the methyl donor S-adenosyl methionine (SAM). In this reaction, EDTA is used to chelate magneisum ions and thus inactivate the restriction endonucleases that are also present in the lysate [141]. In an alternative strategy, the methyltransferase genes from the target organism, which are easily identified bioinformatically, are expressed from a compatible plasmid in the *E. coli* host strain used for plasmid propagation so that plasmid purified from this strain is already methylated. This procedure is known as Plasmid Artificial Modification System (PAMS) [142, 143].

In this section, I describe the construction and evaluation of a PAMS plasmid for pre-methylating plasmids for transformation into *M. thermoacetica*. Two modification genes identified from REBASE were cloned into an arabinose-inducible plasmid, pMETH. Transformation plasmid pMTK1 - constructed to emulate the strategy of a recently reporter transformation protocol for *Moorella* [98] - isolated from this strain was partially resistant to digestion by ApaI, which was identified as an isoschizomer of a *Moorella* endonuclease. This methylating strain was used for the subsequent preparation of the plasmids used in transformation experiments in the proceeding section.

2.4.2 Methods

Plasmid Construction

Strains and plasmids used in this study are shown in Table 2.3, and primers are shown in Table 2.4. pBAD33 was used as the base vector for expression of methyltransferases Moth_1672MS and Moth_2281M. To construct pBAD1672, the methylation and specificity domains of Moth_1672 were amplified from genomic DNA (gDNA) of *M. thermoacetica* using primers SalI-RBS-1672M-F and OE-1672M-R, and OE-RBS-1672S-F and HindIII-1672S-R, respectively. The resulting fragments were joined via Overlap Extension PCR, digested with SalI and HindIII, and ligated into pBAD33 digested with the same enzymes. To construct pBAD2281, the methylase Moth_2281 was amplified from gDNA using SacI-RBS-2281-F and XbaI-2281-R, digested, and ligated into pBAD33 digested with SacI and XbaI. pMETH1, which expresses all three genes, was constructed by ligating the SacI-2281-XbaI fragment into pBAD1672 digested with the same enzymes.

The transformation plasmid pMTK1 was constructed in two stages. First, a 1kb fragment upstream of *pyrF* (Moth_0883) was amplified from gDNA using EcoRI-pyrF-UP-F and EcoRI-pyrF-UP-R, digested with EcoRI, and ligated into pUC19 digested with the same vector to generate pUC19pyrF-UP. Sanger sequencing was used to verify the correct orientation of insertion. Next, a 1kb fragment downstream

Name	Description	Reference
Strains		
E. coli DH10β	Cloning strain	NEB
Plasmids		
pUC19	Cloning vector. pColE1-MCS-bla	NEB
pBAD33	Arabinose-inducible expression. p15A_pAra-cat	[144]
pUC19pyrF-UP	pUC19 with 1kb homology to pyrF 5' end	This study
pMTK1	pUC19pyrF-UP with 1 kb homolgoy to pyrF 3' end	This study
pBAD1672	pBAD33:Moth_1672MS	This study
pBAD2281	pBAD33:Moth_2281M	This study
pMETH1	pBAD33:Moth2281M-1672MS	This study

Table 2.3: Strains and plasmids used in *Moorella* restriction analysis

of the same locus was amplified using primers SalI-pyrF-DOWN-F and HindIII-pyrF-DOWN-R, digested with SalI and HindIII, and ligated into pUC19pyrF-UP linearized with the same enzymes. All constructs were verified via Sanger sequencing.

Preparation of Crude Lysate from *M. thermoacetica*

Moorella thermoacetica was grown to OD_{660} 1-2 (approximately) in YTF initially at pH 6.8. Cells were cooled for 30 minutes on ice, and centrifuged at 4°C for 15 minutes at 3,750 RPM. The supernatant was removed and cell pellets were stored at -80°C until used. Pellets were resuspended in 0.5 mL of Extraction Buffer per 100 mL of culture. On ice, the cell suspension was sonicated with 5 cycles of 10 seconds of sonication and 20 seconds of rest with a Micronson Ultrasonic Cell Disruptor, XL2000 at setting 10. The extract was centrifuged at 4°C for 15 minutes at 13,400 RPM and the supernatant was recovered. Per 200 μL of supernatant, 800 μL of ice cold saturated ammonium sulfate was added, followed by incubation on ice for 30 minutes, and centrifugation at 13,400 RPM for 15 minutes. The supernatant was removed and the pelleted protein was resuspended with 50 μL of Storage Buffer per 200 μL of sonicated extract. The Extraction Buffer contained Tris-HCl (50 mM, pH 8.0), glycerol (5% v/v), NaCl (100mM), PMSF (1 mM), DTT (2 mM), and lysozyme. Storage Buffer contained Tris-HCl (25 mM, pH 8.0) and Glycerol (50%). *DNA Digest with M. thermoacetica Crude Extracts* For a standard 25 μL assay, the following were

Primer Name	Sequence
Methylation vectors	
SacI-RBS-2281-F	AATAATGAGCTCAGGAGGAATAACATATGCGT CGAATATATTCCG
XbaI-2281-R	CAGGACTCTAGATCAAAAGGATACTTCATCAC
SalI-RBS-1672M-F	AGTAAGGTCGACAGGAGGATATAAAATGACGG AAAACACCAATATGGA
OE-1672M-R	ATGTTATTCCTCCTGCGGACTTACCCATTCAGC CAGCCTC
OE-RBS-1672S-F	GTCCGCAGGAGGAATAACATATGGGTAAGGAA GTTAATGAGGTG
HindIII-1672S-R	ATTGCTAAGCTTTCACTGCCTAAGGGCATCTTC
Transformation vector	
EcoRI-pyrF-UP-F	TAATAAGAATTCCCCTACCTCTCCAAGATTACC
EcoRI-pyrF-UP-R	ATATAAGAATTCGCAGGCCAGAAGCCCTTAAA
SalI-pyrF-DOWN-F	TCAAATGTCGACTAACTTCGGCCTGCTTTCAT
HindIII-pyrF-DOWN-R	TAGTCCAAGCTTATGCACCTTCCGATTTAGGT

Table 2.4: Primers used in *Moorella* restriction analysis

added: 1x NEB Buffer #4, BSA, 200 to 400 ng of plasmid, and 1 uL of protein extract. A 10 μL layer of mineral oil was added to prevent evaporation. The mixture was incubated for 4 to 18 hours at 55°C. The reaction was visualized on a 1% agarose gel or the Agilent BioAnalyzer in the BioMicroCenter (MIT). Digestion of plasmid DNA with commercially available enzymes (NEB) was performed according to the manufacturer's instructions.

Protection Assay

Plasmid pMTK1 was transformed into DH10β with or without pMETH1. Plasmid DNA was isolated after overnight growth in LB supplemented with or without 1% L-arabinose to induce expression of the methyltransferases, and digested at 37°C with PciI to linearize both plasmids. ApaI was then added to the digestion mixture and the reaction carried out for a further hour before products were analyzed via electrophoresis on a 1% agarose TAE gel. NEB 2-log ladder was used to verify fragment lengths.

2.4.3 Results and Discussion

Determination of *M. thermoacetica* recognition sequence

Digestion of the transformation vector pMTK1, which is designed to knock out the *pyrF* gene in *M. thermoacetica* and confer resistance to 5-FOA, by crude extracts of *M. thermoacetica* revealed a single distinct band corresponding to the size of the plasmid, suggesting this vector carries precisely one site recognized by *Moorella* (Figure 2-7). To identify this site, the linearized vector was further digested using other restriction
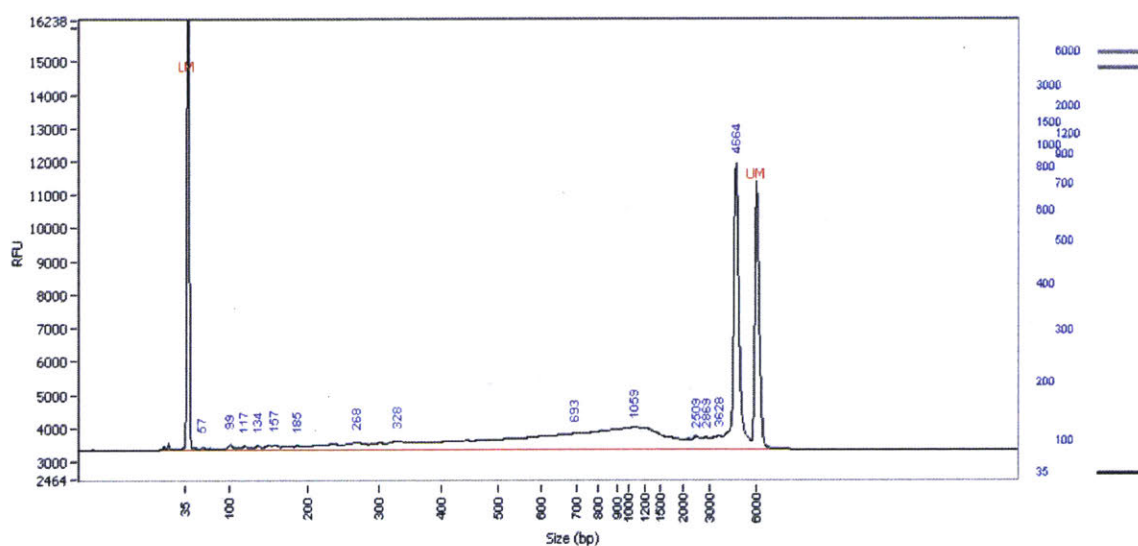
Figure 2-7: Digest of pMTK1 with Crude Extracts of *M. thermoacetica*

enzymes with cut sites on the plasmid, so that the size of the resulting double digest fragment could be used to narrow down the position of the original cut site (Figure 2-8). Analysis of these fragments suggested the *Moorella* cut site lay betwen base pairs 810 and 1279. Amplification of this region by PCR, followed by digestion with crude extracts and size determination by BioAnalyzer, more precisely identified the cutsite location, and identified ApaI as a likely isoschizomer of the *Moorella* enzyme. In support of this, digestion of the PCR fragment with ApaI produced bands of almost identical size to the digestion product of the crude lysate (Figure 2-9). To provide further confirmation of the cut site, intact plasmid was digested by crude extract, blunt-ended with Klenow, and religated. Sequencing of the ApaI site in plasmids

66
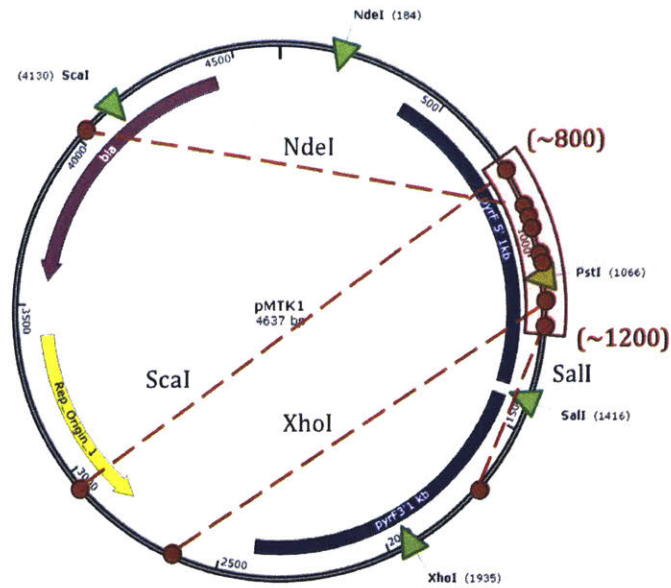
Figure 2-8: **Isolation of Cut Site Region by Sequential Digest.** Red circles are potential cut site positions based on fragment sizes obtained from digestion with commercial nucleases (green triangles).

derived from transformed cells revealed the expected missing bases, confirming the specificity of the endonuclease.

Protecting plasmid from ApaI digestion by PAMS

Having identified the specificity of at least one endonuclease in *Moorella*, we sought a cognate methyltransferase to provide protection against its activity. Inspection of the genome of *M. thermoacetica* reveals the presence of multiple putative methyltransferase enzymes (Figure 2-10). Of these, only two could be definitively identified as DNA methyltransferases based on the presence of a a cognate endonuclease downstream. Moth_2280-2281 encodes a putative Type II methylase/endonuclease, and Moth_1672 encodes a putative Type I system. Following the PAMS strategy [143], the methyltransferase regions of these operons were cloned into an arabinose-inducible plasmid with a p15a origin to generate pMETH1. This vector was designed to be compatible with the shuttle vector for transformation of *Moorella*. The transformation plasmid pMTK1 was isolated from *E. coli* DH10β with or without the methylating
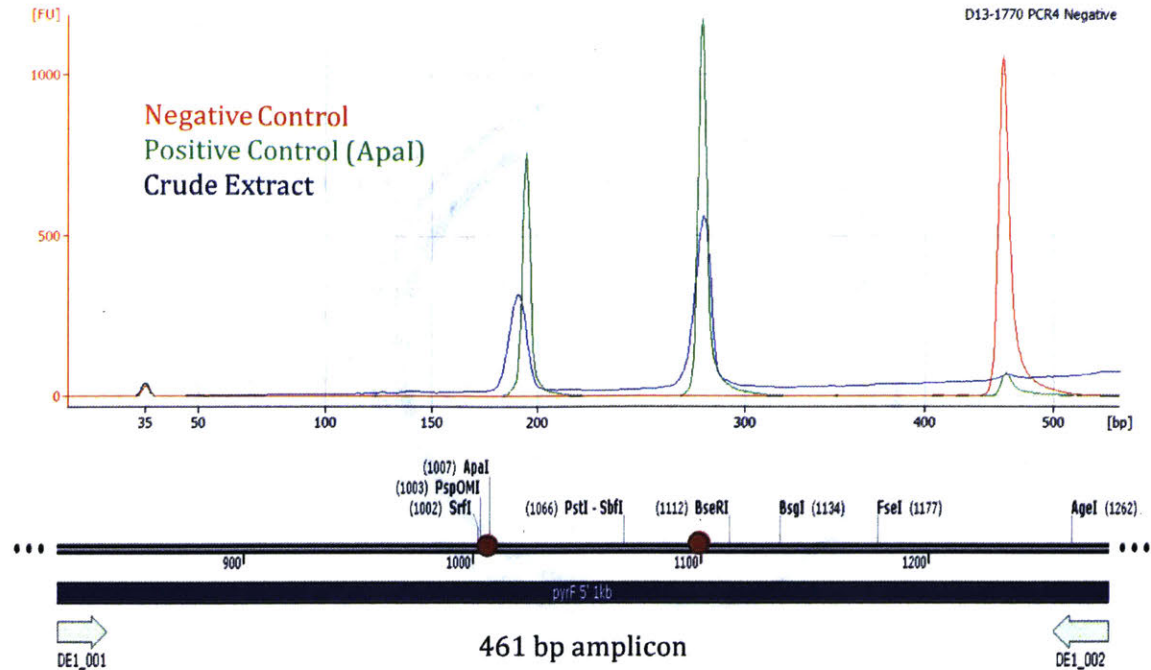
Figure 2-9: **Precise Identification of *Moorella* Cut Site**. The 461bp region containing the cut site as identified by sequential digest was amplified by PCR, digested with crude lystates from *M. thermoacetica*, and analyzed by BioAnalyzer to precisely determine product size and identify cut site location. Red: undigested control. Blue: amplicon digested with crude lysate. Green: amplicon digested with ApaI.

plasmid pMETH1. DH10β was chosen as a cloning host because it does not consume arabinose, which ensures the concentration of the inducer remains constant throughout growth. Both plasmids were linearized by PciI to avoid the presence of multiple isoforms during electrophoresis. After linearization, the reaction was separated into two aliquots, and ApaI was added to one aliquot, and the reaction allowed to proceed for another hour. Since ApaI was identified as an isoschizomer of an endonuclease from *Moorella*, we hypothesized that premethylation of pMTK1, which contains a single ApaI site, should retard the rate of digestion by this enzyme. Indeed, as shown in Figure 2-11, the majority of the plasmid thus isolated remained undigested (Columns 6 and 7), compared to the plasmid isolated from a strain lacking pMETH1, which was completely digested (Column 5). Interestingly, the plasmid was protected regardless of whether arabinose was added. This probably indicates leaky expression from this promoter in rich LB medium in the absence of glucose. Protection of pMTK1 was

68
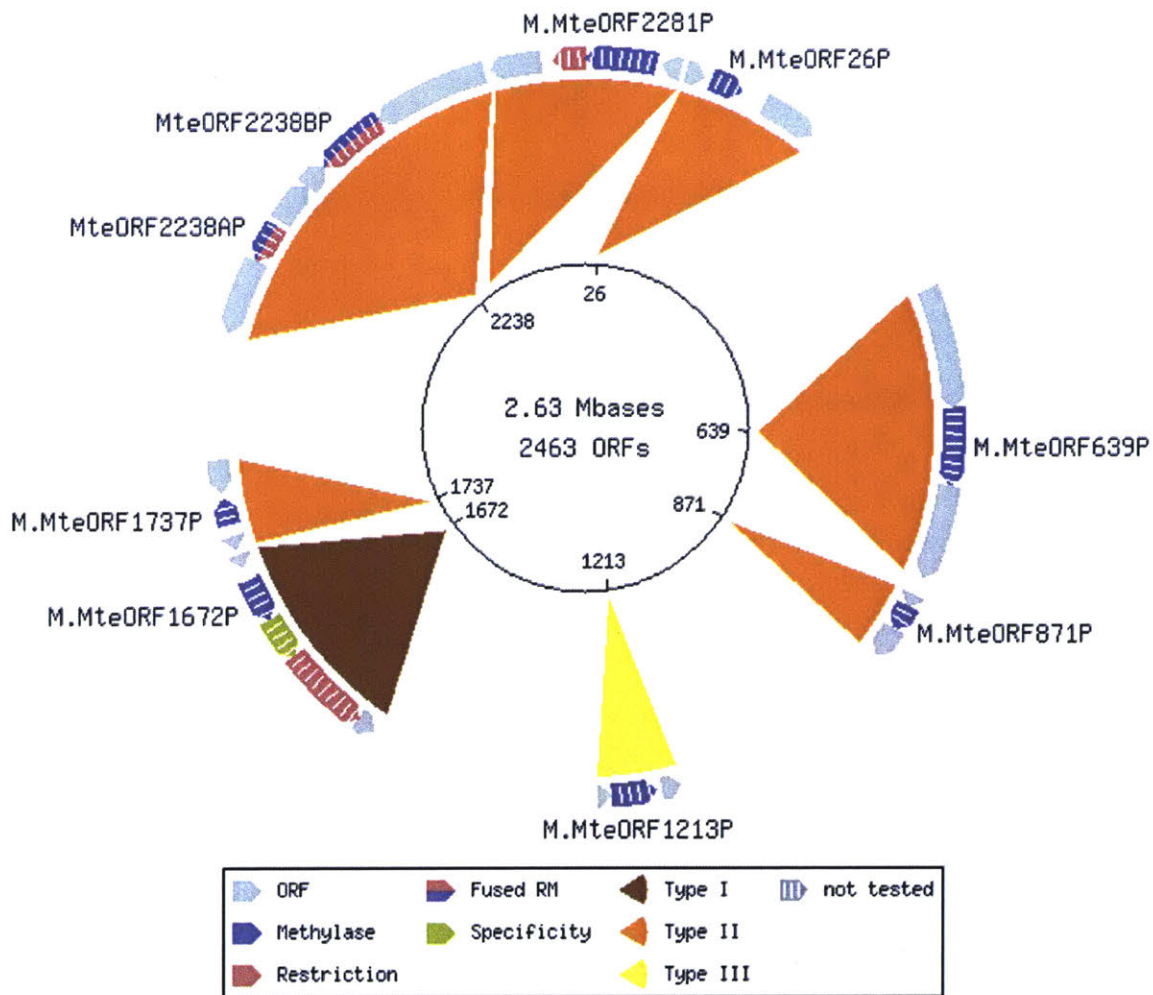
Figure 2-10: Methyltransferases and Endonucleases in *M. thermoacetica*

not complete, as indicated by the faint digestion bands in Columns 6 and 7. The reason for this is unclear, but could be due to low activity of the methyltransferases under these conditions. Regardless, that the majority of the plasmid remained intact confirms the active expression of at least one of these enzymes.

2.4.4 Conclusions

In this section, a plasmid was generated to allow methylation of co-purified vectors for transformation into *M. thermoacetica*, and the activity of at least one of the methyltransferases was verified by showing protection against ApaI digestion. The activity of the second could not be verified, as the cognate recognition site is presently
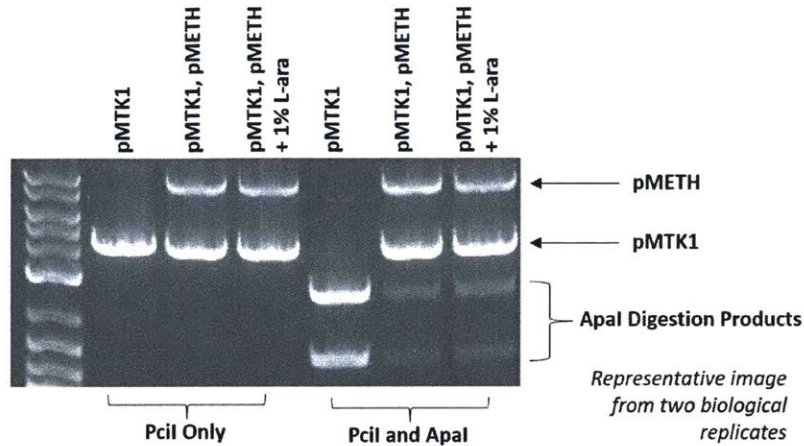
Figure 2-11: **Protection of pMTK1 from ApaI via premethylation.** Column 1: NEB 2-log ladder. Columns 2-4: Plasmids treated with PciI. Columns 5-7: Plasmids treated with PciI and ApaI

unknown. The ApaI methyltransferase is not commercially available, therefore the strategy developed here for pre-methylation is the best option in cases where the ApaI recognition site cannot be conveniently deleted from the transformation vector. As this strategy may also provide protection against the other endonuclease, its use is recommended for purification of all vectors designed for *Moorella* transformation, regardless of whether they contain an ApaI site.

2.5   Transformation of *Moorella* with Replicative and Integrative Plasmids

2.5.1   Introduction

With the ability to deliver plasmid DNA to *Moorella* confirmed, and a system in place for partially protecting that DNA from digestion, we turned out attention to the design of vectors and selectable markers. To maximize our chances of success, we investigated both replicating vectors that could exist autonomously from the chromosome, and suicide vectors designed to integrate cargo into the genome via homologous recombination. After determining the MIC of several antibiotics and

70

comparing these against resistance genes whose proteins had been reported to work well in thermophiles, we settled on kanamycin as the antibiotic to work with. Due to the ubiquitous use of 5-fluoroortic acid (5-FOA) as a counter-selectable marker in *Clostridial* genetic engineering [145], and the publication during this work of a transformation protocol for *M. thermoacetica* using this marker [98], one of the integration vectors was targeted to the *pyrF* locus. The locus of the endonuclease Moth_2280 was chosen as a second integration site, in an effort to improve future transformation efficiencies. In all, five vectors were constructed, and purified from a strain containing pMETH1 (as described in Section 2.4), and electroporated into *M. thermoacetica*. In this section, I describe the construction of these vectors, the results of their transformation into *Moorella thermaocetica*, and analysis of the resulting recombinant strains.

2.5.2 Methods

Plasmid Construction

Strains and plasmids used in this study are shown in Table 2.5, and primers are shown in Table 2.6. Plasmid pMTBW1 was constructed in three steps. First, a 500bp region upstream of the endonuclease Moth_2280 was amplified using EcoRI-2280-5-F and Xma-2280-5-F, digested, and ligated into pUC19 linearized with EcoRI and Xma to generate pUC19-2280UP. Next, the 500 bp region downstream of the same endonuclease was amplified using PstI-2280-3-F and HindIII-2280-3-R, digested, and ligated into pUC19-2280Up digested with the same enzymes, generating pUC19-2280. To generate pMTBW1, a DNA fragment containing the codon optimized thermostable kanamycin resistance gene from pUCG18 [133] downstream of the putative GAPDH promoter from *M. thermoacetica* was synthesized with flanking XmaI and BamHI sites (BlueHeron Biotech, see A.2 for full sequence), digested, and ligated into pUC19-2280 linearized with XmaI and BamHI. To clone GAPDH:kanR into the replicative vectors pUCG18 and pIKM1, the fragment was amplified from pMTBW1 with primers SacI-kanR-F/R, digested, and ligated into vectors linearized with the same restriction enzymes. pQexp-kanR was generated in the same manner, but using the BamHI

version of the same primers. All constructs were verified by Sanger sequencing.

Cell Growth and Competent Cell Preparation

M. thermoacetica was cultivated and prepared for electroporation as described in Section 2.3. To determine the MIC, N-SCM medium was supplemented with various concentrations of antibiotic, transferred to Hungate tubes, and inoculated at 1% (v/v) with WT cells in late exponential phase. To determine the working concentration of 5-FOA, a minimal version of N-SCM was used in which yeast extract and tryptone were omitted. 5-FOA was added in powder form at various concentrations and dissolved by vigorous agitation at 55°C, before inoculation at 1% (v/v) with WT cells grown in the same medium. For antibiotics and 5-FOA, resistance was assessed by measuring culture turbidity after 7 days by spectrophotometer at 600 nm.

Transformation, Outgrowth and Selection

Electroporation of *M. thermoacetica* was performed as described in Section 2.3, but with 2μg plasmid DNA. Cells were recovered in 10 mL N-SCM or N-SCM-Minimal at 55°C with agitation for 24-72 hours post-electroporation, then 0.5 mL was transferred into anaerobic vials with 5 mL semi-solid medium (N-SCM or N-SCM-Minimal amended with 0.3% agar), supplemented with 2 mg mL^{-1} 5-FOA and 10 μg mL^{-1} uracil or 200 μg mL^{-1} kanamycin. Vials were capped and incubated anaerobically without shaking at 55°C. Colonies typically formed within 4 days, and were inoculated into liquid media by carefully pipetting the colony. After growth to an approximate OD of 0.1, potential transformants were cryopreserved in anaerobic vials by combining 1 mL of culture with 1 mL of 50% glycerol, capping the vials with rubber septa and hungate caps, and storing them immediately at -80°C.

Genotype Analysis

Genomic DNA was extracted from saturated liquid cultures using the DNeasy Ultra-Clean Microbial Kit (MOBIO). Insertion into the *pyrF* locus was evaluated using

72

Name	Description	Reference
Strains		
E. coli DH10β	Cloning strain	NEB
M. thermoacetica ATCC 39073	Wild-type strain	ATCC
Plasmids		
pUC19	Cloning vector. pColE1-MCS-bla	NEB
pMETH1	Methylation vector	Section 2.4
pMTK1	Integration into *pyrF* locus	Section 2.4
pUCG18	*Geobacillus thermoglucosidasius* shuttle vector	[133]
pIKM1	*Thermoanaerobacteria sp.* shuttle vector	[146]
pQexp	*Clostridium ljungdahlii* shuttle vector	[91]
pUCG18-kanR	pUCG18 with GAPDH:kanR	This study
pIKM1-kanR	pIKM1 with GAPDH:kanR	This study
pQexp-kanR	pQexp with GAPDH:kanR	This study
pUC19-2280UP	pUC19 with 500bp homology region upstream of Moth_2280	This study
pUC19-2280	pUC19-2280UP with 500bp downstream of Moth_2280	This study
pMTBW1	pUC19-2280 with GAPDH:kanR	This study

Table 2.5: Strains and plasmids used for *Moorella* transformation

Primer Name	Sequence
EcoRI-2280-5-F	ATTATTGAATTCATTACCCTGTTGCCGTGGCC
XmaI-2280-5-R	ATTATACCCGGGACACGATGGTTTCGCGCCTC
PstI-2280-3-F	ATTATTTTAGAACGCCGGTACCCCAA
HindIII-2280-3-R	ATTATTAAGCTTATTACCCAGGAAAAAGCCGC
SacI-kanR-F	TAATAAGAGCTCACCATTTGTGTTGAATAGATAGT
SacI-kanR-R	TAATAAGAGCTCCTAAAAGGGGATACGTTTCGAGA
BamHI-kanR-F	TAATAAGGATCCACCATTTGTGTTGAATAGATAGT
BamHI-kanR-R	TAATAAGGATCCCTAAAAGGGGATACGTTTCGAGA
pyrF-F1	GTCCTCCAGTCCCTGAAGTC
pyrF-R1	CGATTTGCCTGGTATCCCTG
pyrF-mod-F	GCTTCTGGCCTGCGAATTC
pyrF-mod-R	AGCAGGCCGAAGTTAGTCGA
kanF	CTTGCTGGATTATGCCTCCC
kanR	GTTTCTCACTGTCGCTCAACTG

Table 2.6: Primers used in *Moorella* transformation

PCR and combinations of the pyrF primers listed in Table 2.6. Presence of the kanR gene was evaluated using the primers kanF and kanR.

2.5.3   Results and Discussion

Working concentrations of antibiotics and 5-FOA

Four antibiotics were evaluated for their ability to prevent growth of wild-type *M. thermoacetica*: Erythromycin, Chloramphenicol, Thiamphenicol, and Kanamycin. Exponential phase cells were subcultured at 1% (v/v) at antibiotic concentrations between 0.1 and 60 μg mL^{-1}, and the effectiveness was evaluated after 7 days based on culture turbidity. The results are shown in Figure 2-12. All the chosen antibiotics were effective at reasonable concentrations. As kanamycin is the most thermostable of these antibiotics [133], it was chosen for further investigation, and we cloned a codon-optimized thermostable kanamycin resistance [133] under the control of the putative *M. thermoacetica* GAPDH promoter into each of the replicating vectors to be screened. 5-FOA itself is not toxic. Its decarboxylation via OMP decarboxylase, an enzyme in the uracil biosynthesis pathway encoded by *pyrF*, produces the antimetabolite 5-fluorouracil (5-FU), which is highly toxic. Deletion of the *pyrF* gene

Antibiotic concentration (µg/ml)

| 0.1 | 0.2 | 0.4 | 0.8 | 1.5 | 3 | 4 | 5 | 6 | 8 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 |

Key

Strong growth*		$OD_{600} > 0.5$
Low growth		$0.05 < OD_{600} \leq 0.5$
No growth		$O_{D600} \leq 0.05$
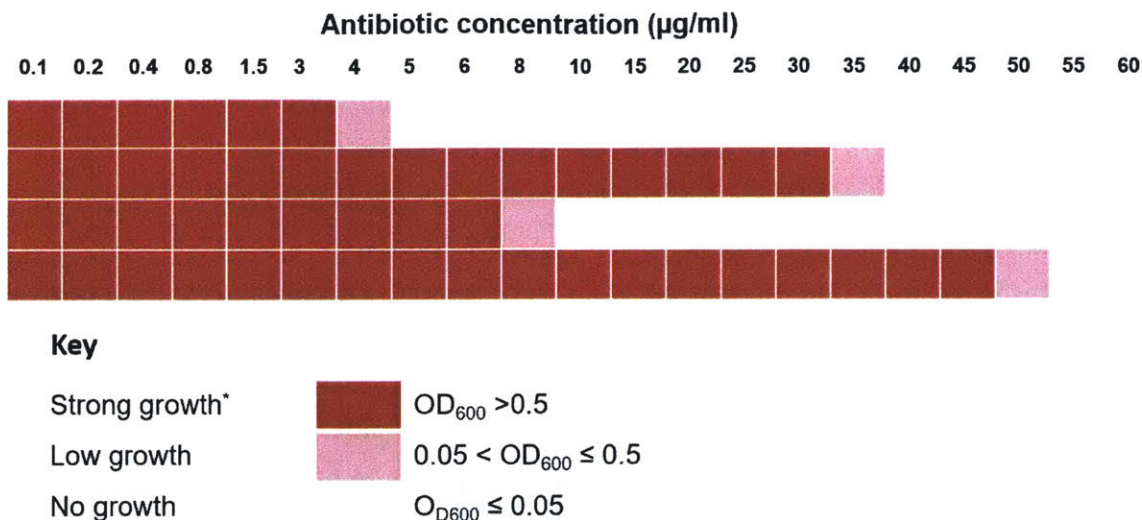
Figure 2-12: **Evaluation of antibiotics for _M. thermoacetica_.** Top to Bottom: Erythromycin, Chloramphenicol, Thiamphenicol, Kanamycin.

therefore provides resistance against 5-FOA, but also converts the strain into a uracil auxotroph. This marker is highly useful for multi-step strain construction, because the presence can be selected for via uracil prototrophy, and the absence selected for by 5-FOA. Following a similar protocol as for the antibiotics above, we tested the MIC of 5-FOA in _Moorella_. 5-FOA is available both in powder form and as a 100 mg mL^{-1} solution in DMSO (Zymo Research). In preliminary experiments, it was found that DMSO in the solution became toxic at concentrations lower than the 5-FOA, and we therefore switched to using solid 5-FOA despite the inconvenience of dissolving the relatively insoluble powder. Concentrations between 0 and 3 mg mL^{-1} were screened, and 2 mg mL^{-1} was found to be sufficient to completely prevent growth (Figure 2-13). Two vectors reported to function in other Gram positive thermophiles were chosen as promising candidates to screen in _M. thermoacetica_: pUCG18 from _Geobacillus thermoglucosidasius_, and pIKM1 from _Thermoanaerobacteria sp.._. In addition, we screened pQexp, which had recently been reported to function in the related aceto-gen _Clostridium ljungdahlii_. To restrict differences between these vectors as much as possible to their replication machinery, we cloned the same GAPDH:kanR resistance cassette into each one. After construction and purification of the various vectors from the methylating _E. coli_ strain, multiple attempts were made to transform them into

5-FOA concentration (mg / mL)

| 1 | 1.5 | 2 | 2.5 | 3.0 |

Key

Strong growth* $OD_{600} > 0.5$

Low growth $0.05 < OD_{600} \leq 0.5$

No growth $O_{D600} \leq 0.05$
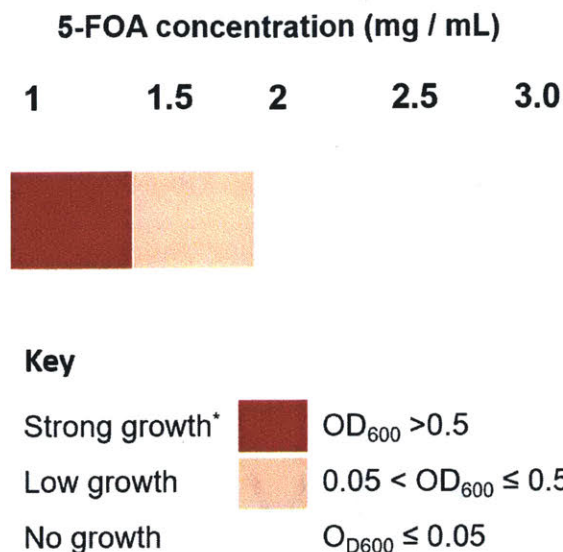
Figure 2-13: Evaluation of FOA inhibition of *M. thermoacetica* growth

M. thermoacetica over the course of approximately 6 months. In these experiments, aliquots of cells from recovery media were inoculated into selective media at several time points: typically 8 hours, 24 hours, 48 hours, and 72 hours post-electroporation. Negative controls (no plasmid) were always included to rule out contamination or breakdown of kanamycin or 5-FOA. The majority of experiments yielded no colonies. On two occasions, resistant colonies were obtained: One after transformation with pMTK1, targeting deletion of the *pyrF* gene, and one with the replicating plasmid pQexp-kanR. Analysis of these transformants is described below.

Analysis of the *pyrF* locus following transformation with pMTK1

In one experiment, FOA-resistant colonies arose after transformation with pMTK1 pre-methylated with pMETH1. 10 of these colonies were picked, grown in selective media, and genomic DNA extracted for PCR analysis. Wild-type cells, as well as purified pMTK1 plasmid were used as controls. The *pyrF* locus was amplified using primers pyrF-F1 and pyrF-R1, and size analyzed via gel electrophoresis (Figure 2-14). Deletion of the *pyrF* gene should yield a band of 726 bp, whereas the unmodified gene yields a 1.5 kb amplicon. Plasmid and WT controls amplified as expected. For all the colonies tested, a band corresponding to the full-length product was detected,

Figure 2-14: PCR analysis of *pyrF* locus in potential transformants

indicating that complete deletion of the gene had not occurred. To further investigate this difference between phenotype and genotype, the amplicons were purified and sequenced. In 8 out of the 10 colonies, the same 36-bp in-frame deletion was detected (Figure 2-15). Though distal from the active site of the cognate enzyme, this 12 AA deletion is likely the reason for the loss of decarboxylase function. In the remaining 2 colonies, the *pyrF* sequence was intact. While resistance of 5-FOA is therefore unexplained in these clones, the lack of the targeted deletion in any of the analyzed clones suggests the ineffectiveness of the homologous recombination-based strategy used in this experiment.

Figure 2-15: Sequencing of *pyrF* mutants

Analysis of transformants resulting from electroporation of pQexp-kanR

In another experiment, electroporation of pQexp-kanR led to kanamycin-resistant colonies. To verify resistance was conferred by the vector, and not a spontaneous mutation, a single colony was grown, and DNA isolated. PCR with primers kanF and kanR confirmed the presence of the kanamycin resistance gene (Data not shown), and transformation of chemically competent *E. coli* DH5α resulted in erythromycin-resistant colonies, indicating the plasmid was maintained in circular form. To provide further evidence for autonomous replication, primers were designed that would amplify fragments of pQexp-kanR in overlapping fashion all the way around the plasmid, such that linearization and insertion of the plasmid backbone into the *Moorella*

genome would result in unsuccessful amplification in at least one primer set. Amplification of the extracted DNA with these primer sets resulted in successful amplification in all cases, confirming autonomous replication and the functionality of the pQexp replication origin in *Moorella*. Emboldened by this success, we set out to replicate this result. However, despite multiple attempts over the course of several months, we were unable to establish resistant colonies in a controllable manner. In the majority of experiments, no resistant colonies arose.

2.5.4   Conclusions

With optimized electroporation conditions established, and a method to methylate DNA prior to transformation, we set about constructing and screening several vectors for transformation into *Moorella*. While we were occasionally successful in generating FOA-resistant mutants, full length deletion of *pyrF* was not achieved. Likewise, we showed that pQexp-kanR could be stably maintained as an autonomus replicon, but with limited reproducibility.

The quest for robust transformation of *M. thermoacetica* consumed two years of my PhD work, and resulted in sparse glimmers of success. During the course of these experiments, another group published a transformation protocol based on 5-FOA resistance [98]. Despite incorporating a transformation vector almost identical to theirs (pMTK1) in our panel of plasmids, and pre-methylating vectors via the same PAMS strategy, we were unable to consistenly reproduce this result: The only *pyrF* mutants isolated were partial deletions, not useful for the insertion of heterologous genes. We had occasional success with one of the replicating vectors that we designed (pQexp-kanR), but again, positive results were sporadic. At around the same time, a highly optimized transformation protocol for the related acetogen *C. ljungdahlii* was published [91]. Since the ultimate focus of my thesis work was to be the metabolic engineering of syngas-fermenting organisms, the choice of host was largely immaterial, and the ability to easily reproduce *C. ljlungdahlii* transformation convinced us that efforts should instead be focused on that organism. The flow-cytometry assay for detection of fluorescent plasmid delivery developed during this work is nevertheless a

useful technique that could be harnessed in the development of transformation tools for other organisms.

.

Chapter 3

Development of a CRISPRi System for *Clostridium ljungdahlii*

3.1 Summary

Clostridium ljungdahlii has emerged as an attractive candidate for the bioconversion of synthesis gas (CO, CO_2, H_2) to a variety of fuels and chemicals through the Wood-Ljungdahl (WL) pathway. However, the lack of robust tools for generating gene deletions and inducible knockdowns in this organism limits the ability to both divert carbon flux away from acetate toward higher-value products, and to assign gene function in the poorly annotated genome. To address this need, in this chapter I describe the development of a CRISPRi system for *C. ljungdahlii* that allows for the controlled knockdown of specific genes using the deactivated Cas9 nuclease (dCas9) targeted to the coding region of the gene of interest by a single-guide RNA (sgRNA). Expression of *dCas9* from a lactose-inducible promoter using a variety of different ribosome binding sites was first verified by Western analysis, and expression of the sgRNA from an engineered promoter for *C. cellulolyticum* in the same plasmid was confirmed by RT-PCR. qRT-PCR analysis of strains carrying derivatives of these plasmids with sgRNAs for several genes with proposed roles in energy conservation and carbon flux revealed up to 30-fold downregulation of the target transcript. To address leaky expression of *dCas9* from the lactose promoter, which was found to lead to constitutive

downregulation, we screened the p_{tet} promoter developed for *C. acetobutylicum*, and demonstrated tighter regulation of expression of both a reporter gene (*gusA*) and *dCas9*. The improved system significantly reduced background downregulation of *pta*, resulting in a 5-fold difference in expression between uninduced and inducing conditions. The regulated strain grew much more slowly than the uninduced strain, but interestingly, the loss of PTA activity did not result in increased ethanol yield, suggesting a yet unidentified regulatory system in *C. ljungdahlii* for controlling production of acetate and ethanol. The results demonstrate the utility of the CRISPRi strategy as a tool to interrogate native regulatory networks and divert carbon flux, and highlight the complexities of autotrophic fermentative metabolism which must be overcome to enable industrially revelant production in these strains. As a proof of concept to test the ability of the CRISPRi system to divert flux to heterologous product synthesis in an engineered strain, compatible plasmids were constructed that enabled 3-hydroxybutyrate (3HB) synthesis.

3.2 Introduction

Synthesis gas (syngas) fermentation by anaerobic bacteria is a novel method for the biological production of fuels and chemicals from gaseous substrates such as carbon monoxide (CO), carbon dioxide (CO_2), and hydrogen (H_2). These substrates can be renewably produced via biomass gasification, thus the biological gas-to-liquids (GTL) process represents a sustainable platform that will be important for reducing dependence on fossil fuels. However, the range of products available by syngas fermentation, as exemplified by the commercial processes of LanzaTech, Coskata and INEOS, is currently limited to ethanol by the absence of well-established metabolic engineering tools for the organisms capable of metabolizing the gases in question.

Clostridium ljungdahlii has emerged as a potential platform organism for syngas fermentation. Its genome sequence and preliminary genetic tools that enabled low titers of butanol production were reported in 2010 [90]. More robust genetic tools, including several plasmids with compatible origins of replication and a scheme for gen-

erating knockouts via homologous recombination followed in 2013 [91]. The first inducible promoter followed in 2014 [93], along with an extensive metabolic engineering effort to produce butyrate [95]. The generation of a genome-scale model (GSM) [147] and its combination with spatio-temporal reactor models [148] have allowed insight into the range of products available from syngas. The recent report of CRISPR-cas9 tools for more efficient knockout generation [149], as well as a temperature-sensitive origin of replication and fluorescent protein [92] have further expanded the scale of genetic engineering possible. At the same time, the biochemistry of acetogenesis and ethanol production in this host and its close relative *C. autoethanogenum* is being worked out at the enzymatic level [61, 150], and by targeted knockout of genes with putative roles in these processes [151, 152].

Missing from this growing list of technologies is the ability to make conditional knockouts – or multiple conditional knockouts simultaneously – rapidly and with minimal cloning. Such a tool would enable the interrogation of essential genes without the tedium associated with curing temperature-sensitive plasmids. There is still a lot that we don't understand about the fundamentals of acetogenesis [41], and this knowledge is critical to the rational engineering of strains to produce high titers of value-added products. Improved genetic methods for assessing fundamental questions in metabolism are therefore called for [39]. Inducible downregulation can also support dynamic metabolic engineering strategies for manipulating carbon and electron flux in engineered strains. This is an emerging theme in metabolic engineering [153] that recognizes that, while increasing product yield, static downregulation of pathways that compete for carbon flux but are essential for growth, such as acetate, can limit overall productivity.

Recently, CRISRPi has emerged as a powerful technique for gene downregulation. Leveraging the programmabilty and specificity of the protein-DNA binding mediated by CRISPR-cas9, this approach uses a nuclease-deficient cas9 (dCas9) to block transcription of a gene specified by the N20 region of a single-guide RNA (sgRNA). This approach was first demonstrated in *E. coli* for the control of a fluorescent reporter gene [154]. Since then, it has been multiplexed to allow simultaneous control

of multiple genes, and used to improve naringenin production in engineered *E. coli* [155]. It has also been adapted for use in a variety of other organisms, including *Corynebacterium glutamicum* [156], *Saccharomyces cerevisae* [157], and many others. The implementation of this approach is conceptually simple, as it relies only on the expression of two components, the dCas9 and the sgRNA, which is feasible with the tools already described for *C. ljungdahlii*. *C. ljungdahlii* also does not have a native CRISPR system, further simplifying the development.

In this work, we therefore adapted CRISPRi for use in *C. ljungdahlii*, with the aim of assessing targets whose downregulation lead to increased acetyl-CoA flux to heterologous products. We demonstrate the ability to reduce expression of three enzymes thought to be involved in energy conservation and carbon flux – Phosphotransacetylase (PTA), Aldehyde/alcohol dehydrogenase (AdhE1), and Aldehyde Ferredoxin Oxidoreductase (AOR) – by up to 97%. Screening of a tet promoter for control of *dCas9* expression improved the difference between inducing and non-inducing conditions, and enabled characterization of physiological effect of PTA downregulation. The results establish CRISPRi as a potent strategy for the manipulation of transcription in *C. ljungdahlii*, an important contribution to the genetic toolkit.

3.3 Materials and Methods

3.3.1 Reagents

Unless specified, all chemical reagents were purchased in the highest grade available from Sigma. LB, BACTO Agar, and casamino acids were purchased from BD. Trace Elements (MD-TMS) and Vitamin Solution (MD-VS) were purchased from ATCC.

3.3.2 Strains and Plasmids

Strains and plasmids used in this study are shown in Tables 3.1 and 3.2, respectively. *C. ljungdahlii* PETC was obtained from the American Type Culture Collection (ATCC #55383). *E. coli* DH5α was used for general cloning, and *E. coli* NEBEx-

Name	Description	Reference
C. ljungahlii ATCC 55383	Wild-type strain	ATCC
C. ljungahlii ATCC 55383 Δ*AdhE1*	Wild-type strain with AdhE1 knocked out	[91]
E. coli DH5α	Cloning strain	NEB
E. coli NEBExpress	Plasmid purification strain	NEB

Table 3.1: Strains used in this work

press was used for preparation of plasmids for transformation into *C. ljungdahlii.* Both were purchased as competent cells from NEB. The adhE1 knockout strain [91] was a generous gift from Derek Lovley. The pMTL modular plasmid system [158] was a gift from Nigel Minton. Plasmid pdCas9-bacteria was a gift from Stanley Qi (Addgene #44249) [154]. Genomic DNA from *Clostridium acetobutylicum* 824 was a generous gift from Terry Papoutsakis. pAH2 was a gift from Stephen Melville [159]. pBbS2k-RFP was a gift from Jay Keasling (Addgene plasmid #35330). *R. eutropha* was provided by Anthony Sinskey. Plasmid pJIR750ai was purchased from Sigma.

3.3.3 Cloning

Primers used in the construction of CRISPRi vectors and 3HB vectors are shown in Tables 3.3 and 3.4, respectively. Q5 Polymerase (NEB) was used for all amplifications, and all amplicons were digested with DpnI before PCR purification. Gibson Assembly (GA) Master Mix (NEB) was used for construction of plasmids. All restriction endonucleases were purchased from NEB. Assembled vectors were transformed into DH5α chemically competent cells (NEB), and verified by Sanger sequencing before being transformed into NEBExpress (NEB).

pdCas9x variants were constructed by amplifying dCas9 with dCas9_fwd_X and dCas9_rev using pdCas9-bacteria as template, and assembling the purified fragment with pAH2 digested with PstI and SacI. pEG100 was constructed by amplifying the evoglow gene with primers 100EG_fwd/100EG_rev and assembling with them with the same digested pAH2 backbone. sgRNA under control of the P4 promoter [161] was synthesized as a gBlock by IDT with overhangs to mediate gibson assembly with either pdCas9B or pEG100. The sequence of the PTA GBlock is shown in Appendix

Name	Description	Reference
pAH2	Lactose-inducible shuttle vector. pColE1 (*E. coli*), pIP404 (*C. ljungdahlii*), *catP*, bgaR-P_{bgaL}:gusA	[159]
pdCas9-Bacteria	p15A, *catP*, $p_{LtetO-1}$:dCas9	[154]
pMTL82254	pColE1 (*E. coli*), pBP1 (*C. ljungdahlii*), *ermB*, *catP* reporter	[158]
pBbS2k-RFP	pSC101 tetR pTet-RFP Kn^R	[160]
pCL2	Derivative of pJIR750ai. pColE1 (*E. coli*), pIP404 (*C. ljungdahlii*), *catP*	[91]
pdCas9A	pAH2 with dCas9 from pdCas9-bacteria and RBS from gusA	This work
pdCas9B	pAH2 with dCas9 from pdCas9-bacteria and RBS from *C. ljungdahlii ack*	This work
pdCas9C	pAH2 with dCas9 from pdCas9-bacteria and RBS from *C. ljungdahlii buk*	This work
pEG100	pAH2 with anaerobic fluorescent protein derived from *P. putida* (evoglow-Pp1)	This work
pdCas9B-sgRNApta	pdCas9B with sgRNA for *pta* (CLJU_c12760) driven by *C. cellulolyticum* P4 [161]	This work
pdCas9B-sgRNAadhE1	pdCas9B with sgRNA for *adhE1* (CLJU_c16510) driven by *C. cellulolyticum* P4 [161]	This work
pdCas9B-sgRNAaor2	pdCas9B with sgRNA for *aor2* (CLJU_c20210) driven by *C. cellulolyticum* P4 [161]	This work
pEG-sgRNApta	pEG100 with sgRNA for *pta*	This work
pEG-sgRNAadhE1	pEG100 with sgRNA for *adhE1*	This work
pEG-sgRNAaor2	pEG100 with sgRNA for *aor2*	This work
pBbS2k-2tetO1-RFP	pBbS2k with $p_{miniThl}$tetR and p_{2tetO1}RFP from [162]	This work
pMTL822542tetO1gusA	pMTL82254 with $p_{miniThl}$tetR and p_{2tetO1}gusA	This work
pMTLdCas9-sgRNApta	pMTL82254 with $p_{miniThl}$tetR and p_{2tetO1}dCas9 and sgRNA for *pta* driven by *C. cellulolyticum* P4	This work
p3HB100	pAH2 with *phaA* and *phaB* from *R. eutropha*, *ptb* and *buk* from *C. acetobutylicum*	This work
p3HB130	pAH2 with *phaA* from *R. eutropha*, *hbd* from *C. acetobutylicum*, and *tesB* from *E. coli*	This work
p3HB160	pAH2 with *thl* and *hbd* from *C. acetobutylicum*, and *tesB* from *E. coli*	This work
p3HB131	p3HB131 with optimized RBSs and genes codon optimized for expression in *C. ljungdahlii*	This work
pCL2pta3HB131	pCL2 with putative *pta* promoter from *C. ljungdahlii* and 3HB operon from p3HB131	This work

Table 3.2: Plasmids used in this study

A.3 as an example, and the variable N20 region used for each gene is shown in Table 3.5.

The tet promoter system was constructed in two steps. First, a fragment with the mini P_{thl} promoter from [162] was generated by annealing primers miniPthl_F and miniPthl_R, and a fragment containing p2TetO1 was generated by annealing 2tetO1_F and 2tetO1_R. Annealing was carried out in IDT Duplex Buffer following the manufacturer's recommendation.[1] These fragments were combined in a triple ligation with pBbS2k digested with XbaI and EcoRI to generate plasmid pBbS2k-2tetO1-RFP, where *tetR* expression is driven by the P_{thl} promoter, and RFP is under control of the 2TetO1 promoter. To clone this promoter construct into pMTL82254 along with gusA or dCas9B, the construct was amplified with p2TetO1_fwd/p2TetO1_rev. gusA was amplified from pAH2 with gusA_fwd/gusA_rev, and dCas9 along with the *pta* sgRNA was amplified with dCas9B-SalI_fwd/dCas9B-SalI_rev using pdCas9B-sgRNA-pta as template. These fragments were assembled with pMTL82254 digested with NdeI and MluI to generate constructs pMTL822542tetO1gusA and pMTLdCas 9-sgRNApta, respectively.

p3HB100, 130, and 160 were constructed by amplifying each gene with the relevant primers in Table 3.4, and assembling them with pAH2 digested with PstI and SacI. To generate p3HB131, a codon usage template was generated for *C. ljungdahlii*, and used to optimize codon usage for *phaA* and *tesB* with software provided by BlueHeron. The resulting fragments were synthesized as GBlocks by IDT, amplified along with *hbd* with the listed primers, and again assembeld into the digested pAH2 backbone. pCL2pta3HB131 was constructed to mirror the design of similar constructs in [95]. First, pCL2 was generated from pJIR750ai (Sigma) as described in [91]. Then, the *pta* promoter was amplified from genomic DNA of *C. ljungdahlii* using PTA_Prom_fwd/PTA_Prom_rev. The p3HB131 operon was amplified from plasmid template with primers 131_Op_fwd/131_Op_rev. The purified fragments were assembled with pCL2 digested with KpnI and BamHI.

---

[1]IDT Oligo Annealing Protocol: https://www.idtdna.com/pages/decoded/decoded-articles/pipet-tips/decoded/2012/06/15/annealing-oligonucleotides

Primer Name	Sequence
dCas9 expression in pAH2	
dCas9_fwd_A	aataccatgaagatactgcagaaataaggagatgttaattataatATGGATAAGAAATACTCAATAGGCTTAGC
dCas9_fwd_B	aataccatgaagatactgcaGAAATAAGGAGGAATATTAAACATGGATAAGAAATACTCAATAGGCTTAGC
dCas9_fwd_C	aataccatgaagatactgcaGAAATAGTGGAGGAATGTTAACtATGGATAAGAAATACTCAATAGGCTTAGC
dCas9_rev	tgtaaaacgacggccagtgccgagctgtcgacAACGCAGAAAGGCCCACC
EvoGlow expression in pAH2	
100EG_fwd	aataccatgaagatactgcagaagggaggaaatgaacatgATAAATGCTAAACTTTTACAACTTATG
100EG_rev	tgtaaaacgacggccagtgccgagctcttaatggtgatgatgatggtgATGTTTTGCTTGTCCTTG
tet promoter construction	
miniPthl_F	CTAGTcatcatTCTAACTAACCTCCTAACAACTTAATTATACCCACTATTATTATTTTTATCAATATATTTTATTCTCCAGTGTTATATACTATAGGGGa
miniPthl_R	CCGGTCCCCTATAGTATATAACACTGGAGAATAAAATATATTGATAAAAATAATAATAGTGGGTATAATTAAGTTGTTAGGAGGTTAGTTAGAatgatga
2tetO1_F	CCGGTACTCTATCATTGATAGAGTTTGAAACTCTATCATTGATAGAGTATAATATCTTTGTTCATTTAGAGCGATAAACTTGAATTTG
2tetO1_R	AATTCAAATTCAAGTTTATCGCTCTAAATGAACAAAGATATTATACTCTATCAATGATAGAGTTTCAAACTCTATCAATGATAGAGTA
2tetO1:dCas9B-sgRNApta and gusA	
pMTL82254_rev	CATATGGATACAGCGGCC
p2TetO1_fwd	gcggccgctgtatccatatgTTAAGACCCACTTTCACATTTAAG
p2TetO1_rev	aattcaagtttatcgctctAATGAACAAAGATATTATACTCTATCAATG
dcas9B-SalI_fwd	tttgttcattagagcgataaacttgaatttGAAATAAGGAGGAATATTAAACATGGATAAGAAATACTCAATAG
dcas9B-SalI_rev	cgagatctccatggacgcgtAAGGCGATTAAGTTGGGTAACGC
gusA_fwd	tttgttcattagagcgataaacttgaatttgaaataaggaggaatattaaacATGTTACGTCCTGTAGAAAC
gusA_rev	cgagatctccatggacgcgtTCATTGTTTGCCTCCCTG
pMTL82254_fwd	ACGCGTCCATGGAGATCTC

Table 3.3: Primers used for construction of CRISPRi vectors

Primer Name	Sequence
p3HB100	
phaAB_fwd 1	ataccatgaagatactgcagtttaagggaggaaagcatATGACTGACGTTGTCATCGTATC
phaAB_rev1	ctcctaaatttTCAGCCCATATGCAGGCC
ptb/buk_fwd1	tatgggctgaaaatttaggaggtttaaaatgATTAAGAGTTTTAATGAAATTATCATGAAG
ptb/buk_rev1	ccagtgccgagctcaagcttTTATTTGTATTCCTTAGCTTTTTCTTC
p3HB130	
phaA_fwd	aataccatgaagatactgcagTTTAAGGGAGGAAAGCATATG
phaA_rev	ctaaatttaaaTTATTTGCGCTCGACTGC
hbd_fwd	gcgcaaataaTTTAAATTTAGGGAGGTCTGTTTAATG
hbd_rev	cattttaaacctcctaaatttTTATTTTGAATAATCGTAGAAACCTTTTC
tesB_fwd2	attcaaaataaaaatttaggaggtttaaaATGAGTCAGGCGCTAAAAAATTTAC
tesB_rev	tgtaaaacgacggccagtgccgagctcTTAATTGTGATTACGCATCACC
p3HB160	
thil_fwd	aataccatgaagatactgcagtttaagggaggaaagcatATGAGAGATGTAGTAATAGTAAGTG
thil_rev2	ctaaatttaaaTTAGTCTCTTTCAACTACGAG
hbd_fwd2	aagagactaaTTTAAATTTAGGGAGGTCTGTTTAATG
hbd_rev	cattttaaacctcctaaatttTTATTTTGAATAATCGTAGAAACCTTTTC
tesB_fwd2	attcaaaataaaaatttaggaggtttaaaATGAGTCAGGCGCTAAAAAATTTAC
tesB_rev	tgtaaaacgacggccagtgccgagctcTTAATTGTGATTACGCATCACC
p3HB131	
phaA-EO_fwd	aataccatgaagatactgcagaagggaggaaatgaacATGACAGATGTAGTTATAGTGAG
phaA-EO_rev	tttaatattcctcctttaaaTTATTTTCTTTCTACTGCTAATGC
hbd_fwd	aaagaaaataatttaaaaggaggaatattaaacATGAAAAAGGTATGTGTTATAGG
hbd_rev	aactaacctcctgaaatttTTATTTTGAATAATCGTAGAAACCTTTTC
pCL2ₚₜₐ3HB131	
PTA_Prom_fwd	cagctatgaccatgattacgaattcCATTTGTCAACTATAGATGGTG
PTA_Prom_rev	catctgtcatGTTCATTTCCTCCCTTTAAATTTAAC
131_Op_fwd	ggaaatgaacATGACAGATGTAGTTATAGTG
131_Op_rev	cctgcaggtcgactctagaggatccTTAATTGTGATTCCTCATAACTC

Table 3.4: Primers used for construction of 3HB vectors

Gene Target	N20 Sequence
pta	TTACCTTTTCATTCCCTACA
adhE1	TTTGTTATATATATATTCTG
aor2	TATTTGACGAAGTAGATCCA

Table 3.5: N20 Sequences in sgRNA for *C. ljungdahlii* Genes

3.3.4 Cell Cultivation

E. coli was cultivated in LB medium according to standard procedures, supplemented with appropriate antibiotics (Chloramphenicol, 25 μg mL^{-1}; Erythromycin, 250 μg mL^{-1}). For general cultivation, *C. ljungdahlii* was grown anaerobically on YTF medium [91] (Per liter: Yeast Extract 10g, Bacto Tryptone 16g, NaCl 4g, Fructose 5g, Resazurin 0.5 mL of 0.2% stock solution, Cysteine-HCl 0.3g), supplemented with appropriate antibiotics (Thiamphenicol, 5 μg mL^{-1}; Clarithromycin, 5 μg mL^{-1}). For physiological experiments, qRT-PCR analysis, and enzyme assays, cells were grown heterotrophically on PETC 1754 medium (Per liter: NH$_4$Cl 1g, KCl 0.1g, MgSO$_4$·7H$_2$O 0.2g, NaCl 0.8g, KH$_2$PO$_4$ 0.1g, CaCl$_2$·2H$_2$O 20 mg, Yeast Extract 1g, ATCC Trace Minerals (MD-TMS) 10 mL, ATCC Vitamins (MD-VS) 10 mL, NaHCO$_3$ 2g, Fructose 5g, Resazurin 0.5 mL of 0.2% stock solution, Cysteine-HCl 0.3g). Cells were inoculated from mid-log starter cultures at 2-5% (v/v). *C. ljungdahlii* freezer stocks were prepared by combining an equal volume of culture and 50% glycerol in anaerobic vials and storing at -80°C.

3.3.5 *C. ljungdahlii* Transformation

Electrocompetent cells were prepared as described by Leang et al [91]. Cells were subcultured at least twice in YTF medium, before being inoculated at 2% (v/v) into 200 mL YTF supplemented with 40 mM DL-threonine. After overnight growth to an OD of 0.2-0.3, cells were cooled on ice for 30 minutes, then centrifuged at 7,000 RCF for 10 minutes. Cells were washed twice with ice cold 50 mL SMP buffer (272 mM sucrose, 1 mM MgCl$_2$, 7 mM sodium phosphate, pH 6.0), and resuspended in 180 μL SMP buffer, to which 20μL DMSO was added. 25 μL aliquots were transferred into pre-cooled Eppendorf tubes and stored at -80°C for up to one month. All steps apart

from cell growth were conducted in an anaerobic chamber with an atmosphere of 5% H_2, 20% CO_2, balance N_2 (Coy Labs).

For electroporation, aliquots of competent cells were thawed on ice, and 1-5 μg plasmid added before transferring the mixture to 1 mm electroporation cuvettes. The cells were electroporated with a single exponential pulse (0.625 kV, 600 Ω, 25μF), and 1 mL pre-warmed YTF immediately added. The cells were diluted 10-fold into pre-warmed YTF medium in Hungate tubes, and recovered for 9-12 hours at 37°C, before being diluted 10-fold into pre-warmed YTF medium containing the relevant antibiotics. Successful transformation was indicated by robust growth after 3-5 days, compared to a negative control in which cells were electroporated in the absence of plasmid. To isolate individual colonies, 1 mL of the outgrowth culture at different dilutions was mixed with 24 mL of YTF with 1% agar and poured into petri dishes.

3.3.6   Product Quantification

Extracellular metabolites were analyzed using an Agilent 1200 HPLC system with a G1362A RID detector with isocratic flow of 14 mM sulfuric acid (0.7 mL min^{-1}, 50°C) through a BioRad Aminex HPX-87H ion exchange column (300 mm x 7.8 mm). 500 μL culture was withdrawn and filtered through 0.2 μm syringe filter into HPLC vials, and 10 μL injected into the HPLC. Product identity and concentration was confirmed by comparison to standard curves generated from pure compounds (Sigma).

3.3.7   dCas9 Expression Verification

Cell lysis and protein denaturation were achieved by boiling cell pellets in Laemmli buffer in a thermocycler for 10 minutes. 10-45 μL were loaded in a 4-20% pre-cast SDS gel (BioRad) and separated by electrophoresis at 150V for approximately one hour. dCas9 expression was examined via Western blot with polyclonal rabbit Anti-CRISPR-cas9 (AbCam, #ab204448) as primary, and HRP-conjugated goat Anti-Rabbit (SantaCruz Biotechnolog, #sc2030) as secondary. Detection was performed with Western Lightning ECL substrate (PerkinElmer) and imaged by a Sony A5

camera.

3.3.8   RNA Isolation and Quantification via qRT-PCR

To isolate RNA, 1 mL of culture (approximate OD_{600} 0.6-1.0) was mixed with 2 mL RNA-Protect (Qiagen), and collected by centrifugation at 3,250 RCF for 15 min at room temperature. Cell pellets were stored at -80°C until extraction. Pellets were re-suspended in 100 μL of lysozyme/TE buffer (20 mg/mL lysozyme, 10 mM Tris, 1 mM EDTA, pH 8), vortexed for 10 sec and incubated for 10 min at 37°C, with frequent mixing. RNA was then harvested using an RNeasy Mini Kit (Qiagen). DNA was digested with the DNase I (NEB) for 10 min at 37°C, according to the manufactuer's protocol. The RNA was re-purified with the RNeasy Mini Kit, eluted in RNase free water, quantified by NanoDrop, and stored at -80°C for no more than one month prior to cDNA synthesis.

qRT-PCR was performed with commercially available kits, QuantiTect Reverse Transcription (Qiagen) and RT2 SYBR Green qPCR Master Mix (Qiagen), following the provided protocol. Primers used for RT-PCR analysis in CRISPRi strains are shown in Table 3.6, and those used for 3HB strains are shown in Table 3.7. Reactions were carried out in 96-well PCR plates (VWR) and sealed with BioRad Microseal B seals. For each biological replicate, there were 3 technical replicates and 1 'minus RT' control. qPCR cycles were performed on a BioRad iCycler using iQ3 software. The raw data was analyzed in BioRad iQ5 to determine the Ct number; the Ct number was quantified as copies/pg total RNA with a logarithmic standard curve generated from linear DNA fragments of the gene of interested. These linear fragments contained the whole gene of interest, were generated by PCR, purified with gel electrophoresis and quantified with a NanoDrop. The standard curve allowed for the calculation of the qPCR primer efficiency, which was generally greater than 90%. Melt curves were generated after each qPCR reaction cycle to ensure the specificity of the primers to the target gene.

For non-quantitative RT-PCR, cDNA synthesis was performed as above. Presence of transcript was detected by amplification with Q5 polymerase using 1 μL of cDNA

as template, with annealing temperatures specific to the primer pairs used and 40 total cycles. Products were analyzed by electrophoresis with 2% agarose gels.

3.3.9 Histochemical Glucuronidase Assay

Qualitative assay of glucuronidase activity was conducted with whole cells of *C. ljungdahlii*. A volume of culture equal to 1 OD·mL was spun down at 20,000 RCF and resuspended aerobically in 400 μL 50 mM NaH_2PO_4. Cells were permeabilized by addition of 25 μL chloroform, and 5 μL X-gluc stock solution (50 mg mL^{-1} in DMF) was added at the same time. Cells were incubated with periodic shaking at 37°C, and color development documented by photography after approximately 1 hour.

3.3.10 Crude Lysate Preparation and PTA Assay

To prepare crude lysates, *C. ljungdahlii* strains were grown in 100 mL PETC medium with relevant antibiotics and inducing compounds until an approximate OD of 0.7-0.8. Cultures were spun down at 7,000 RCF for 10 minutes, resuspended in 7 mL ice-cold lysis buffer (20 mM Tris-HCl pH 7.5, 0.1 mM PMSF, 25 U mL^{-1} Benzonase) and lysed aerobically by three passages through a High Pressure Homogenizer (EmulsiFlex-C5). The soluble fraction was separated by centrifugation at 20,000 RCF for 20 minutes at 4°C, and kept on ice until use.

PTA activity was measured in the acetyl-phosphate forming direction by measuring the liberation of coenzyme A (CoA) by Ellman's reagent at 412 nm. The assay mixture consisted of 50 mM Tris-HCl (pH 7.5), 50 mM KCl, 5 mM KH_2PO_4, 1 mM Ellman's Reagent (from a 10 mM stock in DMSO), 1 mM acetyl-CoA, and was initated by the addition of crude lysate. Activity was normalized to the protein concentration of the lysate, as determined by BCA assay.

Primer	Sequence	Purpose
dCas9		
dCas9-q-F1	AAACCCTATTAACGCAAGTGGA	qRT-PCR
dCas9-q-R1	ATGACAAAGCAATGAGATTCCC	
pAH2-F	TAAAAGATTAATTTACATATTAACATTTAATTATGG G	DNA Standard
pAH2-R	AAGGCGATTAAGTTGGGT	
gusA		
gusA-q-F1	AAGTGTGGGTCAATAATCAGGA	qRT-PCR
gusA-q-R1	CCAGTTCAGTTCGTTGTTCAC	
gusA_fwd	ATGTTACGTCCTGTAGAAAC	DNA Standard
gusA_rev	TCATTGTTTGCCTCCCTG	
pta		
PTA-q-F1	AGGTGCGGTTCATACTACAG	qRT-PCR
PTA-q-R1	GGATTTACAGCACAATCAGCA	
PTA-FULL-F1	CTACTCCATTTAGAAGTGCCA	DNA Standard
PTA-FULL-R1	AAATTGAACCGTCCATTCCT	
adhE1		
AdhE1-q-F1	TGGTTAGGTCAGGTGGATTTG	qRT-PCR
AdhE1-q-R1	CCTTGGGATGAAGGTGTGTTA	
AdhE1-FULL-F1	CCCAATTTAGCATACTAGGCA	DNA Standard
AdhE1-FULL-R1	TGTCTTCTACAATTCCCATTCC	
aor2		
AOR2-q-F1	GGATTAGATACCATTACAGCAGGA	qRT-PCR
AOR2-q-R1	GTCTCCAAATCCTTCTCTAAGTC	
AOR2-FULL-F1	AAGATGGAACAGTTTGCTTGG	DNA Standard
AOR2-FULL-R1	GCAGGTGTTACAGGATCGTC	
sgRNA-pta		
PTA-T1-q-F1	TTACCTTTTCATTCCCTACAGTTTTAG	RT-PCR
gRNA-q-R1	AAGCACCGACTCGGTGCCAC	
recA		
recA-q-F1	ATGTGGATGCCATTTCAACA	RT-PCR
recA-q-R1	TGCCACCGTAGTCTTACCTG	

Table 3.6: Primers Used in RT-PCR Analysis of CRISPRi Strains

Primer	Sequence	Purpose
phaA		
phaA-q-F1	CACGATCAAGAGCTATGCCA	qRT-PCR
phaA-q-R1	CCGTTCACATTGACCTTGGA	
hbd		
hbd-q-F1	GCAATTCAAATGAGCCTTCC	qRT-PCR
hbd-q-R1	AAATAGCAATTCTGCCCGT	
tesB		
tesB-q-F1	GAGTGAAGATTTAGGTTTACGCCA	qRT-PCR
tesB-q-R1	TGTGAAACGAATGTACCAGCC	
pta		
pta-q-F1	GCTCCAGGTACATCAGTAGTTTC	qRT-PCR
pta-q-R1	TCACTATCTGGGCATGGATTTAC	

Table 3.7: Primers Used in RT-PCR Analysis of 3HB Strains

3.4 Results and Discussion

3.4.1 Western analysis and RT-PCR confirm expression of dCas9 and sgRNA

CRISPRi requires two components: deactivated cas9 (dCas9) that acts as a transcriptional blockade, and a single-guide RNA (sgRNA) to direct the nuclease to the targeted site on the genome [154]. We therefore sought promoters to drive expression of these two pieces independently. Since some of the genes being targeted for downregulation may be essential, we wanted expression of dCas9 to be tightly controlled by an inducible promoter. When this work began, there was only one such promoter characterized for *C. ljungdahlii* - the lactose-inducible promoter from pAH2 that consists of the regulator-promoter pair $bgaR$-P_{bgaL} [159, 93]. We therefore cloned the dCas9 from *S. pyogenes* into pAH2 with three different RBS sequences to generate pdCas9A-C, and assayed for inducible expression by Western Blot (Figure 3-1). Full-length dCas9
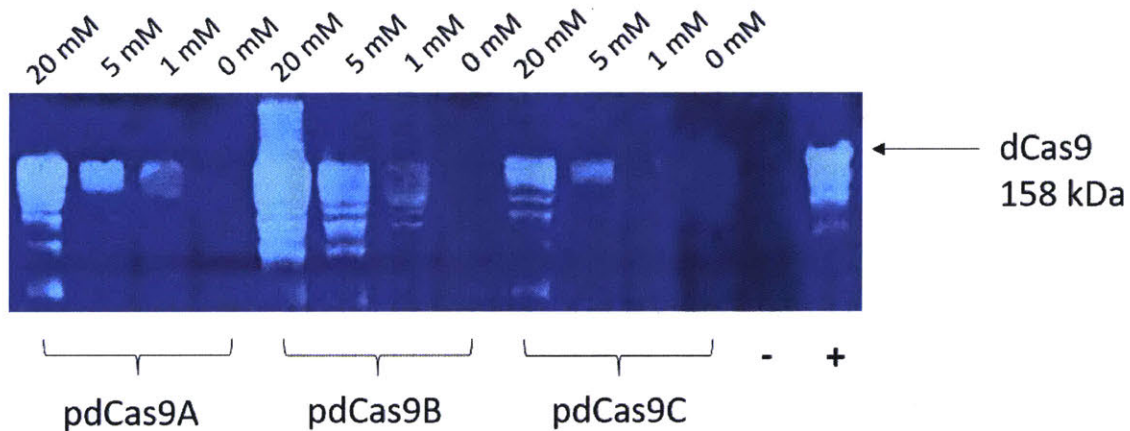
Figure 3-1: **Western Analysis of dCas9 Expression from pAH2 in *C. ljungdahlii*.** Expression of dCas9 was induced with varying concentrations (0-20 mM) lactose in early exponential phase, and expression was analyzed after 8 hours. pdCas9A-C represent three different RBS sequences: A, original sequence in pAH2. B, *C. ljungdahlii ack*. C, *C. ljungdahlii buk*. *E. coli* carrying pdCas9-bacteria induced (+) or uninduced (-) with aTc was used as a control. Rabbit anti-Cas9 polyclonal antibody was the primary, and HRP-conjugated goat anti-rabbit was the secondary antibody.

was detected in all constructs, and the intensity increased with increasing concentra-

tion of lactose. Some background expression was detectable in the absence of lactose, compared to the uninduced *E. coli* control. Cells carrying pdCas9B showed the highest level of expression, therefore this strain was chosen for implementation of the full CRISPRi system. That expression of dCas9, a large 158kDa protein, was detectable even with suboptimal codon usage bias was encouraging.

The ideal promoter for sgRNA expression is a small, constitutive one with well-defined transcription start site, so that many sgRNAs can be multiplexed without significantly increasing plasmid size, and so that the 5' end of the RNA sequence does not have extra bases that may impede target recognition. In *C. ljungdahlii* there are no well characterized constitutive promoters. Therefore we chose to use an engineered promoter that had been used for successful expression of sgRNA in the related organism *Clostridium cellulolyticum* [161]. To facilitate multiplexing, the sgRNA expression construct was designed using a BioBricks strategy with the enzymes XhoI, SalI, and XbaI (Figure 3-2a). Here, multiple sgRNA fragments can be built up by ligating an XhoI-XbaI fragment into an existing construct opened with SalI-XbaI, generating a fused XhoI-Sal site and allowing multiple iterations. An
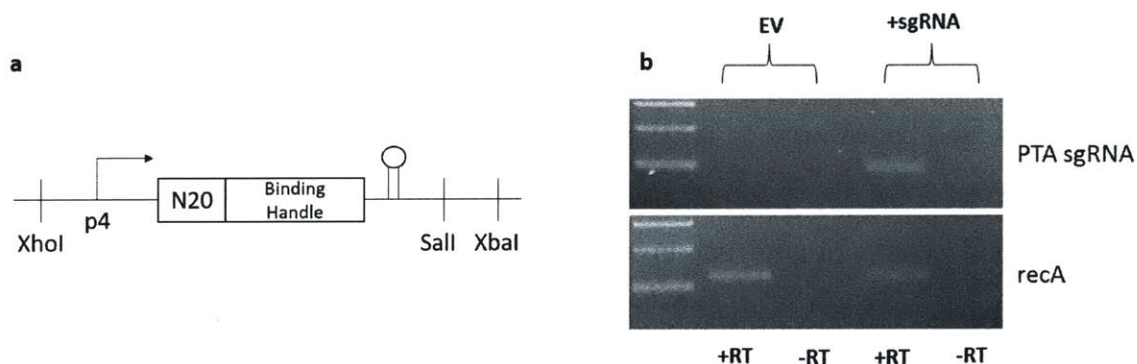
Figure 3-2: **Western Analysis of dCas9 Expression from pAH2 in *C. ljungdahlii*.** a) A BioBricks strategy using XhoI, SalI, and XbaI was designed to allow the iterative addition of multiple sgRNA expression units to multiplex gene control. b) RT-PCR verification of expression of sgRNA from P4 promoter. Top panel, primers for PTA sgRNA. Bottom panel, control primers for recA. +RT = with reverse transcriptase (RT). -RT = gDNA control with no RT step. EV = empty vector control (pdCas9B). +sgRNA = pdCas9B-sgRNA-pta.

sgRNA construct with an N20 region for *pta* (Table 3.5) was synthesized as a GBlock

and cloned into pdCas9B, to generate pdCas9B-sgRNA-pta. The N20 sequence was chosen by searcing for the PAM site (NGG) on the non-template strand closest to the beginning of the coding sequence, that showed no similarity to other regions of the genome. Expression of the sgRNA was assessed qualitatively by RT-PCR on cells in late exponential phase (Figure 3-2b). The expected band was clearly present only in cells carrying pdCas9B-sgRNA-pta, and only after reverse transcriptase treatment, confirming expression of the sgRNA.

3.4.2 The CRISPRi System Downregulates *pta*, *adhE1* and *aor2* to Different Extents

Having confirmed that both components of the CRISPRi system were successfully expressed, we next asked if the full system could mediate knockdown of *C. ljungdahlii* genes. Three target genes were chosen, that have putative roles in carbon flux and energy conservation: *pta*, *adhE1*, and *aor2* (Figure 3-3). *pta* encodes phosphotransacetylase, which catalyzes the conversion of acetyl-CoA to acetyl-phosphate, in the first of two steps from the central intermediate acetyl-CoA to acetate. Acetate production represents a major carbon flux under both heterotrophic and autotrophic conditions, and contributes to ATP generation through substrate-level phosphorylation. Therefore, the disruption of *pta* expression could be expected to have significant physiological effects. In addition, in strains engineered to produce value-added products, flux must be diverted away from acetate synthesis, and *pta* is therefore a logical target for regulatory interference. *ack*, the gene that encodes acetate kinase activity that converts acetyl-phosphate to acetate, forms an operon with *pta*. While we did not target *ack* directly, reduced transcription of the *pta* gene may also lead to reduced *ack* expression.

adhE1 encodes a bifunctional alcohol/aldehyde dehydrogenase that has been shown to contribute to ethanol production, the other major fermentative end-product under heterotrophic conditions [91]. Deletion of this gene reduces ethanol titer to about 15% that of the WT. This gene is one of only two characterized knockouts of a metabolic

gene in *C. ljungdahlii* (the other being the RNF complex [33]), thus we were interested in whether downregulation of this gene could confer the same phenotype. Again, ethanol production represents an undesired side flux in engineered strains, so targeting this gene may be of value in metabolic engineering efforts.

aor2 encodes an aldehyde ferredoxin oxidoreductase. One of several *aor* genes in *C. ljungdahlii*, *aor2* is the most highly expressed under heterotrophic conditions [163]. This enzyme catalyzes the ferredoxin-dependent interconversion of acetaldehyde and acetate. Under autotrophic conditions and high acetate concentration, it is thought to reduce acetate to acetaldehyde as the first step toward ethanol biosynthesis to obviate the toxicity associated with acetate. Importantly, this is the only proposed pathway for autotrophic ethanol production that supports ATP conservation [164]. Under heterotrophic conditions, the thermodynamically favorable oxidation of acetaldehyde to acetate could represent another mode of *pta*-independent acetate production that is still coupled to energy conservation: The reduced ferredoxin generated under these conditions can be reoxidized at the RNF complex with NAD as electron acceptor, leading to generation of proton motive force and approximately 0.5 ATP [41]. Together, these three genes encode enzymes that are thought to carry high flux under a variety of conditions, and thus the ability to regulate their expression could lead to both interesting insights into their functionality, and novel ways to mediate carbon flux to heterologous produts in engineered strains.

We therefore synthesized sgRNA fragments with N20 sequences for each of these genes (Table 3.5), and cloned them into pdCas9B. As controls, we also generated plasmids carrying the sgRNA sequences, but not the dCas9 (pEG110-sgRNAxxx). All plasmids were transformed into *C. ljungdahlii*. Downregulation of target genes was assessed first by qRT-PCR. At early exponential phase (OD_{600} 0.3-0.4), cells carrying CRISPRi strains were induced with 1 mM lactose, and RNA was harvested 8 hours later. The low concentration of lactose was chosen since only a stoichiometric amount of dCas9 is required for downregulation, and overexpression of dCas9 can cause toxic effects unassociated with the regulation of the target gene. Gratifyingly, for all three targeted genes, the mRNA level was significantly reduced only in the
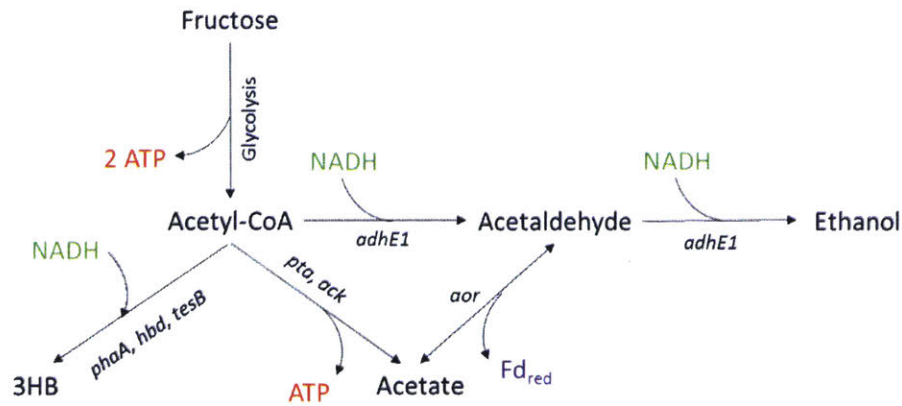
Figure 3-3: Schematic of Metabolic Interconversions Mediated by *pta*, *ahdE1*, and *aor*

presence of both dCas9 and the specific sgRNA (Figure 3-4). The *pta* transcript
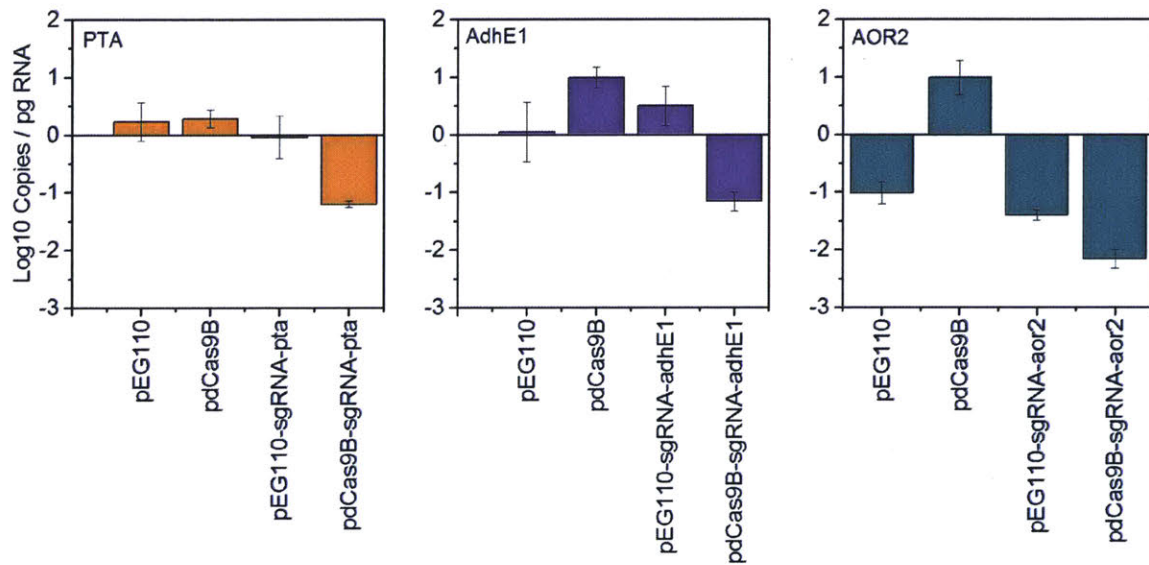
Figure 3-4: **qRT Analysis of CRISPRi Downregulation of *pta*, *adhE1* and *aor2*.** Error bars represent standard deviation of three biological replicates

was most reduced, by a factor of 27, or 96%, relative to the pEG110 control. For *adhE1* and *aor2*, the reduction was 88% and 93%, respectively. We cannot currently explain the aberrantly high level of *aor2* and *adhE1* expression in the pdCas9B strain compared to the other controls, but it could represent a response to the expression of *dcas9*, which is known to be toxic at high levels. That this response is not seen in strains carrying pdCas9B-sgRNA-aor2 is likely due to the repression caused by

100

the CRISPRi system. For all other targets, there was no discernable difference in expression level between the three controls (p=0.05). These results demonstrate that the CRISPRi system functions as expected, and can be easily targeted to a number of different genes in *C. ljungdahlii* with high efficiency.

3.4.3 Metabolic Effects of *pta* and *adhE1* Downregulation

Having demonstrated the power of the CRISPRi system to downregulate genes in *C. ljungdahlii*, we were next interested in the effect of this repression on cell physiology. Specifically, we asked if the growth rate and extracellular product distribution was significantly changed. Given the low expression of *aor2* in control strains and the lack of a clear contribution of the cognate enzyme under heterotrophic conditions, we focused our analysis of *pta* and *adhE1*. Cells carrying either pEG110, pdCas9B-sgRNA-pta, or pdCas9B-sgRNA-adhE1 were grown on PETC medium to OD_{600} 0.3-0.5, and either induced with 1 mM lactose or left untreated. Growth and product formation (ethanol and acetate) was followed over the remainder of the fermentation by spectrophotometer and HPLC (Figure 3-5). Induction only elicited a reduction in growth in the strains carrying pdCas9B-sgRNA-pta, and pdCas9B-sgRNA-adhE1, although in the latter case it was a small reduction and the final ODs were almost identical. Acetate titers were almost identical across strains at around 3-3.5 g L^{-1}, except for the induced strain carrying pdCas9B-sgRNA-pta, which produced roughly one third as much acetate. Given that PTA catalyzes the first step of acetate formation from acetyl-CoA, reduced acetate titer is consistent with the inactivation of this enzyme. Since the production of acetate also leads to ATP production, it is not surprising that the inability to produce acetate also led to a reduction in growth rate and final OD in this strain. We had hypothesized that the reduction in PTA activity would lead to increased flux of acetyl-CoA through the ethanol pathway, thus we were somewhat surprised that this strain produced less ethanol than the control strain (0.13 g L^{-1} vs. 0.25 g L^{-1}). These data are consistent with fermentation results in a recently published Δpta mutant of *C. ljungdahlii* [149], and could simply reflect the overall unhealthiness of the cell due to the reduced ability to produce ATP. Consistent with

101
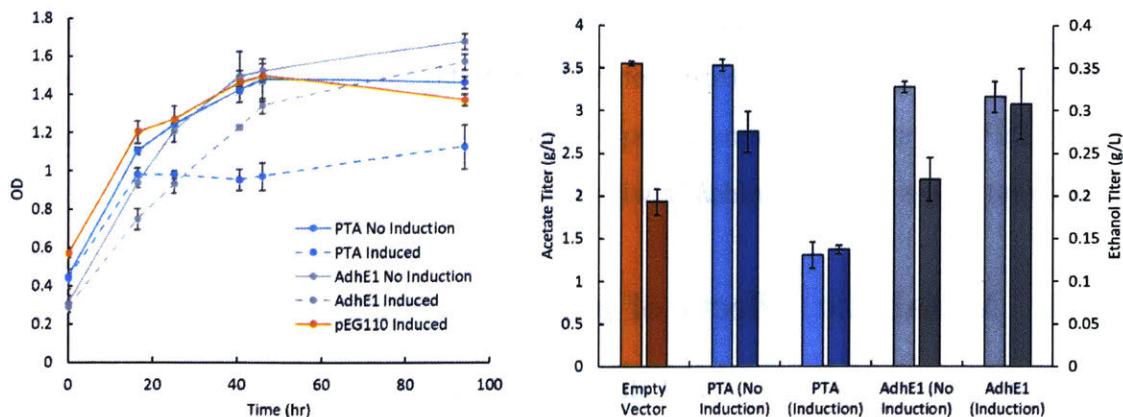
Figure 3-5: **Growth and Production in Regulated Strains.** Left: Growth was followed by OD after induction at early exponential phase. Dashed lines represent strains induced with 1 mM lactose. Solid lines are uninduced strains. Right: Final titers of acetate (left axis, light color) and ethanol (right axis, dark color). Error bars represent standard deviation of three biological replicates.

this interpretation, although the ethanol titer was lower, the yield on fructose was higher for the induced strain compared to the uninduced strain (0.037 g g^{-1} vs 0.019 g g^{-1}).

Interestingly, downregulation of *adhE1* had almost no effect on end-product concentration. Acetate titers were almost identical, and the ethanol titer was actually slightly higher in the repressed strain. One possible explanation for this is that, despite an 88% reduction in mRNA level, there is still sufficient AdhE1 enzyme present to allow the continued production of ethanol, especially if the enzyme is stable *in vivo* such that inactivation of transcription does not lead to a significant loss of activity. Owing to the technical challenge of assaying the oxygen-sensitive AdhE1 enzyme, activity data to support this claim are currently unavailable. Another possibility is that AdhE1 is not the major source of ethanol production under these growth conditions. Although Leang and co-workers reported a significant defect in ethanol production in a $\Delta adhE1$ strain [91], knockout of the same gene in the almost identical strain *C. autoethanogenum* led to only a marginal drop in ethanol titer [151]. This hypothesis is further discussed below in the context of 3HB production. Taken together, these data highlight the ability of the CRISPRi system to effect change at the metabolic level, which can be leveraged to gain fundamental insight into the metabolism of *C.*

ljungdahlii.

3.4.4 Assessing Leaky dCas9 Expression from Lactose Promoter

Given the background dCas9 detected in Western blots (Figure 3-1), we next asked if this was sufficient to lead to downregulation even in the absence of induction, which would significantly reduce the utility of the system for implementing downregulation of critical genes. To examine this possibility, we used qRT-PCR to reanalyze *aor2* downregulation in non-inducing conditions (Figure 3-6a). The repression of *aor2* was
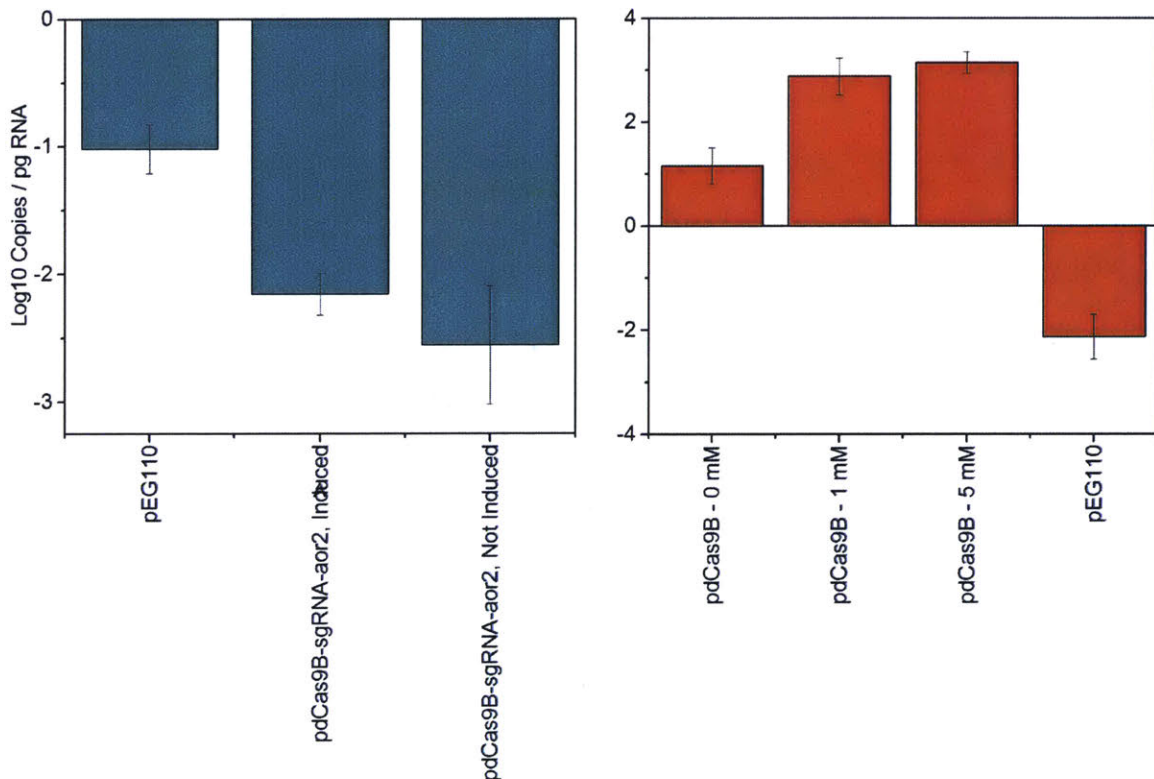
Figure 3-6: **Analysis of Leaky CRISPRi Induction.** Left: qRT-PCR analysis of *aor2* expression with and without induction. Right: qRT-PCR analysis of dCas9 induction at different lactose concentrations. Error bars represent standard deviation of three biological replicates.

equally strong regardless of whether or not lactose was added to the cells, suggesting that low level of dCas9 expression under non-inducing conditions was sufficient. This is reasonable, since theoretically only a single molecule of dCas9 is necessary to bind to

the genomic target and inhibit transcription. qRT-PCR analysis of dCas9 expression under these conditions confirmed that, although there is 52-fold more dCas9 transcript present with 1 mM lactose than 0 mM, there is still a significant amount of dCas9 expression under non-inducing conditions (Figure 3-6b). There was also no difference in transcript level between 1 mM and 5 mM lactose.

This analysis further complicates the interpretation of the results in Figure 3-5. Assuming leaky dCas9 expression was also present in the *pta* and *adhE1* strains, then the similarity between the uninduced controls and the empty vector is surprising, as is the difference between inducing and non-inducing conditions for *pta*. As mentioned before, one explanation for the *adhE1* data could be the physiological insignificance of this gene under these cultivation conditions. An explanation for the *pta* strain is less clear. One possibility is that, since the downregulation is constitutive, the transformed strain could have had multiple generations to adapt to the reduction in *pta* level. After transformation, the culture seed train consists of outgrowth, selection in liquid media, plating on solid media, and three subcultures in PETC medium before inoculating the experiment, during which mutations could have arisen or the native regulatory network otherwise adapted. Full analysis of this hypothesis would require RNAseq to address changes across the genome, but in partial support of this, we measured the expression level of *adhE1* in *pta*-regulated strains (and vice versa), and found a significant degree of *adhE1* downregulation when *pta* was targeted (Figure 3-7). This co-regulation, coupled with incomplete abolishment of PTA activity, could explain the similarities between the empty vector control and the uninduced *pta* strain. An explanation for the difference between inducing and non-inducing conditions in this strain remains elusive. Attributing this to the burden of *dCas9* expression is likely incorrect, since the same effect would be expected in the *AdhE1* strain.

3.4.5   Establishing a Stringent Promoter for *dCas9* Expression

These results clearly demonstrate that leaky *dCas9* expression confounds the interpretation of physiological results obtained with the CRISPRi system. We therefore sought an alternative promoter to drive *dCas9* that would have lower background.
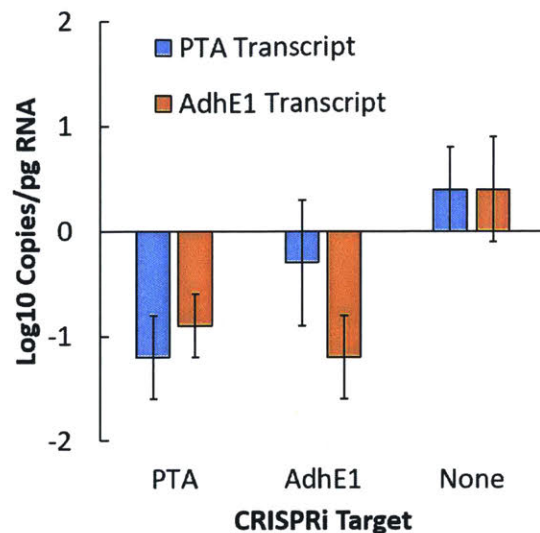
Figure 3-7: **Co-Regulation of *pta* and *adhE1*.** Transcript level of *pta* (blue) or *adhE1* (orange) when different genes were targeted, compared to the empty vector control. Error bars represent standard deviation of three biological replicates.

The tetracycline promoter (p_{tet}) is known to be tightly controlled, and background expression can be further tuned by modulation of the number of repressor binding sites and their position relative to the core promoter region [165]. Recently, a p_{tet} promoter was described for the related microbe *Clostridium acetobutylicum*, whose low background expression enabled the promoter to drive the expression of genes toxic to this organism [162]. To test whether this promoter could function in *C. ljungdahlii*, first we used it to drive expression of the glucuronidase reporter, encoded by *gusA* [159]. The so-called '2tetO1' promoter was cloned upstream of *gusA* in pMTL82254, generating pMTL822542tetO1gusA as described in the methods section. Cells containing this plasmid or pAH2, which encodes *gusA* under control of the lactose inducible promoter, were grown in YTF medium under inducing (100 ng mL^{-1} aTc or 20 mM lactose) or non-inducing conditions. GUS activity was qualitatively assessed after overnight growth, in permeabilized cells (Figure 3-8). Strong GUS activity is clearly visible for both strains upon induction, and is much stronger for pAH2 than for the new 2tetO1 promoter. Gratifyingly, under non-inducing conditions 2tetO1 showed almost no blue color, compared to the clearly visible blue in pAH2. To obtain a more quantitative comparison of these strains, *gusA* transcript
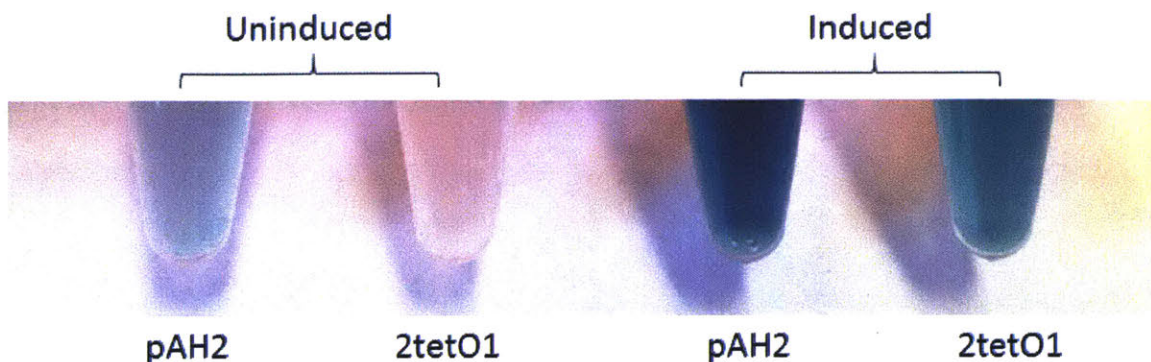
Figure 3-8: **Qualitative Comparison of GUS Activity with tet and Lactose Promoters.** Activity of GUS was measured by cleavage of X-gluc, which generates an insoluble blue color, in cells permeabilized with 25 μL chloroform.

was determined by qRT-PCR (Figure 3-9). Under non-inducing conditions there was
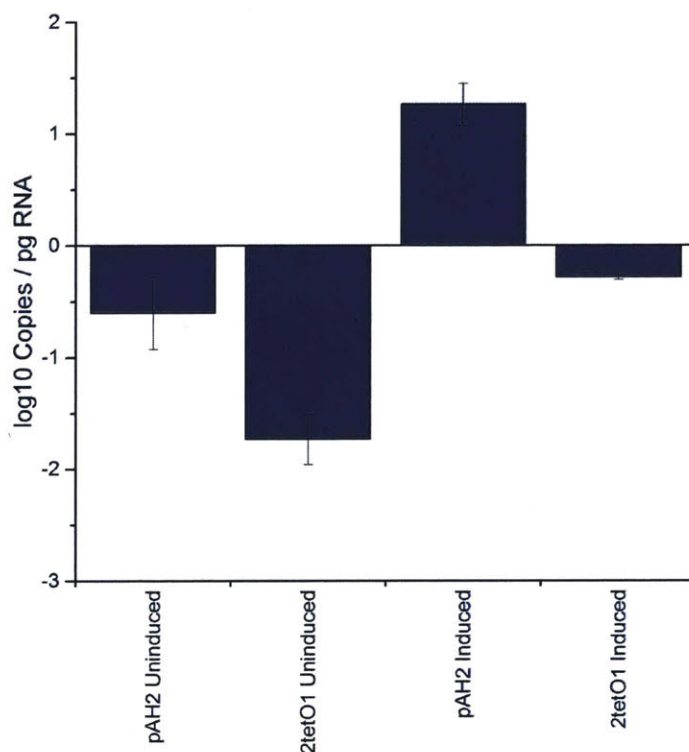
Figure 3-9: **Quantitative Comparison of GUS Activity with tet and Lactose Promoters.** qRT-PCR analysis of *gusA* transcript after overnight growth with or without induction. Error bars represent standard deviation of three biological replicates.

14-fold more *gusA* transcript from pAH2 than from pMTL822542tetO1gusA, showing

106

Plasmid	aTc (ng mL^{-1})	OD at RNA Harvest
pMTL822542tetO1gusA	0	0.61±0.03
pMTL822542tetO1gusA	100	0.85±0.03
pMTLdCas9-sgRNApta	0	0.59±0.01
pMTLdCas9-sgRNApta	100	0.41±0.02

Table 3.8: OD at RNA Harvest in p$_{tet}$ Strains

much tighter control of transcription in the new promoter. When the latter strain was induced with 100 ng mL^{-1} aTc the transcript level increased 28-fold, but was only marginally higher (2-fold) than in uninduced pAH2. However, since the baseline dCas9 expression from pAH2 is sufficient to mediate downregulation, comparable expression from 2tetO1 should enable the same extent of regulation.

To assess the impact of tighter control of *dCas9* expression, dCas9 and the sgRNA for *pta* were placed under control of the tet promoter, generating plasmid pMTLdCas9-sgRNApta. Cells carrying this plasmid or pMTL822542tetO1gusA were grown in PETC medium. To minimize the risk that residual PTA produced before induction of the CRISPRi system would mask a phenotype in the regulated strain, in this experiment, cells were induced at inoculation, with 100 ng mL^{-1} aTc. Uninduced controls were included for each strain. Growth and production were assessed throughout the fermentation, and RNA was harvested at mid-exponential phase (Table 3.8) to assess *pta* expression. The results are shown in Figure 3-10. qRT-PCR revealed that the induced strain carrying pMTLdCas9-sgRNA-pta showed a 37-fold reduction in *pta* transcript compared to the strains carrying pMTL822542tetO1gusA, which is similar to the level of repression seen with pdCas9B-sgRNA-pta, despite a much lower level of *dCas9* expression. This confirms the expectation that, above a critical threshold concentration, increasing amounts of dCas9 do not affect the extent of downregulation. Gratifyingly, the uninduced strain carrying pMTLdCas9-sgRNA-pta showed only 6-fold reduction in *pta* expression. Though this is still a significant downregulation, the new system allows for differential downregulation in inducing and noninducing conditions, whereas the previous lactose-based iteration did not (Figure 3-6). Induction in the *gus*-expressing strains caused a slight growth defect, likely due to the known toxicity of aTc in *Clostridia* [162, 166], but the final OD was similar to the
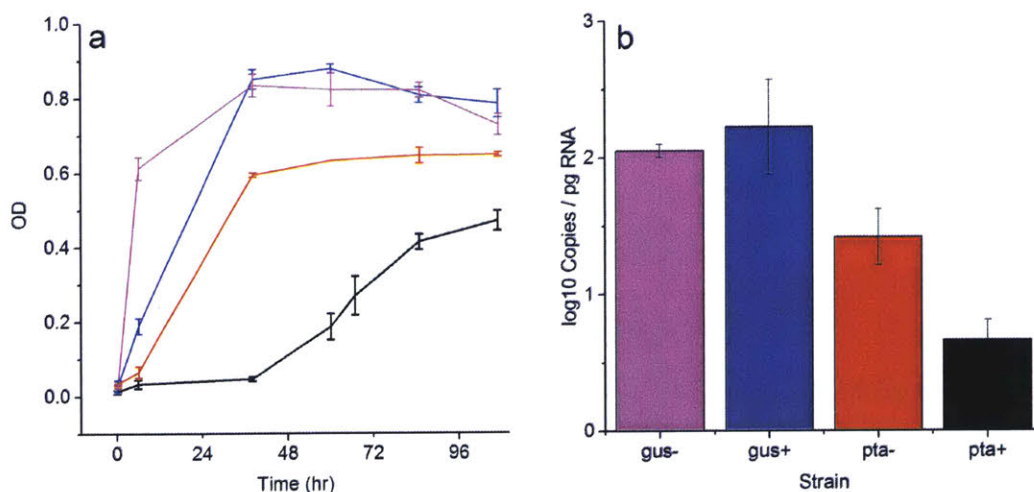
Figure 3-10: **Analysis of Strains Carrying tet-based CRISPRi System.** a) Growth. b) qRT-PCR analysis of *pta* transcript. -: no induction, +: 100 ng mL^{-1} aTc added at inoculation.

uninduced control. In the uninduced *pta* strain, growth was further retarded, and the cells reached a lower final OD than the *gus* strains (0.6 vs 0.85). This could either be due to the burden of *dCas9* expression, or the response to a partial knockdown of *pta*. In contrast, cells carrying pMTLdCas9-sgRNA-pta and induced with aTc showed a dramatic growth defect, with virtually no growth detected until approximately 60 hours. The final OD of these cells was lower still, reaching a maximum after 107 hours of 0.4. The growth defect observed in this experiment is similar to that observed in the lactose-induced system post-induction.

Taken together, these data show that the revised CRISPRi system is more controllable than the prototype. The stringent control of expression by the 2tetO1 promoter results in very low background expression of *dCas9*, allowing for differential target downregulation upon induction with aTc.

3.4.6 Establishing a Heterologous 3HB Pathway

Having established a CRISPRi system for regulating gene expression, we were next interested in whether this system could be used to divert flux to heterologous products in engineered strains of *C. ljungdahlii*. To answer this question, we needed a heterologous pathway that competed for acetyl-CoA flux to serve as a testbed to examine the

effect of downregulation. Here we chose 3-hydroxybutyric acid (3HB). 3HB is an industrially relevant compound used as a building block for the production of optically active vitamins, antibiotics, pheromones, and flavor compounds [167]. Biosynthesis of 3HB begins with the condensation of two molecules of acetyl-CoA to form acetoacetyl-CoA, and is shown in Figure 3-11. This intermediate is then reduced with either NADH or NADPH to generate the (S)- or (R)-entantiomer, respectively. Cleavage of the CoA ester affords 3HB either by direct hydrolysis, or through the promiscuous action of phosphotransbutyrylase (PTB) and butyrate kinase (BUK). The latter pathway functions only with the (R)-entantiomer, and produces one molecule of ATP per molecule of 3HB. The former pathway is not energy-conserving.
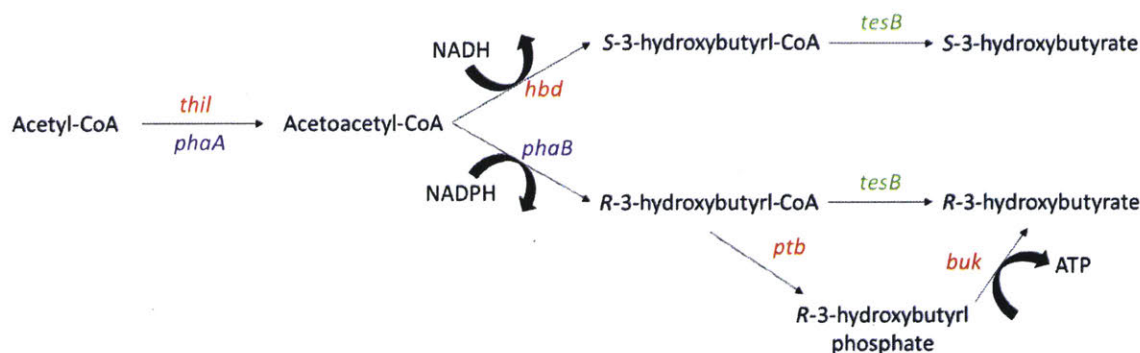
Figure 3-11: **Pathways for 3HB Production.** Red genes are derived from *C. acetobutylicum*. Green genes are from *E. coli*, and purple from *R. eutropha* H16.

Synthetic operons expressing the relevant enzymes were constructed and placed under control of the lactose-inducible promoter through Gibson assembly, to generates plasmids p3HB100, p3HB130, and p3HB160. p3HB100 expresses the NADPH-dependent 3-hydroxybutyryl-CoA dehydrogenase from *Ralstonia eutropha* and *ptb* and *buk* from *Clostridium acetobutylicum*, and should therefore produce the (R)-entantiomer. p3HB130 and 160 both express the NADH-dependent 3-hydroxybutyryl-CoA dehydrogenase from *C. acetobutylicum* and the thioesterase (*tesB*) from *E. coli*, and thus produce (S)-3HB. They differ in the thiolase enzyme used for the condensation of acetyl-CoA: p3HB130 uses the *phaA* gene from *R. eutropha*, whereas p3HB160 expresses the thiolase *thl* from *C. acetobutylicum*. All three plasmids were transformed into *C. ljungdahlii*, grown in PETC medium, and induced with 20 mM

lactose^2 at early exponential phase (OD_{600} 0.3-0.4), after which growth and production was followed until cessation of growth.

The results are shown in Figure 3-12. Cells carrying the empty vector control (pAH2) produced no detectable 3HB, whereas all the other strains produced small amounts, ranging from 0.7 mM (p3HB160) to 2.8 mM (p3HB130). This represents a maximum of 14% of the total carbon in fermentation product. The majority of the carbon flux went toward the native products acetate and ethanol. Acetate production was similar throughout all the tested strains. Ethanol production was also similar, except with cells carrying p3HB160, which produced almost twice as much as any other strain. The reason for this is unclear.

---

^{2}HPLC analysis confirmed that lactose used as inducer was not metabolized
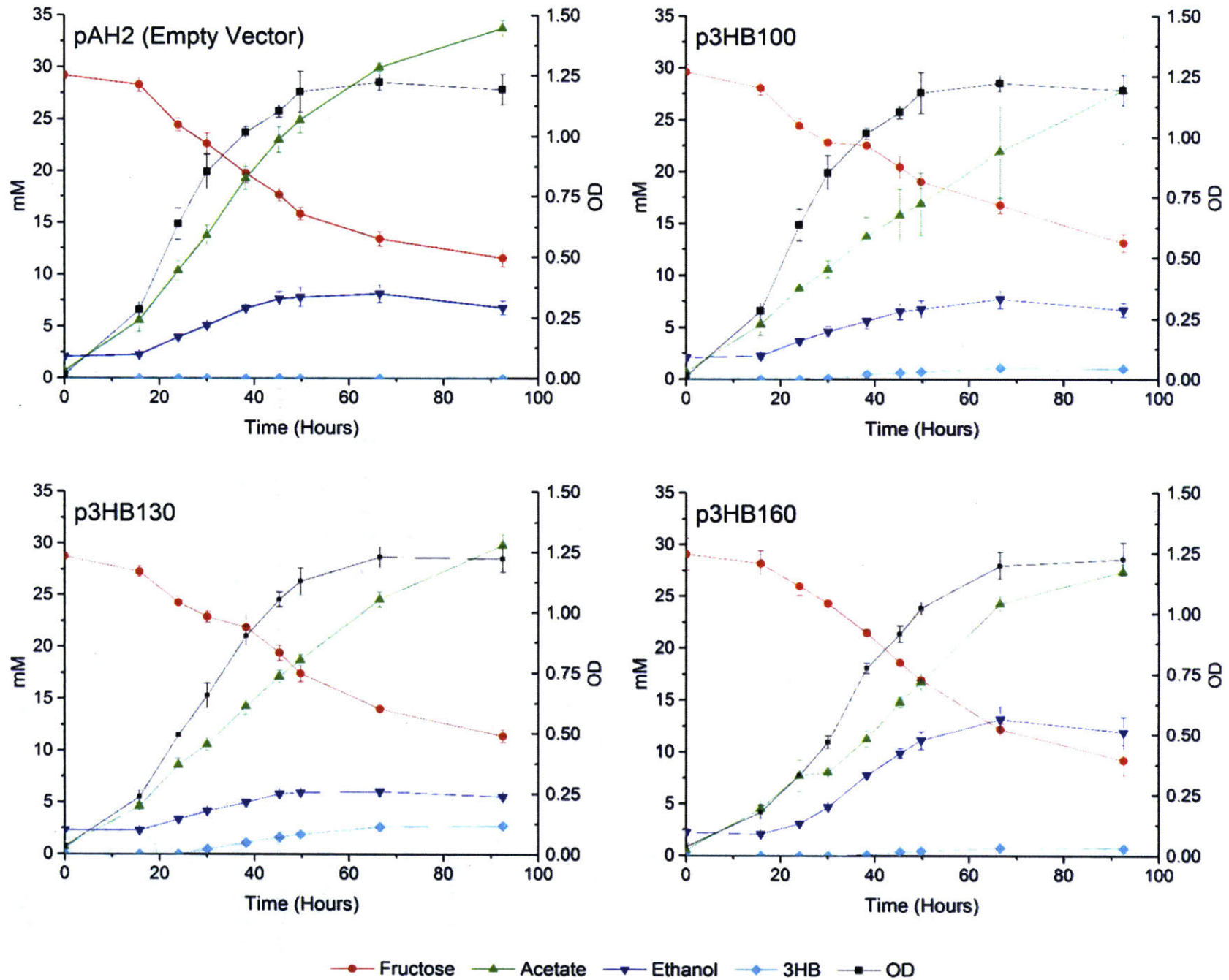
Figure 3-12: **Production of 3HB in pAH2 Derivatives.** Error bars represent standard deviations of three biological replicates.

To understand the basis of the low 3HB titer in these strains, several lines of investigation were followed: First, qRT-PCR analysis was used to confirm the expression of the heterologous genes (Figure 3-13). This analyis revealed that the plasmid-borne pathway genes were expressed at a level roughly 100-fold higher than the genomic *pta* gene, confirming strong activity of the lactose promoter. Next, we analyzed codon
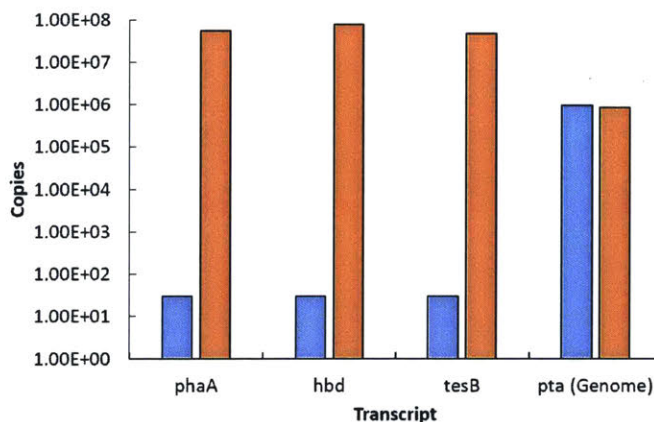
Figure 3-13: **qRT-PCR Analysis of 3HB-Producing Strains.** qRT-PCR was used to assess the relative transcript abundance of 3HB pathway genes (*phaA*, *hbd* and *tesB*) compared to the native gene *pta*. Blue bars: pAH2 (empty vector control). Orange bars: p3HB130 (best producing strain).

usage bias in the heterologous genes. In general, genes from *R. eutropha* and *E. coli* were found to preferentially use rare codons in *C. ljungdahlii*, whereas genes from *C. acetobutylicum* were well-adapted. A sample analysis of *phaA* from *R. eutropha* is shown in Figure 3-14, clearly indicating a large number of rare codons near the beginning of the coding sequence, which is thought to lead to reduced translation [168]. To overcome any barrier to expression linked to poor codon usage bias, we synthesized codon-optimized variants of *phaA* and *tesB*, and assembled these with *hbd* from *C. acetobutylicum*. We took advantage of the gene synthesis to simultaneously encode ribosome binding sites and 5' spacer regions that had previously been shown to be effective in *C. ljungdahlii* [95] (Table 3.9). The resulting plasmid (p3HB131) contains the same genes as p3HB130, but with optimized codon usage and RBS sequences. p3HB131 performed approximately 40% better than p3HB130, producing 3.8 mM 3HB, compared to 2.7 mM (Figure 3-15), showing the effectiveness of the optimized

Gene	RBS Source	Sequence
phaA	*C. ljungdahlii pta*	AAGGGAGGAAATGAAC*ATG*
hbd	*C. ljungdahlii ack*	AGGAGGAATATTAAAC*ATG*
tesB	*C. ljungdahlii thl*	AAATTTCAGGAGGTTAGTTAGA*ATG*

Table 3.9: **RBS Sequences in p3HB131.** Start codon denoted by italicized lettering.

operon.
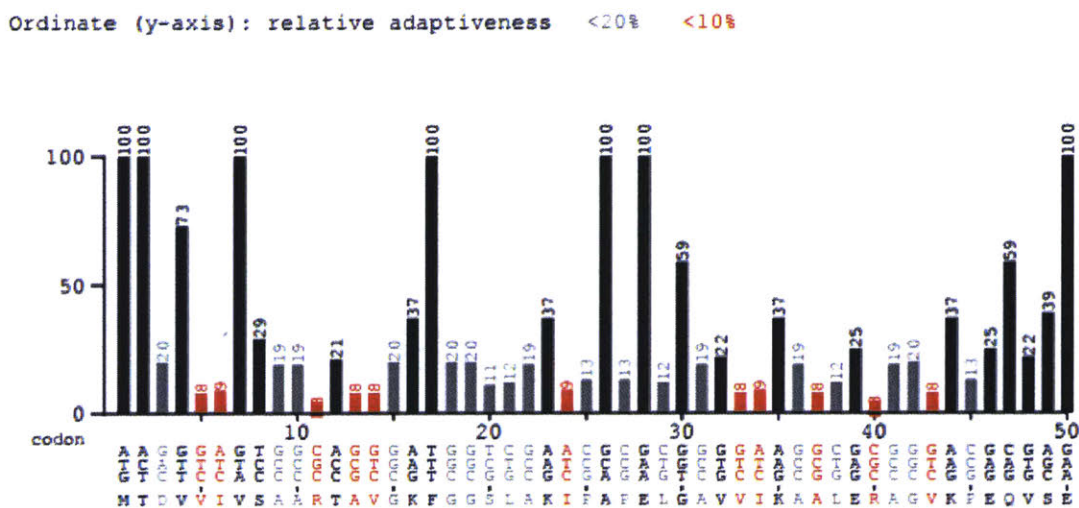
Figure 3-14: **Codon Usage Analysis of *phaA*.** Codon usage was analyzed by the Graphical Codon Usage Analyser [169] (http://gcua.schoedl.de/). The x-axis represents the first fifty codons in *phaA* beginning at the start codon, and the y-axis shows the relative adaptiveness, ranging from 0 (rare) to 100 (common). Red bars represent extremely rare codons (<10%) and gray bars represent rare codons (<20%).

In all our strains up to this point, ethanol was a major byproduct. Since NADH is necessary for 3HB biosynthesis in p3HB131, we reasoned that knocking out *adhE1* may both reduce ethanol production and also increase the availability of NADH for 3HB production (although the preference of AdhE1 for NADH/NADPH has not yet been determined). We therefore acquired the *adhE1* mutant [91], transformed it with p3HB131, and assessed its fermentation characteristics. While the final ethanol yield was reduced by a factor of two, 3HB production remained unchanged, and carbon was instead diverted to acetate production (Figure 3-15). Similar results were observed in heterologous butyrate production in *C. ljungdahlii* [95]. Interestingly, in our hands,

113
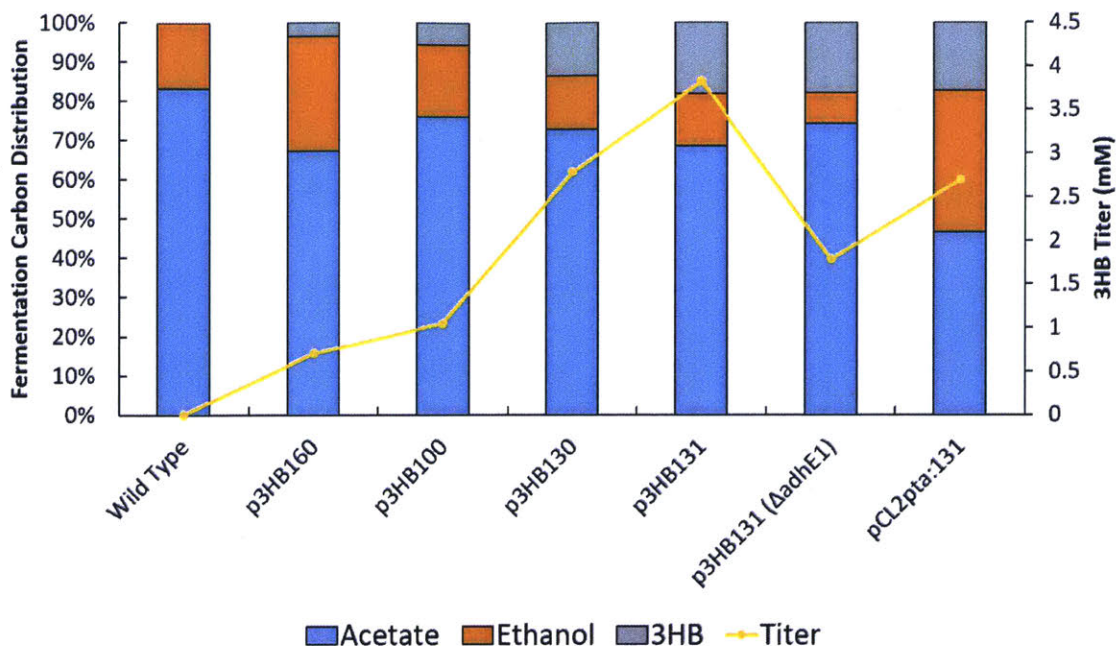
Figure 3-15: **Comparison of 3HB Production Metrics in Engineered Strains.**
Yellow line: 3HB titer (right axis). Stacked bars: Final distribution of fermentation
carbon (left axis). Blue = acetate, Orange = ethanol, Gray = 3HB. Data points
represent averages of three biological replicates.

knockout cells containing the empty vector pAH2 showed no difference in the ratio
of acetate:ethanol compared to the WT strain with the same plasmid. This result
contrasts with the findings in [91], but is in agreement with the lack of phenotype
associated with CRISPRi-mediated knockdown of *adhE1* described above, and casts
doubt on the role of AdhE1 in ethanol production under the conditions used here.

To reduce the complexity of the 3HB experiments and eliminate the need for in-
duction with lactose, which would not be economically viable in an industrial process,
we tested the ability of the *pta* promoter from *C. ljungdahlii* to drive 3HB biosyn-
thesis. This promoter was cloned upstream of the p3HB131 3HB operon into the
plasmid pCL2, in a manner similar to that described previously [95]. This plasmid
has the same origin of replication and resistance marker (*catP*) as pAH2, but has
the added advantage of being significantly smaller. WT cells transformed with the
plasmids pCL2pta131 produced 3HB without the need for induction (Figure 3-15).
The titer was lower than with the strain carrying p3HB131 (2.6 mM vs 3.8 mM),

but the yield was the same. Interestingly, this strain diverted a significantly greater fraction of the carbon flux to ethanol formation. A possible explanation for this could be the reduced burden of expression with the weaker *pta* promoter, that causes less stress to the cells and reduces the need for ATP from acetate production.

The end result of these engineering efforts is two plasmids, one inducible, and one constitutive, for the production of 3HB. These plasmids are compatible with the pMTL82254 vector used in the CRISPRi system, and will therefore enable the future assessment of effect of targeted knockdown of specific genes on 3HB production.

3.5 Conclusions and Outlook

The ability to generate conditional deletions is useful both for probing gene functionality, and for manipulating carbon flow to improve the production of heterologous compounds. The CRISPRi system described in this chapter can fulfill this need. Critical to the success of this approach is the tight regulation of *dCas9* expression by the tet promoter that ensures controllable downregulation and minimizes the risk of adaptation before physiological characterization. Even with the low background of the tet promoter, some downregulation was still observed in the absence of induction. This is a general problem with CRISPRi technologies, where only a very small amount of dCas9 is needed to cause repression. One potential solution would be to control sgRNA expression from the same inducible promoter, such that even with background levels of dCas9, the low expression level of the sgRNA would limit repression. In *C. ljungdahlii*, where the native homologous recombination machinery is still intact, the repeated use of the same promoter can lead to unexpected recombination events, therefore this approach should be used cautiously. Another intriguing possibility would be to make *dCas9* expression autoregulated by using a sgRNA to target *dCas9*, and therefore further reduce the background protein level.

For each of the genes examined in this work, RNA downregulation was greater than 88%. That high repression of *pta* did not improve the ethanol yield in WT cells was surprising. A possible explanation is that the rate of ethanol synthesis is

controlled by the availability of reducing equivalents (NADH/NADPH), such that changes in the expression level of PTA relative to AdhE1 have minimal impact on this flux. Metabolomic analysis of the regulated strain would provide insight into this. An alternative possibility is that *C. ljungdahlii* has a native regulatory mechanism to balance the production of acetate and ethanol according to cellular demands. Some support for this comes from the finding that in the *pta* downregulated strain, *adhE1* was also significantly downregulated. This result highlights the potential of the CRISPRi system to aid in the probing of regulatory networks.

The CRISPRi system is highly scalable, such that construction of sgRNA libraries in combination with next-generation sequencing can provide unparalleled insight into these questions. For this to be implementable in *C. ljungdahlii*, the chief barrier is the low transformation efficiency, which currently limits library size. Transformation in this work was accomplished via electroporation protocols reported to give efficiencies of $10^4 \mu g^{-1}$ [91] (compared to $10^{10} \mu g^{-1}$ for *E. coli*)[3]. Recently, other groups working with *C. autoethanogenum* have begun to use bacterial conjugation for DNA delivery [166]. Though efficiencies were not reported, conjugation has been the method of choice for the generation of transposon libraries [170], and may well be suitable for large-scale CRISPRi approaches in *C. ljungdahlii*.

---

[3]NEB DH10β product spec sheet

Chapter 4

Evaluation of anaerobic methanotrophy for the conversion of methane to liquid fuel

4.1 Summary

It has been estimated that the proven natural gas reserves in the U.S. could provide enough energy to meet the country's transportation energy demand for the next 50 years [11]. However, the infrastructure to use natural gas directly (either as CNG or LNG) has not been developed, and so natural gas must first be converted to a liquid fuel. Chemical conversion processes suffer from high capital costs and low yields, which has led to significant interest in developing alternative processes for the high-efficiency conversion of methane to a drop-in transportation fuel [101]. Various microorganisms have evolved to grow on methane using a variety of different enzymes and pathways, thus with the application of metabolic engineering, it is conceivable that a strain capable of producing a liquid fuel from methane could be developed. Before embarking on this undertaking, it is critical to establish the inherent potential and limitations of the various pathways to select the most attractive candidate for development. To this end, in this chapter we develop stoichiometric models to estimate

theoretical yields for the conversion of methane to n-butanol via the canonical aerobic methane pathways based on Methane Monooxygenase (MMO), and the less understood anaerobic pathways based on Methyl Coenzyme M Reductase (MCR). These yields, combined with thermodynamic and kinetic consideration of the pathway steps involved, are used to gauge the relative merits of each pathway. Ultimately, we conclude that while the energy efficiency of methane oxidation via MCR is higher than by MMO, other features of the anaerobic pathway preclude improvements in the overall carbon and energy yield compared to the aerobic pathway. This, combined with the superior genetic tractability, physiological understanding, and biochemical characterization of aerobic methanotrophs, suggests that these organisms and their pathways should be the focus of efforts to engineer the biological conversion of methane to liquid fuel.

4.2 Introduction

Over the last decade, widespread implementation of hydraulic fracturing (fracking) has increased the availability of unconventional natural gas in the United States. On an energy basis, this resource could meet the demands of the transportation industry for the next 50 years [5]. However, conversion of natural gas to a fuel of sufficient energy density remains problematic. Chemical conversion through the capitally intensive Fischer-Tropsch synthesis is economically viable only at large scale [6], and incompatible with the distributed nature of the resource [7]. Compression to LNG and CNG is an expensive endeavor, further mired by a lack of refueling infrastructure [8]. In the absence of an economically viable alternative, up to 32% of the natural gas produced in some areas of the country is currently flared as a waste product [9]. A low-capital process for converting methane to a liquid fuel, compatible with existing engine design, can therefore be a disruptive technology, with the potential to tap this underutilized resource and reduce the environmental impact of fracking [10]. A biotechnological process using an engineered microbe to produce fuel from methane could meet this need [11]. Biological processes are generally less capital-

intensive than chemical alternatives, and provide other benefits in their use of ambient temperatures and pressures, and their avoidance of harsh solvents and energy-intensive separation operations. However, to date, no commercial process exists for the bioconversion of methane to liquid fuels. In recognition of this market potential, and the burgeoning biochemical and genetic understanding of aerobic and anaerobic methanotrophic metabolism [171, 172], the DOE's Advanced Research Projects Agency-Energy (ARPA-E) division announced a $30MM program for research into this area, 'Reducing Emissions using Methanotrophic Organisms for Transportation Energy (REMOTE)' [173].

Various microorganisms have evolved enzyme systems to metabolize methane. With the application of metabolic engineering then, an organism capable of producing a liquid fuel from methane could be developed. Before embarking on such an undertaking, the inherent potential and limitations of the chosen pathway must be established to maximize the likelihood of developing a commercially viable process. Selecting an organism, pathway, and fuel product from the huge diversity in the published literature is no easy feat and depends upon many factors. Principal among these in the case of fuel production, where the margins are typically low and substrate wastage must be avoided, is the expected yield - defined as the total percentage of chemical energy input as methane captured in the final product. This was highlighted in a recent perspective [101], in which the authors proposed several hypothetical pathways for methane bioconversion. With sufficient knowledge of the biochemical steps connecting substrate to product, it is possible to calculate the maximum theoretical yield for a given pathway. These simple estimations are inevitably higher than could be achieved experimentally, but are a useful starting point for making decisions about what organism and pathways to investigate for a particular process, and have been used extensively for this purpose [174, 175]. The robustness of the estimations is improved by balancing the important cofactors (ATP, NADH, etc.), and relating their production back to the substrate through stoichiometric equations. Including the ATP maintenance requirements of the organism further improves the confidence in the estimation, and provides a link between the process yield and the rate of sub-

strate conversion. In this chapter, I use this approach to estimate theoretical yields for the conversion of methane to a fuel via the canonical aerobic methane pathways (based on MMO) and the less understood anaerobic pathways for methane oxidation (based on MCR), considering a number of permutations within those categories.

The specific motivation for this comparison was to understand whether using an MCR-based process could lead to a higher energy yield than an MMO-based process. Based purely on comparison of the two enzymes, this should be the case: As shown in Figure 4-1, the oxidation of methane to methanol via MMO requires stoichiometric reducing equivalents, generating water as a byproduct and effectively wasting 50% of the energy content of the methane. By contrast, oxidation of methane to methyl-coenzyme M by MCR requires no input of reducing power, and the electrons removed from methane are conserved in the thiol of coenzyme B. However, analyzing a single enzyme in the pathway is insufficient - understanding the efficiency of the process requires a global analysis of all the reactions involved in methane metabolism, motivating the development of such a model in this work. Butanol was chosen as the target molecule as it is a preferred biofuel, but in principle the methodology could be used to calculate the yield for any molecule for which the biosynthetic steps are characterized. This chapter begins with a review of the pathways and enzymes involved in these conversions, and then presents the results of the yield calculations. These results, which are purely derived from stoichiometry, are then discussed in the context of additional thermodynamic and kinetic considerations.

4.3   Results and Discussion

4.3.1   Aerobic Methane Oxidation

The pathways in aerobic methane metabolism are shown in Figure 4-2. The key enzyme in aerobic methane oxidation is Methane Monooxygenase (MMO), which catalyzes the oxygen-dependent oxidation of methane to methanol, with reduction of

a. Aerobic activation of methane via MMO

$$CH_4 + O_2 \rightarrow CH_2O + H_2O$$

b. Anaerobic activation of methane via MCR and HDR

$$CH_4 + Y_{ox} + X \rightarrow CH_3\text{-}X + Y_{red}$$
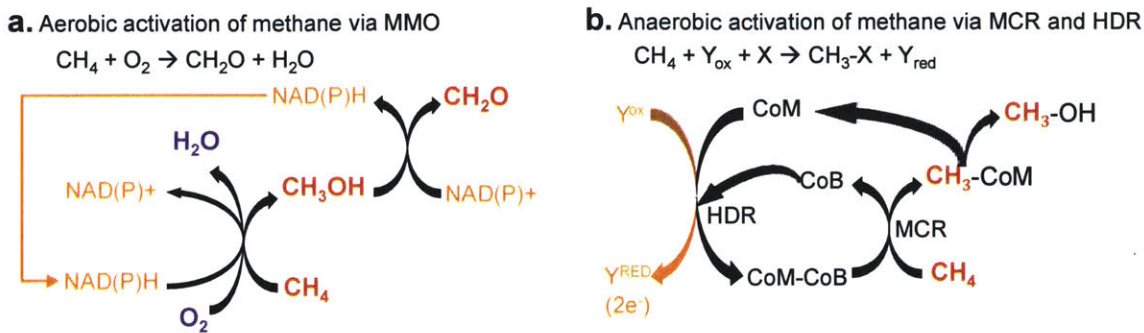
Figure 4-1: **Enzymes Catalyzing the Oxidation of Methane.** Left: Methane Monooxygenase (MMO). Right: Methyl Coenzyme M Reductase (MCR). Y represents an unknown electron carrier in anaerobic methanotrophy
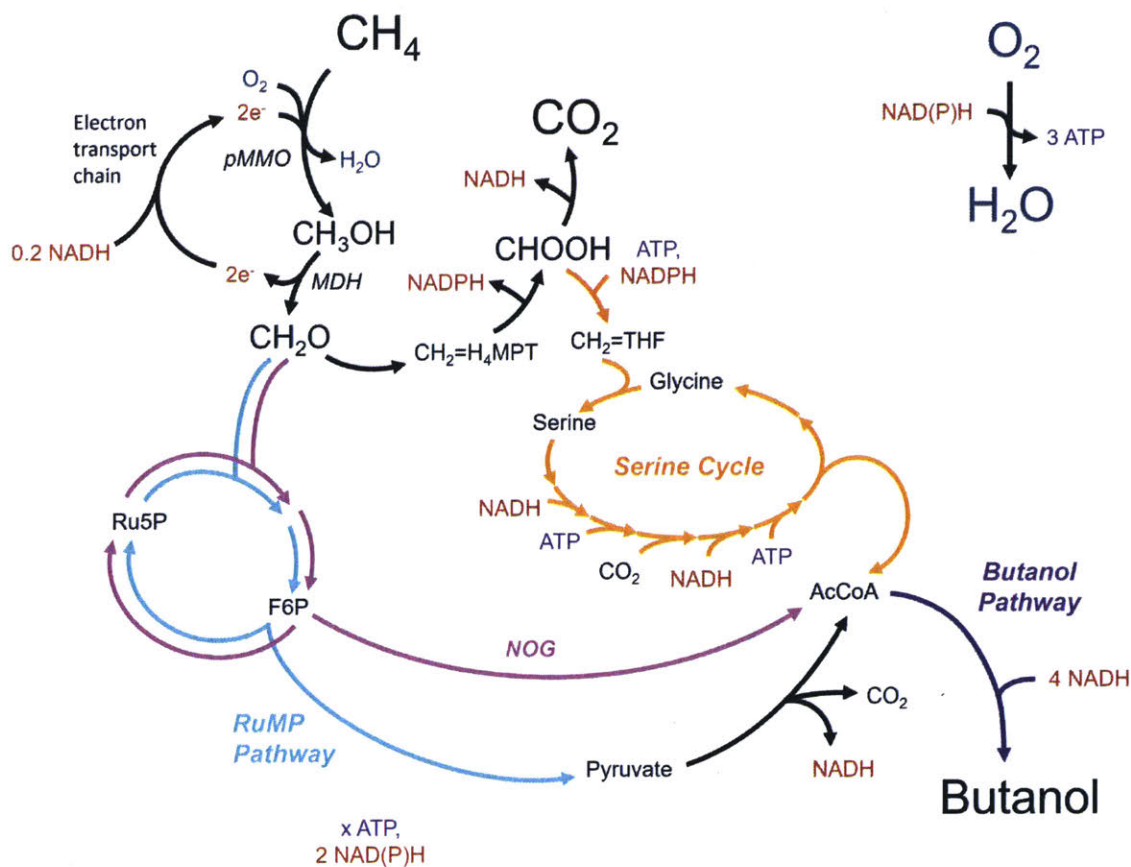
Figure 4-2: Pathways in Aerobic Methane Metabolism

the second oxygen atom to water:

$$CH_4 + O_2 + AH_2 \rightarrow CH_3OH + H_2O + A$$

121
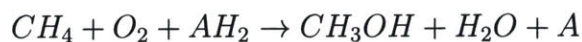
There are two variants: the soluble version (sMMO) which uses NADH as electron donor, and the membrane bound version (pMMO). Methanol is then oxidized by a PQQ-dependent methanol dehydrogenase (MDH) to formaldehyde. Though the electron donor for pMMO is unknown, experimental growth yield data and CH_4/O_2 values support the existence of an electron transport chain between MDH and the pMMO [80, 67].

In Type I methanotrophs, formaldehyde represents a node at the split point of biosynthesis (through the ribulose monophosphate (RuMP) pathway) and energy conservation (through complete oxidation to CO_2 via the NAD-dependent enzymes formaldehyde dehydrogenase and formate dehydrogenase.) In the RuMP pathway, formaldehyde is condensed with ribulose-5-phosphate (Ru5P) through an aldolase enzyme, then isomerized to fructose-6-phosphate. Three turns of this cycle and re-arrangements carried out by various combinations of enzymes of glycolysis and the pentose phosphate pathway regenerate the Ru5P acceptor and produce a 3-carbon intermediate that can be used for biosynthesis. Several variants of the RuMP pathway are known, that differ in their requirements for NADH and ATP, and which produce 3-carbon outputs at different oxidation states [64, 103]. In the models reported here, all these pathways are normalized to the production of acetyl-CoA. Oxidation of pyruvate to acetyl-CoA via pyruvate dehydrogenase results in the loss of carbon as CO_2, which lowers the overall carbon yield. Recent work in the Liao lab [176, 177] has demonstrated the potential of a synthetic pathway called non-oxidative glycolysis (NOG) to bypass this loss, converting two molecules of formaldehyde to acetyl-CoA without involvement of NADH or ATP.

By contrast, Type II methanotrophs use the Serine Cycle for biosynthesis, with formate as the node between biosynthesis and energy conservation [65]. Formate is reduced in an ATP-dependent manner to the formaldehyde oxidation level with NADPH and tetrahydrofolate (THF). In the serine cycle, carbon enters from formaldehyde and CO_2. The pathways for regenerating the serine vary based on the methanotroph, and have only been recently fully elucidated and demonstrated [66].

Butanol biosynthesis requires the condensation of two molecules of acetyl-CoA,

followed by their stepwise reduction with 4 molecules of NADH. These reducing equivalents are primarily derived from dissimilatory metabolism, with some being provided in the RuMP pathway. In our calculations, we include an ATP maintenance requirement. This ATP, as well as that needed in the serine cycle, is generated by reoxidation of NADH in the electron transport chain, where we assumed a P/O ratio of 3 ATP per mol NADH [178]. Some additional ATP is provided in the Emden-Myerhof-Parnas (EMP) variant of the RuMP pathway.

4.3.2 Butanol Yields from Aerobic Methane Oxidation
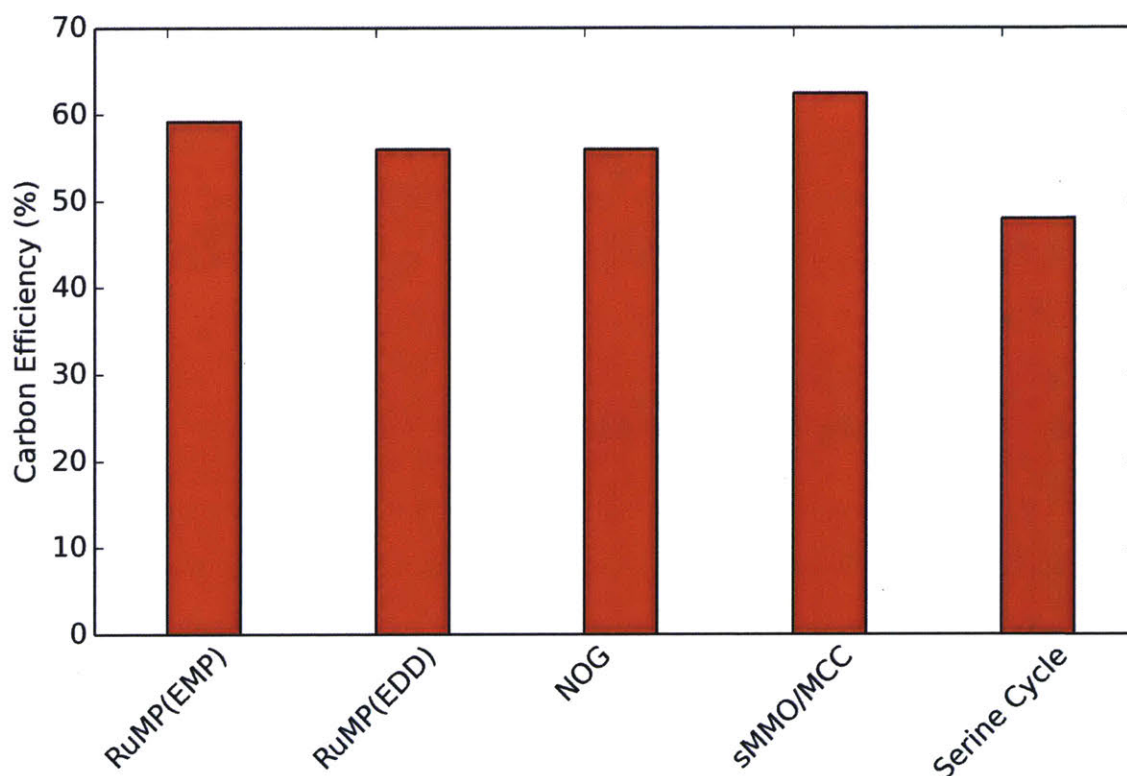
Figure 4-3: Maintenance-Adjusted Butanol Yields With Aerobic Methane Oxidation

The carbon conservation efficiencies for the production of butanol using the aerobic pathways discussed above are shown in Figure 4-3. Use of the serine cycle results in the lowest efficiency of 48%, due to the increased requirement for ATP and NADH for carbon fixation. The two natural variants of the RuMP pathway result in yields

123

ranging from 56-59%. These pathways differ only in the amount of ATP produced, thus the small difference in yield between these two scenarios is in agreement with the established notion that methanotrophs are generally reducing equivalent-limited, not ATP-limited. Interestingly, use of non-oxidative glycolysis (NOG) to convert formaldehyde to acetyl-CoA quantitatively, does not result in improved carbon yields compared to the natural variants. This is because this assimilation pathway does not generate any NADH or ATP, and therefore more formaldehyde has to be oxidized to provide these cofactors for butanol production and maintenance. Indeed, in the NOG case, the model predicts that 44% of the methane is completely oxidized, vs. 11% in the RuMP(EMP) case. As direct electron flow from MDH to pMMO is leaky, and requires additional NADH from formaldehyde oxidation [80], one could consider the alternative strategy of using the sMMO (which requires NADH), and the NAD-dependent methanol dehydrogenase characteristic of the Gram-Positive methylotroph *Bacillus methanolicus* [73]. This changes the net NADH cost of methane oxidation from 0.2 to 0 per mol methane oxidized, resulting in an improved yield of 62% when NOG is used.

4.3.3 Anaerobic Methane Oxidation

The exact details of anaerobic methane oxidation are unknown, though it is thought to proceed at least in part by a reversal of the canonical methanogenesis pathway [171]. In support of this, all of the required genes have been found in ANME-2 genome drafts [180], and a thermodynamically consistent model based on these reactions was recently put forward [179], that serves as the basis of our analysis (Figure 4-4). By this model, methane is first oxidized to methyl-coenzyme M (CH_3-S-CoM) by methane coenzyme M reductase (MCR), with its electrons transferred to the thiol of coenzyme B (CoB-SH). The methyl group is then transferred to the carrier tetrahydromethanopterin (H_4MPT), driven by the import of two sodium ions (Na^+), and the liberated CoM-SH is reoxidized with CoB-SH by heterodisulfide reductase (HDR) to form the mixed disulfide CoM-S-S-CoB, the cosubstrate for MCR. The electron acceptor for the HDR-catalyzed reaction is hypothesized to be the membrane-bound elec-
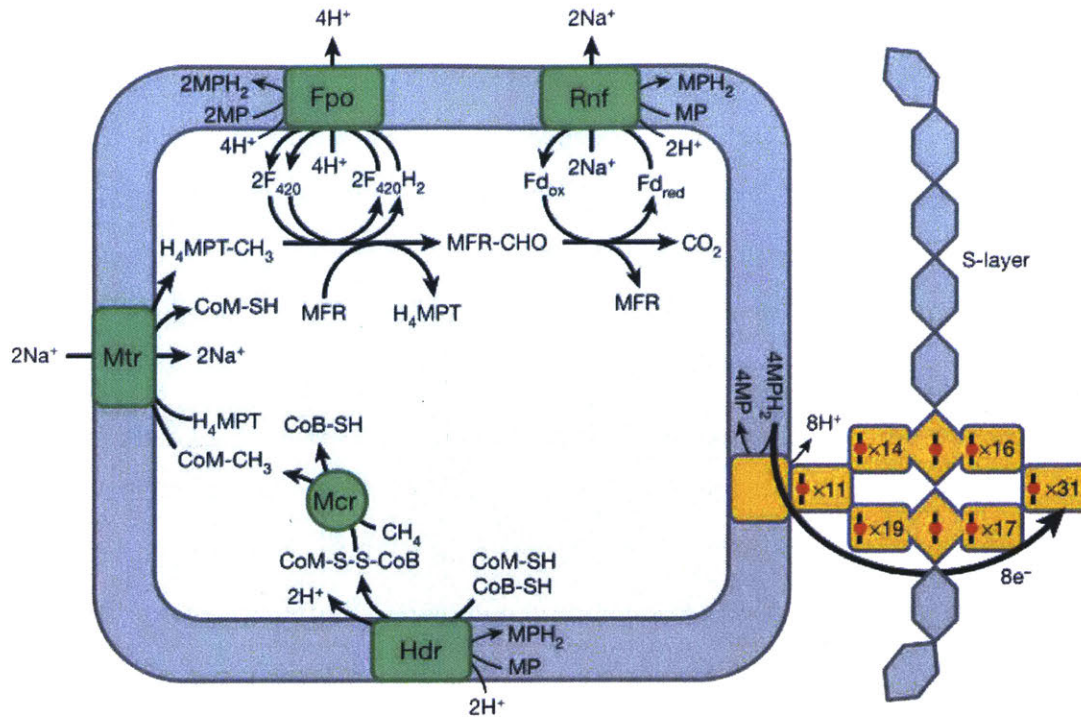
Figure 4-4: Pathways in Anaerobic Methane Metabolism, as presented in [179]

tron carrier methanophenazine ($MP + 2H^+ + 2e^- \rightarrow MPH_2$). In the methanogenic direction, reduction of CoM-S-S-CoB with MPH_2 has been shown to be coupled to the electrogenic translocation of two protons from the cytoplasm [181]. This is presumed to be scalar proton transport: Two protons are consumed in the cytosol to form the thiol groups of CoM-SH and CoB-SH, and the oxidation of MPH_2 occurs on the periplasmic side of the membrane, releasing two protons into the periplasm, while the electrons are funneled back through the cytochrome to the active site of the HDR on the cytoplasmic face of the membrane. Under standard conditions this reaction is not exergonic enough to allow proton translocation, thus it is hypothesized that a higher intracellular concentration of CoM-S-S-CoB, driven by the activity of MCR, provides the necessary drivig force. In the reverse direction, we must therefore assume that reduction of MP with CoM-SH and CoB-SH by membrane-bound HDR is coupled to the net *import* of two protons.

CH_3-H_4MPT is then oxidized in a stepwise manner to the formate level by the successive action of Mer and Mtd, with the electrons from each step transferred to

125

the soluble carrier F_{420}, and the carbon transferred to another carrier, methanofuran (MFR). Reoxidation of reduced F_{420} cofactor with MP as electron acceptor by the membrane-bound Fpo catalyzes the export of two protons, acting as a coupling site for energy conservation. The final oxidation to CO_2 is sufficiently exergonic to reduce ferredoxin (Fd). Reoxidation of the Fd at the membrane-bound RNF complex, again with MP as the oxidant, is coupled to the export of two to four Na^+ ions. The impact of this stoichiometry is examined later.

The overall result of complete methane oxidation is therefore the reduction of four molecules of MP, and the net export of approximately two protons. ATP is generated via the downhill flow of these protons through an ATPase, which is assumed here to require four protons to catalyze the conversion of ADP $\rightarrow$ ATP. The reduced MPH_2 must be reoxidized to allow metabolism to continue. In sulfate-linked AOM, ANME organisms exist primarily in aggregates with sulfur-reducing bacteria (SRB), leading to the hypothesis of direct intercellular electron transport [182]. Consistent with this notion, recent evidence has demonstrated the presence of a network of extracellular cytochrome 'wires' on the ANME surface [179, 183]. Reoxidation of MPH_2 on the outer face of the membrane with electron transfer to the cytochrome results in non-electrogenic proton translocation, and is thus not coupled to energy conservation.

AOM has also been reported with nitrate [172], nitrite [184], manganese and iron [185] as terminal electron acceptors. Nitrite-linked AOM is actually thought to use pMMO and other pathways from aerobic methanotrophy, with the required oxygen being produced intracellularly via a hypothetical NO dismutase [184], and is therefore not considered here. Nitrate-linked AOM has been demonstrated with ANME-2d [172], and will be discussed below for its potential to provide ATP through respiration. Iron-dependent anaerobic methane oxidation has been claimed in an engineered *Methanosarcina acetivorans* [186], and several possible mechanisms put forth [187, 188, 189]. The differences between these mechanisms and the one used in our analysis, and their effects on the calculated process yield, are discussed below. Electron transfer to managanese is not well understood, and not considered here.

Biosynthesis in ANMEs most likely starts from acetyl-CoA. The source of the

methyl group is CH_3-H_4MPT, and the CO moiety is derived from the ferredoxin-dependent reduction of CO_2. Both the reduction and synthase reactions are carried out the ACS/CODH complex. To model the conversion of acetyl-CoA into biomass, we take the approach used in [175], using the generic biomass equation $CH_{2.08}O_{0.53}N_{0.24}$, an ATP yield of 10.7 gDCW mol ATP^{-1}, and F_{420} as the source of reducing equivalents.

4.3.4 Model Validation and Maintenance Energy

To validate the proposed anaerobic methane oxidation metabolism, first the biomass yield was calculated in the absence of a fuel pathway. The simulation predicted that 9.7% of the methane flux would go toward biomass, with the remainder being fully oxidized to CO_2. Experimentally, it has been demonstrated that only 1% is captured as biomass [190, 191]. One possible source of the overestimation is the omission of a maintenance term. Though a maintenance coefficient has not been reported for ANME consortia, one can be estimated by introducing a non-zero ATP accumulation term in the model, such that the predicted biomass yield agrees with the experimentally observed value. Here, using the value 0.45 mol ATP mol CH_4^{-1} resulted in the correct biomass yield. Using a doubling time of 7 months, and a biomass yield of 0.6 gDCW mol CH_4^{-1} [191], this corresponds to a maintenance coefficient of 0.093 mmol ATP $gDCW^{-1}$ hr^{-1}. This value is two orders of magnitude lower than *E. coli* (7 mmol ATP $gDCW^{-1}$ hr^{-1}), and thus does not represent an inefficiency of ANME metabolism, but simply reflects the fact that, for a given maintenance coefficient, a lower specific substrate uptake rate results in a higher proportion of the substrate being used for ATP generation, as opposed to biosynthesis. As the assignment of this maintenance coefficient is somewhat speculative, in the results that follow we use the maintenance-free and maintenance-adjusted models as upper and lower bounds for the calculated yields.

The carbon conversion efficiencies for a variety of different metabolic scenarios are shown in Figure 4-5. Using the base case AOM model described above, the efficiency ranges from 3.2-33%, with or without maintenance, respectively. The large
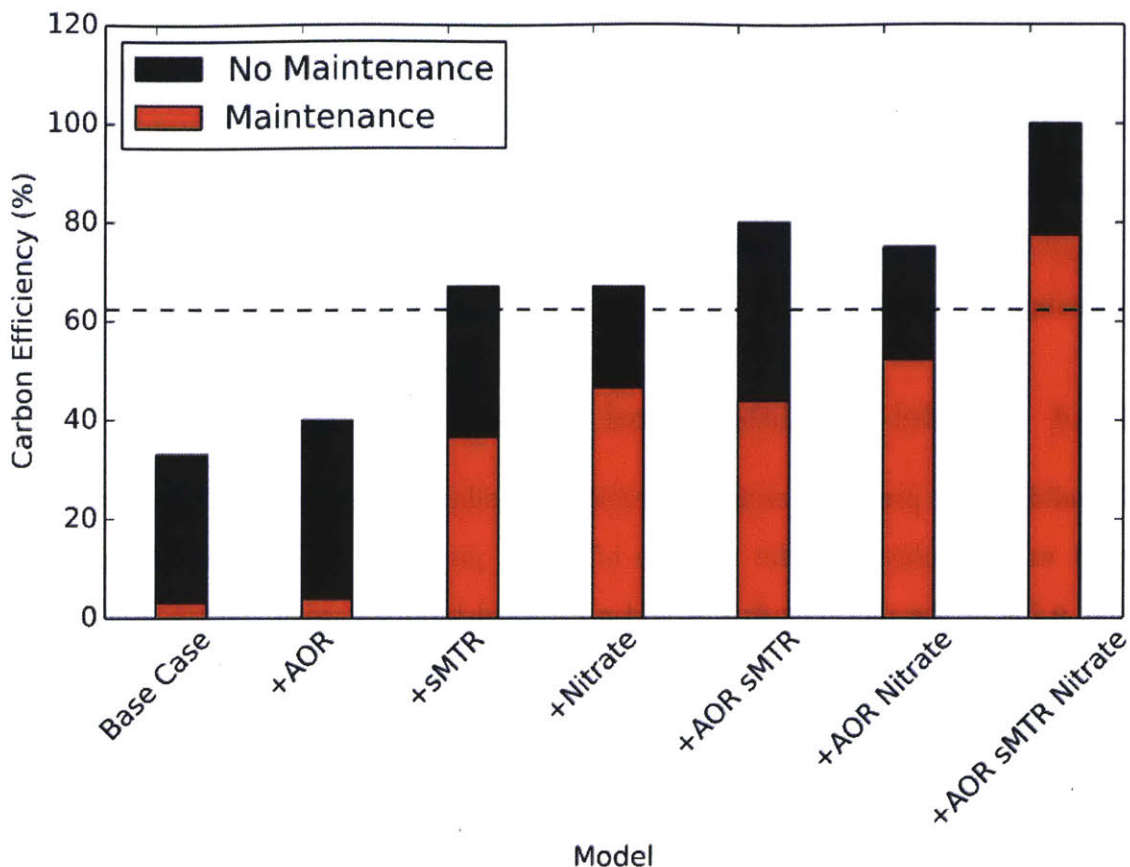
Figure 4-5: Butanol Yields With Anaerobic Methane Oxidation. Dotted line represents maximum yield achievable with MMO

difference between these upper and lower bounds suggests that the yield is limited by ATP production. We therefore examined alternative pathway and enzyme choices that could bolster ATP production. First, in the butanol pathway we considered implementing ferredoxin-dependent direct reduction of butyric acid to butyraldehyde, catalyzed by an aldehyde oxidoreductase (AOR). This allows the production of one ATP from the conversion of butyryl-CoA to butyric acid via phosphotransbutyrylase and butyrate kinase, compared to the prototypical pathway in which butyryl-CoA is directly reduced to butyrylaldehyde without ATP production [175]. This resulted in efficiencies of 4-40%, a modest improvement.

Second, one of the most costly steps in the reversal of methanogenesis is the methyltransferase (MTR) cascade, converting CH_3-S-CoM to CH_3-H_4MPT. This reaction as written is endergonic, and is driven by the influx of sodium ions down a

gradient. As the sodium ion gradient can alternatively be used for ATP production (via an H^+/Na^+ antiporter), this reduces the amount of ATP that can be produced. If the membrane bound, sodium-dependent MTR enzyme were replaced with the soluble variant, the transformation could be accomplished without the sodium ion gradient, allowing greater ATP production per mol methane. This increases the efficiency range dramatically, to between 37% and 67%, with or without maintenance. However, without involvement of the sodium gradient, the ΔG of the reaction is $+30$ kJ mol^{-1} [192]. This would only be thermodynamically feasible if the ratio of CH_3CoM to CH_3-H_4MPT were kept above 180,000:1. While there are instances of reactions this endergonic under standard conditions occurring in nature, the preceding reaction, conversion of CH_4 to CH_3CoM by MCR is also endergonic, with a ΔG of $+30$ kJ mol^{-1}. Combined, these two steps represent a practically insurmountable thermodynamic barrier of $+60$ kJ mol^{-1} [26].

The thermodynamics of anaerobic methane oxidation coupled to sulfate reduction are close to the thermodynamic limit of life. However, using an alternative electron acceptor such as nitrate or ferric iron could increase the amount of energy that could be captured in ATP from methane oxidation (Table 4.1). In the case of nitrate, the reduction of one mol of nitrate to nitrite on the cytoplasmic side of the membrane by a reductant at the redox potential of methanophenazine can be coupled to the export of at most two protons [193, 194]. To accomplish this, nitrate must first be imported to the cell. The most efficient way to do this would be via nitrate/nitrite antiport catalyzed by NarK [195]. Thus, the oxidation of four mol of methanophenzine could theoretically be coupled to the export of a maximum of four protons. This drastically increases the efficiency above the base case, to between 68% and 83%. Excitingly, these yields are above those predicted for aerobic methanotrophy and, in combination with the other modifications mentioned above, allow for 100% conversion of methane to butanol. A way to directly couple the reduction of ferric iron to ATP conservation is less clear. Iron-dependent growth is not well characterized, but in all cases the reduction is carried out extracellularly. As such, to maintain electroneutrality both the protons and electrons must cross the membrane, in a process that is not electrogenic

and thus cannot be used for ATP production by the ATPase.

There are a number of caveats to respiring nitrate in this approach: First, the nickel co-factor F_{430} of MCR is labile, with a redox potential of $-0.6V$ between the Ni(I) form and the inactive Ni(II) form, and is slowly oxidized even under strict anaerobic conditions. The rate of this inactivation has been shown to increase with increasing redox potential of the enzyme environment [196]. The nitrate/nitrite redox potential is $+0.43V$, therefore F_{430} would likely oxidize rapidly in the presence of intracellular nitrate. Since the reactivation of F_{430} is ATP-dependent [197], this would offset the benefit of respiring nitrate. Because of this, it has been hypothesized that MCR could not function in the presence of nitrate [198]. How then do the ANME-2d couple AOM to nitrate reduction? The nitrate reductase contains a twin arginine translocase (TAT) signal peptide for translocation to the outer cell membrane, suggesting that nitrate is reduced extracellularly, without influencing the redox state of the cytoplasm. By this model, nitrate reduction would not be energy conserving, which is consistent with the observation that ANME-2d do not have significantly faster growth rates or higher yields than ANME catalyzing sulfate-linked AOM [172], which would be expected if ATP could be derived from the process. So, it is highly unlikely that MCR-catalyzed AOM coupled to nitrate respiration could be implemented.

Given that the anaerobic methane oxidation is ATP limited, the butanol yield is highly sensitive to the stoichiometry of sodium translocation by the RNF complex. In the base case model, we assumed a value of two sodium ions. Under standard conditions, the reduction of methanophenazine (E^0 = -0.165 V) by reduced ferredoxin (E^0 = -0.41 V)[1] gives a free energy change of -47.3 kJ mol^{-1}, according to

$$\Delta G^0 = -nF\Delta E^0 = -nF(E^0_{acceptor} - E^0_{donor}) \tag{4.1}$$

where n is the number of electrons transferred and F is the Faraday constant. The movement of two sodium ions across the membrane has a ΔG^0 of 30 kJ mol^{-1} [192],

---

[1]A table of midpoint potentials used in this thesis is provided in the appendix

Electron Acceptor	$E^0_{acceptor}$ (V)	Overall Reaction	ΔG^0 (kJ mol^{-1})
Oxygen (O_2)	0.816	$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$	-815
Fe^{3+}	0.771	$CH_4 + 8Fe^{3+} + 2H_2O \rightarrow CO_2 + 8Fe^{2+} + 8H^+$	-780
Mn^{4+}	1.23	$CH_4 + 4Mn^{4+} + 2H_2O \rightarrow CO_2 + 4Mn^{2+} + 8H^+$	-1134
Nitrate (NO_3^-)	0.36	$CH_4 + 4NO_3^- \rightarrow CO_2 + 4NO_2^- + 2H_2O$	-463
Sulfate (SO_4^{2-})	-0.22	$CH_4 + SO_4^{2-} \rightarrow CO_2 + S^{2-} + 2H_2O$	-15.4

Table 4.1: Electron acceptors associated with methane oxidation

thus the free energy change of the RNF-catalyzed reaction could theoretically enable the transport of three sodium ions. Ferredoxin is thought to be overreduced in a number of organisms, and values as low as -0.5V have been reported for its potential. Under these conditions, the RNF reaction is associated with a ΔG of -64.6 kJ mol^{-1}, and could therefore support the translocation of as many as four sodium ions. We therefore compared the butanol yield expected for each of these stoichiometries, using the AOR-dependent butanol pathway (Figure 4-6). As expected, increasing the number of sodium ions translocated led to increases in butanol yield, up to a maximum of 30% or 57%, with or without maintenance. This value is still below the maximum attainable with aerobic methanotrophy.
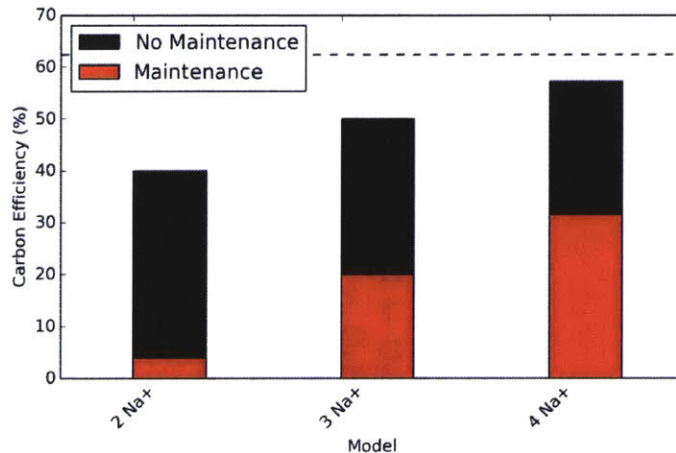
Figure 4-6: Butanol Yields With Varying RNF Stoichiometry

Besides the MTR, the next most costly step is the HDR-catalyzed reoxidation of CoM-SH and CoB-SH to the mixed disulfide CoB-S-S-CoM, which results in the import of two protons. Recently, a soluble bifurcating HDR was characterized that could couple the oxidation of two $F_{420}H_2$ to the endergonic reduction of Fd and the exergonic reduction of the CoB-S-S-CoM [189]. In principle, one could imagine using this enzyme in reverse to complete the pathway, and thus bypass the membrane-bound HDR and avoid the import of protons. This is shown schematically in Figure 4-7. By this pathway, complete oxidation of methane results in the net export of six protons, a significant boost compared to the base case scenario. This results in a favorable carbon conversion efficiency of 100% in butanol production. However,

132

once again there exists a problem with pathway thermodynamics: In this direction, the ΔG of the soluble HDR reaction is 10.6 kJ mol^{-1}, even with highly reduced ferredoxin. For net flux in the forward direction, this value must be negative. In principle, this could be achieved by keeping the ratio of the reactants (CoB-SH and CoM-SH) high compared to the product (CoM-S-S-CoB). However, as stated before, the MCR-catalyzed reaction is endergonic, and requires the opposite ratio to enable forward-driven flux. Thus, as with the replacement of membrane-bound MTR with soluble MTR, the replacement of the membrane-bound HDR with soluble HDR leads to a thermodynamically infeasible situation.
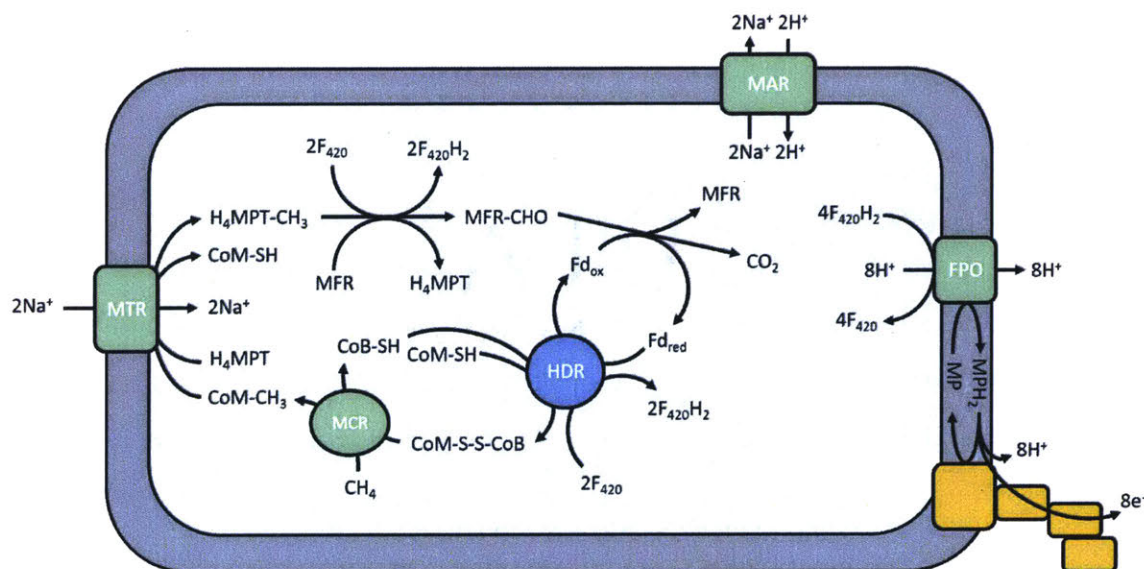
Figure 4-7: Anaerobic Methane Oxidation Using Soluble HDR

Other models for anaerobic methane oxidation coupled to fuel production have been proposed. In their work, Nazem-Bokaee et al calculate yields for a variety of compounds from methane using a refined Genome Scale Model (GSM) of *Methanosarcina acetivorans*, modified to account for extracellular iron respiration [187]. These extraordinary yields arise from the assignation of electron bifurcation to the RNF complex, an enzyme which they posit to catalyze the methanophenazine-dependent reduction of ferredoxin (endergonic) and ferric iron (exergonic). To the best of our knowledge, there exists no biochemical evidence to support the ability of RNF to catalyze this bifurcation, so this result should be treated with caution. Indeed, in re-

porting the discovery of the soluble HDR mentioned above, the same group removed this reaction in the updated model [189]. This new model is similar to ours, except that reoxidation of reduced F_{420} is mediated by the soluble HDR instead of by FPO. This results in the need to re-oxidize two molecules of CoM-SH and CoB-SH per methane oxidized by the membrane-bound HDR. Based on the discussion above, this is costly because each oxidation is connected to the import of two protons. Indeed, when we calculated butanol yields by this proposed pathway, we found the cell incapable of producing butanol due to an ATP deficit. It should be pointed out that in their model, these authors show MP reduction by HDR on the cytoplasmic side of the membrane, which is incongruous with the observation described above that in the methanogenic direction this enzyme is associated with energy conservation.

4.4   Conclusions

The calculations presented above show that, in the absence of pathway modifications that we have argued to be thermodynamically infeasible, the maintenance-adjusted yields achievable by anaerobic methane oxidation are *significantly* lower than those from aerobic methane conversion. This is interesting from a theoretical standpoint, given that the thermodynamic calculations suggest that similar amounts of energy could be extracted from methane with oxygen as with nitrate or iron (Table 4.1). This may be due to an inherent penalty associated with using an extracellular electron acceptor: In aerobic methanotrophy, the electrons removed from methane are retained intracellularly through the reduction of oxygen. By contrast, in anaerobic methanotrophy, they are deposited outside of the cell. To preserve electroneutrality, this transfer of electrons must be accompanied by the transfer of the same number of protons. If the protons are transported separately from the electrons, this is energetically costly because of the opposing proton-motive force, and results in the synthesis of less ATP. This has been documented for *Geobacter sulfurreducens* growing with ferric iron as the electron acceptor instead of fumarate [199]. If the protons are transferred together in a neutral molecule (as in reduced methanophenazine), there is no

direct cost for transport, but the reduction of the extracellular acceptor is associated with a decrease in pH, e.g.

$$MPH_2 + 2FeCl_3 \rightarrow 2FeCl_2 + MP + 2HCl \tag{4.2}$$
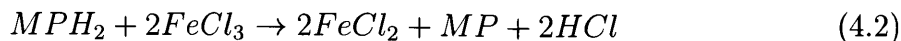
The acidification of the media results in increased cost for the electrogenic export of subsequent protons. Regardless of the exact mechanism, an important conclusion that can be drawn from this work is that thermodyanmic evaluation of a process, in the absence of a proposed pathway, is not a good indicator of theoretical yield.

Another important finding is that, although MCR allows more efficient conservation of the electrons from methane than MMO, this efficiency is nullified by the requirement for the input of energy to drive downstream reactions that are uphill, such as the MTR reaction. The benefit of MCR becomes even more questionable in light of the fact that its redox-sensitive cofactor, F_{430} prohibits the use of highly oxidized internal electron acceptors for energy conservation. One could make the assertion that the maintenance coefficient used in this work is an overestimation. However, even the maintenance-free yields are lower than the aerobic ones, with the exception of the cases using the soluble methyltransferase system— which brings the additional problem of two consecutive reactions with ΔG^0 values greater than $+30kJ$ mol^{-1}, and nitrate respiration— the feasibility of which is questionable due to F_{430} oxidation.

The astute reader will have realized that the arguments presented up to this point have drawn almost exclusively on support from principles of thermodynamics, and that kinetic considerations have been ignored. This is deliberate. One can always argue that through protein engineering, adaptive evolution, etc. the kinetics of a process can be improved. By contrast, the thermodynamic constraints of a pathway cannot be obviated by engineering (short of using an entirely different pathway with different chemistry). Thus, by casting our arguments in the context of the immutable thermodynamic constraints of the pathways, we avoid the counterargument that the scope of our analysis is limited by the current state of the art in metabolic engineer-

ing of ANMEs. This is not to say that kinetic considerations are not also important. A particularly glaring difference between anaerobic and aerobic methane oxidation is the rates of growth and methane oxidation that have been observed thus far. Methanotrophs of the gamma proteobacteria have growth rates on the order of 0.3 hr^{-1}, or a doubling time of 2.3 hours [2]. By contrast, growth rates of anaerobic methanorophs (ANMEs) are on the order of months. This may be due to the low activity of MCR in the reverse direction, which has been measured at 11.4 nmol min^{-1} mg^{-1} [200]. We have argued that yield is a critical parameter in establishing the economic viability of a biofuel process, but so too is productivity. So, in addition to the yield constraints, the expected slow rate of methane oxidation in an anaerobic setting provides further motivation to develop an aerobic process.

In light of these results, the utility of the extensive efforts and resources that will be needed to implement the anaerobic system heterologously seems questionable. This is especially true given the need for stoichiometric supplementation of sulfate or nitrate as electron acceptors, which adds an additional raw material cost of $450 or $850 per ton methane converted (based on ICIS historical pricing) in the absence of a cost-effective method for their recycling. This, coupled with the relative simplicity of working with aerobic methanotrophs and significant advances that are continually being made in understanding their metabolism [80, 12], manipulating their genetics [105], and engineering them for biochemical production [201], suggest that an aerobic system is a more viable candidate for development.

4.5   Theory and Detailed Model Equations

Aerobic Pathways

pMMO and MDH^{PQQ}

$$CH_4 + O_2 + A_{H_2} \rightarrow CH_3OH + H_2O + A \tag{4.3}$$

$$CH_3OH + PQQ \rightarrow CH_2O + PQQ_{H_2} \tag{4.4}$$

[2]http://methanotroph.org/

136

$$PQQ_{H_2} + 0.2NADH \rightarrow A_{H_2} \qquad (4.5)$$
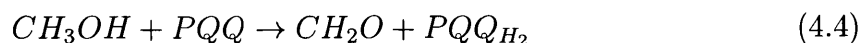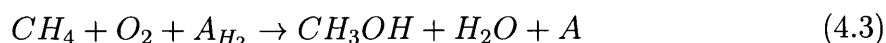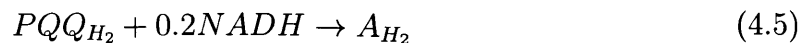
The combination of these three steps yields the overall stoichiometry:

$$CH_4 + O_2 + 0.2NADH + A \rightarrow CH_3OH + H_2O \qquad (4.6)$$

The additional 0.2 NADH represents leaky electron flow between MDH and pMMO, as discussed in [80].

sMMO and MDH^{NAD}

$$CH_4 + O_2 + NADH \rightarrow CH_3OH + H_2O \qquad (4.7)$$

$$CH_3OH \rightarrow CH_2O + NADH \qquad (4.8)$$
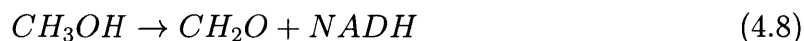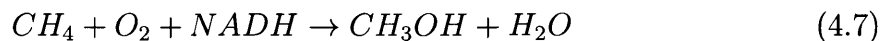
The sum of these equations gives the overall stoichiometry:

$$CH_4 + O_2 \rightarrow CH_2O + H_2O \qquad (4.9)$$

RuMP Pathway

Formaldehyde is either oxidized to CO2 to provide reducing equivalents, or assimilated via the following equations:

$$CH_2O \rightarrow CO_2 + 2NADH \qquad (4.10)$$

$$3CH_2O \rightarrow AcCoA + CO_2 + xATP + 2NADH \qquad (4.11)$$
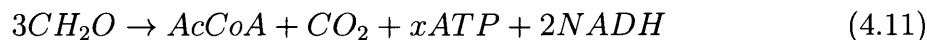
In the EMP variant, $x = 1$. In the EDD variant, $x = 0$.

Serine Cycle

Formaldehyde is transferred to the methyl carrier H$_4$MPT via the formaldehyde-activating enzyme (FAE), then oxidized in an NADP-dependent reaction to formate:

$$CH_2O + H_4MPT \rightarrow CHOO^- + NADPH \qquad (4.12)$$

137
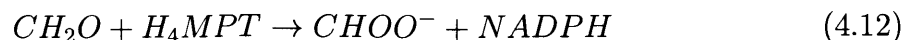
Formate is then either oxidized to produce reducing equivalents, or reduced in an ATP-dependent reaction to the formaldehdye level for incorporation via the serine cycle.

$$CHOO^- \rightarrow CO_2 + NADH \tag{4.13}$$

$$CHOO^- + ATP + NADPH + THF \rightarrow CH_2 = THF \tag{4.14}$$

$$CH_2 = THF + CO_2 + 3ATP + 2NADH \rightarrow AcCoA \tag{4.15}$$
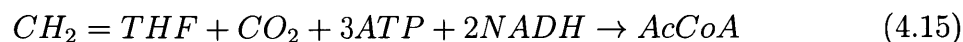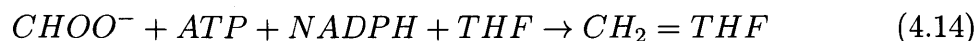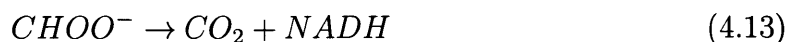
The combination of Equations 4.12, 4.14 and 4.15 gives us the stoichiometry of the serine cycle starting from formaldehyde to directly compare to the RuMP pathway:

$$CH_2O + CO_2 + 3ATP + 2NADH \rightarrow AcCoA \tag{4.16}$$

Non-oxidative Glycolysis

In addition to the natural formaldehyde assimilation pathways described above, we also consider the synthetic Non-oxidative glycolysis (NOG) pathway designed and demonstrated in the Liao lab [176, 177]. This pathway allows for the stoichiometric conversion of formaldehyde to acetyl-CoA, with no carbon loss:

$$2CH_2O \rightarrow AcCoA \tag{4.17}$$

Butanol Biosynthesis

Under aerobic conditions, NADH is the electron donor for butanol synthesis from acetyl-coA:

$$2AcCoA + 4NADH \rightarrow C_4H_{10}O \tag{4.18}$$
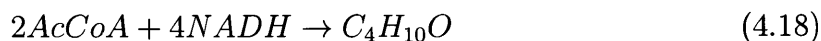
Electron Transport Chain

Apart from that generated in the EMP variant of the RuMP pathway, ATP is produced in the electron transport chain, with an assumed P/O ratio of 3:

$$NAD(P)H + O_2 \rightarrow 3ATP + H_2O \tag{4.19}$$
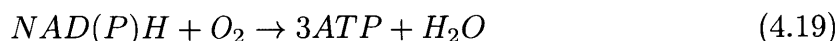
Maintenance

The maintenance requirement is estimated based on the *E. coli* maintenance coefficient (m) of 7 mmol ATP gDCW^{-1} hr^{-1}. This is converted to a per methane basis $(\tilde{m})$ via:

$$\tilde{m} = \frac{m}{q_s} = \frac{mY}{\mu} \tag{4.20}$$

Based on a doubling time (μ) of 0.37hr^{-1} and yield (Y) of 1.2 gDCW gCH$_4^{-1}$, we derive a maintenance coefficient of 0.369 mmol ATP gCH$_4^{-1}$.

Model Closure

The equations described above are converted to matrix-vector form $\mathbf{Sv} = \mathbf{b}$, where $\mathbf{S}$ is the stoichiometry matrix and $\mathbf{b} = \mathbf{0}$ except for the row corresponding to the ATP maintenance requirement, and the row corresponding to the substrate, for which a basis of 1 mol CH$_4$ is used. The solution vector $\mathbf{v}$ therefore represents the coefficients of the linear combination of the equations that describes the conversion of 1 mol CH$_4$. From this, the carbon efficiency is calculated as

$$C_{eff} = 4 \left(\frac{mol\ C_4H_{10}O\ produced}{mol\ CH_4\ consumed} \right) (100\%) \tag{4.21}$$

Anaerobic Pathways

For simplicity, generally only the reduced form of each electron carrier is shown in the equations below

Methane Oxidation

$$CH_4 + CoM\text{-}S\text{-}S\text{-}CoB \rightarrow CH_3\text{-}CoM + CoB\text{-}SH \tag{4.22}$$

$$CH_3\text{-}CoM + xNa^+_{out} \rightarrow CH_3\text{-}H_4MPT + CoM\text{-}SH + xNa^+_{in} \tag{4.23}$$
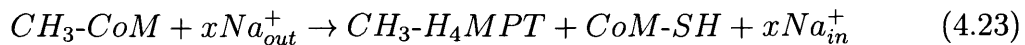
x = 2 for membrane-bound MTR, or 0 for soluble variant, as discussed in the text.

$$CH_3\text{-}H_4MPT \rightarrow CH_2 = H_4MPT + F_{420}H_2 \tag{4.24}$$

$$CH_2 = H_4MPT \rightarrow CH \equiv H_4MPT + F_{420}H_2 \qquad (4.25)$$

$$CH \equiv H_4MPT \rightarrow CHO\text{-}MFR \qquad (4.26)$$

$$CHO\text{-}MFR \rightarrow CO_2 + Fd_{red} \qquad (4.27)$$

Cofactor Regeneration

$$CoM\text{-}SH + CoB\text{-}SH + MPH_2 + 2H_{out}^+ \rightarrow CoM\text{-}S\text{-}S\text{-}CoB + 2H_{in}^+ \qquad (4.28)$$

$$Fd_{red} + MP + 2Na_{in}^+ \rightarrow MPH_2 + Fd_{ox} + 2Na_{out}^+ \qquad (4.29)$$

$$F_{420}H_2 + MP + 2H_{in}^+ \rightarrow MPH_2 + F420 + 2H_{out}^+ \qquad (4.30)$$

$$MPH_2 + A + xH_{in}^+ \rightarrow MP + AH_2 + xH_{out}^+ \qquad (4.31)$$
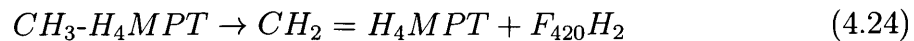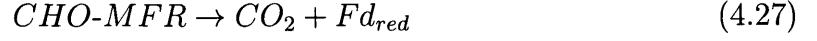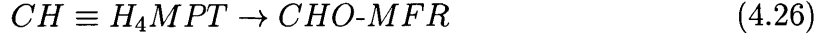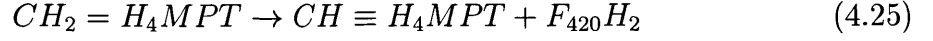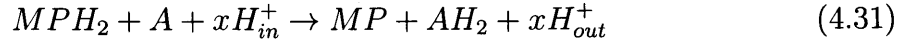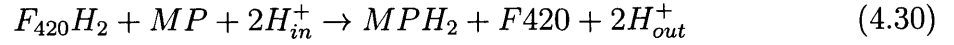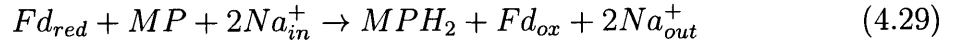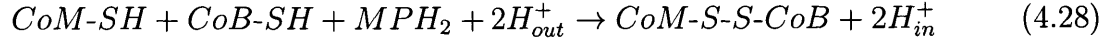
In the case of sulfate reduction, A is extracellular cytochrome, and $x = 0$. With nitrate reduction, A is nitrate, and $x = 1$.

ATP Production

ATP is generated through the membrane-bound ATPase, with an assumed stoichiometry of 4H$^+$/ATP:

$$4H_{out}^+ + ADP \rightarrow 4H_{in}^+ + ATP \qquad (4.32)$$

ATP can also be generated indirectly from the sodium gradient via a H$^+$/Na$^+$ antiporter, with an assumed stoichiometry of 1:1

$$H_{out}^+ + Na_{in}^+ \rightarrow H_{in}^+ + Na_{out}^+ \qquad (4.33)$$
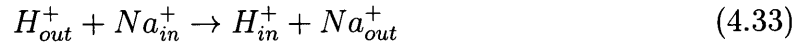
Biosynthesis

Carbon is assumed to be assimilated through the Wood-Ljungdahl pathway, with CO generated by ferredoxin-dependent reduction of CO_2.

$$CO_2 + Fd_{red} \rightarrow CO \qquad (4.34)$$

$$CH_3\text{-}H_4MPT + CO \rightarrow AcCoA \tag{4.35}$$
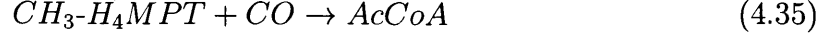
As discussed in the text, biomass production is assumed to follow the stoichiometry

$$0.5AcCoA + 4ATP + 0.3Fd_{red} \rightarrow 1C_{mol}\ Biomass \tag{4.36}$$
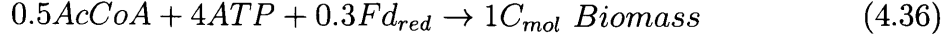
Anaerobic butanol synthesis is most ATP-efficient when crotonyl-CoA reduction is coupled to ferredoxin reduction using electron bifurcation [35]. The overall stoichiometry for butanol production, assuming $F_{420}H_2$ as the source of reducing power, is therefore

$$2AcCoA + 5F_{420}H_2 \rightarrow C_4H_{10}O + Fd_{red} + xATP \tag{4.37}$$
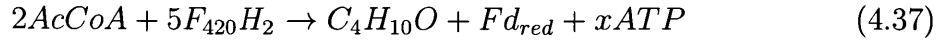
where x is 0, except when the AOR-dependent pathway is used, as described in the text.

Maintenance and Model Closure

The anaerobic model is solved in the exact same way as described for the aerobic models, except with a different maintenance coefficient. As values are not available for ANME cultures, we consider both the case of no maintenance, and a value of 0.45 that results in a biomass yield in line with those reported in the literature.

Chapter 5

Engineering *E. coli* to Metabolize Methanol

5.1 Summary

Methanol is an attractive substrate for biochemical production, but is not widely used in industrial biotechnology, largely because the microbes capable of metabolizing methanol remain poorly characterized and are not amenable to the scale of metabolic engineering required to develop a commercial process. One approach to overcome this barrier is to engineer the methanol assimilation pathway into a microbe that is already well characterized, and for which there is significant expertise in metabolic engineering for the production of biochemical products. To that end, in this chapter I describe the engineering of the canonical host *E. coli* to metabolize methanol. First, various isoenzymes catalyzing the three heterologous reactions were screened, and a combination was determined that allowed the robust incorporation of ^{13}C methanol into resting cells. These strains could also direct methanol into the heterologous product 3-hydroxybutyrate. Second, a combination of theoretical analysis and experiments with various labeled tracers was used to establish the primary limitations in the pathway. This work revealed that insufficient ribulose 5-phosphate (Ru5P) leads to the accumulation of formaldehyde, which reduces pathway flux by reducing the driving force for methanol oxidation. In strains where Ru5P regenera-

tion was addressed, either by the feeding of pentose precursors, or the inhibition of glycolysis by a combination of chemical and genetic techniques, the pathway becomes limited by the activity of the first enzyme, methanol dehydrogenase. The insight gained in this work forms the foundation for the design of metabolism control strategies to rewire Ru5P regeneration, and motivates efforts to evolve kinetically superior MDH variants, ultimately enabling the growth of *E. coli* on methanol as the sole carbon and energy source.

5.2 Introduction

One of the principle motivations for industrial biomanufacturing is the desire to increase the sustainability of the production of chemicals that have typically been produced from non-renewable petroleum resources by producing them instead from renewable feedstocks. Up to this point, however, most biotechnological processes have relied on easily accessible sugars from plants such as corn or sugarcane as a substrate. While these simple sugars can be produced renewably by photosynthesis, their increased use in biofuels and biochemicals production may lead to disruption in food supply. This so-called 'food vs fuel' debate has led to increased interest in finding other renewable feedstocks for biomanufacturing. Methanol represents a promising candidate for biochemical production [15]. Its current price is comparable to that of glucose, despite being more energy-dense, with 50% more electrons per carbon. Methanol can be produced renewably from municipal solid wastes and biogas through syngas as an intermediate, as expemlified by Enerkem[1]. It can also be produced using renewable electricity from CO_2, for instance in the process of CRI in Iceland[2]. Conversion of methane to methanol as a starting material for value-added processes has also been proposed as a way to monetize stranded natural gas, the availabillity of which has increased dramatically due to the technological advances in hydraulic fracturing, and a large fraction of which is currently flared in the absence

---

[1]http://enerkem.com/
[2]http://carbonrecycling.is/

of an economically viable alternative. For these reasons, the concept of a 'Methanol Economy', advocated by George Olah, has received considerable attention [14].

The use of methanol in bioprocessing presents several challenges. Though it has been used extensively for the production of single-cell protein (SCP), until recently, gaps in the fundamental understanding of the biochemistry of methylotrophy and the lack of robust tools for metabolic engineering in methylotrophs limited the ability to engineer these organisms to produce alternative products. Genetic tools for methanotrophs have improved significantly in recent years [105], and our knowledge of the biochemical pathways has increased [12], but the ability to divert carbon flux to heterologous products remains limited [201, 111]. Because of these difficulties, significant interest has arisen in importing methanol assimilation into the more tractable host *E. coli*. This concept, dubbed 'synthetic methylotrophy' [202], aims to combine the advantages of methanol as an attractive carbon substrate for fermentation with the robust engineering tools and range of downstream heterologous products available in *E. coli*.

The first step in methanol metabolism in native organisms is the oxidation of methanol to formaldehyde (Figure 5-1). In Gram negative methylotrophs, methanol is oxidized in the periplasm by a PQQ-dependent methanol dehydrogenase. The electrons are transferred into the electron transport chain by way of a cytochrome, and there contribute to ATP production. Methylotrophic yeast use an alcohol oxidase to convert methanol to formaldehyde while simultaneously reducing molecular oxygen to peroxide. Because of the toxicity of peroxide, this process occurs in special intracellular vesicles called peroxisomes, where catalase is expressed to disproportionate the peroxide to water and oxygen. In both of these cases, the electrons derived from methanol are unavailable for the biosynthesis, limiting the theoretical yield of a variety of reduced molecules. In Gram positive methylotrophs, typified by the thermophile *B. methanolicus*, methanol oxidation is linked to NAD reduction in the cytosol. This feature makes this enzyme advantageous for synthetic methylotrophy, because NADH is a universal electron carrier in *E. coli* and therefore increases the theoretical yield by allowing for more of the energy of methanol to be conserved in the product. However,

this benefit comes with a cost: NAD-dependent methanol oxidation is unfavorable ($\Delta G^0 = +34.2$ kJ mol^{-1}), imposing an uphill thermodynamic barrier in the first step of the pathway.
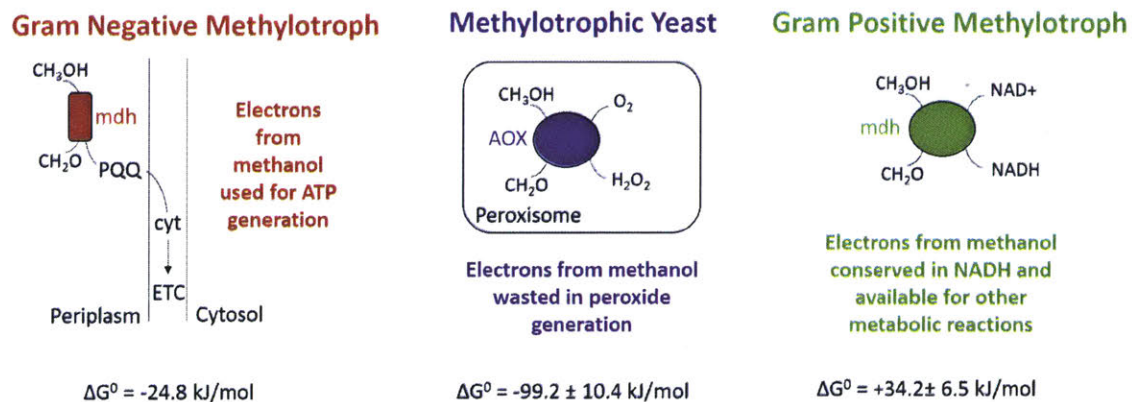
Figure 5-1: Enzymes for the Oxidation of Methanol to Formaldehyde

In native methylotrophs, formaldehyde acts as a node between assimilation into biomass, and further oxidation to provide energy for the cell. Because of the toxicity of formaldehdye, these processes are highly co-ordinated and the concentration of formaldehyde tightly regulated. The two best-known pathways for formaldehdye assimilation are the ribulose monophosphate (RuMP) pathway [64], and the serine cycle [66]. The latter incorporates carbon at the redox level of formaldehyde and CO_2 and is energetically costly. In the RuMP pathway (Figure 5-2), formaldehyde is condensed with ribulose 5-phosphate through an aldolase, hexulose phosphate synthase (HPS), to generate d-arabino-3-hexulose-6-phosphate. This molecule is then isomerized via phosphohexulose isomerase (PHI) to provide fructose 6-phosphate (F6P). Several variants of the regenerative portion of the RuMP pathway are known, and each uses or generates varying amounts of the cofactors ATP and NADH in a cycle involving enzymes of the non-oxidative pentose phosphate pathway to regenerate Ru5P from F6P to allow further formaldehyde assimilation. The FBP aldolase / transaldolase (FBP/TAL) version of the pathway is the most energetically efficient [64], and uses enzymes found in *E. coli*. This pathway therefore represents an ideal starting point for engineering a methylotrophic *E. coli*, as it requires the expression

of only three heterologous enzymes: MDH, HPS, and PHI.
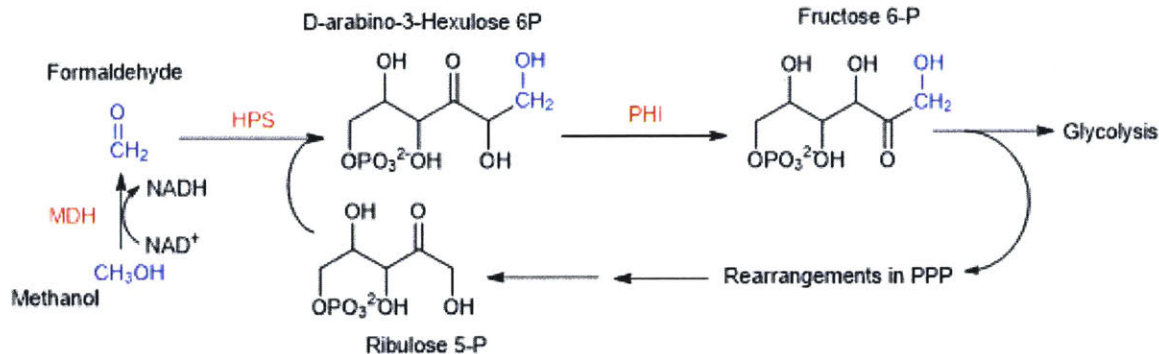
Figure 5-2: Schematic Representation of the Ribulose Monophosphate (RuMP) Pathway

Several groups have attempted this before. In the seminal work in this field, Müller and co-workers demonstrated the ability of resting *E. coli* cells expressing the heterologous enzymes to incorporate ^{13}C methanol into central carbon metabolites [203]. Recently, Whitaker and co-workers showed that supplementation of methanol during growth on yeast extract could allow a similarly engineered strain to reach higher cell densities than with yeast extract alone, and could produce the heterologous product naringenin with a significant amount of the carbon derived from methanol [204]. To date, however, no group has been able to demonstrate growth of *E. coli* on methanol as a sole carbon source. Similar efforts in *Corynebacterium glutamicum* have met with the same results [205]. Various hypotheses for this have been put forth, chief among these being the relatively poor kinetics and thermodynamics of heterologously expressed MDH. This has led to efforts to find and engineer more active variants of this enzyme [206], and a strategy involving the co-localization of MDH and HPS through a SH3-ligand interaction pair to prevent the build-up of formaldehyde [207]. However, the effects of these modifications on the incorporation of methanol into central metabolism have not been characterized. More broadly, there has been no systematic attempt to quantitatively address the limitations of the pathway. Such an analysis is critical in the design-build-test workflow, but is complicated by the slow rate of methanol oxidation and its volatility, which together make measuring the rate of methanol consumption difficult.

To address this need, in this work we engineered strains with similar performance to the previously reported strains, and then developed quantitative assays to systematically characterize several hypothesized limitations, with the aim of establishing the most effective targets for further metabolic engineering. Our analysis revealed the depletion of Ru5P as the primary limitation to high methanol flux, followed by the kinetics of MDH. These results provide direction for subsequent work in this field. In addition, the assays developed here are generalizable, and can be used to evaluate the removal of these bottlenecks in improved strains in the future.

5.3 Methods

5.3.1 Strains and Plasmids

Strains and plasmids used in this study are shown in Table 5.1. MG1655(DE3) was the base strain for methanol metabolic engineering [208]. *E. coli* DH5α was used as a general cloning host, and BL21(DE3) was used for the purification of proteins. Both were purchased as competent cells from NEB. pCas9 and pTargetF were gifts from Sheng Yang (Addgene plasmids 62225 and 62226) [209].

Table 5.1: Strains and plasmids used in *E. coli* methanol assimilation

Description	Reference
Cloning strain	NEB
Protein expression and purification	NEB
F^- λ^- *ilvG rfb-50 rph-1* (DE3)	[208]
MG1655(DE3) with gshB knocked out	This work
MG1655(DE3) with frmA knocked out	This work
MG1655(DE3) with frmA and rpe knocked out	This work
MG1655(DE3) with frmA and rpiA knocked out	This work
MG1655(DE3) with frmA and rpiA and rpiB knocked out	This work
ColE1 lacI pT7 Kn^R	Novagen
pET28 carrying *B. methanolicus* MGA3 mdh and act, *M. gastrii* hps and phi	This work
pET28 carrying *B. methanolicus* MGA3 mdh2 and act *M. gastrii* hps and phi	This work
pET28 carrying *B. methanolicus* MGA3 mdh2, *M. capsulatus* hps and phi	This work
pET28 carrying *B. methanolicus* MGA3 mdh2 and act, *M. capsulatus* hps and phi	This work
pETMEOH140 with optimized RBS for hps and phi	This work

pETMEOH140 with adhA from *C. glutamicum* replacing *B. methanolicus* mdh2	This work
pETMEOH300 with optimizd RBS for hps	This work
pETMEOH300 with *B. methanolicus* hps in place of *M. capsulatus* hps	This work
pETMEOH320 with *G. stearothermophilus* mdh in place of adhA	This work
pETMEOH320 with *C. necator* mdh2 V4-1 [210] in place of adhA	This work
pET28 carrying with *C. necator* mdh2 V4-1	This work
pET28 carrying hps from *M. capsulatus*with N-terminal HIS tag	This work
pET28 carrying phi from *M. capsulatus*with N-terminal HIS tag	This work
pET28 carrying rpiA from *E. coli*with N-terminal HIS tag	This work
p15a lacI pT7 Cm^R	Novagen
pACYC-Duet1 carrying *E. coli* rpe, tktA, talB, rpiA	This work
pACYC-Duet1 carrying *E. coli* glpX	This work
pACYC-Duet1 carrying *S. pyogenes* NADH oxidase (nox)	This work
pACYC-Duet1 carrying catalytically inactviated *S. pyogenes* NADH oxidase (nox)	This work
pACYC-Duet1 carrying *R. eutropha* phaA, *C. acetobutylicum* hbd, and *E. coli* tesB	This work
pMB1 aadA sgRNA-pMB1	[209]
repA101(Ts) kan P$_{cas}$-cas9 P$_{araB}$-Red lacI^q P$_{trc}$-sgRNA-pMB1	[209]
pTargetF with homology for *frmA*	This work
pTargetF with homology and N20 for *frmA*	This work

5.3.2　Reagents

Unless specified, all chemical reagents were purchased in the highest grade available from Sigma. DIFCO M9 salts and LB, BACTO Agar, and casamino acids were purchased from BD. Methanol-free formaldehyde (16% v/v) was purchased as 1 mL ampules from ThermoFisher (Catalog # 28906), and dilutions for cellular assays were prepared fresh each day. Trace Elements (MD-TMS) and Vitamin Solution (MD-VS) were purchased from ATCC. Isotopically labeled substrates were purchased in the highest grade available from Cambridge Isotope Laboratories.

5.3.3　Cell Growth

LB was used in cloning and for growing cells for protein purification. For all other studies, cells were grown in M9$^+$ medium (DIFCO M9 amended with 1% MD-TMS and MD-VS) as follows: three colonies were inoculated from fresh transformation plates into M9$^+$ medium supplemented with the desired carbon source, 0.1 % casamino acids, and the relevant antibiotics (Kanamycin, 50 μg mL^{-1}; Chloramphenicol, 25 μg mL^{-1}) and grown overnight. In the morning, fresh M9$^+$ medium containing the desired carbon source and antibiotics but no casamino acids was inoculated at 1-2% (v/v) from the overnight cultures, grown at 37°C, and then treated as described below for the relevant assays.

5.3.4　Cloning

Heterologous genes were codon-optimized for *E. coli* and synthesized as GBlock fragments by IDT. All plasmids were assembled via Gibson assembly using primers shown in Tables 5.2, 5.3, 5.4, 5.5, and 5.6.*E. coli* DNA fragments were amplified using Q5 Polymerase (NEB) and 0.5 μL culture as a template in a 30 μL reaction. Deletion of *gshB* was accomplished by the method of Datsenko and Wanner [211]. All other deletion strains were generated using the CRISPR-Cas9 method described by Jiang et al [209], with the modification of curing the pCas9 plasmid at 42°C. The pTarget vectors for the knockouts are listed in Table 5.1. These vectors were constructed

in two steps: First, Gibson assembly was used to insert approximately 500 bp homology regions upstream and downstream of the target gene into pTargetF digested with XhoI. The homology regions were designed to make the exact deletion described in the KEIO collection [212]. Next, the N20 region for the specific gene was incorporated by amplifying the vector containing the homology region with a forward primer containing this sequence (xxx-sg-F) and a common reverse primer (pTarget-sg-R), and assembling the linear fragment after DpnI digest and purification. To generate a triple mutant of *nox*, the gene was amplified in two segments with nox-up_fwd/nox-up_rev and nox-dn_fwd/nox-dn_rev, which encode the mutations in the GA homology regions. These fragments were then assembled with pACYCDuet-1 amplified with pACnox_rev/pACnox_fwd.

5.3.5 Flux Balance Analysis

FBA simulations were carried out in MATLAB 2016b with the CobraToolbox (v 2.0) and the Gurobi optimization package, using the core *E. coli* model iAF1260 [213], amended with the heterolous reactions catalyzed by MDH, HPS, and PHI, and their associated metabolites. For modeling growth on ribose, substrate uptake was assumed to be ATP-dependent, and a reaction corresponding to ribose kinase was added to allow conversion of ribose to ribose-5-phosphate. Methanol uptake was assumed to be mediated by passive diffusion. A complete list of the reactions added to the model is found in Appendix A.4. To compare flux maps across different substrates, growth was first simulated with 10 mmol $gDCW^{-1}$ hr^{-1} glucose, and a growth rate of 0.86 hr^{-1} was determined. For growth on methanol and ribose, the substrate uptake rates were varied until the growth rate was within 2% of the glucose growth rate, resulting in a rate of 42 mmol $gDCW^{-1}$ hr^{-1} methanol, and 11 mmol $gDCW^{-1}$ hr^{-1} ribose.

5.3.6 Resting Cell Experiments

Resting cell experiments were performed essentially as described by Muller et al [214]. Briefly, cells were grown in in $M9^+$ medium until early exponential phase (OD_{600} 0.5),

Fragment	Source	Primer	Primer Sequence
pETMEOH100			
Bm MGA3 mdh	GBlock	N/A	
Bm MGA3 act	GBlock	N/A	
Mg HPS, PHI	GBlock	N/A	
Backbone	pET28	pET28_fwd	CCCGGGAGGAGGAATTAAC
		pET28_rev	GGTATATCTCCTTCTTAAAGTTAAAC
pETMEOH110			
Bm MGA3 mdh2	GBlock	N/A	
act, HPS, PHI	pETMEOH100	pET100_rev	GGTATATCTCCTTCTTAAAGTTAAAC
		pET100_fwd	CCCGGGAGGAGGAATTAAC
pETMEOH140			
Bm MGA3 mdh2	pETMEOH110	mdh2F-140	actttaagaaggagatataccATGACTAACACCCAGTCTG
		mdh2R-140	actttaagaaggagatataccATGACTAACACCCAGTCTG
Mc HPS, PHI	GBlock	N/A	
Backbone	pET28	pET28_fwd	CCCGGGAGGAGGAATTAAC
		pET28_rev	GGTATATCTCCTTCTTAAAGTTAAAC
pETMEOH150			
mdh2 and act	pETMEOH110	mdh2actF-150	actttaagaaggagatataccATGACTAACACCCAGTCTG
		mdh2actR-150	ggcgggccatGGTTAATTCCTCCTAATAATTTATTTG
Mc HPS and PHI	GBlock	N/A	
Backbone	pET28	pET28_fwd	CCCGGGAGGAGGAATTAAC
		pET28_rev	GGTATATCTCCTTCTTAAAGTTAAAC
pETMEOH160			
mdh2	pETMEOH140	pET140_rev	tgtatacctcctattataTTACATGGCGTTTTTGATGATC
		pET140_fwd	CGAATTCGAGCTCCGTCG
Mc HPS	pETMEOH140	hps(mc)F160	atcaaaaacgccatgtaatataataggaggtatacaATGGCCCGCCCGTTAATC
		hps(mc)R160	gttaatacctcctaatattaaTCATGCGCTTGCAGCTAC
Mc PHI	pETMEOH140	phi(mc)F160	gcgcatgattaatattaggaggtattaacATGCATCAGAAACTCATTATC
		phi(mc)R160	cggagctcgaattcgTCATTCCAAATTAGCGTGAC

Table 5.2: Primers used for construction of initial methanol assimilation vectors pETMEOH100-160

Fragment	Source	Primer	Primer Sequence
pETMEOH300			
Cg adhA	GBlock	N/A	
Mc HPS, PHI	pETMEOH140	pET140_fwd	ATTATTAGGAGGAATTAACCATGG
		pET140_rev	GGTATATCTCCTTCTTAAAGTTAAAC
pETMEOH302			
Cg adhA	GBlock	N/A	
Mc HPS	pETMEOH140	hps_mc(302)_fwd	gtgtggcaattcgctactaaagaacaaaacaagctaaggaggtaagaaaATGGCCCGCCCGTTAATC
		hps_mc(302_rev	ggttaattcctcctggatccTCATGCGCTTGCAGCTAC
Mc PHI	pETMEOH140	pET140_fwd	ATTATTAGGAGGAATTAACCATGG
		pET140_rev	GGTATATCTCCTTCTTAAAGTTAAAC
pETMEOH320			
adhA, PHI	pETMEOH300	pET300_fwd	GGATCCAGGAGGAATTAAC
		pET300_rev	TTAGTAGCGAATTGCCAC
Bm HPS	GBlock	N/A	
pETMEOH400			
Gs mdh	GBlock	N/A	
HPS, PHI	pETMEOH320	pET320_fwd	CGAAGAACCAAGAAACGAG
		pET320_rev	GGTATATCTCCTTCTTAAAGTTAAAC
pETMEOH500			
Cn mdh2 4-1	GBlock	N/A	
HPS, PHI	pETMEOH320	pET320_fwd	CGAAGAACCAAGAAACGAG
		pET320_rev	GGTATATCTCCTTCTTAAAGTTAAAC

Table 5.3: Primers used for construction of second generation methanol assimilation vectors pETMEOH300-500

Fragment	Source	Primer	Primer Sequence
pACRuMP			
Ec rpe	gDNA	rpe_fwd	ctttaataaggagatataccATGAAACAGTATTTGATTGCC
		rpe_rev	attgtatacctcctattataTTATTCATGACTTACCTTTGC
Ec tktA	gDNA	tktA_fwd	gtaagtcatgaataatataataggaggtatacaATGTCCTCACGTAAAGAG
		tktA_rev	tgttaatacctcctaatattAATTACAGCAGTTCTTTTGC
Ec talB	gDNA	talB_fwd	agaactgctgtaattaatattaggaggtattaacATGACGGACAAATTGACCTC
		talB_rev	tgttaatacctccttattaaTTACAGCAGATCGCCGATC
Ec rpiA	gDNA	rpiA_fwd	ggcgatctgctgtaattaataaggaggtattaacATGCGTGTGAAATTTCATAC
		rpiA_rev	cggtttctttaccagactcgagTCATTTCACAATGGTTTTGAC
Backbone	pACYCDuet	pDuet_fwd	CTCGAGTCTGGTAAAGAAAC
		pDuet_rev	GGTATATCTCCTTATTAAAGTTAAACAAAATTATTTC
pACnox			
Sp nox	gBlock	N/A	
Backbone	As above		
pACglpX			
Ec glpX	gDNA	glpX_fwd	ctttaataaggagatataccATGAGACGAGAACTTGCC
		glpX_rev	cagcggtttctttaccagacTCAGAGGATGTGCACCTG
Backbone	As above		
pAC3HB			
Re phaA	gDNA	phaA_fwd	tttaactttaataaggagatataccATGACTGACGTTGTCATC
		phaA_rev	attgtatacctcctattataTTATTTGCGCTCGACTGC
Ca hbd	gDNA	hbd_fwd	cgagcgcaaataatataataggaggtatacaATGAAAAAGGTATGTGTTATAGG
		hbd_rev	gttaatacctcctaatattaaTTATTTTGAATAATCGTAGAAACCTTTTC
Ec tesB	gDNA	tesB_fwd	ttattcaaaataattaatattaggaggtattaacATGAGTCAGGCGCTAAAAAATTTAC
		tesB_rev	tcgcagcagcggtttctttaccagaTTAATTGTGATTACGCATCACC
Backbone	As above		
pACnoxM3			
nox 5' frag	pACnox	nox-up_fwd	ttgtaggcgcgaatcacgcagcaACCGCGTGCATCAACACG
		nox-up_rev	cgccgatgtatgcACCACCGACCACTGCGATAC
nox 3' frag	pACnox	nox-dn_fwd	ggtcggtggtgcaTACATCGGCGTGGAACTG
		nox-dn_rev	ttatcgtaaacagtagcgcatgcACCTACTGCGTACACACC
Backbone	pACnox	pACnox_f	gcaTGCGCTACTGTTTACGATAAC
		pACnox_r	tgcTGCGTGATTCGCGCCTAC

Table 5.4: Primers used for construction of auxiliary expression vectors

Fragment	Source	Primer	Primer Sequence
pETCNmdh *Cn* mdh2 4-1	pETMEOH500	CN4-1_fwd	ctttaagaaggagatataccatgcatcatcatcatcatcacACCCACCTGAACATCGCT AATC
		CN4-1_rev	ttgttagcagccggatctcaTTACATCGCCGCAGCGAA
Backbone	pET28	pET28_fwd	TGAGATCCGGCTGCTAAC
		pET28_rev	gtgatgatgatgatgatgcatGGTATATCTCCTTCTTAAAGTTAAACAAAA TTATTTC
pETrpiA *Ec rpiA*	gDNA	pETrpi_fwd	ctttaagaaggagatataccatgcatcatcatcatcatcacCGTGTGAAATTTCATACC AC
		pETrpi_rev	cagtggtggtggtggtggtgctcgagTCATTTCACAATGGTTTTGAC
Backbone	As above		
pEThps *Mc hps*	pETMEOH140	hps(MC)_fwd	ctttaagaaggagatataccatgcatcatcatcatcatcacGCCCGCCCGTTAATCCAG
		hps(MC)_rev	caagcttgtcgacggagctcTCATGCGCTTGCAGCTAC
Backbone	As above		
pETphi *Mc phi*	pETMEOH140	phi(MC)_fwd	tttaagaaggagatataccatgcatcatcatcatcatcacCATCAGAAACTCATTATCG ATAAAATC
		phi(MC)_rev	ttgtcgacggagctcTCATTCCAAATTAGCGTGAC
Backbone	As above		

Table 5.5: Primers used for construction of vectors for protein purification

Fragment	Source	Primer	Primer Sequence
pTarget-frmA			
Upstream homology	gDNA	frmA_UP_fwd	tattaccctgttatccctacTTCCTTCTGCCGCCCGCT
		frmA_UP_rev	ttacggttcgCATCTCTCGCTCTTCCTCAATATGGTAAT AGATTC
Downstream homology	gDNA	frmA_DOWN_fwd	gcgagagatgCGAACCGTAATTCGTTACTG
		frmA_DOWN_rev	tgatggagctgcacatgaacACTAAATCCGGCAGCTCG
Backbone	pTargetF	Digest (XhoI)	
pTarget-frmA-N20			
N20	pTarget-frmA	frmA-sg1-F	gtcctaggtataatactagtCTAATTAAAGTCACCCATACgtttta gagctagaaatagc
		pTarget-sg-R	ACTAGTATTATACCTAGGACTGAG

Table 5.6: Primers used for generating deletion of *frmA*

and induced with IPTG. After 2 hours, cells were harvested by centrifugation at room temperature, washed once with M9$^+$ medium with no carbon source, and resuspended again in carbon-free M9$^+$ medium to an OD$_{600}$ of 1. Various substrates were then added as described in the text, and cells were incubated for a further 15 minutes at 37°C before the addition of methanol. Cells were then sampled periodically for internal metabolites, and supernatants were analyzed for formaldehyde and other extracellular compounds, as described below.

5.3.7 Formaldehyde Analysis

Formaldehyde concentration was measured in the supernatants assayed by the Nash reaction [215] modified for 96-well plate format. 200 μL cells were spun down for 1 minute at 13,000 RPM, before transferring 125 μL supernatant to a well. The plate was kept on ice until all samples had been collected. 125 μL Nash reagent (5 M NH$_4$OAc, 50 mM acetylacetone) was added to each well, the plate was incubated at 37°C for one hour, and absorbance was read on a SpectraMax M2e spectrophotometer at 412 nm. A standard curve was prepared fresh daily from 16% formaldehyde (methanol-free) vials in the range from 200 μM to 0 μM. To measure formaldehyde labeling, supernatant samples were derivatized with 2,4-DNPH and analyzed by LC-MS/MS, by a modification of the protocol by Zwiener et al [216]. Briefly, 20 mg 2,4-DNPH-HCl (TCI America) was dissolved in 15 mL of acidifed water/acetonitrile (HCl:H$_2$O:ACN 2:5:1) and washed twice with hexanes to remove impurities. 150 μL of the reagent was mixed with 50μL of sample, and incubated at room temperature for one hour. For LC-MS analysis, an Agilent 1100 HPLC with a Zorbax SB-C18 column (3.5 μm, 2.1x100mm) was connected to a SCIEX API 4000 Triple Quadrupole MS running in negative MRM mode. The mobile phase was 50:50 Acetonitrile:Water with a flow rate of 0.3 mL min^{-1} and the run was isocratic, with 10 μL injection. The MS source parameters were as follows: Collision Gas, 6; Curtain Gas, 40; Gas 1, 20; Gas 2, 80; IonSpray Voltage, -3000V; Temperature, 600°C. Optimal compound parameters were determined by infusion of a formaldehyde-DNPH standard (Sigma), and are as follows: Declustering Potential, -36V; Entrance Potential, -10V; Collision

Energy, -22V; Collision Cell Exit Potential, -10V). The m/z transition for unlabeled derivatized formaldehyde was 209.0/181.0.

5.3.8 Intracellular Metabolite Extraction and Quantification

Intracellular metabolites were extracted and isotope labeling patterns analyzed as described in [217]. Briefly, 1-2 mL of liquid culture was filtered through a 0.45 μm nylon filter, washed with 10 mL room-temperature water, and then the filter was transferred into a 50 mL falcon tube containing 5 mL of extraction solution (40:40:20 acetonitrile:methanol:water) at -20°C. After 30 minutes, the filter was removed, the samples were centrifuged, and the supernatants dried overnight under air. The next morning, dried metabolites were resuspended in 150 μL water, centrifuged at 13,000 RPM for 20 minutes, and injected into the LC-MS. Labeling distributions were corrected for natural abundance using the software IsoCor [218]. For quantification of internal metabolites, the extraction solution was spiked with ^{13}C internal standards that were previously generated by growing *E. coli* on U-^{13}C-labeled glucose, and area of unlabeled peaks was normalized to the area of fully labeled peak, as described in [219].

5.3.9 Extracellular Metabolite Quantification

Extracellular metabolites were analyzed using an Agilent 1200 HPLC system with a G1362A RID detector with isocratic flow of 14 mM sulfuric acid (0.7 mL min^{-1}, 50°C) through a BioRad Aminex HPX-87H ion exchange column (300 mm x 7.8 mm). 500 μL culture was withdrawn and filtered through 0.2 μm syringe filter into HPLC vials, and 10 μL injected into the HPLC.

5.3.10 Crude Lysate Preparation

For small-scale anaylsis of protein in crude lysates, 1 mL of cell culture (OD$_{600}$ 0.5-1) was spun down, and the supernatant removed by aspiration. Pellets were frozen, and then lysed with 100 μL BPER-Complete at room temperature for 15 minutes. The

soluble and insoluble fraction were separated by centrifugation at 20,000 RCF for 10 minutes at 4°C.

5.3.11 Quantification of MDH Solubility

To estimate the amount of MDH in the soluble and insoluble fraction, first the amount of MDH in each fraction was calculated by image analysis after SDS-PAGE. 10 μL of the soluble fraction were mixed with Laemmli buffer. The insoluble fraction was resuspended by vigorous pipetting in 100 μL BPER, and 10 μL mixed with Laemmli buffer. Samples were boiled for 5 minutes in a thermocycler, before being run on a Mini-PROTEAN TGX gel (12%, 15-well comb) at 150V for approximately 1 hour. Bands were visualized with InstantBlue gel stain (Expedeon) and size was compared to BioRad Kaleidoscope pre-stained standards. Bands were identified and quantified using the software AlphaImager HP. To quantify the total amount of protein in the soluble and insoluble fraction, first the insoluble fraction was re-solubilized using 8M guanidinium hydrochloride (Gu-HCl), and then diluted 3-fold in water to reduce the Gu-HCl concentration. Protein concenration was then determined by BCA Assay (Pierce). The amount of soluble MDH was then calculated as

$$
\begin{aligned}
\%MDH_{soluble} &= \frac{[MDH_{sol}]}{[MDH_{sol}] + [MDH_{insol}]} \\
&= \frac{[Protein_{sol}](\%Protein_{sol,MDH})}{[Protein_{sol}](\%Protein_{sol,MDH}) + [Protein_{insol}](\%Protein_{insol,MDH})}
\end{aligned}
$$

$$(5.1)$$

5.3.12 Enzyme Assays

All enzyme asasys were conducted at 37°C in 96-well plates using a SpectraMax M2e (Molecular Devices). Each well contained 10 μL of crude lysate in a total of 250 μL reaction mixture. Enzyme activities were normalized to the total soluble protein concentratino as measured by BCA assay.

160

MDH

Methanol dehydrogenase activity was measured by following the methanol-dependent reduction of NAD^+ at 340 nm. The assay consisted of 100 mM glycine-KOH (pH 9.5), 5 mM $MgSO_4$, and 1 mM NAD, and was ininitated by the addition of 500 mM methanol.

HPS and PHI

Hexulose-6-phosphate synthase (HPS) and phosphohexulose isomerase (PHI) activity were measured in a coupled assay as described in [177]. For HPS, the assay consisted of 50 mM Tris-HCl (pH 7.5), 5 mM MgCl2, 0.2 mM $NADP^+$, 5 mM ribulose 5-phosphate, excess PHI, PGI, ZWF, and RPI, and was initiated by the addition of formaldehyde (5 mM). For PHI, the assay was the same but with excess HPS instead of PHI. Enzyme velocities were calculated by monitoring $NADP^+$ reduction after the reaction reached steady state.

ACT

Nudix hydrolase activity of ACT was measured in a coupled assay with ADP-ribose as a substrate [220]. Mononucleotide phosphate relased by ACT was hydrolyzed to orthophosphate by Calf Intestinal Alkaline Phosphatase (NEB), and measured by Malachite green phosphate assay (Cayman Chemical, #10009325) after termination of the reaction with sulfuric acid. The reaction was carreid out in 100 mM glycine-KOH (pH 9.5), 5 mM $MgSO_4$, and were initiated by the addition of 250 nmol ADP-ribose.

NOX

NADH oxidase activity was measured by following the oxidation of NADH at 340 nm. The assay consisted of 50 mM Potassium Phosphate Buffer (pH 7.0), and was initiated by the addition of 200 μM NADH [221].

161

5.3.13 Large-Scale HIS-tagged Protein Purification

BL21(DE3) cells containing expression plasmids were grown overnight in LB medium, then subcultured at 1% (v/v) into LB medium in baffled flasks, and grown to early expontential phase (OD 0.5). Expression was induced with IPTG at 0.1 mM, and growth continued for 3-5 hours. Cultures were centrifuged, pellets taken up in ice cold extraction buffer (50 mM Tris-HCl (pH 7.5), 25 U mL^{-1} Benzonase (Sigma), 0.1 mM PMSF), and lysed by three passages through a High Pressure Homogenizer (EmulsiFlex-C5). The soluble fraction was separated by centrifugation at 20,000 RCF for 20 minutes at 4°C. HIS-tagged proteins were purified using Ni-NTA His-Bind Resin (EMD Millipore #70666) following the manufacturer's instructions. Imidazole was removed and protein concentrated by 5 passages through a 3K MWCO Microsep centrifugal filter (Pall Laboratory #MCP003C41) with ice cold Tris-HCl (20 mM, pH 7.5).

5.4 Results and Discussion

5.4.1 Establishing Methanol Assimilation in *E. coli*

To generate *E. coli* strains capable of metabolizing methanol, we cloned candidate genes for each of the heterologous enzymes (MDH, ACT, HPS, and PHI) as an artificial operon into the T7 expression vector pET28. The first plasmid, pETMEOH100, contained the major *mdh1* and *act* from *B. methanolicus MGA3*, and the *hps* and *phi* from *M. gastri*, since the latter two enzymes had previously been shown to be functionally expressed in *E. coli* [112]. To test the functionality of the engineered strain, MG1655(DE3) Δ*gshB* cells carrying this plasmid or the empty vector control were grown in M9^{+} medium with 0.4% glucose, and induced with 1 mM IPTG at early exponential phase. The deletion of *gshB* results in the absence of glutathione synthesis (Supplementary Figure A-3), which is required for the native detoxification system and may reduce flux into central metabolism in WT cells. After 2 hours of further growth, cells were spun down, washed, and re-suspended in the same medium

but with no carbon source. After 15 minutes, 50 mM ^{13}C methanol was added, and samples were taken for metabolite labeling at 0, 30, 60, 90 and 120 minutes. As shown in Figure 5-3, there was no detectable incorporation of methanol in this strain. We hypothesized that this could be due to low MDH activity. Switching to Mdh2 in pETMEOH110, which was shown to have an activity 1.5-fold higher than Mdh1 *in vitro* [73], led to a small improvement, with trace detection of methanol incorporation after approximately 90 minutes (Figure 5-3). That significantly increased MDH activity led to only a small change in methanol incorporation suggested that perhaps the downstream enzymes were rate-limiting. The HPS from *M. gastri* has a markedly higher k_{cat} than MDH, but the K_m for formaldehyde is high, at 2.96 mM [112], such that it could be relatively inactive with the low formaldehyde concentration generated by a slow MDH. We therefore looked for HPS candidates with a lower K_m. A detailed characterization of HPS and PHI from *M. capsulatus* [222], and HPS from *B. methanolicus* [223], showed that both of these enzymes had significantly lower K_m values than the isoenzyme from *M. gastri* (Table 5.7). As *M. capsulatus* is mesophilic, with an optimum growth temperature of 37°C, we reasoned its enzymes may function better in *E. coli* than the enzymes from the thermophilic *B. methanolicus*, and therefore cloned codon-optimized versions of these genes together with *mdh2* and *act*, to generate pETMEOH150. As depicted in Figure 5-3, cells carrying this plasmid showed substantial incorporation of ^{13}C methanol. The time scale and total fraction of carbon labeled is similar to that reported by Müller et al [214] despite using a much lower concentration of methanol (50 mM vs. 1 M), and shows that the pathway for methanol assimilation is functional in *E. coli*. Further, the presence of M+2 isotopes of phosphoenolpyruvate shows that there is some regeneration of Ru5P through the non-oxidative pentose phosphate pathway.

All the plasmids constructed up to this point expressed the MDH activator protein. Currently, there is no data to support a physiological role for this protein in methanol metabolism by *B. methanolicus*. To probe the utility of ACT in *E. coli*, we generated plasmid pETMEOH140, which is identical to pETMEOH150 but which lacks the *act* gene. As shown in Figure 5-4, expression of the activator protein conferred no
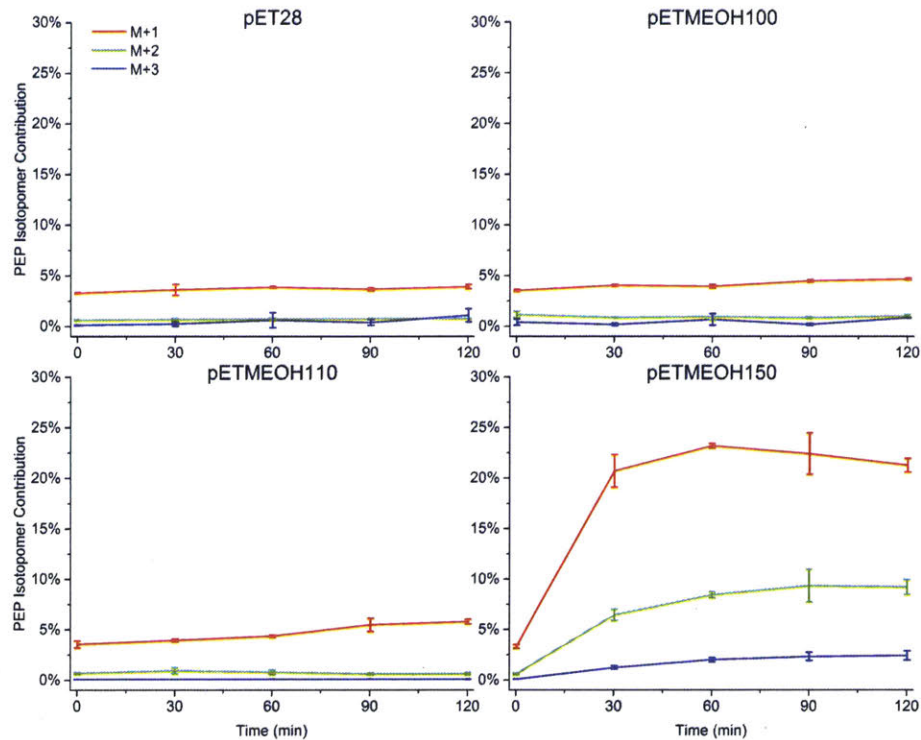
Figure 5-3: $^{13}CH_3OH$ Incorporation in Resting Cells of MG1655(DE3) $\Delta gshB$ Carrying pET28, pETMEOH100, pETMEOH110, and pETMEOH150

Organism	K_m	Reference
M. gastri	Formaldehyde: 2.96 mM Ru5P: 1.49 mM	[112] [224]
M. capsulatus	Formaldehyde: 0.49 mM Ru5P: 0.083 mM	[222]
B. methanolicus	Formaldehyde: 0.00147 mM Ru5P: 0.0045 mM	[223]

Table 5.7: Kinetic properties of HPS variants

improvement in ^{13}C label incorporation, despite confirming activity of the protein in crude lysates (Supplementary Figure A-4), and a 10-fold higher MDH activity in crude lysates of cells containing pETMEOH150 compared to pETMEOH140 (Figure 5-5). These data could indicate either that NUDIX hydrolase-mediated activation of MDH is only relevant *in vitro*, or that the activation is unsuccessful *in vivo* in *E. coli*.

164
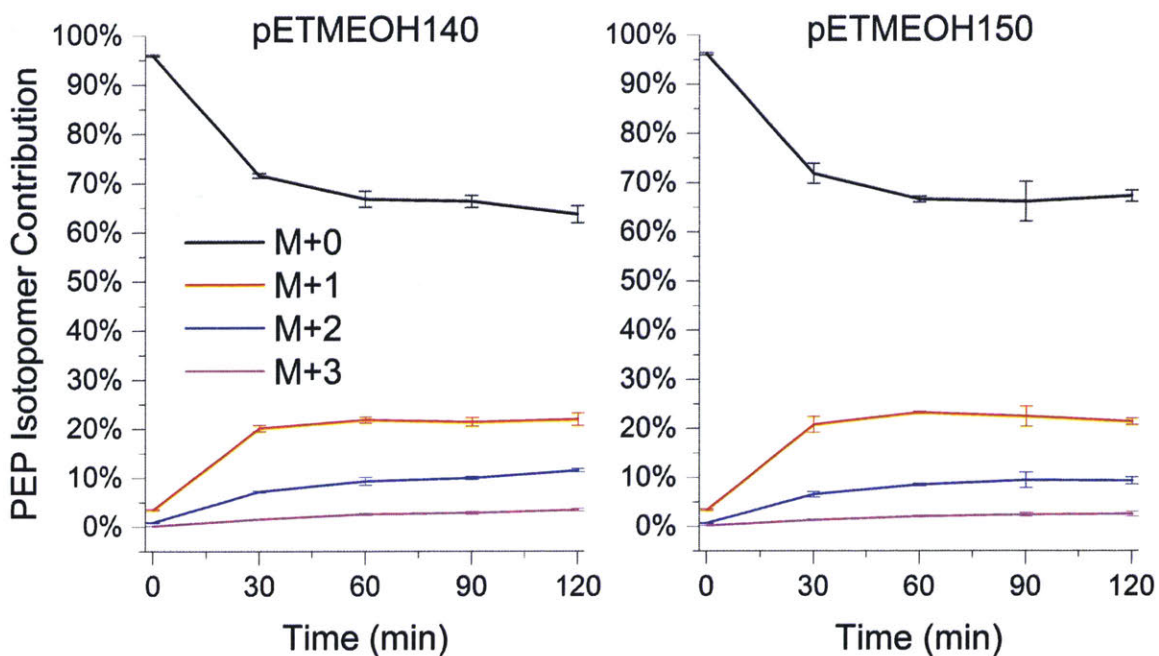
Figure 5-4: $^{13}CH_3OH$ Incorporation in Resting Cells of MG1655(DE3) $\Delta gshB$ Carrying pETMEOH140 and pETMEOH150

	pET28 (Control)	pETMEOH140	pETMEOH150 (+ act)	pETMEOH160
■ 3 hour Induction	1.07	53.29	555.99	31.55
■ Overnight Induction	3.77	38.13	305.44	45.86
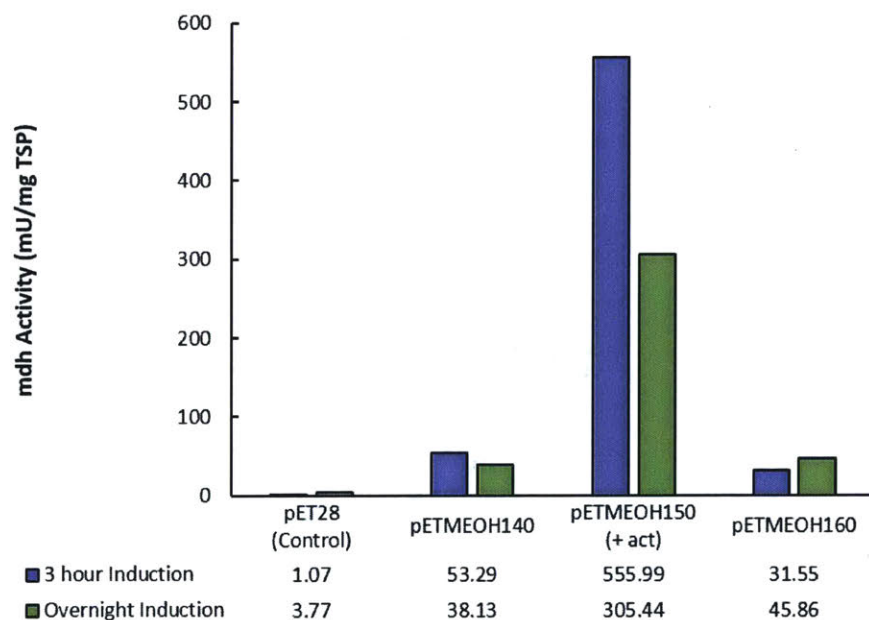
Figure 5-5: MDH Activity in Crude Lysates of pETMEOH140 and pETMEOH150

Given that an improved HPS led to the most significant improvement in methanol assimilation, we asked if a further increase in HPS activity could further improve

pathway flux. We used the Ribosome Binding Site (RBS) calculator [225] to design an optimized binding site for HPS and PHI to maxmize their expression level, and cloned these sites upstream of the respective genes in pETMEOH140 to generate pET-MEOH160. Assays of crude lysates of MG1655(DE3) carrying these vectors revealed a roughly 9-fold increase in HPS expression level in pETMEOH160 (Table 5.8) To si-

Construct	Specific Activity (U mg^{-1})
pETMEOH140	1.04±0.05
pETMEOH160	9.04±4.7

Table 5.8: Activity of HPS in Crude Lysates of pETMEOH140 and pETMEOH160

multaneously investigate the effect of the glutathione knockout, ^{13}C labeling analysis of these constructs was conducted in WT cells, enabling the additional comparison between pETMEOH140 in WT and $\Delta gshB$ cells. The results are shown in Figure 5-6. Despite the large increase in HPS activity, no significant increase in labeling was detected between cells carrying pETMEOH140 and pETMEOH160, suggesting that HPS activity does not control flux. Comparing WT:pETMEOH140 to $\Delta gshB$:140 (Figure 5-4), we notice a small but significant decrease in the total labeling in the WT cells, suggesting the native detoxification pathway competes with the RuMP for methanol flux.

5.4.2   Identification of Bottlenecks in Engineered Strains

Despite the fact that cells showed robust incorporation of ^{13}C methanol, indicating pathway activity, we were consistently unable to obtain methanol-dependent growth. This same phenomenon was observed by others in this field [214, 226], and likely indicates either a limitation in the pathway, or the inability of the strain to adapt to an atypical carbon source. We focused on the first issue, and set about analyzing potential bottlenecks in methanol assimilation.
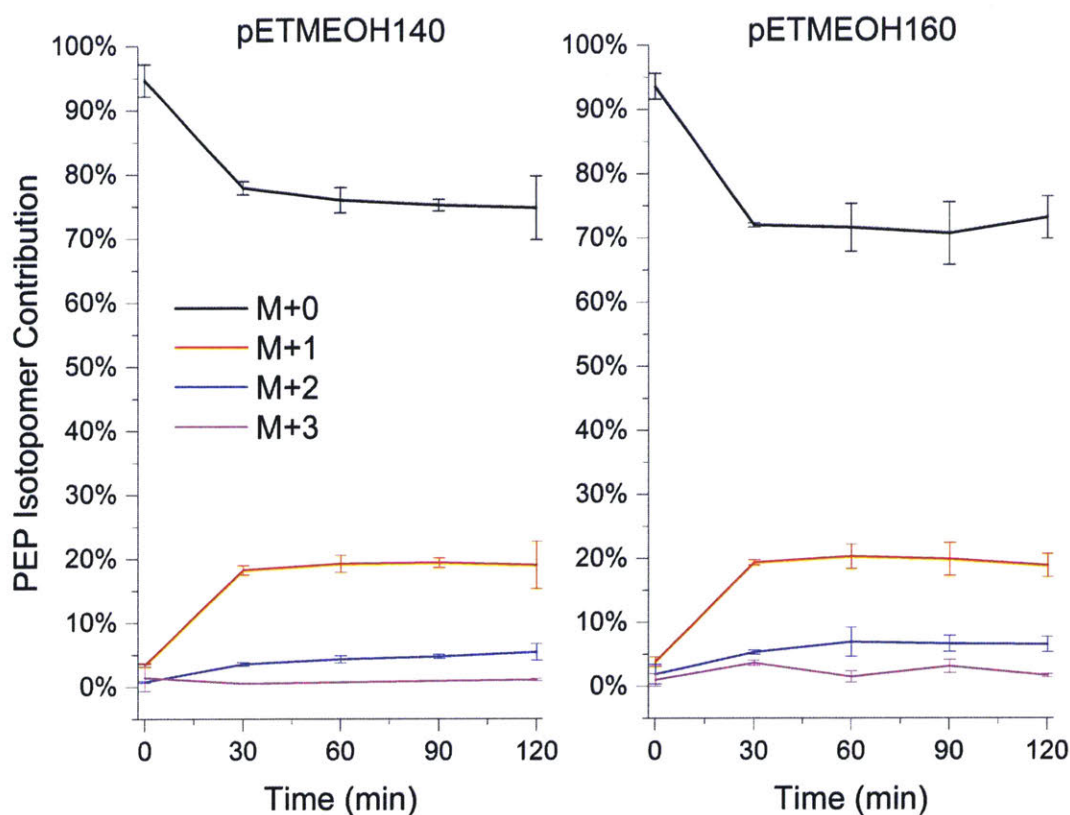
Figure 5-6: $^{13}CH_3OH$ Incorporation in Resting Cells of MG1655(DE3) Carrying pET-MEOH140 and pETMEOH160

Quantifying Methanol Uptake

To analyze pathway limitations, it is critical to quantify the rate of methanol uptake, or methanol assimilation into central metabolism. For methanol, which is highly volatile, metabolized relatively slowly, and shows high variability in replicate HPLC injections, measuring uptake rate by analyzing supernatant concentration presents a challenge. We therefore sought to analyze methanol assimilation rate using the high sensitivity of ^{13}C labeling. Up to this point, all labeling experiments had been conducted in media containing no additional carbon source beyond methanol, with unlabeled carbon presumably being derived primarily from glycogen mobilized in response to starvation conditions. Since ^{13}C labeling data provide relative fluxes, without accurate quantification of the metabolism of the unlabeled substrate, the la-

beling fractions seen in these experiments tell us very little about the absolute rate of methanol assimilation into central metabolism. We therefore revised the procedure for assessing methanol incorporation, by adding 50 mM ^{13}C methanol at the time of induction to cells actively growing on glucose, which allowed us to compare the rate of methanol incorporation to the measured rate of glucose uptake. Preliminary experiments were conducted to determine the time point for metabolite extraction, such that the cells had sufficient time to reach isotopic and metabolic steady state, but not so long that glucose was exhausted and growth ceased. The optimal time was found to be three hours post-induction. Additional plasmids carrying other variants of MDH, HPS, and PHI were generated for this experiment to enable a broader comparison of enzymes. FBP labeling results and the glucose consumption rates for each strain are shown in Figure 5-7. The glucose consumption rate varied from 2 mM hr^{-1}
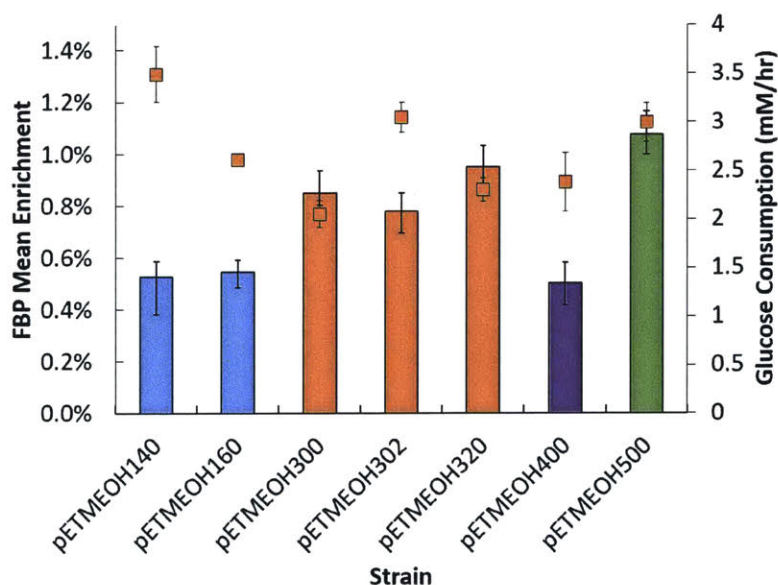
Figure 5-7: **FBP Labeling and Glucose Consumption in Various Methanol-Metabolizing Strains.** Bars indicate FBP labeling, and points indicate glucose consumption rate. Error bars are standard deviations from three biological replicates.

to 3.5 mM hr^{-1} (cell growth and glucose consumption were both linear after IPTG induction). The high variability of glucose consumption prevented robust comparison of the different pathway variants, but in all cases the mean isotopic enrichment of FBP was never more than 1%. This indicates roughly a 100-fold lower rate of assimilation

of methanol than glucose, or 20-35 μM hr^{-1}. This slow rate explains a) why the cells were unable to growth with methanol alone, as this rate is insufficient to generate even the ATP needed for maintenance, and b) why measuring uptake via HPLC has been challenging, given that the rate of evaporation in flasks (approximately 1 mM hr^{-1}) is 50-fold higher than the rate of methanol consumption. The strain carrying pETMEOH500, which expresses an evolved variant of $mdh2$ from $R.$ $eutropha$ [210], showed the highest FBP enrichment and close to the highest glucose consumption rate, suggesting this vector is a promising candidate for further work, and further suggesting that MDH may limit pathway flux, since this is the variant with highest in $vitro$ MDH activity (Table 5.9). Plasmids pETMEOH300, 302, and 320 all carry the $adhA$ from $C.$ $glutamicum$ [227], and differ only in the expression of HPS and PHI: pETMEOH300 and pETMEOH302 both contain genes from $M.$ $capsulatus.$ pETMEOH302 contains the optimized ribosome binding sites used in pETMEOH160 that leads to a roughly 9-fold increase in HPS activity compared to pETMEOH300. pETMEOH320 expresses HPS and PHI from $B.$ $methanolicus.$ The similarity of labeling between all these strains provides further support that the activity of HPS and PHI do not control pathway flux.

It is important to note that the ^{13}C labeling approach may underestimate the rate of methanol assimilation. These experiments were conducted in WT cells with the endogenous formaldehdye detoxification pathway present, in which formaldehyde is oxidixed to fomate and not incorporated into central metabolism. However, formate production was not detected by HPLC. Further, as described above, the difference in labeling between $\Delta gshB$ and WT cells carrying pETMEOH140 was small. In these experiments, FBP was analyzed because it is one of the most abundant metabolites in $E.$ $coli$ [219], whereas with our instrumentation F6P labeling was difficult to accurately measure. The possibility exists that a significant fraction of the F6P was consumed via the oxidative pentose phosphate pathway (oxPPP), in which case the labeled carbon from methanol would be lost as CO_2. To assess this possibility, we grew cells on 1-^{13}C-glucose and analyzed the labeling pattern of 3PG. Complete catabolism of 1-^{13}C-glucose via glycolysis results in retention of the label, leading to

50% of 3PG being singly labeled, whereas catabolism by oxPPP results in complete loss of label. In our strains, 3PG was almost exactly 50% labeled 5-8 (compared to a Δpgi control which can only metabolize glucose through oxPPP and showed almost completely unlabeled 3PG), indicating almost non-existent flux into the oxPPP and validating the choice of FBP as a good proxy for F6P labeling. In addition to
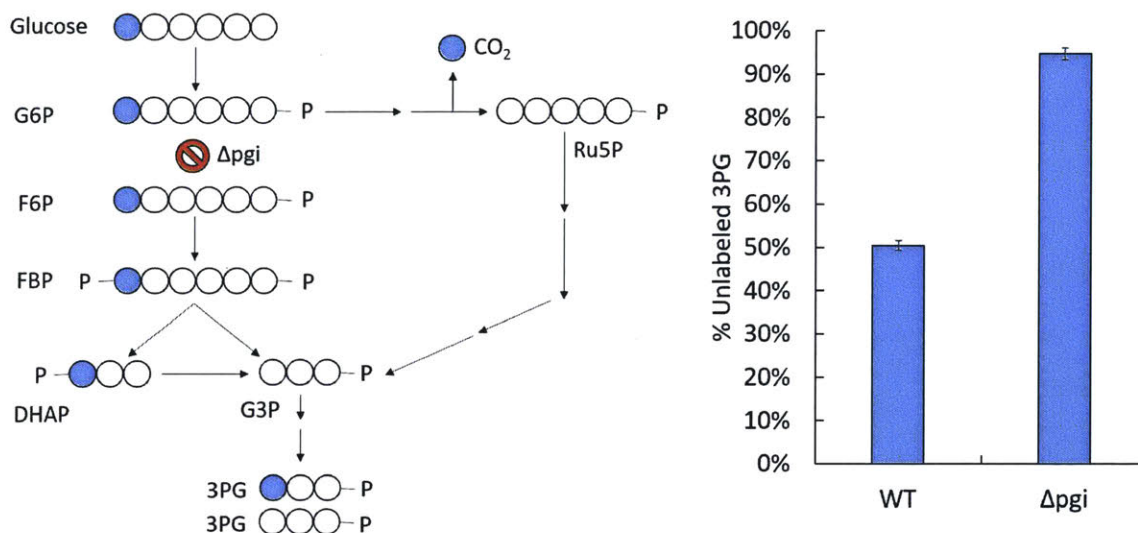
Figure 5-8: **Analysis of oxPPP Flux in Methanol Strains Using 1-^{13}C-Glucose.** Left: Schematic of glucose metabolism in WT and Δpgi cells. Blue circles represent ^{13}C, and white represent ^{12}C. In WT cells, glucose metabolized through glycolysis retains label, whereas glucose metabolized through oxPPP loses label. In Δpgi, glucose is metabolized primarily through oxPPP, resulting in almost complete lack of labeling of 3PG. Right: Experimental 3PG labeling in WT and Δpgi cells. WT cells show approximately 50% labeling, indicating glucose is almost exclusively catabolized through the EMP pathway. In the Δpgi control, 3PG is almost completely unlabeled, as most glucose is catabolized through oxPPP (with the remainder likely mobilized through the Entener-Doudoroff pathway).

asessing internal metabolite labeling, we also examined labeling in cells engineered to produce 3-hydroxybutyrate (3HB). These strains carried the methanol assimilation plasmids described above, in addition to the plasmid pAC3HB, which encodes the three genes required for S-3HB production, *phaA* from *Ralstonia eutropha* H16, *hbd* from *Clostridium acetobutylicum* 824 and *tesB* from *E. coli* MG1655 under control of the lac/T7 promoter system. 3HB production titer and isotopic enrichment was

measured 24 hours post-induction (Figure 5-9). All strains carrying methanol as-
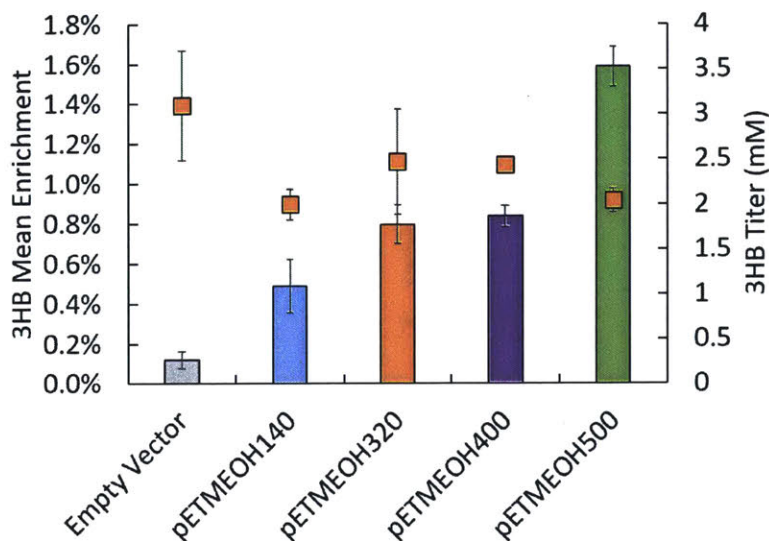
Figure 5-9: **Analysis of 3HB Production and Labeling in Methanol Strains.** Bars indicate 3HB mean isotopic enrichment, as measured by LC-MS/MS after 2-NPH derivatization. Orange squares indicate 3HB titer measured by HPLC-RID. Error bars are standard deviations of three biological replicates.

similation plasmids produced similar titers of 3HB (2-2.5 mM). The empty vector control produced a slightly higher amount (3 mM), probably due to the lower burden of heterologous expression in this strain. Mean isotopic enrichment was the highest with methanol plasmid pETMEOH500, at 1.6%. As expected, no ^{13}C incorporation was seen in the empty vector control. These results are in good agreement with the intracellular labeling, and indicate pETMEOH500 as the most promising vector for enabling methanol assimilation. They also demonstrate the ability to channel methanol into extracellular products, and not just central carbon metabolites. The major conclusion from both the intracellular and extracellular carbon labeling is that, under these conditions, methanol is slowly consumed and contributes very little of the total cellular carbon.

Theoretical Analysis to Guide Experimental Design

Having identified the slow incorporation of methanol into central metabolism as a limiting factor, we next sought to understand the mechanism behind this limitation.

171

Only by understanding the problem can we design effective engineering solutions to improve pathway flux. In principle, any of the three heterologous enzymatic steps could be rate-limiting. The specific activity of MDH2 *in vitro* is two orders of magnitude lower than HPS or PHI (Tabel 5.9), making this a likely limitation. However,

Enzyme	Specific Activity (U mg^{-1})	Reference
B. methanolicus MDH2	0.09	[73]
C. glutamicum AdhA	0.29	[206]
C. necator MDH2 4-1	0.32	[206]
M. capsulatus HPS	69	[222]
M. capsulatus PHI	1560	[222]

Table 5.9: Kinetic Comparison of Methanol Pathway Enzymes

the *in vitro* activity of the heterologously expressed MDH is similar to that of the same enzyme in lysates of the native host [214], so estimations of methanol flux based on the kinetic parameters of MDH must be treated with caution. In addition, the NAD-dependent oxidation of methanol is thermodynamically unfavorable at standard conditions, with an estimated[3] $\Delta G^{0'}$ of $+34.2$ kJ mol^{-1}. In order to sustain net forward flux, ΔG must be <0, per:

$$\Delta G = -RTln\left(\frac{J^+}{J^-}\right) \tag{5.2}$$

where J^+ is the forward flux, and J^- the reverse flux. This can be achieved by modulating the intracellular concentrations of the reactants and products, according to:

$$\Delta G = \Delta G^{0'} + RTlnQ \tag{5.3}$$

where Q is the reaction quotient, defined as:

$$Q = \frac{[CH_2O][NADH]}{[CH_3OH][NAD^+]} \tag{5.4}$$

Functionally, this means that the concentration of formaldehyde must be kept low, and the NAD/NADH ratio high, otherwise insufficient driving force will lead to slow

---

[3]Value taken from eQuilibrator [228] software: http://equilibrator.weizmann.ac.il/

172

methanol oxidation, regardless of the kinetic capacity of the enzyme. This concept is captured quantitatively in the Flux-Force Efficacy (FFE) parameter [229], defined as:

$$FFE = \frac{e^{\frac{-\Delta G}{RT}} - 1}{e^{\frac{-\Delta G}{RT}} + 1} \tag{5.5}$$

This parameter can take on values from 1 (fully irreversible, forward direction), to -1 (fully irreversible, reverse direction), and describes the thermodynamic driving force behind a given reaction. To understand the effect of the build-up of formaldehdye on the driving force of the MDH-catalyzed reaction, in Figure 5-10 the FFE is plotted against the concentration of formaldehyde, for a given NAD/NADH ratio, methanol concentration, and temperature. Various estimates for the NAD/NADH ratio in growing cells have been made, and the most reliable reported is 31.3 [219]. The entropy of the reaction (ΔS^0) is 0.22 kJ mol^{-1} K^{-1} [4], thus the favorability of the forward direction increases with increasing temperature, according to:

$$\Delta G = \Delta H - T\Delta S^0 \tag{5.6}$$

From Figure 5-10, it is clear that, with 50 mM methanol, the unfavorable thermodynamics require a very low formaldehyde concentration of 2 μM to enable forward flux. The situation is slightly improved at 500 mM methanol, where net flux is forward until the formaldehyde concentration reaches 22 μM at 37°C. These results highlight the necessity for a very high flux through the HPS reaction to keep the formaldehyde concentration low. Increasing the temperature leads to substantial increases in the permissible formaldehyde concentration, representing a benefit for the thermophilic *B. methanolicus*. Still, at 50 mM methanol (a concentration where this organism grows well), the formaldehyde concentration must remain below 10 μM. This could explain why the HPS from *B. methanolicus* evolved to have a much lower K_m than that for *M. capsulatus* (Table 5.7), where methanol oxidation is coupled to the highly thermodynamically favorable reduction of pyrroquinoline quinone (PQQ, $\Delta G^{0'}$=-24.8 kJ mol^{-1}). One could argue that the comparatively low temperature optimum for

---

[4]Details of this calculation are provided in Appendix A.5

173
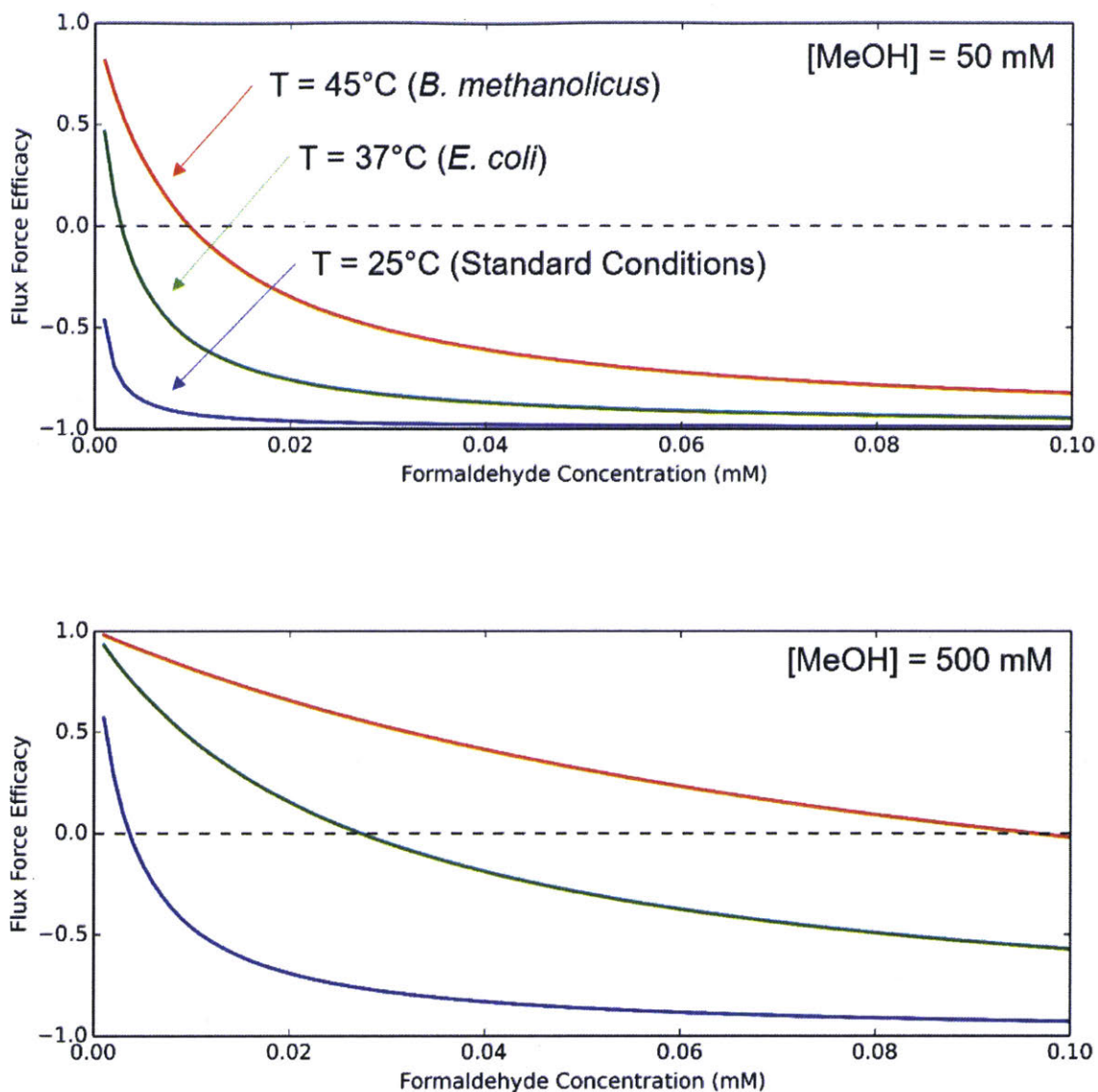
Figure 5-10: **Flux-Force Efficacy as a Function of Formaldehyde Concentration.** Flux-force efficacy (FFE) is plotted against formaldehyde concentration for three different temperatures. Red: T = 45°C, representative of the growth temperature of *B. methanolicus*. Green: T = 37°C, representing *E. coli*. Blue: T = 25°C, standard conditions, corresponding to $\Delta G^{0'} = +34.2$ kJ mol^{-1}. The NAD/NADH ratio was set at 31.3, according to [219]

E. coli poses an insurmountable thermodynamic barrier because of the reduced driving force available at this temperature compared to the optimal temperature for *B. methanolicus*. However, it has been demonstrated that *B. methanolicus* can grow at 37°C, albeit with a lower specific growth rate (0.14 hr^{-1} vs 0.4 hr^{-1}) [76]. The message from this analysis is that, even though MDH kinetics are poor compared to

those of HPS, insufficient flux through the HPS-catalyzed reaction could lead to a thermodynamic limitation. As the rate of this reaction is dependent not only on HPS kinetics and formaldehyde concentration, but also on the concentration of the aldol partner ribulose 5-phosphate (Ru5P), high activity of HPS alone may be insufficient to drive the necessary flux if this compound becomes limiting. This concept of thermodynamic vs kinetic control of flux is illustrated in Figure 5-11, in which the effect of either a 2-fold increase in V_{max} (dashed line) or a 2-fold decrease in formaldehyde concentration (blue, green, red lines) on the rate of methanol oxidation is compared at different values of ΔG. In the kinetically controlled regime, where the system is far from equilibrium, flux is insensitive to changes in the formaldehyde concentration. Close to equilibrium, in the thermodynamically controlled regime, a 2-fold reduction in formaldehyde concentration can easily lead to a 10-fold flux improvement. The three curves represent hypothetical enzymes with different K_m values for formaldehyde, moving from low (blue) to high (red). When the K_m is close to the concentration of formaldehyde, the effect of reducing formaldehyde concentration has both a thermodynamic effect and a kinetic effect, by reducing competition with methanol for the enzyme active site. This plot illustrates why it is critical to understand whether flux is under kinetic or thermodynamic control: In the kinetic regime, engineering efforts designed to improve MDH kinetics will be most effective. In the thermodynamic regime, approaches designed to reduce formaldehyde (or NADH) concentration will be more effective. Thus far we have identified MDH kinetics, MDH thermodynamics, HPS kinetics, and Ru5P concentration as potential limitations, while ignoring HPS reversibility, and PHI kinetics and reversibility. As shown in Table 5.9, PHI is 100-fold more active than HPS, making it highly unlikely that this enzyme could be rate-limiting. The HPS and PHI reactions are also significantly forward driven, with a K_m of 2.5×10^4 and 188, respectively. However, there is evidence that methanogens use the HPS-catalyzed reaction in the reverse direction as a means to produce pentose phosphates for nucleotide biosynthesis [232, 233]. To examine HPS reversibility, we measured formaldehyde labeling in resting cells experiments with ^{13}C methanol, and found greater than 95% M+1 enrichment, showing there is very little HPS activity
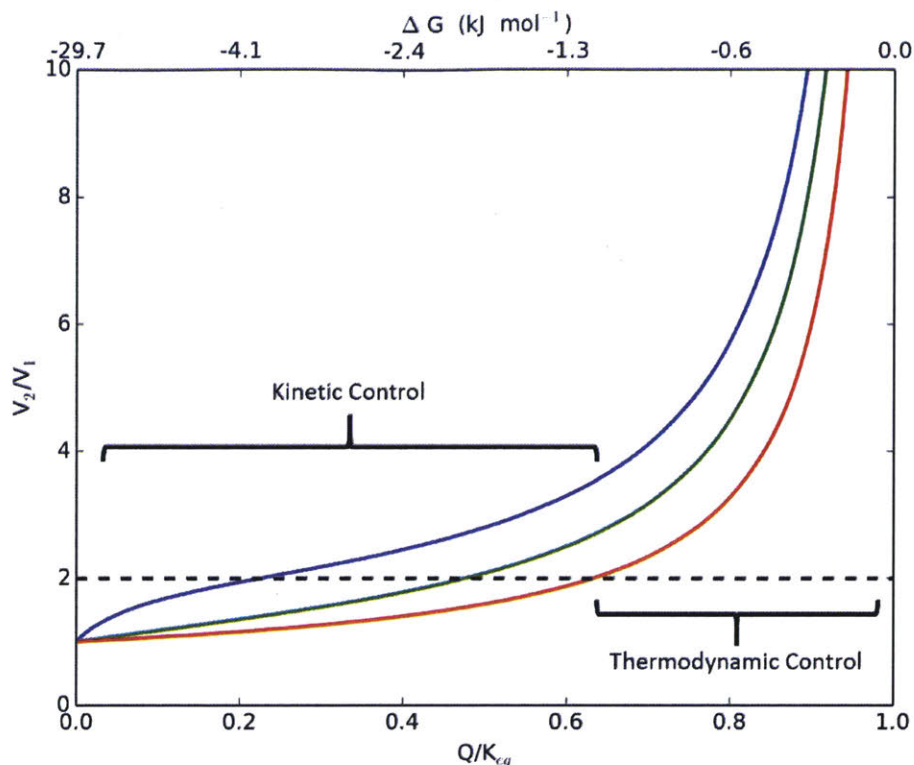
Figure 5-11: **Kinetic and Thermodynamic Control of MDH Flux.** The effect of doubling V_{max} on MDH velocity (dashed line) and the effect of reducing formaldehyde concentration by a factor of 2 (solid lines) on MDH velocity are compared as a function of the position relative to equilibrium (Q/K_{eq}) or ΔG. MDH was modeled using Michaelis-Menten kinetics for two noncompeting substrate-product couples as described in Appendix A.7 and [230]. K_{eq} was calculated from a $\Delta G^{0'}$ of -34.2 kJ mol^{-1}. MDH was assumed to have the following K_m parameters: NAD, 7.6 mM; NADH, 1 mM; MeOH, 25 mM; Formaldeyde, 0.001 mM (Blue), 0.01 mM (Green) or 0.1 mM (Red). Intracellular concentrations were as follows: MeOH, 500 mM; NAD, 2.6 mM; NADH, 0.083 mM [231]. Formaldehyde was varied from 0.0001 mM to 0.027 mM to provide the Q/K_{eq} values shown on the x-axis.

in the cleavage direction (Figure 5-12). In light of the irreversibility of HPS, as long as there is sufficient PHI activity, forward flux is not dependent on the reversibility of this isomerization. In designing experiments to probe pathway limitations, we therefore limit ourselves to the four potential limitations listed above.
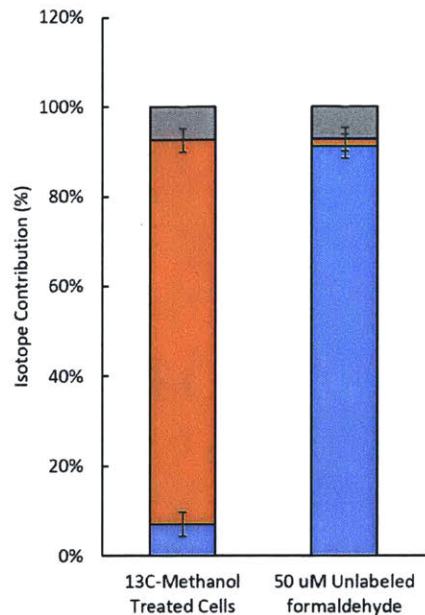
Figure 5-12: **Labeling of Formaldehyde in Resting Cells with ^{13}C Methanol.**
Left: Resting cells of MG1655(DE3) ΔfrmA containing pETMEOH500, treated with
250 mM ^{13}C methanol. Right: 50 μM standard of unlabeled formaldehyde. Formalde-
hyde labeling was measured by LC-MS/MS after derivatization as described in the
Methods. Blue: M+0/M+0, Orange: M+1/M+0, Gray: M+1/M+1, where the first
number is the Q1 (parent) mass, and the second number the Q3 (daughter) mass.

Ru5P Concentration Limits Formaldehyde Assimilation in Resting Cells

Having identified several possible pathway limitations, we set out to experimen-
tally validate them as targets. We started with the concentration of ribulose 5-
phosphate. Insufficient availability of Ru5P, the co-substrate of formaldehyde for
the HPS-catalyzed reaction, would be predicted to lead to increased build-up of
formaldehyde. Previous work [214] established that formaldehyde could be mea-
sured in *E. coli* strains engineered for methanol metabolism when the glutathione-
dependent formaldehyde dehydrogenase encoded by *frmA* was deleted. We therefore
generated this knockout in our MG1655(DE3) strain, and assayed formaldehyde levels
under various conditions. As expected, expression of just the methanol dehydroge-
nase (MDH) led to steady levels of approximately 55 uM formaldehyde after addition
of 250 mM methanol in resting cells (Figure 5-13). Surprisingly, this level was not
significantly reduced in cells expressing the full pathway (including HPS and PHI),
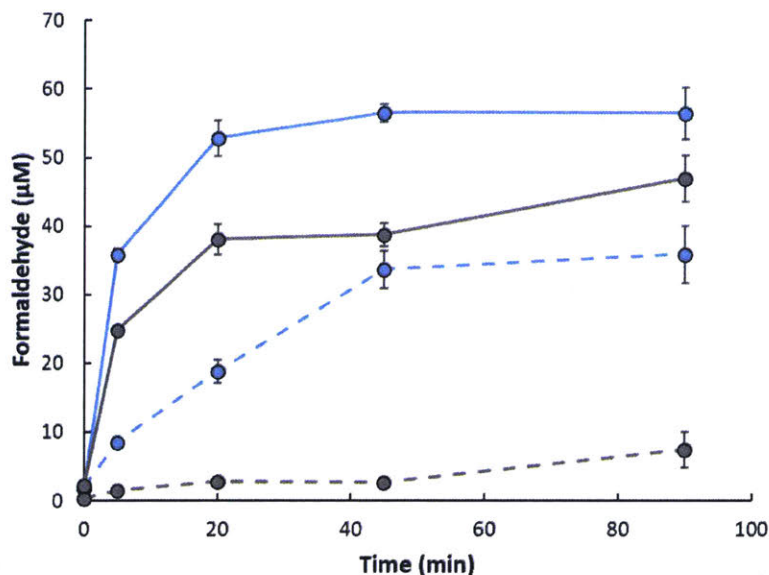
Figure 5-13: **Formaldehyde Concentration in Strains Carrying Methanol Pathway Plasmids.** Formaldehyde was determined by NASH assay in MG1655(DE3) *frmA* cells carrying various plasmids for methanol assimilation and with various substrates: Blue: *C. necator* mdh2 4-1 only. Gray: *C. necator* mdh2 4-1, HPS and PHI from *B. methanolicus*. Solid lines: No additional substrate. Dashed lines: 6 g L^{-1} xylose.

despite high levels of enzyme activity. We hypothesized that Ru5P levels might be low under these conditions. Addition of xylose to the cells 15 minutes before methanol addition resulted in much lower formaldehyde levels of 5 uM. Analysis of intracellular metabolites under these conditions showed the pool size of Ru5P was roughly 8-fold higher in the presence of xylose than in the untreated cells, in agreement with our hypothesis (Figure 5-14). Since a methanol-consuming strain cannot rely on xylose as a precursor, we sought to identify enzymatic steps whose up- or down-regulation could lead to increased flux through the ribulose monophosphate (RuMP) cycle and thus maintain high concentrations of Ru5P. To do this, we performed FBA using a model amended with the methanol assimilation pathway to compare the flux profiles during growth on glucose, ribose, and methanol. Glucose was chosen because resting cells survive by consuming glycogen, metabolizing the storage compound in the same manner as glucose. Ribose was chosen because cells grown on pentoses express more highly the enzymes of the non-oxidative pentose phosphate pathway,
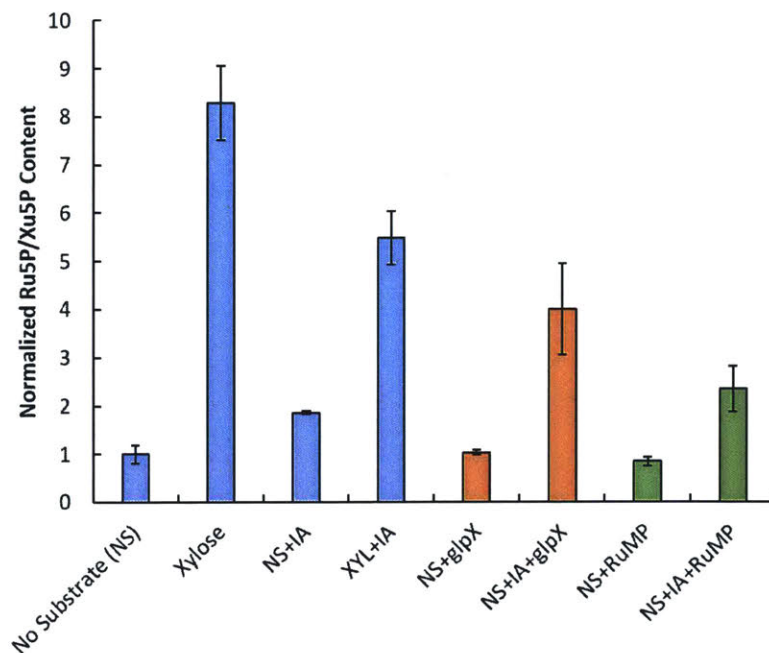
Figure 5-14: **Normalized Ru5P Concentration in Various Strains.** Ru5P/Xu5P were measured by LC-MS/MS after extraction of metabolites. ^{13}C standards spiked into the extraction solution were used to correct for sample degradation, and all concentrations are normalized to the concentration in the No-Substrate (NS) control. The peaks for Ru5P and Xu5P could not be fully resolved, thus we report the total concentration of these two compounds.

which are required for the generation of Ru5P. The analysis revealed two nodes, or branch points in metabolism, where flux varied significantly between methanol and the other substrates: Fructose-6-phosphate (F6P) and Glyceraldehyde-3-phosphate (G3P) (Figure 5-15). During growth on glucose and ribose, the majority of F6P is consumed through glycolysis, with 94% and 98% of the flux going through phosphofructokinase (PFK), respectively. By contrast, during growth on methanol, only 31% of the flux is predicted to be carried by PFK, with the majority diverted into the RuMP pathway via transaldolase (TAL) and transketolase (TKT). Catalyzing one of the few irreversible steps of glycolysis, PFK is a highly regulated enzyme, subject to allosteric activation by AMP, and inhibition by ATP and phosphoenolpyruvate (PEP). The K_m for F6P for PFK1 is 0.16 mM, compared to 1.1 mM for TKT. Similarly, during growth on glucose and ribose, 97% and 82% of the G3P flux is through glyceraldehyde-3-phosphate dehydrogenase (GAPDH), compared to 29% for growth
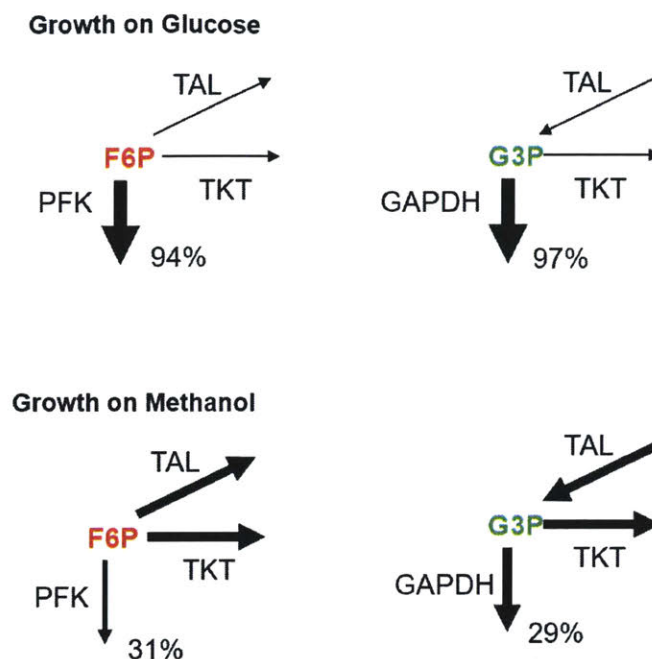
Figure 5-15: **FBA Comparison of Fluxes at the G3P and F6P Nodes in Glucose- and Methanol-Grown Cells.** Thickness of arrows represents flux through enzymatic step. TAL: Transaldoase, TKT: Transketolase, PFK: Phosphofructokinase, GAPDH: Glyceraldehyde-3-phosphate dehydrogenase.

on methanol. To probe whether reducing flux through GAPDH could increase the concentration of Ru5P, we repeated the resting cell experiment, but treated the cells with 1 mM iodoacetate (IA), a specific inhibitor of GAPDH, prior to addition of methanol. Controls were included to ensure the IA had no effect on either the MDH or the HPS and PHI. Cells treated with IA had significantly increased pool sizes of some upper glycolytic intermediates, notably DHAP, G3P and FBP (Figure 5-16), but only limited increases in pentose phosphates (Figure 5-14). PEP and other lower glycolytic intermediates dropped to almost undetectable levels, demonstrating the efficacy of the inhibition. Despite these changes, the concentration of formaldehyde remained essentially unchanged (Figure 5-17). We also tested whether overexpression of the RuMP pathway enzymes could enable Ru5P regeneration. The four *E. coli* genes *rpe*, *tktA*, *talB* and *rpiA* were assembled into a synthetic operon and cloned into the pACYC-Duet1 vector, which is compatible with the pET plasmids for methanol pathway gene expression. Cells co-transformed with these plasmids were measured
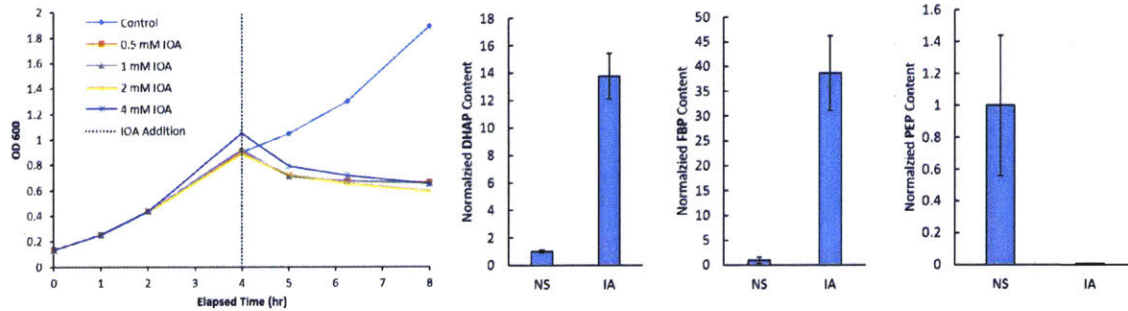
Figure 5-16: **Effect of Iodoacetate Treatment on Cell Growth and Metabolite Concentration.** Left figure: Growth of MG1655(DE3) with addition of various concentrations of Iodoacetate (IA) at 4 hours. Untreated cells continued exponential growth, whereas even 0.5 mM IA was sufficient to completely inhibit growth. Right three figures: Normalized metabolite concentrations in cells with no treatment, or 1 mM IA. DHAP: Dihydroxyacetone phosphate, FBP: Fructose 1,6-Bisphosphate, PEP: Phosphoenolpyruvate. NS = no carbon substrate, IA = no carbon substrate, 1 mM iodoacetate.
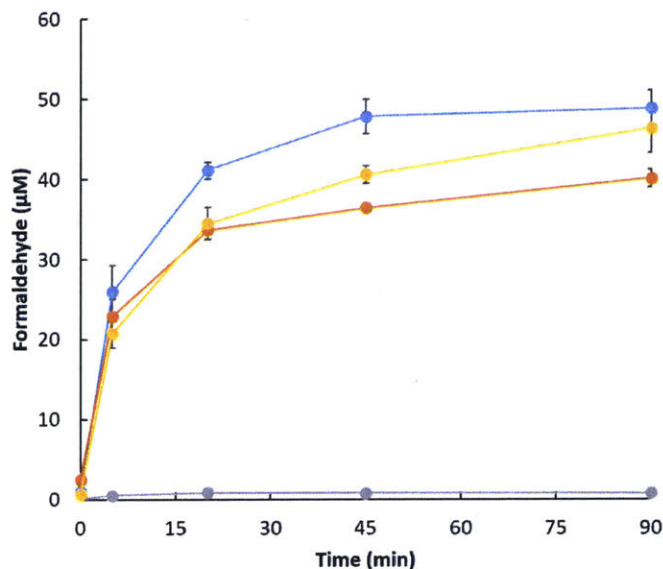
Figure 5-17: **Effect of Iodoacetate Treatment on Methanol Incorporation.** Time course of formaldehyde concentration. Blue: MDH Only, with 1 mM IA. Yellow: Full pathway, with 1 mM IA. Gray: Full Pathway, with 1 mM IA and 6 g L^{-1} xylose. Orange: Full pathway with no IA.

for Ru5P concentration and formaldehyde assimilation, with or without iodoacetate (Figure 5-18). Overexpression of the RuMP without iodoacetate inhibition made no change to the Ru5P concentration compared to the control (Figure 5-14), and

only slightly reduced the formaldehyde concentration, from 56 to 48 μM. In the presence of iodoacetate, the Ru5P concentration was no different with or without RuMP pathway overexpression (Figure 5-14). The steady state formaldehyde concentration was reduced to 40 μM, a modest decrease of 30%. Since the conversion of F6P to
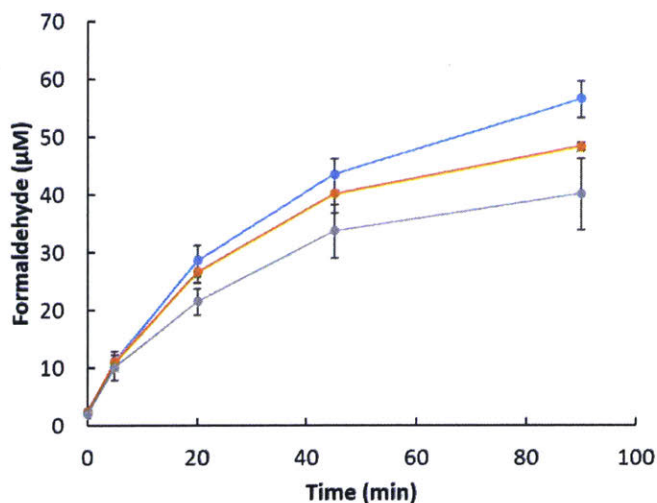
Figure 5-18: **Effect of RuMP Pathway Overexpression on Methanol Incorporation.** Time course of formaldehyde concentration. Blue: pETMEOH500. Orange: pETMEOH500 with pACRuMP. Gray: pETMEOH500 with pACRuMP and 1 mM IA.

FBP by PFK is irreversible, we reasoned that, in the absence of a gluconeogenic FBPase, increases in DHAP, G3P and FBP would be unlikely to lead to increases in F6P, one of the entrance metabolites of the RuMP pathway. We therefore explored the co-expression of the minor FBPase encoded by *glpX* to attempt to equilibrate F6P and FBP. The choice of this FBPase was motivated by its lack of regulation by AMP, compared to the major *E. coli* FBPase, and lack of activity on sedoheptulose bisphosphate [234]. Gratifyingly, cells overexpressing FBPase and treated with IA showed dramatic 7-fold increase in F6P, and a 4-fold increase in Ru5P concentration (Figure 5-14), and a significant reduction of the formaldehyde level in the absence of any substrate (Figure 5-19) of 72%. The labeling in glycolytic intermediates was also significantly greater than in the control strain without *glpX*, as exemplified by FBP (Figure 5-19), with a mean isotopic enrichment of 27.5$\pm$2.2%, compared to 15.2$\pm$0.5%.
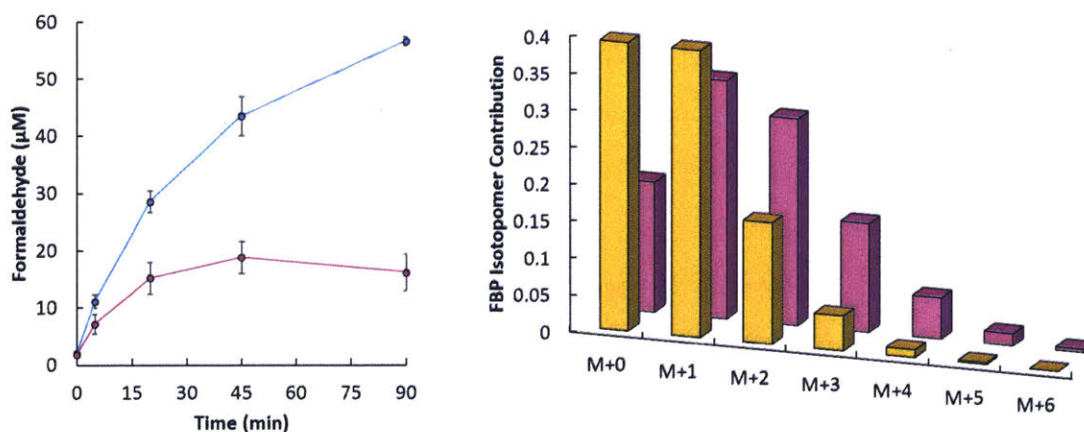
Figure 5-19: **Effect of glpX Overexpression on Methanol Assimilation.** Left: Time course of formaldehyde concentration. Blue: Full pathway, with pACglpX. Pink: Full pathway, with pACglpX and 1 mM IA. Right: FBP isotopomer distribution. Pink: Full pathway with pACglpX and 1 mM IA. Yellow: Full pathway with 1 mM IA.

Given the dramatic effect of *glpX* overexpression, and the fact that homologs of GlpX act as SBPases in some methylotrophic and photosynthetic organisms, we sought additional confirmation that *E. coli glpX* was not an SBPase. To test this, we purified a HIS-tagged version of the enzyme, and examined its activity on SBP in enzyme assays. Since SBP is not commercially available, this substrate was prepared enzymatically by the aldol reaction between E4P and DHAP catalyzed by purified *E. coli* FbaA. Upon incubation of E4P and DHAP with FbaA and GlpX, a clear S7P peak became visible by LC-MS/MS, along with complete depletion of DHAP (Figure 5-20). These data show that, in contrast to previous studies [234], *glpX* from *E. coli* is active as an SBPase. Further examination of metabolite levels revealed that, upon addition of iodoacetate, levels of intracellular SBP in cells not expressing *glpX* rose by a factor of 5.5 (Figure 5-21). Expression of *glpX* under the same conditions restored SBP to baseline levels. Further, in cells expressing *glpX*, levels of S7P were dramatically elevated regardless of whether iodoacetate was added or not (Figure 5-21). These results provide *in vivo* evidence for the activation of the SBPase variant of the RuMP pathway in cells treated with iodoacetate and expressing *glpX*. Iodoacetate treatment leads to increases in G3P concentration, and *glpX* expression

increases S7P. Together, we hypothesize that this provides increased driving force for the transketolase reaction that produces R5P and Xu5P, which are in turn converted to Ru5P. At this point, we cannot conclude how much of the increase in Ru5P concentration is due to this pathway compared to the originally hypothesized FBP-F6P equilibration. A combination of transaldolase deletion (which would inactivated the transaldolase variant of the RuMP pathway) and well designed labeling experiments would answer this question.
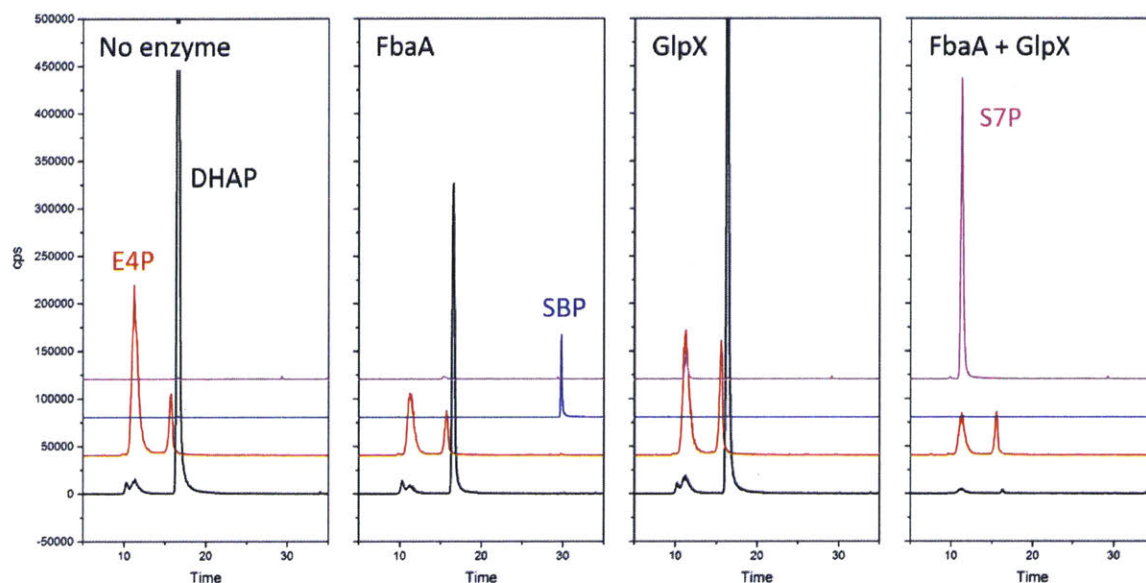
Figure 5-20: **LC-MS/MS Analysis of *E. coli* GlpX SBPase Activity.** Chromatograms of enzyme assay mixtures with various enzymes and substrates. Black trace - DHAP, Red trace - E4P, Blue Trace - SBP, Pink Trace - S7P.

Highly Active Formaldehyde Assimilation Ensures Methanol Oxidation is not Thermodynamically Limited

In the theoretical analysis in section 5.4.2, it was shown that steady state formaldehyde accumulation above a critical threshold concentration of approximately 20 μM would dissipate the driving force for methanol oxdiation via MDH. However, in the experiments described above, concentrations in the range of 50 μM were routinely measured. The discrepancy between these two values is most likely due to error associated with the estimation of $\Delta G^{0'}$, which is derived from group contribution theory
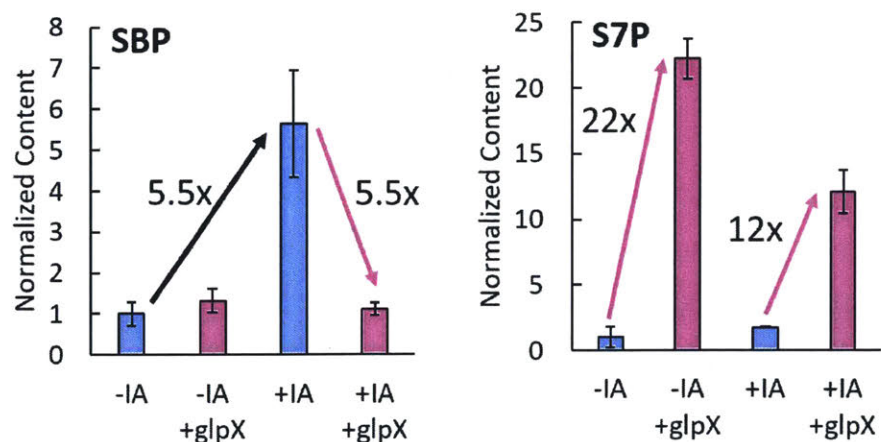
Figure 5-21: **Effect of GlpX on SBP and S7P Concentrations.** IA = iodoacetate. Error bars represent standard deviation from triplicate biological replicates.

and has not been measured experimentally [228]. For example, a reduction in $\Delta G^{0'}$ of 2 kJ mol^{-1} increases the critical formaldehyde concentration to 60 μM at 500 mM methanol. Regardless, the proximity of the measured formaldehyde concentration to equilibrium suggests that methanol oxidation in our strain may be thermodynamically constrained.

To probe this possibility experimentally, we generated a plasmid for the overexpression of the NADH oxidase from *S. pyogenes* (pACnox). This enzyme catalyzes the oxygen-dependent oxidation of NADH:

$$H^+ + NADH + \frac{1}{2}O_2 \rightarrow H_2O + NAD^+ \tag{5.7}$$
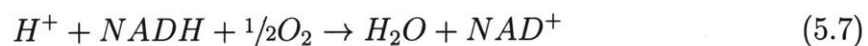
and has been used extensively to modulate the NAD/NADH ratio in engineered *E. coli* strains for a number of applications [221, 235, 236]. We hypothesized that, if methanol oxidation were thermodynamically constrained, increasing the NAD/NADH ratio by co-expression of Nox with MDH would lead to higher levels of formaldehyde at steady state. To minimize variation in MDH expression between the experimental strain and the control strain due to the burden of NOX overexpression, we generated a plasmid carrying an active site mutant of NOX (pACnoxM3) to use in the control. Assays of NOX activity in crude lysates of $\Delta frmA$ cells carrying pETCNmdh/pACnox and

pETCNmdh/pACnoxM3 showed that NADH oxidation was significantly elevated in the former strain (1.2 ± 0.2 U mg^{-1}), and virtually undetectable in the latter (Figure 5-22, left). Formaldehyde production in these strains after treatment with 250 mM methanol was measured as before. As shown in Figure 5-22 (right), the steady state concentration was approximately 32% higher in the strain expressing active NOX, suggesting that, in the absence of the formaldehyde assimilation pathway, methanol oxidation is under thermodynamic control, and the so-called steady state formaldehyde concentration is more likely the equilibrium formaldehyde concentration.
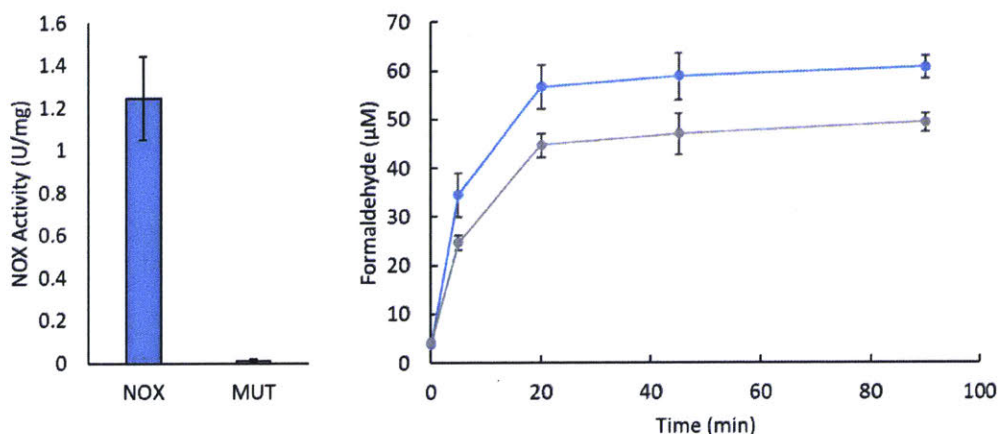
Figure 5-22: **NOX Activity and Impact on Formaldehyde Concentration.** Left: Measurements of NOX activity. NADH oxidation was measured at 37°C in 50 mM Sodium Phosphate Buffer (pH 7.5). Bars are mean oxidation rate of three biological replicates after subtraction of background oxidation from empty vector control strain, and error bars represent one standard deviation. Right: Formaldehyde timecourse in NOX$^+$ and NOX^{mut} cells. Blue: NOX$^+$. Gray: NOX^{mut}. Error bars represent standard deviation from three biological replicates.

Having shown that in the absence of a formaldehyde assimilation pathway, formaldehyde rapidly reaches its equilibrium concentration, we wanted to understand to what extent this would reduce methanol flux in strains expressing the full pathway. The most direct way to answer this would be to calculate the ΔG under these conditions. This would require an accurate value of $\Delta G^{0'}$, and knowing the intracellular concentrations of all the reactants and products. While we can measure NAD and NADH, accurately measuring methanol and formaldehyde is more difficult. We can reasonably assume that formaldehyde diffuses freely across the membrane, given that we can

measure its appearance in the supernatant very shortly after its production begins intracellularly. However, it is unlikely that the methanol concentration in the supernatant is the same as that intracellularly - at 250 mM, methanol would be 2.5-fold more concentrated than glutamate, the metabolite with highest measured concentration in *E. coli* [231]. To avoid the uncertainty in $\Delta G^{0'}$ and methanol concentration in the calculation, we can use the differences in the concentrations of the relevant measurable metabolites between cells expressing only MDH, and cells expressing the full pathway, treated in exactly the same way, to determine ΔG as follows:

$$\Delta G_1 = \Delta G^{0'} + RTlnQ_1 \tag{5.8}$$

$$\Delta G_2 = \Delta G^{0'} + RTlnQ_2 \tag{5.9}$$

where the subscript 1 refers to MDH-only, and subscript 2 refers to the full pathway. If we assume that the system is at equilibrium in the MDH-only case, then $\Delta G_1 = 0$, then $\Delta G^{0'} = -RTlnQ_1$, and

$$\Delta G_2 = -RTlnQ_1 + RTlnQ_2 = RTln\left(\frac{Q_2}{Q_1}\right) \tag{5.10}$$

Since the methanol concentration is the same for both cases, we have

$$\Delta G_2 = RTln\left[\frac{\left(\frac{[NADH]}{[NAD]}\right)_2 [CH_2O]_2}{\left(\frac{[NADH]}{[NAD]}\right)_1 [CH_2O]_1}\right] \tag{5.11}$$

If the NAD/NADH ratio is the same in both cases, this reduces to

$$\Delta G_2 = RTln\left[\frac{[CH_2O]_2}{[CH_2O]_1}\right] \tag{5.12}$$

In reality, the system is likely not exactly at equilibrium with MDH only, such that

$\Delta G_1 \leq 0$. Propagating this change through the analysis leads to

$$\Delta G_2 \leq RT ln \left[\frac{\left(\frac{[NADH]}{[NAD]} \right)_2 [CH_2O]_2}{\left(\frac{[NADH]}{[NAD]} \right)_1 [CH_2O]_1} \right] \tag{5.13}$$

and

$$\Delta G_2 \leq RT ln \left[\frac{[CH_2O]_2}{[CH_2O]_1} \right] \tag{5.14}$$

Thus, by measuring the difference in concentration of formaldehyde, and the NAD/NADH ratio with and without HPS and PHI, we can place an upper bound on ΔG_2.

As shown in Figure 5-23, the NAD/NADH was unchanged between cells expressing MDH and cells expressing the full pathway, allowing the use of equation 5.14 for ΔG calculation. This likely reflects the limited contribution of methanol oxidation to the total NADH pool due to the comparatively low activity of the enzyme. Using this method, the ΔG and reversibility of the MDH reaction was determined for a variety of different conditions described in the previous section for the regeneration of Ru5P, by measuring the formaldehyde concentration in cells expressing MDH or the full pathway. The results of the analysis are shown in Figure 5-23. Addition of xylose, which was previously shown to reduce formaldehyde concentration the most, resulted in a maximum ΔG of -4.0 kJ mol^{-1}, or a minimum ratio of forward to reverse flux of 4.8. Thus, under these conditions, the MDH reaction is significantly forward driven, using at least 83% of the kinetic capacity of the enzyme. The use of FBPase encoded by *glpX* in resting cells led to minimum flux ratio of 2.8, or the use of 73% of the kinetic capacity. Though slightly less forward driven than the xylose case, this was achieved without the addition of an additional substrate, which is important for the eventual sustained growth of *E. coli* on methanol. In cells with no substrate, not expressing *glpX*, the minimum flux ratio is 1.2. Because it is uncertain that formaldehyde production was limited by equilibrium in the MDH-only, we cannot necessarily conclude that there is very litlte forward flux, but it seems likely given the effect of NOX expression under these conditions. The critical finding here is that the primary indicator of forward flux in this pathway is the concentration of Ru5P.

When this compound is adequately regenerated, we can be certain that the pathway is forward driven, whereas at low concentration the pathway is likely thermodynamically controlled.

This assessment appears to contradict the findings of Price at al, who showed increased methanol consumption by co-localizing MDH with HPS in a strategy designed to rapidly consume formaldehyde to ensure a strong driving force [207]. In that study, however, the authors used the HPS from *Mycobacterium gastrii*. This enzyme has a very high K_m for formaldehyde (Table 5.7). In our hands (Figure 5-3), and in *Corynebacterium glutamicum* [205], this results in almost no detectable ^{13}C label incorporation in strains expressing this enzyme. We suggest that the reason for the rate improvement in the co-localization study is actually a lowering of the effective K_m of the HPS, a known phenomenon in substrate-channeling enzyme complexes [237, 238].
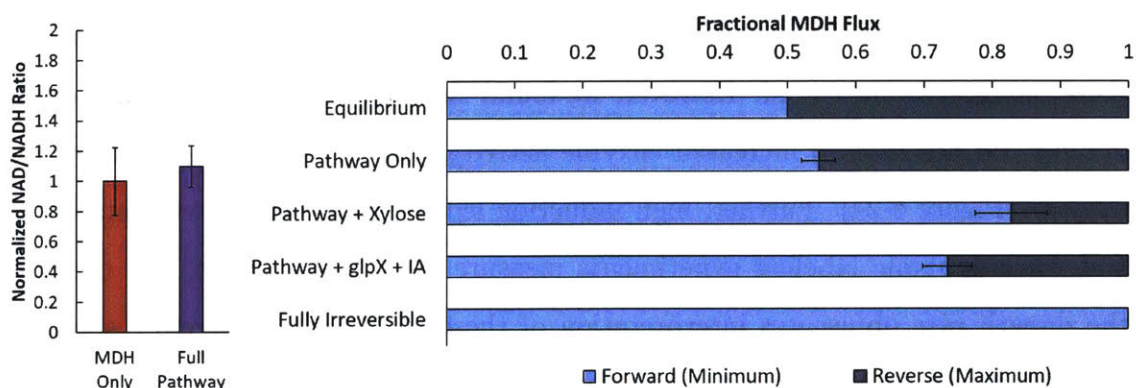
Figure 5-23: **Calculation of *in vivo* MDH Reversibility Under Various Conditions.** Left: Measurement of NAD/NADH ratio in resting cells MG1655(DE3) *frmA*, normalized to the MDH-only strain. Right: Fractional MDH Flux in forward and reverse directions. Estimates for forward flux represent minimum bounds, and estimates for reverse flux are maxima. Error bars represent standard deviation of three biological replicates.

Kinetic Isotope Effect Experiments Highlight Flux Control Exerted by MDH

We next asked whether MDH activity limits incorporation of methanol into metabolism, under conditions that are not thermodynamically limiting. Calculations based on the *in vitro* measured rates of this enzyme would suggest this to be the case, but there is so far no *in vivo* evidence to support this claim. Given that MDH supports fast growth on methanol in *B. methanolicus* despite the reported kinetics, this question is important to address. The typical method to probe this would be to vary the level of MDH expression, and assess the impact on the rate of methanol consumption, the level of formaldehyde produced, and the incorporation of ^{13}C label into central metabolic intermediates. As mentioned above, the comparatively slow rate of methanol consumption compared to evaporation, coupled with high variance in HPLC-based quantification of methanol, has precluded accurate assessment of the rate of methanol consumption directly. Formaldehyde production and intracellular labeling can be accurately quantified, so we designed an experiment where the level of MDH expression would be controlled by the amount of inducer (IPTG), and both of these parameters measured. However, we found that, above 50 μM IPTG, increased inducer did not lead to increased enzyme activity, due to the poor solubility of the enzyme (Figure 5-24). Further, high induction led to significant growth retardation, and since dynamic labeling experiments measure comparative rates, any change in the growth rate obfuscates interpretation of these results. We therefore sought an alternative way to modulate the activity of MDH without affecting growth.

Kinetic isotope effects (KIEs), where the substitution of a heavy atom for a light atom changes the rate of an enzymatic reaction, have been well documented for dehydrogenases where the relvant hydrogen atom is substituted for a deuterium atom [239]. We reasoned that MDH would also exhibit a KIE if the methyl hydrogens were replaced with deuterium (Figure 5-25), and further reasoned that using various mol fractions of protonated methanol (CH_3OH) and deuterated methanol (CD_3OD) would allow us to titrate the activity of MDH without changing the expression level

190
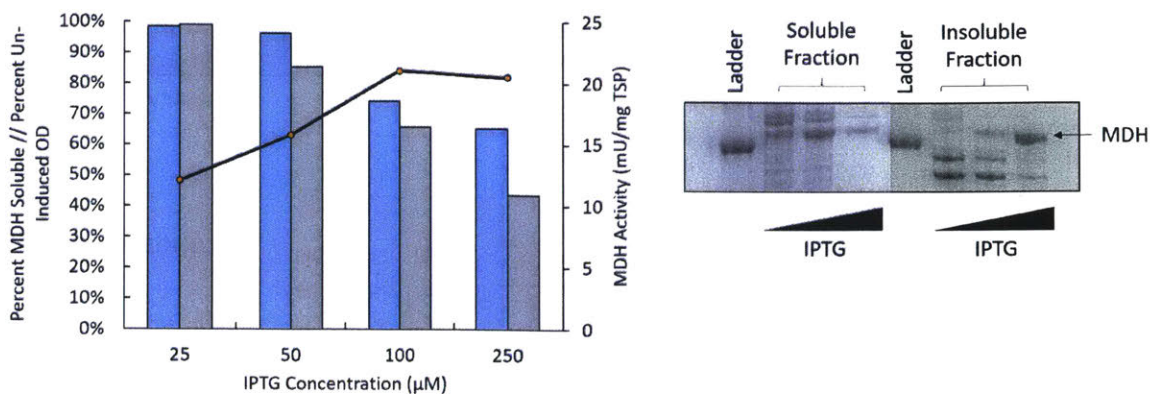
Figure 5-24: **Titration of MDH Activity with IPTG.** Left: Measurements of MDH activity, solubility and cell growth upon induction with varying concentration of IPTG. Blue bars: Percentage of MDH in soluble fraction. Gray bars: OD at harvest (3 hours post-induction), normalized to 0 μM IPTG treatment. Black line: MDH activity in crude lysate, normalized to total soluble protein. Right: SDS-PAGE gel used for quantification of MDH solubility as described in the Methods.

or otherwise impacting cell physiology. In this way, the effect of MDH activity on pathway flux could be assessed by measuring the rate of ^{13}C incorporation in FBP in cells treated with different mol fractions of $^{13}CD_3OD$ and $^{13}CH_3OH$. Critical to this approach, MDH is the only step between methanol and FBP where a KIE would be predicted (Figure 5-25). To test the validity of this strategy, we first demonstrated the ability to modulate the rate of methanol oxidation using purified MDH (Figure 5-26). Increasing the mol fraction of CH_3OH at constant total methanol concentration resulted in a linear increase in reaction rate at both high (500 mM) and low (50 mM) total methanol. Then, using the same mixtures but with additional ^{13}C label ($^{13}CD_3OD$ and $^{13}CH_3OH$), we compared the rate of methanol assimilation into central carbon metabolism in strains expressing the complete methanol pathway, supplemented with xylose to remove the bottleneck of Ru5P regeneration. In this experiment, the labeling of FBP is complicated by the deuterium atoms. Complete retention of deuterium from formaldehyde would lead to the FBP being labeled in increments of three mass units ($^{13}CD_2$). However, the deuterons can be lost in exchange flux between the structural hexose-phosphate molecules, resulting in labeling from M+1 to M+3 depending on the magnitude of these fluxes. Since these higher

191
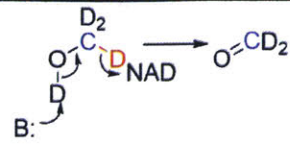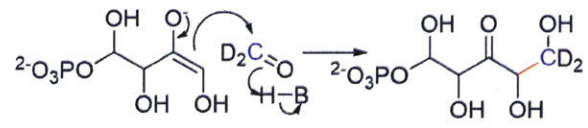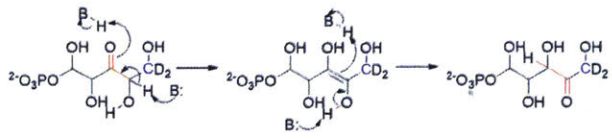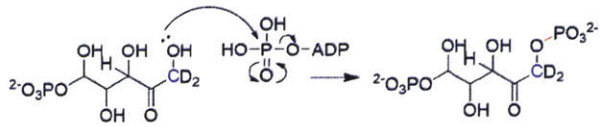
Enzyme	Bond Rearrangements	Isotope Effect
Methanol Dehydrogenase (MDH)		Primary
Hexulose Phosphate Synthase (HPS)		Secondary
Phospohexulose Isomerase (PHI)		None
Phosphofructokinase (PFK)		None

Figure 5-25: **Kinetic Isotope Effects on Enzymatic Steps Between Methanol and FBP**

isotopes can also be generated during the RuMP cycle, it is impossible to resolve the relative contribution of these pathways to the overall labeling, as shown in Figure 5-27. However, completely unlabeled FBP (M+0) can only result from xylose, and thus the total percentage of FBP that is labeled:

$$\%FBP_{Labeled} = 100\% \left(\frac{\sum_{i=1}^{\infty} FBP_{M+i}}{\sum_{i=0}^{\infty} FBP_{M+i}} \right) \tag{5.15}$$

is an indicator of the relative flux of methanol compared to xylose. LC-MS/MS analysis of FBP showed a strong linear correlation between MDH activity, as controlled by the mol fraction of $CH_3OH:CD_3OD$, and FBP labeling (Figure 5-28). These reults demonstrate that flux from methanol to formaldehyde is kinetically controlled by methanol dehydrogenase, and suggests that improvements to MDH activity, either through addressing expression level, e.g. through improving solubility, or by bioprospecting for or evolving more catalytically active variants, will improve pathway flux as long as Ru5P concentration is maintained at a level sufficient to ensure that methanol oxidation is under kinetic, not thermodynamic control.
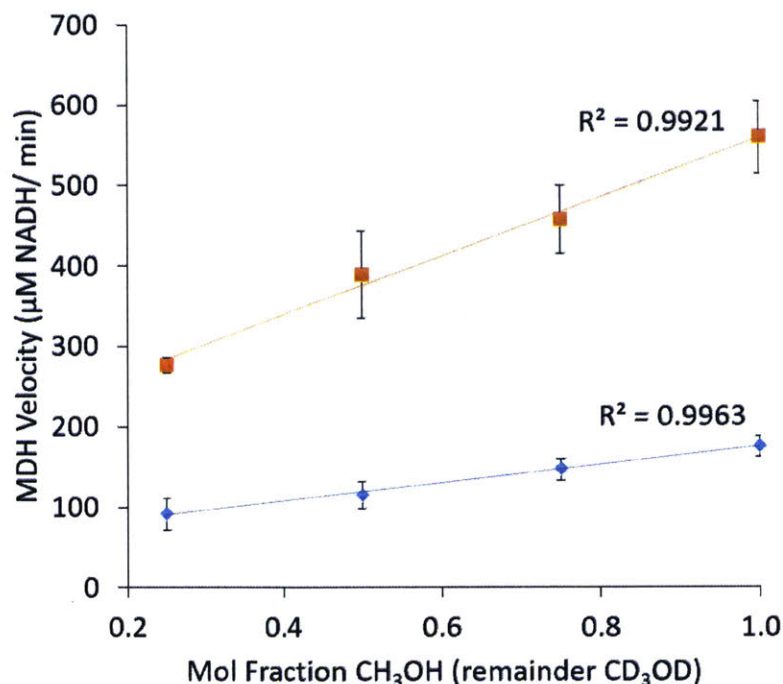
Figure 5-26: **Titration of MDH Activity with Different Mol Ratios of CH₃OH and CD₃OD.** MDH velocity was determined in assays with purified protein (*C. necator* mdh2 4-1) by following the reduction of NAD^+ at 340 nm with different molar ratios of $CH_3OH:CD_3OD$ at constant total methanol concentration. Orange: 500 mM total methanol. Blue: 50 mM total methanol. Error bars represent standard deviation of three technical replicates.

5.5   Conclusions

In this work, we engineered *E. coli* to metabolize methanol, and examined the pathway limitations that have contributed to the elusivity of synthetic methylotrophy. Screening of various isoenzymes of MDH, HPS and PHI led to robust ¹³C-methanol incorporation in resting cells, and co-expression of the genes required for 3-hydroxybutyrate synthesis allowed label to accumulate in the heterologous product. Analysis of steady state formaldehyde levels in resting cells demonstrated that, in the absence of a source of pentose phosphates, formaldehyde accumulates to almost equilibrium levels, reducing net forward flux through the pathway and potentially inhibiting cell function through toxicity. Increasing the concentration of Ru5P, independent of overexpressing enzymes of the ribulose monophosphate pathway, was
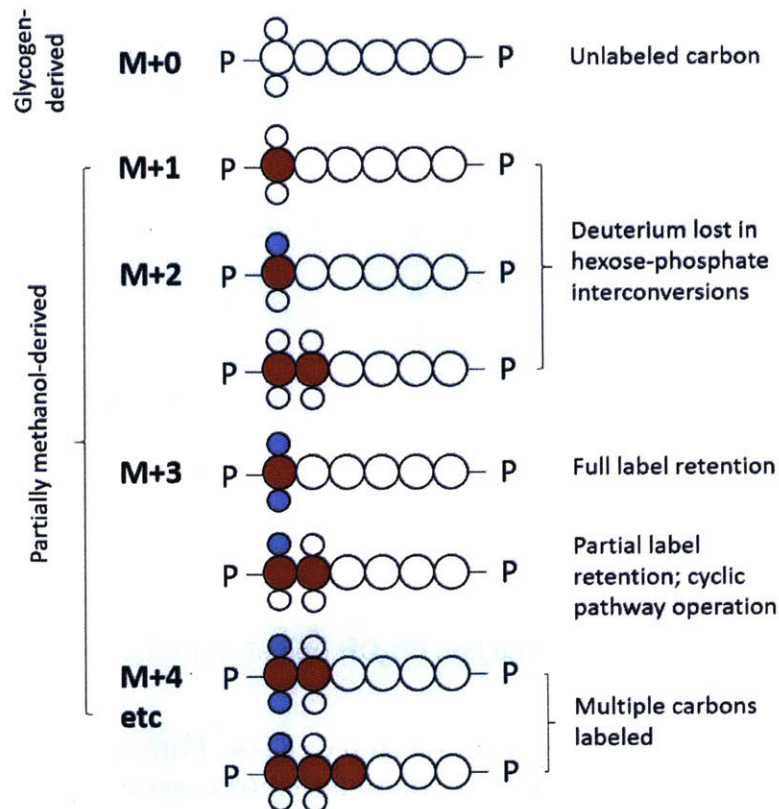
Figure 5-27: **Expected FBP Labeling Patterns due to Incorporation of CD₃OD.** Large and small unfilled circles represent ^{12}C atoms and ^{1}H atoms in FBP, respectively. Red circles indicate ^{13}C atoms and blue circles represent ^{2}H atoms.

sufficient to reduce formaldehyde concentration and ensure a high ratio of forward to reverse flux. FBA analysis of the predicted network topology during growth on methanol and glucose suggested targets for improving Ru5P regeneration by reducing flux leakage through glycolysis. Iodoacetate-mediated inhibition of GAPDH was insufficient to restore Ru5P titers, but the combination of IA with overexpression of *E. coli glpX* led to a 4-fold increase in Ru5P concentration, and a concomitant reduction in formaldehyde concentration and increase in pathway flux as evidenced by ^{13}C labeling analysis. Further analysis showed that this phenotype was at least partially due to activation of the SBPase variant of the RuMP pathway. These results validate the G3P node as a target for genetic strategies to control the division of flux between glycolysis and the pentose phosphate pathway, and suggest that the SBPase variant may be the preferred pathway for Ru5P regeneration in *E. coli*. We then used kinetic
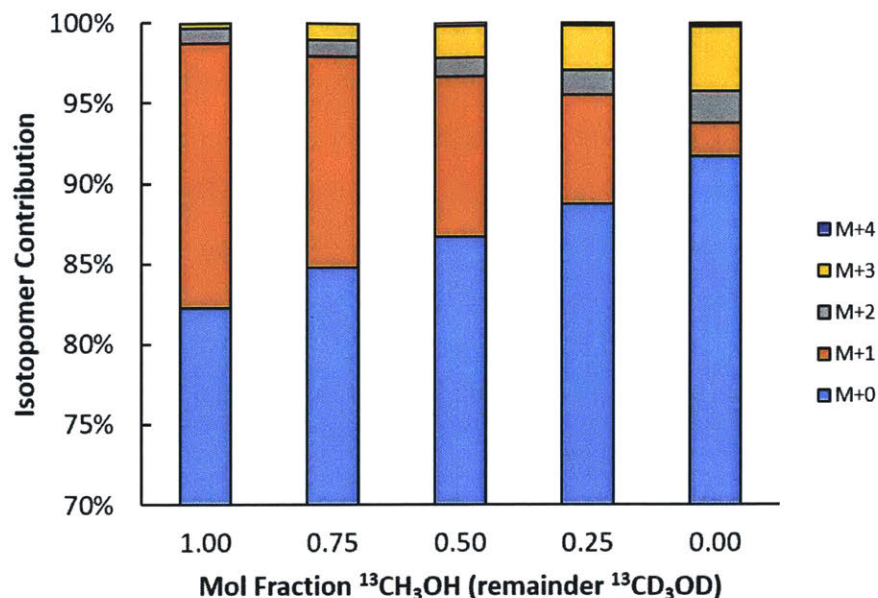
Figure 5-28: **FBP Labeling Patterns in Cells Fed with Various Molar Ratios of $^{13}CH_3OH$ and $^{13}CD_3OD$.** FBP labeling was assessed by LC-MS/MS and corrected for natural abundance. Bar heights represent the mean isotope contribution from three biological replicates.

isotope effects combined with labeling anlaysis to titrate MDH activity *in vivo* and demonstrate that pathway flux is controlled by MDH kinetics. Together, these results highlight Ru5P regeneration as the primary target for future metabolic engineering, followed by MDH activity, because in the absence of sufficient Ru5P to drive forward flux, the relative improvement of pathway flux by improved MDH catalysis will be small compared to the benefit that could be achieved by making the pathway more irreversible.

Beyond the conclusions presented here, the methodologies developed in this work to probe pathway limitations will have broad utility in measuring progress as methanol-consuming strains are improved. These approaches also highlight the benefit of tools not typically associated with metabolic engineering, but rather with chemical biology, for pathway engineering - principally the use of chemical inhibitors (iodoacetate) to modulate pathway flux, and the deliberate introduction of kinetic isotope effects instead of genetic modulation to probe rate-controlling enzymes. As our ability to leverage new high-throughput, low-cost technologies from synthetic biology to overcome

pathway obstacles begins to surpass our ability to identify these hurdles, the need for novel, rapid and versatile tools to identify these limitations becomes paramount, and should draw on a wide range of tools from disciplines beyond metabolic engineering. As metabolism is fundamentally a series of chemical processes, the tools from chemical biology should play a major role in these analyses.

Chapter 6

Development and Application of a Formaldehyde Biosensor

6.1 Summary

Formaldehyde is a prevalent environmental toxin, and a key intermediate in single carbon metabolism. The ability to monitor formaldehyde concentration is therefore of interest for both environmental monitoring and for metabolic engineering of native and synthetic methylotrophs, but current methods suffer from low sensitivity, complex workflows, or the requirement for expensive analytical equipment. Here we develop a formaldehyde biosensor based on the FrmR repressor protein and cognate promoter of *E. coli*, coupled to either luciferase or GFP. Optimization of the native repressor binding site and regulatory architecture enabled detection at levels as low as 3 μM. We then used the sensor to benchmark the *in vivo* activity of several NAD-dependent methanol dehydrogenase (MDH) variants, the rate-limiting enzyme that catalyzes the first step of methanol assimilation. In support of MDH directed evolution using this assay, we examined the potential for cheaters in a mixed population, and developed a strategy to prevent cross-talk by using glutathione as a formaldehyde sink to minimize intercellular formaldehyde diffusion. Finally, we show the utility of the reporter in balancing expression of MDH and the formaldehyde assimilation enzymes HPS and PHI in an engineered *E. coli* strain to minimize formaldehyde build-up while also

reducing the burden of heterologous expression. This biosensor offers a quick and simple method for sensitively detecting formaldehyde, and has the potential to be used as the basis for directed evolution of MDH and dynamic formaldehyde control strategies for establishing synthetic methylotrophy.

6.2 Introduction

Formaldehyde is a highly toxic chemical, and is classified as a Group 1 human carcinogen by the International Agency for Research of Cancer (IARC) [240]. As a potent electrophile, its toxicity stems from its ability to react rapidly with nucleophilic components of DNA, RNA and proteins, leading to protein and DNA damage in the form of crosslinking [241]. Somewhat paradoxically, formaldehyde is also a ubiquitous intermediate in one-carbon metabolism across the tree of life, from archaeal methanogens to bacterial methanotrophs and methylotrophic yeasts. In methanotrophs, methane is first oxidized to methanol by methane monooxygenase (MMO), and then to formaldehyde by methanol dehydrogenase (MDH) [22]. Methylotrophic yeasts such as *Pichia pastoris* convert methanol to formaldehyde using an FAD-linked alcohol oxidase (AOX) [242], and gram-positive methylotrophs typified by *Bacillus methanolicus* perform the same conversion using an NAD-linked MDH [243]. In all these organisms, formaldehyde acts as a branch point between further oxidization to CO_2 for energy conservation, and incorporation into biomass via the serine cycle, ribulose monophosphate (RuMP) pathway, or the xylulose-5-phosphate (Xu5P) pathway [22, 244, 243]. Besides methylotrophs, formaldehyde is present at low levels in all organisms as a result of demethylation reactions [245]. Because of its cytotoxicity, the intracellular formaldehyde concentration must be tightly controlled, which has led to the evolution of a variety of highly coordinated metabolic strategies for detoxifying formaldehyde [246]. The need to keep the concentration of formaldehyde low while supporting high flux places an even more stringent burden on methylotrophs that rely on formaldehyde metabolism for growth.

Recently, significant interest has arisen in using methylotrophs as platform hosts

for metabolic engineering, largely driven by the abundance of natural gas in the U.S. [12, 101]. Methanol can also be produced renewably from a variety of sources, and as a non-agricultural commodity does not contribute to the 'Food vs. Fuel' debate. In addition, its general toxicity is attractive in preventing bioprocess contamination, and the current price is comparable to that of glucose despite carrying 50% more electrons per carbon [15]. Genetic engineering and physiological understanding of methylotrophs still lags in comparison to typical metabolic engineering hosts, creating interest in importing methanol assimilation into the prototypical synthetic biology host *E. coli*. This concept, dubbed 'synthetic methylotrophy' [202], aims to combine the advantages of methanol as an attractive carbon substrate for fermentation with the robust engineering tools and range of downstream heterologous products available in *E. coli*. The seminal work in this area established that the expression of an NAD-dependent MDH from *B. methanolicus*, along with the heterologous enzymes of the RuMP pathway, hexulose phosphate synthase (HPS) and phosphohexulose isomerase (PHI), was sufficient to enable methanol incorporation in resting cells, as measured by ^{13}C-methanol labeling experiments [203]. Recent work has shown that, in similarly engineered cells with a more active MDH variant from *Geobacillus stearothermophilus*, supplementation of cells growing on yeast extract with methanol could improve biomass production by 40% [226]. To date, no group has managed to achieve growth of *E. coli* with methanol as a sole carbon substrate. One possible reason for this is the unbalanced production and consumption of formaldehyde in an organism that has not evolved to tolerate high formaldehyde flux.

The ability to easily measure intracellular formaldehyde could therefore provide basic insights into the regulation of formaldehyde metabolism in native methylotrophs, as well as aid in the development of a synthetic methylotroph. However, typical methods are limited by low sensitivity, cumbersome workflows, or the requirement for expensive instrumentation. In the work conducted thus far on synthetic methylotrophy, formaldehdye has been measured in culture supernatants using the Nash assay, taking adavantage of the fact that formaldheyde can diffuse rapidly across the cell membrane [203, 226, 210]. Here, formaldehyde reacts spontaneously with

2,4-pentanedione and ammonium acetate to form the dihydropyridine 3,5-diacetyl-1,3-dhydrolutidine (DDL), which is quantified by abosbrance spectroscopy at 412 nm [215]. Due to the low sensitivity of this assay, in order to detect formaldehdye production in engineered cells, the native detoxification pathway had to first be eliminated [203]. The assay is also limited to small numbers of samples in a kinetic expeirment due to the need to separate cells from supernatant before analysis. The gold standard for environmental formaldehdye quantification in both air and water samples involves derivitization with 2,4-dinitrophenyl hydrazine (2,4-DNPH), followed by HPLC to separate the various carbonyl derivatives before quantification via UV [241]. While highly sensitive, this technique suffers from the same bottleneck of requiring cell separation before analysis, and the additional challenges of low throughput and high cost due to the requirement for HPLC separation.

Recently, there has been tremendous interest in developing genetically encoded biosensors for monitoring the concentration of a multitude of different compounds [247, 248, 249]. These sensors offer several advantages compared to traditional methodolgies: Since the signal is often a fluorescent protein such as GFP, sensor read-out can be determined easily using widely available instrumentation without the need for separating cells from their media, and is therefore adaptable for high-throughput sampling. In addition, because GFP is relatively stable, the fluorescence signal represents an integral of the substrate concentration over time, allowing for significantly increased sensitivity compared to single time-point measurements.

The native glutathione-dependent formaldehyde detoxification pathway in *E. coli* provides a convenient architecture for a formaldehdye biosensor. In this pathway, formaldehyde reacts spontaneously with the nucleophilic cysteine residue of glutathione to form the hemiacetal *S*-(hydroxymethyl)-glutathione. This adduct is enzymatically oxidized to *S*-formylglutathione by *S*-(hydroxymethyl)-glutathione dehydrogenase encoded by the *frmA* gene. Hydrolysis of this species by *S*-formylglutathione hydrolase, encoded by *frmB*, liberates glutathione and produces formate, which is much less toxic than formaldehyde [250]. Expression of *frmA* and *frmB* is controlled by a repressor protein *frmR*, expressed in the same operon. In the absence of formalde-

hdye, FrmR binds to the promoter region, preventing transcription. In the presence of formaldehyde, the nucleophilic Cys36 of the FrmR is thought to react with formaldehdye and cause a conformational change in the protein that results in dissociation from the promoter and allows transcription to occur [251]. Addition of 0.25 mM formaldheyde is sufficient to induce an approximately 100-fold increase in *frmA* transcript after 30 minutes [252]. Previously, Tralau and coworkers employed a GFP-linked biosensor based on this regulatory system to detect formaldehdye produced during the oxidation of dimethylglycine in *E. coli* [253].

Here we report significant advances in the development of the formaldehdye biosensor, and the application of the sensor to the ongoing efforts to engineer synthetic methylotrophy. First, we developed a variant of the sensor where *frmR* was expressed from the same plasmid as the signal protein (either GFP or luciferase), under the orthogonal p_{tet} promoter to remove the negative feedback of the original system and thus increase sensitivity. Second, the sensitivty was further improved and the dynamic range shifted toward physiologically relevant formaldehyde concentrations by making systematic mutations in the *frm* promoter region. To demonstrate the utility of the refined biosensor, we then showed that the reporter could detect formaldehyde production in engineered *E. coli* strains without having to first delete *frmA*, in contrast to previous work [253, 203, 226].

In addition to reporting concentrations, biosensors have emerged as powerful tools for directed evolution [254, 255]. When the product of an enzyme can be linked in a concentration-dependent manner to a fluorescent signal, FACS can be used to select for high-activity variants from a library of mutants. In previous work establishing synthetic methlotrophy, it has been shown that the kinetic properties of NAD-dependent MDH are poor compared to the other heterologous enzymes, making methanol oxidation the likely rate-limiting step [203, 226]. It is therefore of interest to evolve MDH variants with more attractive kinetic parameters [210]. To this end, we first verified that our reporter could detect differential MDH activity by comparing variants known to have differing *in vivo* properties. Selection of improved variants in directed evolution is often hampered by the presence of 'cheaters' - cells in which the reporter

is activated despite not displaying the desired phenotype. Since formaldehyde can diffuse between cells, we hypothesized that in a mixed population, selecting based on high GFP reporter output might lead to significant 'cheating'. This was examined, and we devised a strategy to minimize cross-talk by addition of exogneous glutathione to the medium.

Beyond directed evolution of enzymes, biosensors are gaining interest in synthetic biology and metabolic egnineering for their ability to actuate a dynamic metabolic response to the presence of the target analyte in a concentration-dependent manner [118, 153]. Notable examples of this include the DSRS system for fatty acid biosynthesis [256], and the quorum-sensing regulation of PFK activity for glucaric acid production [257]. For this approach to work, the dynamic range of the sensor must be tuned to the expected concentration range. To examine whether this was the case here, as a proof of principle we tested the ability of our biosensor to discriminate between formaldehyde concentrations in cells induced to different extents for graded expression of MDH (formaldheyde production) and HPS and PHI (formaldehdye consumption), and showed that we could indeed use this approach to optimize the relative expression level of the heterologous enzymes.

6.3 Methods

6.3.1 Reagents

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

6.3.2 Strains and Plasmids

Bacterial strains and plasmids used in this study are listed in Table 6.1. *E. coli* DH5α was used as a cloning host. Luciferase-based assays and MDH comparisons were carried out using *E. coli* S1030 [258] with or without *frmA* knocked out, as described in the text. Pathway optimization assays were conducted in *E. coli* MG1655(DE3) [208]. pBbS2k-RFP was a gift from Jay Keasling (Addgene plasmid #35330). pETM6-mCherry was a gift from Mattheos Koffas (Addgene plasmid #66534)

6.3.3 Cloning

Primers used for PCR amplification of cloning fragments are shown in Table 6.2. Q5 Polymerase (NEB) was used for all amplifications, and all amplicons were digested with DpnI before PCR purification. Gibson Assembly (GA) Master Mix (NEB) was used for construction of plasmids. All restriction endonucleases were purchased from NEB. Assembled vectors were transformed into DH5α chemically competent cells (NEB), and verified by Sanger sequencing. The *frmA* deletion in S1030 was made using the protocol described by Datsenko and Wanner [211]. Cell growth for cloning purposes was carried out using DIFCO LB media supplemented with appropriate antibiotics (Kanamycin, 50 μg mL^{-1}; Carbenicillin, 50 μg mL^{-1}; Spectinomycin, 50 μg mL^{-1}). Solid media was prepared with the addition of 1.5% BACTO agar. Plasmids and PCR products were purified using Qiagen kits.

To construct GFP versions of the reporter plasmids (pTR47m4-GFP and pTR59-GFP), the corresponding luciferase-carrying plasmids were linearized with HindIII, then the backbone amplified with primers pTR47_fwd and pTR47_rev. Super-folder GFP was amplified from pTrc-sGFP [259] with primers pTR-sGFP_fwd and pTR-sGFP_rev. Fragments were assembled via GA. To clone different variants of MDH into the expression plasmid pTR48, first pTR48 was amplified with primers pTR48_fwd and pTR48_rev, treated with DpnI, and purified. MDH variants were codon optimized for *E. coli* and synthesized as gBlocks, with 5' extension gagctaaggaggaaaaaaaa and 3' extension aggagtgccgttaattaagt for assembly with the lin-

Name	Description	Reference
Strains		
E. coli DH5α	Cloning strain	NEB
E. coli S1030	F'proA+B+ (lacIZY) zzf::Tn10(Tet^R)lacIQ P$_{N25}$-tetR luxCDE/endA1 recA1 galE15 galK16 nupG rpsL(Str^R) lacIZYA araD139 (ara,leu)7697 mcrA (mrr-hsdRMS-mcrBC) proBA::pir116araE201 rpoZ flu csg ABCDEFG pgaC λ$^-$	[258]
E. coli S1030ΔfrmA	S1030 with glutathione-dependent formaldehyde dehydrogenase *frmA* knocked out	This Work
E. coli MG1655(DE3)	F$^-$ λ$^-$ *ilvG rfb-50 rph-1* (DE3)	[208]
Plasmids		
pTR47	pSC101 Ap^R pTet-frmR pFRM-xluxAB	This work
pTR47m4	pTR47 with pFRM binding site mutation	This work
pTR47m4-GFP	pTR47m4 with sGFP replacing xluxAB	This work
pTR59	pTR47m4 with insulated constitutive promoter Pro4 controlling frmR	This work
pTR59-GFP	pTR59 with sGFP replacing xluxAB	This work
pTR48	ColE1 araC pBad Sp^R	This work
pTR48mdh1	pTR48 carrying mdh1 from *Bacillus methanolicus MGA3*	This work
pTR48mdh2	pTR48 carrying mdh2 from *Bacillus methanolicus MGA3*	This work
pTR48mdh2act	pTR48 carrying mdh2 and act from *Bacillus methanolicus MGA3*	This work
pTR48adhA	pTR48 carrying adhA from *Corynebacterium glutamicum*	This work
pTR48bstmdh	pTR48 carrying mdh from *Geobacillus stearothermophilus*	This work
pTR48cnmdhWT	pTR48 carrying WT mdh2 from *Cupriavidus necator*	This work
pTR48cnmdh4-1	pTR48 carrying evolved mdh2 CT4-1 from *Cupriavidus necator*	This work
pTR48mCherry	pTR48 carrying mCherry	This work
pET28	ColE1 lacI pT7 Kn^R	Novagen
pTrc-sGFP	pBR322 lacI^Q pTrc-sGFP Ap^R	[259]
pET-sGFP	pET28 carrying pT7-sGFP	This work
pETcnmdh4-1	pET28 carrying mdh2 CT4-1	This work
pETM6-mCherry	ColE1 lacI pT7-mCherry Ap^R	[260]
pBbS2k-RFP	pSC101 tetR pTet-RFP Kn^R	[160]
pETsGFP-pTet-RFP	pET28 carrying pT7-sGFP and pTet-RFP	This work
pETMEOH560	pET28 carrying pT7-mdh2 CT4-1 and pTet-hps-phi from *Bacillus methanolicus MGA3*	This work

Table 6.1: Strains and plasmids used in this study

ear fragment of pTR48. To clone *act* downstream of *mdh2* in pTR48-mdh2act, pTR48mdh2 was amplified with act_pTR48mdh2_rev and act_pTR48mdh2_fwd. A ribosome binding site was encoded upstream of *act* by amplifying the gene with primers RBS_act_fwd and RBS_act_rev. The WT *C. necator mdh2* was generated from the evolved variant CT4-1 in one step by amplifying the region between the substitions as one fragment (originalCNMDH_fwd and originalCNMDH_rev), and the rest of the plasmid as a second fragment (pTR48CN41_fwd and pTR48CN41_rev), and encoding the substiutions on the homology regions of the primers. Both fragments were DpnI digested, purified, and assembled. mCherry was cloned into pTR48 by amplifying pETM6-mCherry with primers mCherry_fwd and mCherry_rev, and assembling with with the pTR48 fragment. The optimized sequences for *mdh1, mdh2, act, adhA* and *C. necator mdh2* 4-1 are shown in Supplementary Information. The sequence for the *mdh* from *G. stearothermophilus* was taken from Whitaker et al [226]. pET-sGFP was constructed by digesting pET28 with NcoI and SacI, and assembling the purified product with the PCR fragment generated from the amplification of sGFP from pTrc-sGFP with primers pET_sGFP_fwd and pET_sGFP_rev. pET-sGFP-pTet-RFP was generated by linearizing pET-sGFP with PpuMI, and assembling the purified product with the PCR fragment amplified from pBbS2k-RFP with tetRFP_fwd and tetRFP_rev. This assembly results in sGFP and RFP expressed from opposite strands, to minimize any potential for read-through transcription. To generate pETcnmdh4-1, pET28 backbone was amplified with primers pET28_fwd and pET28_rev, and CNmdh4-1 was amplified from pTR48cnmdh4-1 with primers CN401_fwd and CN401_rev. The two fragments were assembled via GA. pET-MEOH560, which is based on pETcnmdh4-1 and additionally expresses HPS and PHI under control of the Tet promoter, was constructed by linearizing pETcnmdh4-1 with PpuMI, then assembling it with the Tet fragment amplified from pBbS2k-RFP with primers pTet_fwd and pTet_rev, and a GBlock containing a synthetic operon with optimized HPS and PHI from *B. methanolicus* (Supplemental Information).

Primer	Sequence
pTR47_fwd	ACGGAGCCAATGTACGCA
pTR47_rev	TTTTTTTTCCTCCTTACTGCACC
pTR-sGFP_fwd	cagtaaggaggaaaaaaaaATGAGCAAGGGCGAAGAG
pTR-sGFP_rev	cgtacattggctccgtTTACTTATAGAGTTCATCCATGCC
pTR48_fwd	ACTTAATTAACGGCACTCC
pTR48_rev	TTTTTTTTCCTCCTTAGCTCG
act_pTR48mdh2_fwd	CGGCACTCCTCAGCAAAT
act_pTR48mdh2_rev	ggttaattcctcctcccgggTTACATGGCGTTTTTGATGATC
RBS_act_fwd	agatcatcaaaaacgccatgtaacccgggaggaggaattaaccATGGGTAAACTGTTTGAAGAAAAAC
RBS_act_rev	atatttgctgaggagtgccgTTATTTGTTTTTCAGTGCTTCTTG
originalCNMDH_fwd	ccctgcgtgaccctcttcggtccgggctgcgcgcgcgaaacgggcgcgCGCGCCAGATCACTCGGG
originalCNMDH_rev	gattaatggcgtgcaacgccagtcgacgatcgcCATCTTCACATGATTGCTCGAATTAGTGATG
pTR48CN41_fwd	gcgATCGTCGACTGGCGTTGC
pTR48CN41_rev	cgcgcccgtttcgcgcgcGCAGCCCGGACCGAAGAG
mCherry_fwd	gagctaaggaggaaaaaaaaATGGTTTCAAAAGGCGAAG
mCherry_rev	aggagtgccgttaattaagtTTATTTGTACAGTTCATCCATAC
pET_sGFP_fwd	ctttaagaaggagatataccATGAGCAAGGGCGAAGAG
pET_sGFP_rev	cttgtcgacggagctcTTACTTATAGAGTTCATCCATGCC
tetRFP_fwd	atcgtgctcctgtcgttgaggacccTTAAGACCCACTTTCACATTTAAG
tetRFP_rev	aacccgccagcctagccgggtcctTATAAACGCAGAAAGGCC
pET28_fwd	TGAGATCCGGCTGCTAAC
pET28_rev	GGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTC
CN401_fwd	ctttaagaaggagatataccATGACCCACCTGAACATCG
CN401_rev	ttgttagcagccggatctcaTTACATCGCCGCAGCGAA
pTet_fwd	atcgtgctcctgtcgttgaggacccTTAAGACCCACTTTCACATTTAAG
pTet_rev	GCAATTCCATATGTATATCTCCTTCTTAAAGATCTTTTG

Table 6.2: Primers used in this study

6.3.4 Luciferase-Based Reporter Assays

For all reporter assays, a minimum of three individual colonies from transformation plates were grown overnight in M9$^+$ medium (DIFCO M9 supplemented with 0.2% glucose, 0.1% casamino acids, and 1% (v/v) each of trace mineral and vitamin solution). In the morning, cultures were diluted 100-fold into fresh medium and grown until early exponential phase (OD$_{600}$ approximately 0.4) supplemented with anhdyrotetracycline (ATc) to induce frmR expression. For direct testing of formaldehyde-responsive or non-specific induction, potential substrates were added directly to 200 μL of sample in a black, clear-bottom 96-well plate (Corning) and analyzed for optical density at 600nm and luciferase activity at 37°C on a Tecan Infinite Pro M1000 plate reader.

6.3.5 GFP-Based Formaldehyde Reporter and MDH Assays

GFP reporter assays were conducted almost identically to the luciferase-based ones. After overnight growth of colonies, cells were diluted into fresh medium with appropriate antibiotics in 14 mL culture tubes and grown until early exponential phase. To test the formaldehyde response, cells were then transferred into clear 96-well plates (200 μL per well), and formaldehyde (0-500 μM) was added. The plate was transferred to a SpectraMax M2e (Molecular Devices) spectrophotometer pre-warmed to 37°C, and fluorescence was monitored every 10 minutes. Between reads, the plate was continuously shaken in the spectrophotometer. For MDH-linked assays, after initial growth in culture tubes, cells were induced with 10 mM arabinose and immediately transferred to 96-well plates (200 μL per well), where 50 μL methanol was added to a final concentration of 0-500 mM. Plates were covered with Breathe-Easy sealing membranes (Sigma), and growth continued at 37°C with maximum agitation on a Jitterbug 2.0 (Boekel). Pathway optimization experiments were conducted analogously, except that induction was mediated by IPTG (0-1000 μM) and ATc (0-200 ng mL^{-1}). GFP fluorescence was assessed every hour by spectrophotometer to minimize temperature fluctuations that would affect growth and MDH activity. Excitation

and emission wavelengths were 488 nm and 525 nm respectively, with autocutoff set to 515 nm. Fluorescence was measured from the top after removal of the sealing membrane, with automatic gain. To account for growth differences due to methanol or formaldehyde toxicity, or the expression of different MDHs, fluorescent signal was normalized by dividing by the absorbance of the culture at 600 nm measured by the same spectrophotometer, from which the abosorbance of a media blank was first subtracted.

6.3.6 Flow Cytometry

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

6.4 Results and Discussion

6.4.1 Design and Optimization of the Biosensor

Autoregulated negative-feedback promoters are a means of keeping steady-state levels of gene-expression across varying conditions, providing greater stability to genetic networks [261]. However, in designing a circuit for monitoring the concentration of a compound in a culture, such regulation attenuates the dynamic range of the reporter system, making it harder to discern meaningful differences in signal across broad input ranges. The formaldehyde operon in *E. coli* is an autoregulated negative-feedback promoter regulated by the protein product, FrmR, of the *frmR* gene. Previous work on this promoter and its response to formaldehyde showed relatively limited signal gain in response to toxic concentrations of formaldehyde [253]. We hypothesized

that we could achieve a much higher, dose-dependent response to formaldehyde in an engineered *E. coli* strain were this promoter to be taken out of its autoregulatory context. To show this, we de-coupled expression of *frmR* from its own regulation and placed it under inducible control of the P_{tet} promoter on a synthetically constructed plasmid. We then compared this construct to an analogous, autoregulated synthetic plasmid (Figure 6-1). Both constructs express the luciferase reporter *luxAB* from the *frm* promoter. When FrmR was not induced from the P_{tet} plasmid, some change in signal due to formaldehyde dosing still occurred. This discrepancy between 'on' (formaldehyde present) and 'off' (formaldehyde absent) states for the uninduced-frmR conditions likely reflect a combination of background expression from the p_{tet} promoter as well as the chromosomal copy of *frmR* present in the *E. coli* strains used. Addition of 40ng mL^{-1} ATc to induce *frmR* expression greatly lowered background signal without greatly lowering observed reporter values for the 'on' state. Comparing our version of this de-coupled repressor to an analogous auto-regulated system, we see an order of magnitude improvement in signal-to-background gene expression, giving us the ability to detect formaldehyde at levels as low as 3 μM (Fig. 6-1).
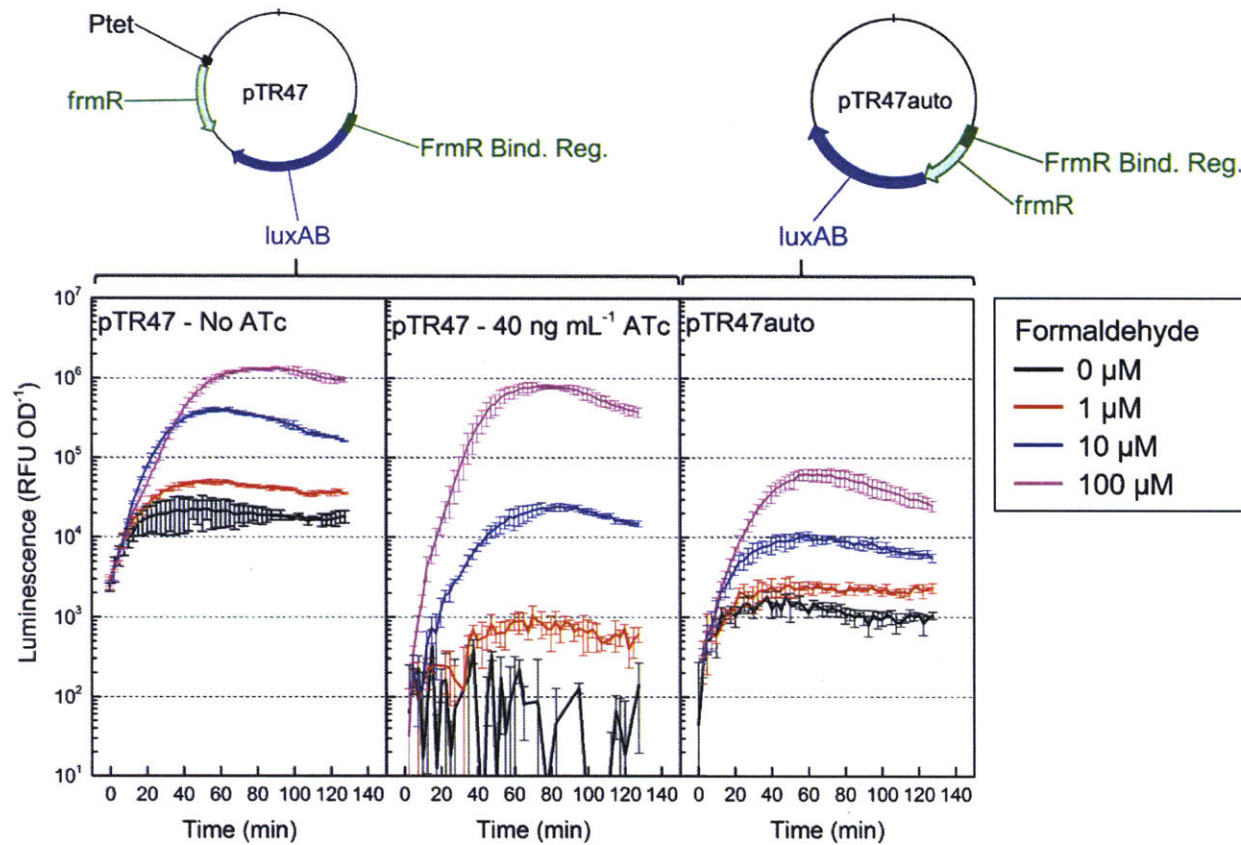
Figure 6-1: **Reporter construct design and formaldehyde response.** Top: The de-coupled plasmid pTR47 controls frmR expression through the P_{tet} promoter, while the pTR47auto plasmid mirrors the native context of the frm operon of *E. coli*. Bottom: Comparison of the coupled (pTR47auto) and de-coupled (pTR47) reporter constructs, with or without anhydrotetracycline (ATc). Error bars show the standard deviation of three biological replicates. RLU, Relative Luminescence Units; OD, Optical Density at 600nm. Breaks in the lines for the 40ng mL^{-1} ATc graph at low concentrations of formaldehyde are due to negative values for luminescence at those time-points.

Despite significant improvements in overall signal over time from formaldehyde treatment, we wanted better separation of signal response to lower formaldehyde concentrations more relevant in cells engineered for methanol assimilation. Precedent in the RncR repressor protein of *E. coli* suggests that repeated cytosine and guanine tracts can induce A-form DNA geometry, likely playing an important role in DNA-protein interactions at binding sites at or around these tracts [262]. In the native sequence, a repeated stretch of cytosine residues has been implicated as a contributing factor to the DNA geometry around the position that would typically bind to recruit the $\sigma70$ portion of the *E. coli* RNA polymerase to initiate transcription of downstream genes [251]. Changing two cytosine residues in one such tract allowed us to observe higher signal at lower formaldehyde doses compared to the wild-type binding sequence, supporting this hypothesis (Figure 6-2). Additionally, the sensitized binding sequence allowed for a much clearer separation of signal trajectories at low formaldehyde doses. Other changes that we tried, such as directly modifying residues in the -35 or -10 sequence of the predicted $\sigma70$ binding sequence, were less effective (data not shown).
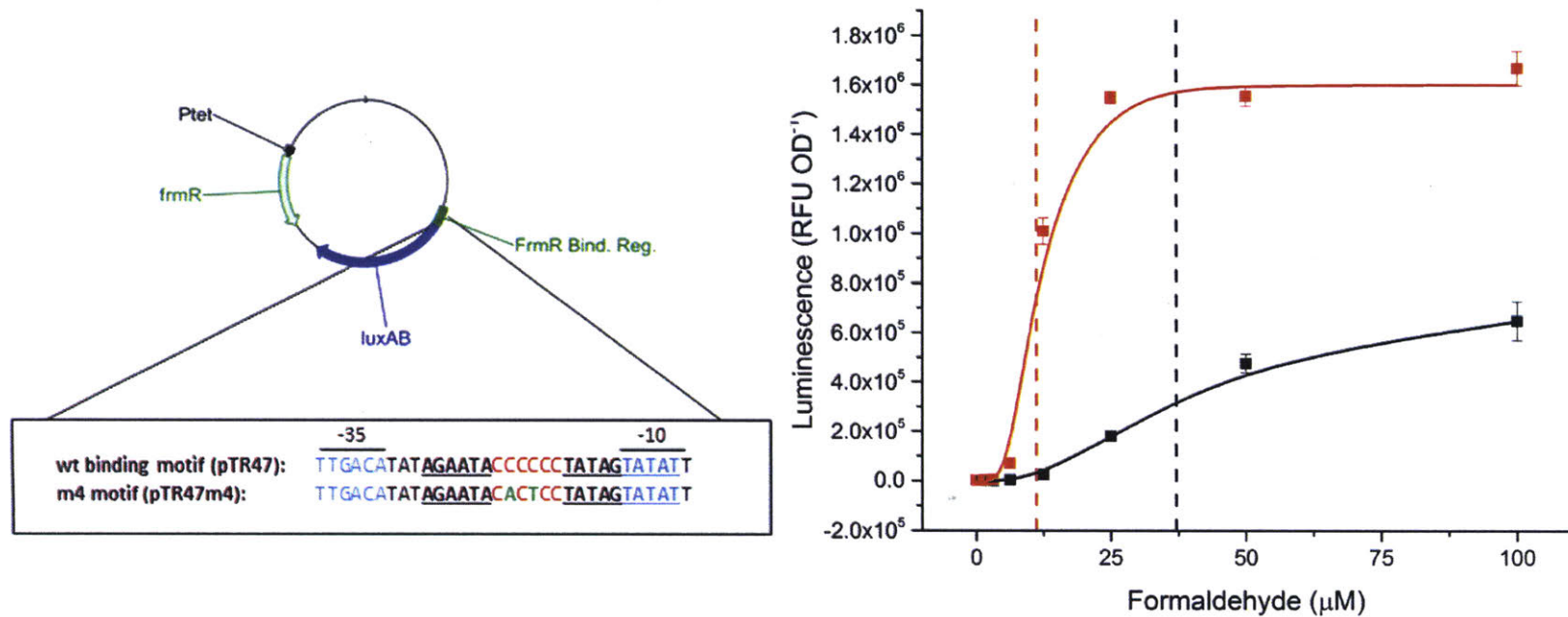
Figure 6-2: **Evaluation of Binding Site Mutants.** Left: Schematic showing the specific differences in the pTR47 and pTR47m4 constructs. Residues in the binding sequence that are underlined show predicted FrmR binding regions. Blue residues are putative $\sigma 70$ binding sequences. Residues in red represent the cytosine repeat tract, and green residues show the mutations introduced in pTR47m4. Right) Comparison of the formaldehyde-based response of cells containing the wild-type (pTR47) and mutant (pTR47m4) binding sites in the de-coupled circuit architecture. Error bars show the standard deviation across three biological replicates. RLU, Relative Luminescence Units; OD, Optical Density at 600nm.

In developing and applying a regulatory circuit, it is crucial that we understand not only its response to a desired substrate, but also its response to similar, but irrelevant molecules. In the context of selections, such off-target activity can lead to cheaters or severe promiscuity. To our knowledge, no data has been reported for the *frm* operon in *E. coli* outside of its initial characterization, which only tested glutathione adducts [252]. To better understand the specificity of the formaldehyde repressor, we tested its response to three other inputs: methanol, acetaldehyde, and propionaldehyde (Fig. 6-3). All three of these molecules showed a response at high doses, but the cutoff for the response was at a much higher concentration than for formaldehyde. Acetaldehyde was the best off-target substrate for turning on the circuit, but was still about 100-fold less effective than formaldehyde based on signals observed at 10mM acetaldehyde compared to 0.1mM formaldehyde. Propionaldehyde was worse still than acetaldehyde, suggesting that as the length of the aliphatic chain of the substrate increases, its ability to bind to FrmR decreases. The activity observed from methanol dosing may itself come from contaminating levels of formaldehyde present in the methanol stock, but whatever the reason for the signal it generated, it needed to be present at about 1000-fold higher concentrations than formaldehyde itself to create an equivalent signal from our sensitized circuit. In all cases tested, the concentrations of off-target substrates required to achieve high signal would likely be prohibitive toward cell growth under standard conditions and are unlikely to be a significant issue in most applications.

6.4.2   Evaluation of a GFP Version of the Biosensor

Luciferase is an excellent reporter in biosensor applications because of its rapid turnover rate, which leads to very high sensitivity. However, for certain applications, coupling the formaldehyde sensor to a fluorescent protein such as GFP could be advantageous. Not every lab has convenient access to luminescence measurements, therefore a GFP version of the reporter may make the sensor more widely usable. To test the feasibility of this approach, we cloned the superfolder GFP [263] in place of the luciferase gene on pTR47m4, generating pTR47m4-GFP. The response to formalde-

213
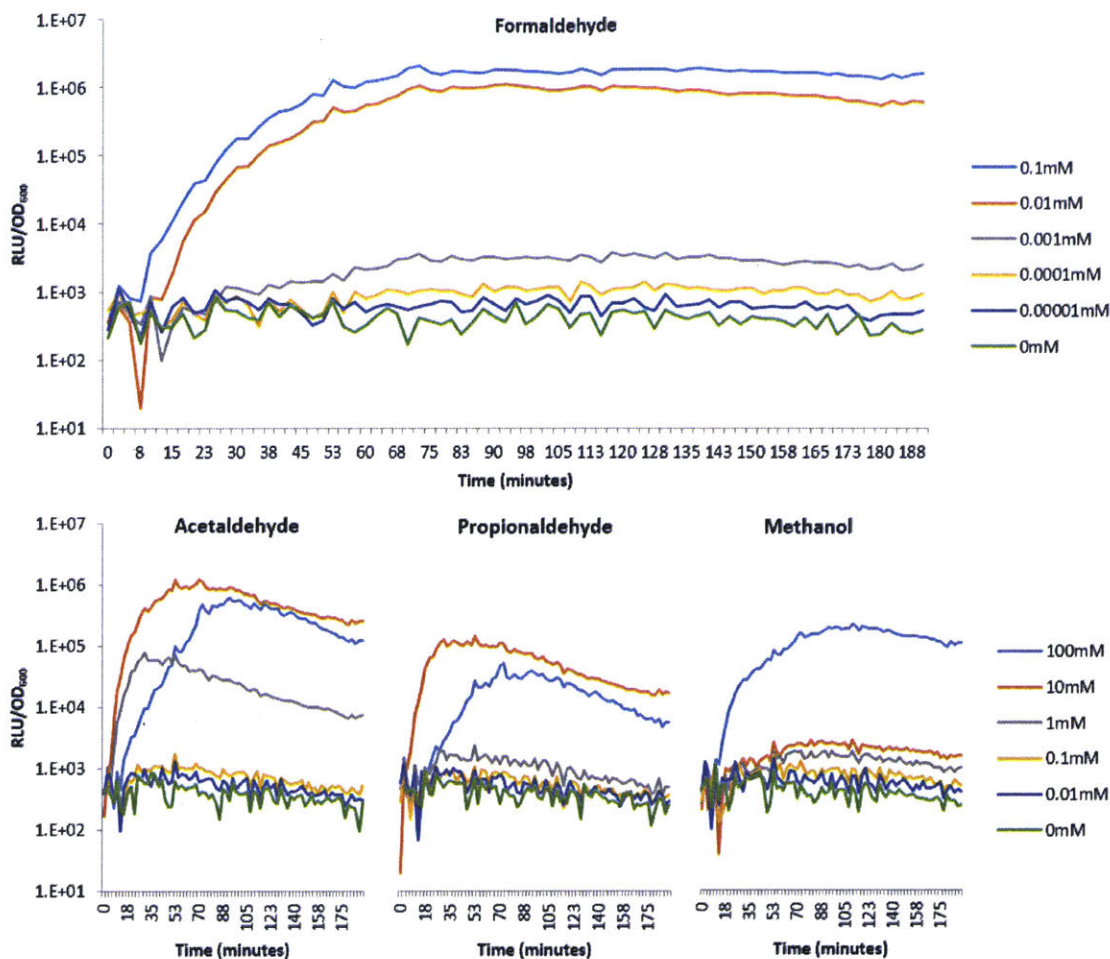
Figure 6-3: **Evaluation of Off-Target Activation.** Comparison of signal from the mutant (pTR47m4) binding site based on dosing of various substrates. Lines show the average of three biological replicates. RLU, Relative Luminescence Units, OD_{600}, Optical Density at 600nm.

hyde addition was tested in exactly the same manner as for the luciferase variant, but with monitoring for GFP fluroescence. The results are shown in Figure 6-4. Compared to the luciferase version, the GFP variant responded well to a similar range of formaldehyde concentrations, but with a lower signal/noise ratio and a slightly higher LOD (10 μM compared to 3 μM). The relationship between the final GFP intensity and formaldehyde concentration was sigmoidal, with a K_A of 21 μM and n of 1.83. Formaldehyde concentrations in *E. coli* engineered for methanol assimilation have generally lower than 100 μM, which is the upper end of the dynamic range of this sensor, and suggests it should be well suited to assessing formaldehyde metabolism

in these strains.
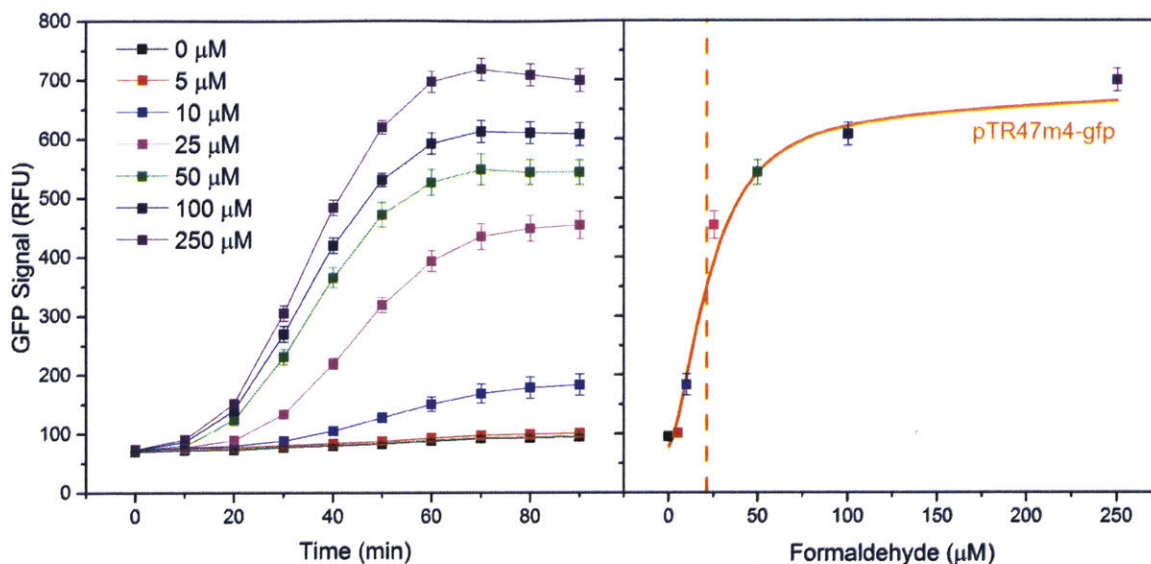
Figure 6-4: **Evaluation of a GFP Reporter Variant.** Left: Time course of the dose-dependent response to formaldehyde. Right: Fitting of a Hill plot to the final GFP signal. Dashed vertical line represents the K_a of the reporter (21 μM). Error bars represent standard deviation of three biological replicates. RFU, raw fluorescence signal (AU).

6.4.3 Comparison of MDH Variants

With a robust and highly sensitive formaldehyde sensor in hand, we set about using it to aid in our lab's ongoing efforts to engineer *E. coli* for methanol assimilation. Prior work in this field has shown the kinetic properties of the first enzyme of the pathway, MDH to be much less favorable than those of the downstream formaldehyde assimilation enzymes HPS and PHI [214, 177, 73, 264], which has led to a search for catalytically more active homologs [226, 210]. The *in vivo* comparison of the performance of these enzymes has typically involved a labor-intensive experimental set-up: After growth and induction of MDH expression in the strains of interest, cultures are pelleted, washed several times, and re-suspended in carbon-free medium. Methanol is then added to the culture, and supernatant samples are taken periodically and assessed for formaldehyde concentration by the NASH assay [215]. Methanol conversion rates are calculated from the initial velocity of formaldehyde production,

215

normalized to the OD of the culture [214]. Because of the low sensitivity of the NASH assay, the endogenous glutathione-dependent formaldehyde dehydrogenase *frmA* must first be knocked out. This process is low-throughput and does not lend itself to the simultaneous comparison of many variants. In their efforts to evolve a better MDH, Wu and coworkers [210] developed a 96-well plate-based assay where the NASH reagent is added directly to the culture upon methanol addition. However, they had to then wait 4 hours for sufficient signal to develop, and separate the supernatant from the cells. Rapid methodologies to assess *in vivo* MDH activity with minimal experimental intervention would therefore be valuable in both the bioprospecting and directed evolution of novel MDH candidates.

To test whether our formaldehyde reporter could meet this need, we cloned several candidate MDHs into pTR48 under control of the arabinose promoter. MDH1 and MDH2 from *Bacillus methanolicus* were chosen as the major isoenzyme expressed during methylotrophic growth and the most active candidate in *E. coli* identified in the seminal work by Muller et al, respectively [214]. To assess the importance of the activator protein (ACT), we also generated a construct that co-expressed MDH2 and ACT. This protein is a nudix hydrolase which, at least *in vitro*, hydrolzes the nicotinamide mononucleotide moiety of the NADH cofactor of MDH, leading to a drastic reduction in the K_m for methanol [70, 220]. A role for this modification *in vivo* has yet to be confirmed, but has been shown *in vitro* for a number of Type III alcohol dehydrogenases [265]. The alcohol dehydroengase *adhA* from *Corynebacterium glutamicum* was also included in the panel, because of the very low reported K_m of 3 mM [227]. We also included the MDH from *Geobacillus stearothermophilus* reported by Whitaker et al [226, 266], and both the WT and evolved variants of MDH2 from *Cupriavidus necator* reported by Wu et al [210]. Finally we included a negative control that expressed mCherry in place of a methanol dehydrogenase. To the best of our knowledge, this is the first time that this broad range of MDH variants has been compared head-to-head. Preliminary work established that, with induction of MDH expression at the same time as methanol addition, 2 hours of further incubation was sufficient for response saturation (Figure 6-5), therefore this timepoint was chosen

for end-point fluorescence analysis. This allowed us to incubate the plates with the assays on a Boekel Jitterbug 2.0, which can achieve much higher agitation rates and oxygen mass transfer than the spectrophotometer, in support of the need to keep the NAD/NADH ratio high to provide a thermodynamic driving force for MDH activity [202].
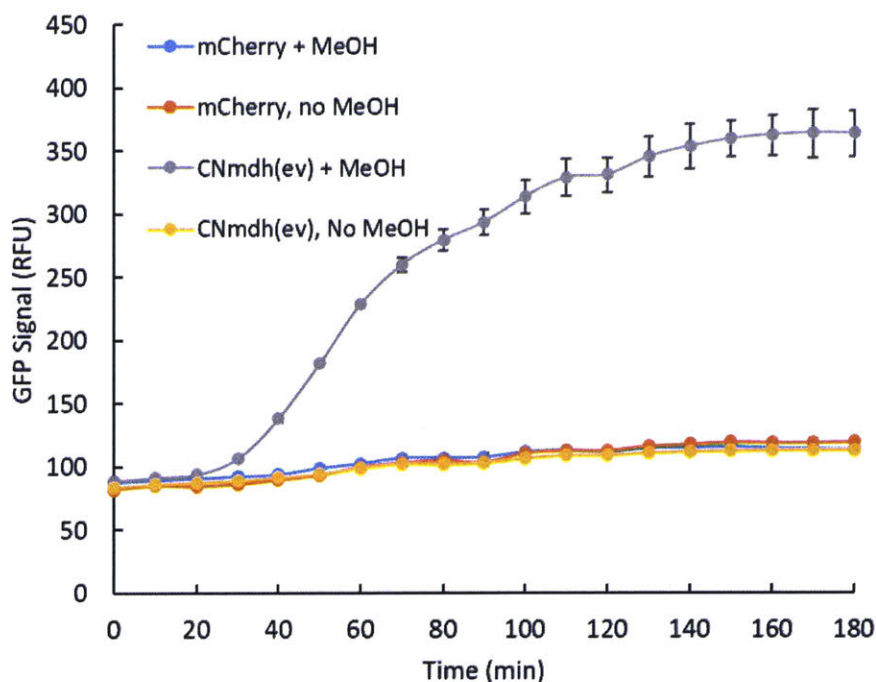
Figure 6-5: **Timecourse of reporter in response to formaldehyde generated *in situ* with MDH.** Cells harboring either mCherry (negative control) or the evolved mdh2 from *C. necator* were incubated with or without 500 mM methanol and fluorescence monitored for 3 hours

After 2 hours, signal was detected for all variants except MDH1 (Figure 6-6). The strongest signal, from the evolved variant of *C. necator* MDH2, showed over a 20-fold increase in signal compared to the control, clearly demonstrating the ability to rapidly detect MDH activity without deletion of the endogenous formaldehdye detoxification pathway. In general, the differences in fluorescence intensity between variants matched previous reports: It was previously shown that Mdh2 from *B. methanolicus* outperforms Mdh1 (30 mU mg^{-1} vs. 1.7 mU mg^{-1}), and that co-expression of Act in *E. coli* does not change the *in vivo* activity [214]. This is reflected in our data

by the 9-fold difference in signal between Mdh2 and Mdh1 at the highest methanol concentration, and the absence of any difference between Mdh2 and Mdh2+act. It should be noted that although the Mdh1 strain showed no GFP signal after 2 hours, fluorescent signal above the control could be seen upon further incubation for a total of 8 hours, long after signals from the other variants had reached saturation, reflecting the extremely slow but still detectable rate of methanol oxidation. Whitaker et al showed that the Mdh from *G. stearothermophilus* outperforms the Mdh2 from *B. methanolicus* at low concentrations (60 mM) of methanol [226]. This is also reflected in our data, where at 50 mM methanol the former shows a fluorescent intensity 2.8-fold higher than the latter. As expected, the evolved variant of the the *C. necator* enzyme outperformed the WT version. Given the reportedly low K_m of the AdhA from *C. glutamicum*, we were suprised to see no improvement over Mdh2 from *B. methanolicus*. In their work, Wu et al purified this enzyme and found a higher K_m than previously reported, of 100 mM [210], which explains this discrepancy. Our assay identified the evolved variant from *C. necator* and the mdh from *G. stearothermophilus* as the most promising MDHs for engineering synthetic methylotrophy, based on their ability to rapidly produce formaldehyde at low concentrations of methanol, but could not distinguish between the two. Taken together, the data presented here show that our assay can be used to detect *in vivo* formaldehyde production by MDH in a WT background and faithfully reconfirms comparisons made in previous literature. From inoculation to assay completion, the procedure requires a total of 5 hours, and only one intervention to add the inducer and methanol. This a significant simplification over the current state of the art, and should enable comparison of different MDH candidates in a high-throughput manner with instrumentation available in most laboratories.

It is interesting to note that, in the WT strain used by Müller to express *B. methanolicus* Mdh2 [214], the steady state formaldehyde level measured by NASH assay after addition of 1M methanol was approximately 10 μM, which is right at the limit of detection of our GFP-based assay, and yet we detect a very strong GFP signal for this strain. The key difference is that the LOD determined above is for
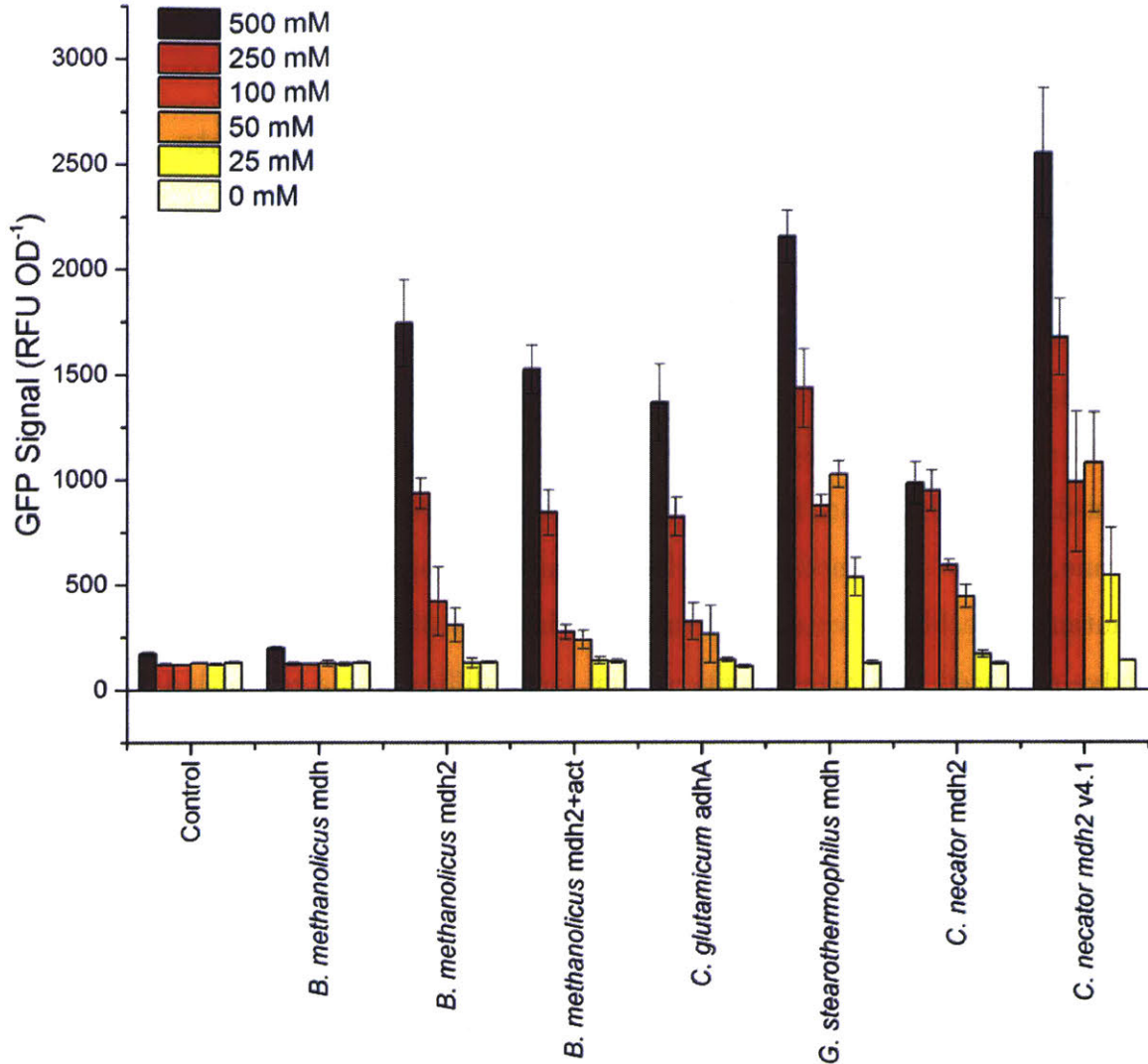
Figure 6-6: **Comparison of MDH variants by GFP formaldehyde reporter.** GFP fluorescence measured 2 hours after induction of MDH expression and addition of various concentrations of methanol. Error bars represent standard deviations for three individual colonies.

the single addition of a bolus of formaldehyde to the culture. Due to the presence of the glutathione detoxification system and the high reactivity of formaldehyde, this formaldehyde is rapidly eliminated, and the continued synthesis of FrmR leads quikcly to deactivation of GFP transcription and low sensitivity. By contrast, in the MDH assay, there is a continual production of formaldehyde, such that de-repression continues and the GFP signal is amplified. Thus the high sensitivity of our reporter system in this context is based on the ability to report essentially the integral of

formaldehyde concentration over time.

6.4.4 Toward Directed Evolution of MDH Variants

Having demonstrated that the formaldehyde biosensor could be used to discriminate between MDH candidates based on their kinetic properties, we were interested in evaluating the potential of the sensor for the directed evolution of novel variants using high-throughput technologies such as FACS or PACE [267]. For these approaches to be successful, the assay must be able to discriminate between candidates within a mixed population. Since formaldehyde is known to diffuse rapidly across the cell membrane, we were concerned that the lack of spatial segregation of high- and low- activity mutants in a library could lead to the enrichment of cheaters, where the formaldehdye produced by a high-activity variant could diffuse into a cell with a low-activity variant and activate the reporter. To assess this possiblity, we co-inoculated a culture with two strains: one carrying pTR48mdh2, and one carrying pTR48mCherry, which expresses mCherry instead of *mdh2*. Both strains contained the reporter plasmid pTR47m4GFP, but only the one with *mdh2* should be able to produce formaldehyde from methanol and activate the reporter. If formaldehyde diffusion can activate the reporter in cells not expressing MDH, we would expect to see GFP signal in cells expressing mCherry after treatment with methanol. If not, the mCherry cells should all show no GFP fluorescence. Cells were grown under the same conditions as before, 10 mM arabinose and 500 mM methanol were added to the culture, and the population was analyzed by flow cytometry after 2 hours. As shown in Figure 6-7, the mCherry+ cells showed mean GFP fluorescence similar to the mCherry- cells, indicating that formaldehdye produced in one cell was able to activate the reporter in another.

In order to prevent this cross-talk, we devised two strategies to reduce the extracellular formaldehyde concentration: 1) addition of NAD-dependent formaldehyde dehydrogenase (FaDH) from *Pseudomonas sp.* and NAD^+, to enzymatically oxidize the formaldehyde to formate; 2) addition of glutathione as a formaldehyde scavenger. Addition of the FaDH and NAD^+ was unsuccessful, possibly due to the breakdown of

the enzyme in the supernatant or unfavorable reaction conditions (data not shown). In contrast, addition at induction of increasing concentrations of glutathione, from 1 to 10 mM, increased the difference in GFP fluroescence between the mCherry$^+$ and mCherry$^-$ cells. At 10 mM, the mCherry$^+$ cells showed only background levels of GFP, indicating complete sequestration of formaldehyde in the supernatant as the glutathione adduct. The addition of high levels of glutathione negatively impacted cell growth, with the OD at analysis of the 10 mM culture roughly half that of the 0 mM control. Despite the growth defect, the GFP fluorescence was roughly the same, indicating that the glutathione had no impact on the activation of the reporter or the activity of MDH. These results clearly show the potential for cheaters in this assay, and establish that glutathione should be added as a formaldehyde sink if this assay is employed for directed evolution of MDH activity.
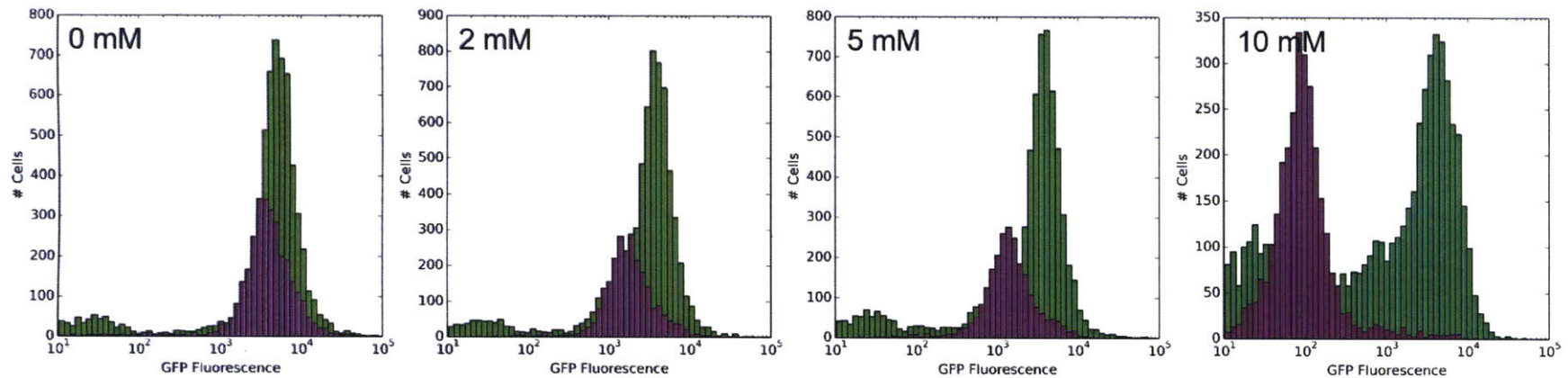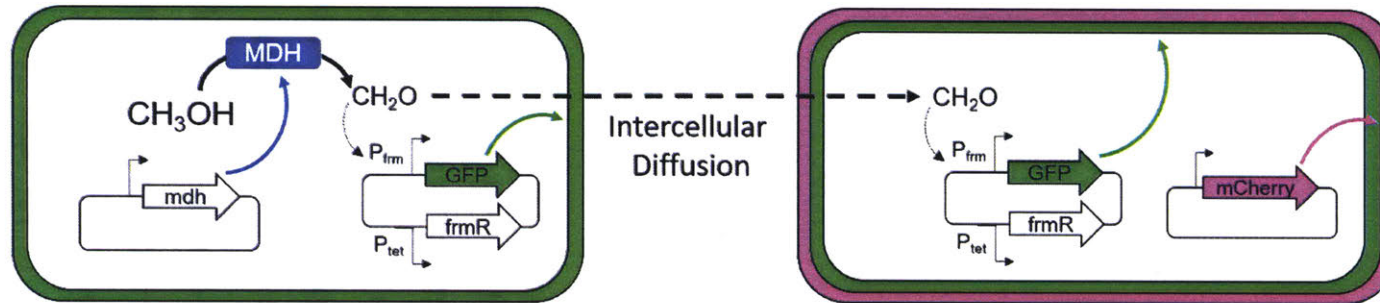
Figure 6-7: **Analysis and prevention of diffusion-mediated cross-talk via FC.** Top: Schematic showing potential false activation of reporter in cells with no mdh (right) by diffusion from cells with active mdh (left) in a mixed culture. Cells that are only GFP$^+$ are true formaldehyde producers, whereas cells that are both mCherry$^+$ and GFP$^+$ are cheaters activated by diffusion. Bottom: Analysis of GFP fluorescence in mdh-containing cells (green bars) and mCherry-containing cells (pink bars) by FC with varying concentrations of glutathione added as an extracellular formaldehyde sink to prevent diffusion.

6.4.5 Biosensor-Based Pathway Tuning

Since formaldehyde is highly toxic, strains engineered to metabolize methanol must carefully balance the production of formaldehyde by MDH with its consumption by HPS and PHI. This control can be engineered at the transcriptional level by varying the induction of the upstream and downstream pathways, and assessing the build-up of formaldehyde under different induction conditions. Here we used the formaldehyde reporter to facilitate high-throughput testing of different induction concentrations. The evolved MDH from *C. necator* was placed under the control of the T7 promoter, and HPS and PHI from *B. methanolicus* as an operon under control of the p_{tet} promoter, in a single vector. Formaldehyde was measured using the biosensor plasmid pTR59GFP, which constitutively expresses *frmR* and therefore has no cross-talk with the tet promoter in the methanol plasmid. Cells were grown on xylose to provide non-limiting levels of Ru5P. Expression was induced at early exponential phase and 500 mM methanol added at induction. GFP signal was measured after 2 hours, and again noramlized to OD_{600}. The results are shown in Figure 6-8. Without induction of either MDH or HPS/PHI, only background GFP signal was detected. Upon induction with IPTG, the GFP signal rose to a maximum of 2500 (14-fold increase) at 50 μM, before falling again at higher induction levels. This fall in signal is likely attributable to the low solubility of MDH: Increased induction leads to no measurable increase in the activity in assays of crude lysates, but a marked increase in the amount of the enzyme in the insoluble fraction, and a significant decrease in cell growth (Figure 5-24). The slow growth under these conditions is the likely cause of reduced GFP synthesis. At 25 μM IPTG, low-level induction of HPS/PHI with 10 ng mL^{-1} is sufficient to reduce the formaldehyde concentration to background levels. At 50 μM IPTG, the higher MDH activity require an increase in HPS/PHI induction to compensate. At 100 μM IPTG and above, even full induction of the downstream pathway with 200 ng mL^{-1} is insufficient to reduce the formaldehyde concentration to background levels. This could be because of two reasons: 1) High level IPTG induction reduces the expression level from the p_{tet} promoter, as shown in

experiments where IPTG and tet controlled GFP and RFP respectively (Figure 6-9), such that the HPS/PHI activity is too low to balance formaldehyde generation. 2) At high levels of T7 induction, cells must produce high levels of nucleotides to support mRNA synthesis. As ribose-5-phosphate is the precursor for nucleotide biosynthesis, this increased level of mRNA production might deplete the pentose phosphate pools, leading to insufficient ribulose-5-phosphate to support formaldehdye assimilation, regardless of HPS activity. Whatever the reason, the results here demonstrate the high sensitivity of the reporter through its ability to detect formaldehyde even in strains engineered for its consumption, and without deletion of the endogenous detoxification system. They also demonstrate how the reporter can be used to optimize the relative expression level of MDH and HPS/PHI to maximize flux while minimizing expression burden. In this case, 50 μM IPTG provides maximum formaldehyde production in the absence of ATc, and the addition of 40 ng mL^{-1} is sufficient to allow its complete consumption through the engineered pathway.

In this experiment, the expression level was statically controlled through chemical induction. An emerging trend in synthetic biology and metabolic engineering is the dynamic control of metabolism, where a sensor of an intermediary metabolite actuates the expression or repression of other pathway genes to control the rate of its consumption, in imitation of the regulation of many natural pathways. In principle, the formaldehyde reporter engineered in this work could be converted into a dynamic sensor-regulatory system, by using the *frm* promoter to drive HPS and PHI expression. This would help eliminate the need for inducers, and could be helpful eventually in a bioreactor setting, where spatial gradients in methanol concentration could require cells to respond dynamically to perturbations in formaldehyde concentration. The results provided here are foundational for this approach, because they show that the dynamic range of the biosensor is well-matched to the range of formaldehdye concentrations detected in cells engineered for methanol assimilation.
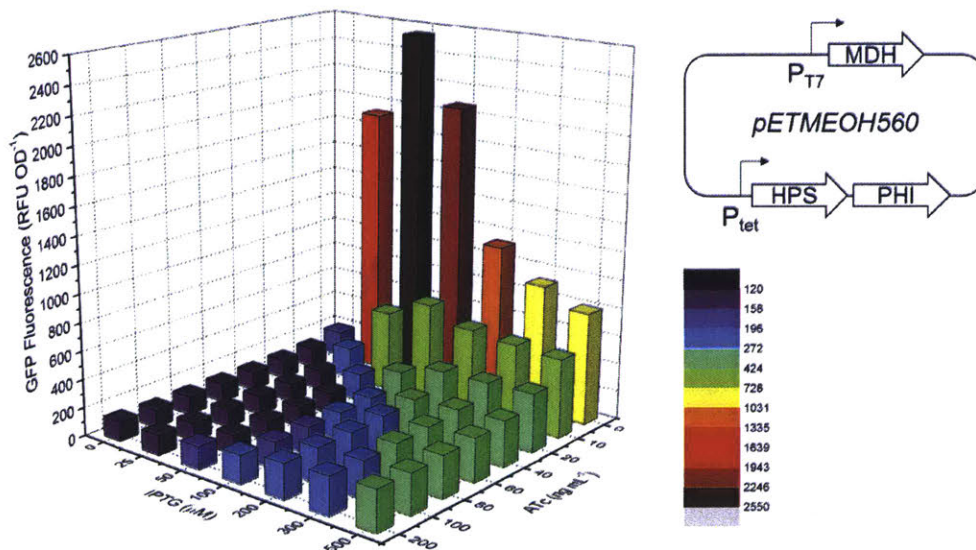
Figure 6-8: **High-throughput pathway balancing monitored by the formalde-hyde reporter.** Normalized GFP fluorescence in cells carrying pETMEOH560 2 hours after induction with various concentrations of ATc and IPTG, and treatment with 500 mM methanol. IPTG induction controls *mdh* expression, while ATc in-duction controls *hps* and *phi* (top right). Bars represent means of three biological replicates. RFU, raw fluorescence units; OD, optical density at 600 nm.
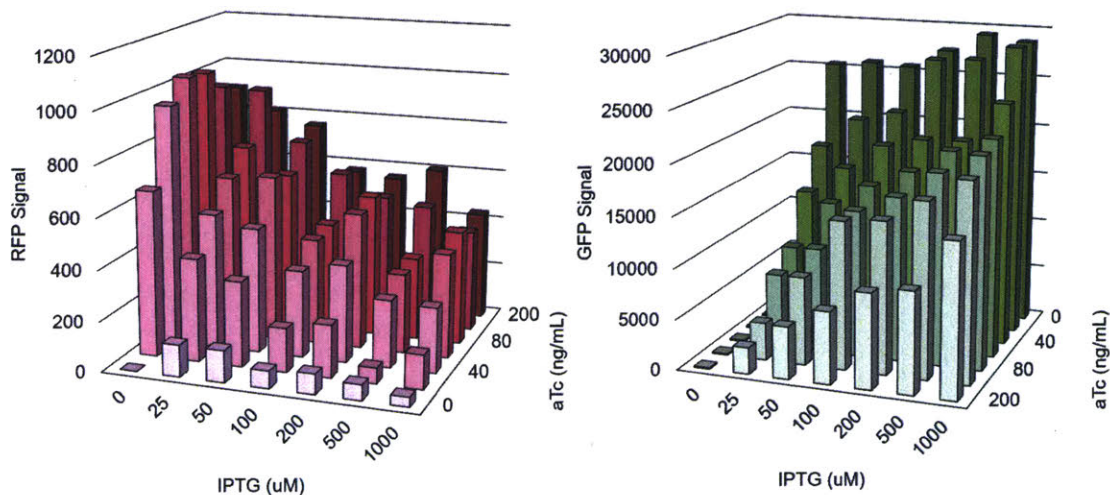
Figure 6-9: **Cross-Talk Associated with IPTG and ATc.** Cells carrying pET-sGFP-tet-RFP were induced with varying concentrations of IPTG and ATc, and RFP fluorescence (left) and GFP fluorescence (right) measured after 2 hours. Bars represent the average of three biological replicates.

225

6.5 Conclusions

In this work, we made significant improvements to a formaldehyde biosensor, and demonstrated its utility in efforts to establish synthetic methylotrophy. Engineering of the promoter binding site and regulatory architecture led to dramatically improved sensitivity over the previous version. Our system could detect the addition of 3 μM exogenous formaldehdye with a 10-fold increase in signal, whereas the previously reported sensor required 2 mM formaldehyde and the addition of Triton to permeabilize the cells. We showed our reporter could discriminate between MDH candidates, and with the addition of glutathione as a formaldehyde sink could detect differences within a mixed population. Finally, we showed that the dynamic range of the sensor matches well with physiological formaldehyde concentrations in strains engineered for methanol assimilation, allowing high-throughput optimization of induction conditions for the upstream and downstream pathway. These proof-of-concept experiments demonstrate the utility of the biosensor, and pave the way for directed evolution of MDH, as well as the implementation of dynamic control strategies in synthetic methylotrophy.

Chapter 7

Conclusions and Future Directions

7.1 Conclusions

Though the development of a reproducible protocol for the transformation of *Moorella thermoacetica* was not achieved, these investigations led to several concrete results: First, a quantitative assay for analyzing the delivery of fluorescently labeled DNA via electroporation was developed. The protocol is applicable to the development of genetic tools for any organism, and has already found utility elsewhere in our group for assessing electroporation of *Thiobacillus denitrificans*, and elsewhere at MIT. Second, the specificity of one of *Moorella's* endonucleases has been uncovered, and its effect on transformation can be evaded either by pre-methylation as demonstrated, or by avoidance of the recognition sequence in shuttle vectors. Finally, it was shown that the plasmid pQexp can replicate autonomously in *M. thermoacetica*.

A CRISPRi system for the related organism *Clostridium ljungdahlii* was developed, and shown to enable reduction of target mRNA levels by greater than 96%. The system was designed to allow easy change of the N20 targeting region and multiplexing of several targets with simple cloning steps, and will therefore find broad utility in the acetogen research community. Identification of leaky dCas9 expression revealed an inherent design problem of the CRISRPi strategy, which in this case was largely offset by the identification of a new, tightly regulated promoter for *C. ljungdahlii*. The physiological effect of *pta* downregulation was examined, and found to

lead to impaired growth and low titers of both acetate and ethanol, suggesting the presence of hitherto unknown regulatory or metabolic mechanisms for the balancing of carbon flux. Strains producing mM quantities of 3HB were generated, and rationally improved by codon optimization and RBS tuning. These plasmids provide a testbed for identifying native genes whose downregulation promotes increased acetyl-CoA flux into a heterologous pathway, a general question in acetogen metabolic engineering.

In light of the availability of cheap natural gas, pathways for the conversion of methane to n-butanol were evaluated based on theoretical yield. It was shown that, while the MCR-catalyzed reaction allows conservation of more energy from methane than the MMO-reaction in this single transformation, other factors in the pathway mean that this does not translate to an increased yield overall. This argument is not dependent on the rate of methane oxidation, so that even if drastic rate increases can be achieved in anaerobic methane oxidation, there is no benefit to using this pathway over the aerobic pathway. Efforts to develop a methane-based bioprocess should therefore show preference for aerobic methanotrophy.

E. coli strains capable of methanol assimilation were generated, and pathway limitations systematically addressed by a combination of theoretical analysis, traditional metabolic engineering approaches including labeled isotope experiments, and chemical biology-inspired approaches. This analysis revealed flux leakage from the RuMP pathway as the primary bottleneck, and strategies were developed to redirect flux into this pathway and enable improved formaldehyde assimilation. With sufficient Ru5P regeneration, it was shown that methanol oxidation proceeded largely irreversibly, counterindicating approaches to modulate pathway thermodyanmics. Further analysis using kinetic isotope effects demonstrated that, with this bottleneck mitigated, flux became limited by the intrinsic kinetics of MDH, motivating directed evolution and bioprospecting to find better candidate isoenzymes for this conversion. The insight and analysis developed in this work will be critical to ongoing efforts to establish synthetic methylotrophy.

A formaldehyde biosensor was developed that allowed μM level detection of this intermediate. In support of the aforementioned need to improve MDH activity, the

sensor was examined as a platform for directed evolution, and a strategy developed to guard against cheaters by limiting intercellular formaldehyde diffusion. The utility of the sensor was further highlighted by the high-throughput screening *in vivo* of MDH candidates, and the tuning of upstream and dowsntream pathway expression to minimize formaldehyde toxicity and heterologous expression burden. The sensor architecture is currently in use in our collaborator's lab to evolved MDH activity via Phage-Assisted Continuous Evolution (PACE), with promising results thus far.

7.2 Recommendations for Future Work

7.2.1 *Moorella* transformation

The quest for robust transformation of *M. thermoacetica* consumed two years of my PhD work, and resulted in sparse glimmers of success. During the course of these experiments, another group published a transformation protocol based on 5-FOA resistance [98]. Despite incorporating a transformation vector almost identical to theirs (pMTK1) in our panel of plasmids, and pre-methylating vectors via the same PAMS strategy, we were unable to consistently reproduce this result: The only *pyrF* mutants isolated were partial deletions, not useful for the insertion of heterologous genes. We had occasional success with one of the replicating vectors that we designed (pQexp-kanR), but again, positive results were sporadic. At around the same time, a highly optimized transformation protocol for the related acetogen *C. ljungdahlii* was published [91]. Since the ultimate focus of my thesis work was to be the metabolic engineering of syngas-fermenting organisms, the choice of host was largely immaterial, and the ability to easily reproduce *C. ljlungdahlii* transformation convinced us that efforts should instead be focused on that organism. The flow-cytometry assay for detection of fluorescent plasmid delivery developed during this work is nevertheless a useful technique that could be harnessed in the development of transformation tools for other organisms.

In the time since I worked on *Moorella* transformation, technology has advanced,

and achieving this goal might now be more feasible. Critically, the immensely popular CRISPR-Cas9 system could be useful for the selection of correct recombinants, as was demonstrated recently by the highly efficacious use of this system in generating several gene knockouts in *C. ljungdahlii* [94, 268] and other Gram positives [161, 269]. SMRT sequencing [138, 140, 139], which was a nascent technology when I worked on this project, is now a convenient service that could identify the remaining endonuclease recognition sites in *Moorella*. Combined with restriction-independent cloning methods, such as Gibson assembly, to rapidly generate transformation vectors missing these sites, this could overcome the barrier posed by DNA degradation. The list of vectors screened could also be expanded to take advantage of the modular system of *Clostridial* vectors developed in the Minton lab [158]. In several experiments with pQexp-kanR, kanamycin resistance arose in control cells that had been mixed with plasmid but not electroporated, leading us to suspect a possible role for natural competence in *Moorella* transformation [270]. Natural competency has been found across a wide array of bacteria, including the related Gram positive thermophile *Thermoanaerobacterium saccharolyticum* [271]. Inspection of the genome revealed *ComEA* and *ComEC* orthologs (Moth_0570 and Moth_0575, respectively), adding credibility to this notion. Further investigation to confirm natural competence and identify the environmental signals (if any) that trigger the ability to take up DNA could be fruitful. Combining all of these resources could lead to significant strides forward in domesticating the organism that formed the basis of the biochemical elucidation of the Wood-Ljungdahl pathway, that has so far remained elusive.

7.2.2   CRISPRi for Metabolic Engineering of *C. ljungdahlii*

The CRISPRi system developed here is an exciting enabling technology for acetogen metabolic engineering, but more extensive characterization should be carried out. In particular, in this thesis the tet-controlled system was fully induced with 100 ng mL^{-1} aTc. Given the toxicity of aTc in *Clostridia* and the low concentration of dCas9 required for downregulation, the effect of various aTc concentrations on repression should be evaluated, with the aim of reducing the toxicity of the system and exploring

the dynamic range of regulation. These experiments would be facilitated by use of an easily assayed reporter. A fluorescent protein functional in *C. ljungdahlii* was recently reported, but the signal was weak, and required development of media with reduced flavin content to eliminate background autofluorescence [92]. Excitingly, a new class of fluorescent proteins derived from phycobiliproteins has been developed whose fluorescence is dependent on the binding of biliverdin, rather than an oxygen-dependent protein backbone oxidative reaction cascade, and are therefore suitable to anaerobic environments [272]. Efforts to express a representative protein, smURFP [273], as a reporter protein for CRISPRi tuning are ongoing. Alternatively, gusA could be used, with assessment of activity conducted in a mid-throughput manner by X-gluc staining in permeabilized cells (as demonstrated in this thesis) to avoid the laborious process of cell lysis, quantitative enzyme assay, and BCA standardization.

Beyond exploring the tunability of the system, the 3HB production developed here should be used to assess the impact of the downregulation of various genes on heterologous production. Here it was shown that *pta* downregulation led to a severe growth defect. It would be interesting to explore whether induction of the knockdown later in fermentation, after the production of sufficient biomass, could improve 3HB titers. Alternatively, other genes with putative roles in carbon flux could be targeted. As there is significant redundancy in the *C. ljungdahlii* genome, it may be that combinations of gene knockdowns are required to generate the desired production phenotype. The CRISPRi strain desigend here is amenable to multiplexing, so this should be feasible. Compared to *E. coli*, genetic engineering in *C. ljungdahlii* is still slow, and transformation efficiencies are low. So while construction of CRISPRi libraries is simple, transforming a library into *C. ljungdahlii* is not yet possible. Methods to improve transformation efficiency would therefore speed development of metabolic engineering in this host and provide access to the advantages of modern high-throughput techniques that have transformed metabolic engineering in prototypical hosts.

7.2.3 Synthetic Methylotrophy

This thesis identified Ru5P regeneration and MDH kinetics as the primary limitations in synthetic methylotrophy in *E. coli*. With GAPDH identified as a target for down-regulation, and activation of the SBPase variant shown to increase Ru5P regeneration, more elaborate genetic strategies can now be readily employed to control expression of these enzymes. This approach will require dynamic metabolic control strategies, since reduced expression of these enzymes during normal growth will result in slow-growth phenotypes. Fortunately, with the wealth of tools available, this will be comparatively simple to achieve. An interesting possibility here is to use the formaldehdye biosensor developed in this thesis to screen in a high-throughput manner genetic modifications to central metabolism that lead to higher Ru5P concentrations in resting cells, and thus reduce the formaldehyde concentration. Experiments to adapt the biosensor to resting cells, which do not have sufficient substrate to synthesize GFP in response to formaldehyde, are underway. The strategy is to take advantage of the diffusivity of formaldehyde, and contain the reporter plasmid in a strain that can consume sucrose, and co-incubate these cells with strains engineered for methylotrophy (and which cannot consume sucrose), in media containing sucrose and methanol as the only carbon sources.

By contrast, evolving an MDH candidate that is 10-fold more active presents a much more difficult technical challenge. We are already running PACE to address this, and have already dramatically improved the activity of MDH2 from *B. methanolicus*. However, there is a more funamdental question here: Given the poor *in vitro* and *in vivo* kinetics of MDH in *E. coli*, how does this same enzyme enable methanol consumption at 1 g g^{-1} hr^{-1} in the native host? The temperature is higher, but the difference between 37°C and 45°C likely corresponds to an increase of at most 2-fold in rate, and *B. methanolicus* can grow at 37°C with a resonable doubling time of approximately 5 hours. We have shown that thermodynamic force, which is greater at higher temperature, is not a bottleneck in *E. coli*, so we cannot use that as the explanation. I believe that the answer to this fundamental question will provide the

quantum leap needed to establish industrially relevant rates of methanol consumption when applied to *E. coli*.

7.2.4 Closing Remarks

The central question posed in this thesis was whether it is better to develop metabolic engineering tools and strategies for strains that already metabolize the desired carbon substrate, or to import these pathways into heterologous hosts with more advanced engineering technologies. After five years gaining experience on both sides of this debate, a universal answer remains elusive. The methanol assimilation pathway consists of only three genes with well described biochemistry, and yet has taken an equal number of years simply to establish the bottlenecks. Adapting the CRISPRi technology to *C. ljungdahlii*, on the other hand, was relatively simple using the basic genetic tools that had already been established, but revealed the depths of our lack of knowledge of metabolic regulation in this organism. Establishing these tools completely from scratch, as with *Moorella thermoacetica*, has been an exasperating process, both for our lab and others. Regardless, the journey has been an interesting one, and tackling the breadth of projects presented in this thesis has introduced me to a wide range of techniques, and an exceptionally broad community of scientists who have shaped the way I view the pursuit of science, and whom I am thankful to consider peers.

Bibliography

[1] Steven Chu and Arun Majumdar. Opportunities and challenges for a sustainable energy future. *Nature*, 488(7411):294–303, aug 2012.

[2] John Valentine, John Clifton-Brown, Astley Hastings, Paul Robson, Gordon Allison, and Pete Smith. Food vs. fuel: The use of land for lignocellulosic 'next generation' energy crops that minimize competition with primary food production. *GCB Bioenergy*, 4(1):1–19, 2012.

[3] Gea Guerriero, Jean Francois Hausman, Joseph Strauss, Haluk Ertan, and Khawar Sohail Siddiqui. Lignocellulosic biomass: Biosynthesis, degradation, and industrial utilization. *Engineering in Life Sciences*, 16(1):1–16, 2016.

[4] M Y Menetrez. An overview of algae biofuel production and potential environmental impact. *Environ Sci Technol*, 46(13):7073–7085, 2012.

[5] Annual Energy Outlook 2013 (Early Release). Technical report, 2013.

[6] The world ' s largest gas-to-liquids plant is now fully online, 2011.

[7] Rajab Khalilpour and I.a. Karimi. Evaluation of utilization alternatives for stranded natural gas. *Energy*, 40(1):317–328, apr 2012.

[8] Golden Rules for the Golden Age of Gas. Technical report, 2012.

[9] Jeff Tollefson. Oil boom raises burning issues. *Nature*, 495(7441):290–1, mar 2013.

[10] David a. Wood, Chikezie Nwaoha, and Brian F. Towler. Gas-to-liquids (GTL): A review of an industry offering several routes for monetizing natural gas. *Journal of Natural Gas Science and Engineering*, 9:196–208, nov 2012.

[11] R. J. Conrado and R. Gonzalez. Envisioning the Bioconversion of Methane to Liquid Fuels. *Science*, 343(6171):621–623, feb 2014.

[12] Marina Kalyuzhnaya, Aaron W Puri, and Mary E Lidstrom. Metabolic engineering in methanotrophic bacteria. *Metabolic engineering*, 29:142–152, mar 2015.

[13] Andrea M Ochsner, Frank Sonntag, Markus Buchhaupt, Jens Schrader, and Julia A Vorholt. Methylobacterium extorquens: methylotrophy and biotechnological applications. *Applied microbiology and biotechnology*, 99:517–534, 2015.

[14] George A. Olah, Alain Goeppert, and G. K Surya Prakash. Beyond Oil and Gas: The Methanol Economy: Second Edition. *Beyond Oil and Gas: The Methanol Economy: Second Edition*, pages 1–334, 2009.

[15] Jens Schrader, Martin Schilling, Dirk Holtmann, Dieter Sell, Murillo Villela Filho, Achim Marx, and Julia a Vorholt. Methanol-based industrial biotechnology: current status and future perspectives of methylotrophic bacteria. *Trends in biotechnology*, 27(2):107–15, mar 2009.

[16] Jonas E N Müller, Tonje M B Heggeset, Volker F Wendisch, Julia a Vorholt, and Trygve Brautaset. Methylotrophy in the thermophilic Bacillus methanolicus, basic insights and application for commodity production from methanol. *Applied microbiology and biotechnology*, 99(2):535–51, jan 2015.

[17] Arren Bar-Even. Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways. *Biochemistry*, 55(28):3851–3863, 2016.

[18] Fung Min, Michael Kopke, Sean Dennis, Fung Min Liew, Michael Köpke, Séan Dennis Simpson, Fung Min, Michael Kopke, and Sean Dennis. Gas Fermentation for Commercial Biofuels Production. *Liquid, Gaseous and Solid Biofuels - Conversion Techniques*, pages 125–173, 2013.

[19] Pradeep Chaminda Munasinghe and Samir Kumar Khanal. Biomass-derived syngas fermentation into biofuels: Opportunities and challenges. *Bioresource technology*, 101(13):5013–22, jul 2010.

[20] Michael Köpke, Christophe Mihalcea, Jason C Bromley, and Séan D Simpson. Fermentative production of ethanol from carbon monoxide. *Current opinion in biotechnology*, 22(3):320–5, jun 2011.

[21] Ryoji Mitsui, Masahiro Omori, and Hideaki Kitazawa. Formaldehyde-Limited Cultivation of a Newly Isolated Methylotrophic Bacterium , Methylobacterium sp . MF1 : Enzymatic Analysis Related to C 1 Metabolism. *Journal of Bioscience and Bioengineering*, 99(1):18–22, 2005.

[22] Ludmila Chistoserdova, Marina G Kalyuzhnaya, and Mary E Lidstrom. The Expanding World of Methylotrophic Metabolism. *Annual Review of Microbiology*, pages 477–499, 2009.

[23] Jan T Keltjens and GD Vogels. Methanogenesis. In James Ferry, editor, *Methanogenesis*, pages 253–303. Chapman & Hall, 1993.

[24] Lorenz Adrian and Frank E Löffler. *Organohalide- Respiring Bacteria*. Springer, 2016.

[25] Yan Kung and Catherine L Drennan. Nickel-Containing Carbon Monoxide Dehydrogenase and Acetyl-CoA Synthase. *One-Carbon Chemistry of Nickel-Containing Carbon Monoxide Dehydrogenase and Acetyl-CoA Synthase*, (10):121–148, 2017.

[26] Arren Bar-Even, Elad Noor, and Ron Milo. A survey of carbon fixation pathways through a quantitative lens. *Journal of experimental botany*, 63(6):2325–42, mar 2012.

[27] Alan G Fast and Eleftherios T Papoutsakis. Stoichiometric and energetic analyses of non-photosynthetic CO2-fixation pathways to support synthetic biology strategies for production of fuels and chemicals. *Current Opinion in Chemical Engineering*, pages 1–16, sep 2012.

[28] Elizabeth Pierce, Gary Xie, Ravi D Barabote, Elizabeth Saunders, Cliff S Han, John C Detter, Paul Richardson, Thomas S Brettin, Amaresh Das, Lars G Ljungdahl, and Stephen W Ragsdale. The complete genome sequence of Moorella thermoacetica (f. Clostridium thermoaceticum). *Environmental microbiology*, 10(10):2550–73, oct 2008.

[29] Amaresh Das and Jeroen Hugenholtz. Structure and function of a menaquinone involved in electron transport in membranes of Clostridium thermoautotrophicum and Clostridium thermoaceticum. *Journal of bacteriology*, 1989.

[30] Michael Köpke, Claudia Held, Sandra Hujer, Heiko Liesegang, Arnim Wiezer, Antje Wollherr, Armin Ehrenreich, Wolfgang Liebl, Gerhard Gottschalk, and Peter Dürre. Clostridium ljungdahlii represents a microbial production platform based on syngas. *Proceedings of the National Academy of Sciences of the United States of America*, 107(29):13087–92, jul 2010.

[31] Anja Poehlein, Silke Schmidt, Anne-Kristin Kaster, Meike Goenrich, John Vollmers, Andrea Thürmer, Johannes Bertsch, Kai Schuchmann, Birgit Voigt, Michael Hecker, Rolf Daniel, Rudolf K Thauer, Gerhard Gottschalk, and Volker Müller. An ancient pathway combining carbon dioxide fixation with the generation and utilization of a sodium ion gradient for ATP synthesis. *PloS one*, 7(3):e33439, jan 2012.

[32] Wolfgang Buckel and Rudolf K Thauer. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation. *Biochimica et biophysica acta*, 1827(2):94–113, feb 2013.

[33] PL Tremblay, Tian Zhang, Shabir Dar, Ching Leang, and Derek R. Lovley. The Rnf Complex of Clostridium ljungdahlii Is a Proton-Translocating Ferredoxin:NAD Oxidoreductase Essential for Autotrophic Growth. *Applied and Environmental Microbiology*, 8(1):1–8, 2013.

[34] Reno Heise, V Muller, and Gerhard Gottschalk. Sodium Dependence of Acetate Formation by hte Acetogenic Bacterium Acetobacterium woodii. *Journal of Bacteriology*, 171(10):5473–5478, 1989.

[35] Fuli Li, Julia Hinderberger, Henning Seedorf, Jin Zhang, Wolfgang Buckel, and Rudolf K. Thauer. Coupled ferredoxin and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from Clostridium kluyveri. *Journal of Bacteriology*, 190(3):843–850, 2008.

[36] Wolfgang Nitschke and Michael J Russell. Redox bifurcations: mechanisms and importance to life now, and at its origin: a widespread means of energy conversion in biology unfolds.... *BioEssays : news and reviews in molecular, cellular and developmental biology*, 34(2):106–9, mar 2012.

[37] Nilanjan Pal Chowdhury, Amr M Mowafy, Julius K Demmer, Vikrant Upadhyay, Sebastian Koelzer, Elamparithi Jayamani, Joerg Kahnt, Marco Hornung, Ulrike Demmer, Ulrich Ermler, and Wolfgang Buckel. Studies on the mechanism of electron bifurcation catalyzed by electron transferring flavoprotein (Etf) and butyryl-CoA dehydrogenase (Bcd) of Acidaminococcus fermentans. *The Journal of biological chemistry*, 289(8):5145–57, feb 2014.

[38] Carolyn E Lubner, David P Jennings, David W Mulder, Gerrit J Schut, Oleg A Zadvornyy, John P Hoben, Monika Tokmina-lukaszewska, Luke Berry, Diep M Nguyen, Gina L Lipscomb, Brian Bothner, Anne K Jones, Anne-frances Miller, Paul W King, Michael W W Adams, and John W Peters. Mechanistic insights into energy conservation by flavin-based electron bifurcation. *Nature chemical biology*, (April), 2017.

[39] Johanna Mock, Shuning Wang, Haiyan Huang, Jörg Kahnt, and Rudolf K Thauer. Evidence for a hexaheteromeric methylenetetrahydrofolate reductase in Moorella thermoacetica. *Journal of bacteriology*, 196(18):3303–14, sep 2014.

[40] Shuning Wang, Haiyan Huang, Haiyan Huang Kahnt, Alexander P. Mueller, Michael Köpke, and Rudolf K. Thauer. NADP-Specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in Clostridium autoethanogenum grown on CO. *Journal of Bacteriology*, 195(19):4373–4386, 2013.

[41] Kai Schuchmann and Volker Müller. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nature reviews. Microbiology*, 12(December), nov 2014.

[42] Arren Bar-even. Biochimica et Biophysica Acta Does acetogenesis really require especially low reduction potential? *BBA - Bioenergetics*, 1827(3):395–400, 2013.

[43] Johannes Bertsch and Volker Muller. Co metabolism in the acetogen acetobacterium woodii. *Applied and Environmental Microbiology*, 81(June):AEM.01772–15, 2015.

[44] Ghasem Najafpour and Habibollah Younesi. Ethanol and acetate synthesis from waste gas using batch culture of Clostridium ljungdahlii. *Enzyme and Microbial Technology*, 38:223–228, 2006.

[45] V Svetlichny, T Sokolova, M Gerhardt, M Ringpfeil, N Kostrikina, and G Zavarzin. Carboxydothermus hydrogenoformans gsn. nov., sp. nove., a CO-utilizing Thermophilic Anaerobic Bacterium from Hydrothermal Environments of Kunashir Island. *System. Appl. Microbiol*, 14:254–260, 1991.

[46] Harold L Drake and Steven L Daniel. Physiology of the thermophilic acetogen Moorella thermoacetica. *Research in microbiology*, 155(10):869–83, dec 2004.

[47] Sara Ramió-pujol, Ramon Ganigué, Lluís Bañeras, and Jesús Colprim. Impact of formate on the growth and productivity of Clostridium ljungdahlii PETC and Clostridium carboxidivorans P7 grown on syngas. *International Microbiology*, 17:195–204, 2014.

[48] S. L. Daniel, T. Hsu, S. I. Dean, and H. L. Drake. Characterization of the H2- and CO-dependent chemolithotrophic potentials of the acetogens Clostridium thermoaceticum and Acetogenium kivui. *Journal of Bacteriology*, 172(8):4464–4471, 1990.

[49] Amaresh Das, Zheng-qing Fu, Wolfram Tempel, Zhi-jie Liu, Jessie Chang, Lirong Chen, Doowon Lee, Weihong Zhou, Hao Xu, Neil Shaw, John P Rose, Lars G Ljungdahl, and Bi-cheng Wang. Characterization of a Corrinoid Protein Involved in the C1 Metabolism of Strict Anaerobic Bacterium Moorella thermoacetica. *PROTEINS: Structure, Function and Bioinformatics*, 67:167–176, 2007.

[50] Matthew a Pritchett and William W Metcalf. Genetic, physiological and biochemical characterization of multiple methanol methyltransferase isozymes in Methanosarcina acetivorans C2A. *Molecular microbiology*, 56(5):1183–94, jun 2005.

[51] Andreas Tschech and Norbert Pfennig. Growth yield increase linked to caffeate reductino in Acetobacterium woodii. *Archives of microbiology*, 137:163–167, 1984.

[52] Regina Bache and Norbert Pfennig. Selective Isolation of Acetobacterium woodii on Methoxylated Aromatic Acids and Determination of Growth Yields. *Archives of microbiology*, 130:255–261, 1981.

[53] Bryan P Tracy, Shawn W Jones, Alan G Fast, Dinesh C Indurthi, and Eleftherios T Papoutsakis. Clostridia: the importance of their exceptional substrate

and metabolite diversity for biofuel and biorefinery applications. *Current opinion in biotechnology*, 23(3):364–81, jun 2012.

[54] Verena Hess, José M González, Anutthaman Parthasarathy, and Wolfgang Buckel. Caffeate Respiration in the Acetogenic Bacterium Acetobacterium woodii: a Coenzyme A Loop Saves Energy for Caffeate Activation. *Applied and Environmental Mcrobiology*, 79(6):1942–1947, 2013.

[55] Kai Schuchmann and Volker Müller. Energetics and Application of Heterotrophy in Acetogenic Bacteria. *Applied and environmental microbiology*, 82(14):4056–4069, 2016.

[56] Alan G Fast and Eleftherios T Papoutsakis. Stoichiometric and energetic analyses of non-photosynthetic CO2-fixation pathways to support synthetic biology strategies for production of fuels and chemicals. *Current Opinion in Chemical Engineering*, 1(4):380–395, nov 2012.

[57] Alan G Fast, Ellinor D Schmidt, Shawn W Jones, and Bryan P Tracy. Acetogenic mixotrophy: novel options for yield improvement in biofuels and biochemicals production. *Current Opinion in Biotechnology*, 33:60–72, jun 2015.

[58] Katy J. Jones, Karen Moore, Christine Sambles, John Love, David J. Studholme, and Stephen J. Aves. Draft Genome Sequences of Achromobacter piechaudii GCS2, Agrobacterium sp. Strain SUL3, Microbacterium sp. Strain GCS4, Shinella sp. Strain GWS1, and Shinella sp. Strain SUS2 Isolated from Consortium with the Hydrocarbon-Producing Alga Botryococcus brauni. *Genome Announcements*, 4(1):e01527–15, 2016.

[59] G.-J. Shen, J.-S. Shieh, a. J. Grethlein, M. K. Jain, and J. G. Zeikus. Biochemical basis for carbon monoxide tolerance and butanol production by Butyribacterium methylotrophicum. *Applied Microbiology and Biotechnology*, 51(6):827–832, jun 1999.

[60] Michael Köpke, Christophe Mihalcea, Fungmin Liew, Joseph H Tizard, Mohammed S Ali, Joshua J Conolly, Bakir Al-Sinawi, and Séan D Simpson. 2,3-Butanediol Production By Acetogenic Bacteria, an Alternative Route To Chemical Synthesis, Using Industrial Waste Gas. *Applied and environmental microbiology*, 77(15):5467–75, aug 2011.

[61] Johanna Mock, Yanning Zheng, Alexander P. Mueller, San Ly, Loan Tran, Simon Segovia, Shilpa Nagaraju, Michael Köpke, Peter Dürre, and Rudolf K. Thauer. Energy conservation associated with ethanol formation from H2 and CO2 in Clostridium autoethanogenum involving electron bifurcation. *Journal of Bacteriology*, (July):JB.00399–15, 2015.

[62] Sang Yup Lee, Jin Hwan Park, Seh Hee Jang, Lars K Nielsen, Jaehyun Kim, and Kwang S Jung. Fermentative Butanol Production by Clostridia. *Biotechnology and Biooengineering*, 101(2):209–228, 2008.

[63] Frank R Bengelsdorf, Anja Poehlein, Bettina Schiel-bengelsdorf, and Rolf Daniel. Genome Sequence of the Acetogenic Bacterium Butyribacterium methylotrophicum DSM 3468T. *Genome Announcements*, 4(6):2013–2014, 2016.

[64] J Rodney Quayle and Thomas Ferenci. Evolutionary Aspects of Autotrophy. *Microbiological reviews*, (June):251–273, 1978.

[65] Gregory J Crowther, George Kosály, and Mary E Lidstrom. Formate as the main branch point for methylotrophic metabolism in Methylobacterium extorquens AM1. *Journal of bacteriology*, 190(14):5057–62, jul 2008.

[66] R. Peyraud, P. Kiefer, P. Christen, S. Massou, J.-C. Portais, and J. A. Vorholt. Demonstration of the ethylmalonyl-CoA pathway by using 13C metabolomics. *Proceedings of the National Academy of Sciences*, 106(12):4846–4851, mar 2009.

[67] DJ Leak and Howard Dalton. Growth yields of methanotrophs. *Applied microbiology and biotechnology*, pages 477–481, 1986.

[68] Ludmila Chistoserdova, Sung-Wei Chen, Alla Lapidus, Mary E Lidstrom, and Sung-wei Chen. MINIREVIEW: Methylotrophy in Methylobacterium extorquens AM1 from a Genomic Point of View. *Journal of Bacteriology*, 185(10):2980–2987, 2003.

[69] Nico Arfman, Lubbert Dijkhuizen, Gudrun Kirchhof, Wolfgang Ludwig, K H Schleifer, Eugenia S Bulygina, Konstantin M Chumakov, Natalya I Govorukhina, Y A Trotsenko, Duncan White, and Et al. Bacillus methanolicus sp. nov., a new species of thermotolerant, methanol-utilizing, endospore-forming bacteria. *Int J Syst Bacteriol*, 42(3):439–445, 1992.

[70] N Arfman, J Van Beeumen, G E De Vries, W Harder, and L Dijkhuizen. Purification and characterization of an activator protein for methanol dehydrogenase from thermotolerant Bacillus spp. *The Journal of biological chemistry*, 266(6):3955–60, feb 1991.

[71] Harm Kloosterman, Jan W Vrijbloed, and Lubbert Dijkhuizen. Molecular, biochemical, and functional characterization of a Nudix hydrolase protein that stimulates the activity of a nicotinoprotein alcohol dehydrogenase. *The Journal of biological chemistry*, 277(38):34785–92, sep 2002.

[72] Andrea M. Ochsner, Jonas E N Müller, Carlos A. Mora, and Julia A. Vorholt. In vitro activation of NAD-dependent alcohol dehydrogenases by Nudix hydrolases is more widespread than assumed. *FEBS Letters*, 588(17):2993–2999, 2014.

[73] Anne Krog, Tonje M B Heggeset, Jonas E N Müller, Christiane E Kupper, Olha Schneider, Julia a Vorholt, Trond E Ellingsen, and Trygve Brautaset. Methylotrophic Bacillus methanolicus Encodes Two Chromosomal and One Plasmid Born NAD(+) Dependent Methanol Dehydrogenase Paralogs with Different Catalytic and Biochemical Properties. *PloS one*, 8(3):e59188, jan 2013.

[74] Jessica Stolzenberger, Steffen N Lindner, Marcus Persicke, Trygve Brautaset, and Volker F Wendisch. Characterization of fructose 1,6-bisphosphatase and sedoheptulose 1,7-bisphosphatase from the facultative ribulose monophosphate cycle methylotroph Bacillus methanolicus. *Journal of bacteriology*, 195(22):5112–22, nov 2013.

[75] Marta Irla, Armin Neshat, Trygve Brautaset, Christian Rückert, Jörn Kalinowski, and Volker F Wendisch. Transcriptome analysis of thermophilic methylotrophic Bacillus methanolicus MGA3 using RNA-sequencing provides detailed insights into its previously uncharted transcriptional landscape. *BMC genomics*, 16(1):73, 2015.

[76] Jonas E N Müller, Boris Litsanov, Miriam Bortfeld-Miller, Christian Trachsel, Jonas Grossmann, Trygve Brautaset, and Julia A. Vorholt. Proteomic analysis of the thermophilic methylotroph Bacillus methanolicus MGA3. *Proteomics*, 14(6):725–737, 2014.

[77] Jonas E N M??ller, Fabian Meyer, Boris Litsanov, Patrick Kiefer, and Julia A Vorholt. Core pathways operating during methylotrophy of Bacillus methanolicusMGA3 and induction of a bacillithiol-dependent detoxification pathway upon formaldehyde stress. *Molecular Microbiology*, 98(6):1089–1100, 2015.

[78] Hiroya Yurimoto, Nobuo Kato, and Yasuyoshi Sakai. Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. *Chemical Record*, 5(6):367–375, 2005.

[79] Joan Lin Cereghino and James M Cregg. Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol Rev*, 24(1):45–66, 2000.

[80] Andrea de la Torre, Aisha Metivier, Frances Chu, Lieve M. L. Laurens, David A. C. Beck, Philip T. Pienkos, Mary E. Lidstrom, and Marina G. Kalyuzhnaya. Genome-scale metabolic reconstructions and theoretical investigation of methane conversion in Methylomicrobium buryatense strain 5G(B1). *Microbial Cell Factories*, 14(1):188, 2015.

[81] Megen A. Culpepper and Amy C. Rosenzweig. Structure and protein-protein interactions of methanol dehydrogenase from Methylococcus capsulatus (Bath). *Biochemistry*, 53(39):6211–6219, 2014.

[82] Rudolf K Thauer, Anne-Kristin Kaster, Henning Seedorf, Wolfgang Buckel, and Reiner Hedderich. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature reviews. Microbiology*, 6(8):579–91, aug 2008.

[83] Gary M King and Carolyn F Weber. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nature reviews. Microbiology*, 5(2):107–18, 2007.

[84] A Bar-Even, E Noor, N E Lewis, and R Milo. Design and analysis of synthetic carbon fixation pathways . *Proceedings of the National Academy of Sciences of the United States of America*, 107(19):8889–8894, 2010.

[85] Tobias J Erb, Patrik R Jones, and Arren Bar-Even. Synthetic metabolism: metabolic engineering meets enzyme design. *Current Opinion in Chemical Biology*, 37:56–62, 2017.

[86] T. Schwander, L. Schada von Borzyskowski, S. Burgener, N. S. Cortina, and T. J. Erb. A synthetic pathway for the fixation of carbon dioxide in vitro. *Science*, 354(6314):900–904, 2016.

[87] Justin B Siegel, Amanda Lee Smith, Sean Poust, Adam J Wargacki, Arren Bar-Even, Catherine Louw, Betty W Shen, Christopher B Eiben, Huu M Tran, Elad Noor, Jasmine L Gallaher, Jacob Bale, Yasuo Yoshikuni, Michael H Gelb, Jay D Keasling, Barry L Stoddard, Mary E Lidstrom, and David Baker. Computational protein design enables a novel one-carbon assimilation pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 112(12):3704–9, 2015.

[88] Bettina Schiel-Bengelsdorf and Peter Durre. Pathway engineering and synthetic biology using acetogens, 2012.

[89] S. Eric Nybo, Nymul Khan, Benjamin M Woolston, and Wayne R Curtis. Metabolic engineering in chemolithoautotrophic hosts for the production of fuels and chemicals. *Metabolic Engineering*, 30:105–120, 2015.

[90] Michael Köpke, Claudia Held, Sandra Hujer, Heiko Liesegang, Arnim Wiezer, Antje Wollherr, Armin Ehrenreich, Wolfgang Liebl, Gerhard Gottschalk, and Peter Dürre. Clostridium ljungdahlii represents a microbial production platform based on syngas. *Proceedings of the National Academy of Sciences of the United States of America*, 107(29):13087–92, jul 2010.

[91] Ching Leang, Toshiyuki Ueki, Kelly P Nevin, and Derek R Lovley. A genetic system for Clostridium ljungdahlii: a chassis for autotrophic production of biocommodities and a model homoacetogen. *Applied and environmental microbiology*, 79(4):1102–9, feb 2013.

[92] Bastian Molitor, Kristina Kirchner, Alexander W. Henrich, Simone Schmitz, and Miriam A. Rosenbaum. Expanding the molecular toolkit for the homoacetogen Clostridium ljungdahlii. *Scientific Reports*, 6:31518, 2016.

[93] Areen Banerjee, Ching Leang, Toshiyuki Ueki, Kelly P Nevin, and Derek R Lovley. *A Lactose-Inducible System for Metabolic Engineering of Clostridium ljungdahlii*. Number February. feb 2014.

[94] He Huang, Changsheng Chai, Ning Li, Peter Rowe, Nigel Peter Minton, Sheng Yang, Weihong Jiang, and Yang Gu. CRISPR/Cas9-based efficient genome

editing in Clostridium ljungdahlii, an autotrophic gas-fermenting bacterium. *ACS Synthetic Biology*, page acssynbio.6b00044, 2016.

[95] Toshiyuki Ueki, Kelly P Nevin, Trevor L Woodard, and Derek R Lovley. Converting carbon dioxide to butyrate with an engineered strain of Clostridium ljungdahlii. *mBio*, 5(5):19–23, 2014.

[96] Melanie Straub, Martin Demler, Dirk Weuster-Botz, and Peter Dürre. Selective enhancement of autotrophic acetate production with genetically modified Acetobacterium woodii. *Journal of biotechnology*, mar 2014.

[97] Sabrina Hoffmeister, Marzena Gerdom, Frank R. Bengelsdorf, Sonja Linder, Sebastian Flüchter, Hatice Öztürk, Wilfried Blümke, Antje May, Ralf Jörg Fischer, Hubert Bahl, and Peter Dürre. Acetone production with metabolically engineered strains of Acetobacterium woodii. *Metabolic Engineering*, 36:37–47, 2016.

[98] Akihisa Kita, Yuki Iwasaki, Shinsuke Sakai, Shinya Okuto, Kazue Takaoka, Tohru Suzuki, Shinichi Yano, Shigeki Sawayama, Takahisa Tajima, Junichi Kato, Naomichi Nishio, Katsuji Murakami, and Yutaka Nakashimada. Development of genetic transformation and heterologous expression system in carboxydotrophic thermophilic acetogen Moorella thermoacetica. *Journal of Bioscience and Bioengineering*, xx(xx):1–6, 2012.

[99] Alexander F Arendsen, Mohsin Q Soliman, and Stephen W Ragsdale. Nitrate-dependent regulation of acetate biosynthesis and nitrate respiration by Clostridium thermoaceticum. *Journal of Bacteriology*, 181(5):1489–1495, 1999.

[100] Peng Hu, Sagar Chakraborty, Amit Kumar, Benjamin M Woolston, Hongjuan Liu, David Emerson, and Gregory Stephanopoulos. Integrated Bioprocess for Conversion of Gaseous Substrates to Liquids. *Proceedings of the National Academy of Sciences*, 113(14):3773–3778, 2016.

[101] Chad A Haynes and Ramon Gonzalez. Rethinking biological activation of methane and conversion to liquid fuels. *Nature chemical biology*, 10(5):331–9, 2014.

[102] Qiang Fei, Michael T Guarnieri, Ling Tao, Lieve M L Laurens, Nancy Dowe, and Philip T Pienkos. Bioconversion of natural gas to liquid fuel: opportunities and challenges. *Biotechnology advances*, 32(3):596–614, 2014.

[103] M G Kalyuzhnaya, S Yang, O N Rozova, N E Smalley, J Clubb, a Lamb, G a Nagana Gowda, D Raftery, Y Fu, F Bringel, S Vuilleumier, D a C Beck, Y a Trotsenko, V N Khmelenina, and M E Lidstrom. Highly efficient methane biocatalysis revealed in a methanotrophic bacterium. *Nature communications*, 4(May):2785, jan 2013.

[104] Janet B. Matsen, Song Yang, Lisa Y. Stein, David Beck, and Marina G. Kalyuzhnaya. Global molecular analyses of methane metabolism in Methanotrophic alphaproteobacterium, methylosinus trichosporium OB3b. Part I: Transcriptomic study. *Frontiers in Microbiology*, 4(APR):1–16, 2013.

[105] Aaron W Puri, Sarah Owen, Frances Chu, Ted Chavkin, David a C Beck, Marina G Kalyuzhnaya, and Mary E Lidstrom. Genetic tools for the industrially promising methanotroph Methylomicrobium buryatense. *Applied and environmental microbiology*, (December), dec 2014.

[106] Bo Hu and Mary E Lidstrom. Metabolic engineering of Methylobacterium extorquens AM1 for 1-butanol production. *Biotechnology for Biofuels*, 7(1):156, 2014.

[107] Frank Sonntag, Cora Kroner, Patrice Lubuta, R??mi Peyraud, Angelika Horst, Markus Buchhaupt, and Jens Schrader. Engineering Methylobacterium extorquens for de novo synthesis of the sesquiterpenoid ??-humulene from methanol. *Metabolic Engineering*, 32:82–94, 2015.

[108] Izumi Orita, Kouta Nishikawa, Satoshi Nakamura, and Toshiaki Fukui. Biosynthesis of polyhydroxyalkanoate copolymers from methanol by Methylobacterium extorquens AM1 and the engineered strains under cobalt-deficient conditions. *Applied Microbiology and Biotechnology*, 98(8):3715–3725, 2014.

[109] Frank Sonntag, Markus Buchhaupt, and Jens Schrader. Thioesterases for ethylmalonyl-CoA pathway derived dicarboxylic acid production in Methylobacterium extorquens AM1. *Applied Microbiology and Biotechnology*, 98(10):4533–4544, 2014.

[110] Frank Sonntag, Jonas E. N. Müller, Patrick Kiefer, Julia a. Vorholt, Jens Schrader, and Markus Buchhaupt. High-level production of ethylmalonyl-CoA pathway-derived dicarboxylic acids by Methylobacterium extorquens under cobalt-deficient conditions and by polyhydroxybutyrate negative strains. *Applied Microbiology and Biotechnology*, pages 3407–3419, 2015.

[111] Ingemar Naerdal, Johannes Pfeifenschneider, Trygve Brautaset, and Volker F. Wendisch. Methanol-based cadaverine production by genetically engineered Bacillus methanolicus strains. *Microbial Biotechnology*, 8(2):342–350, 2015.

[112] Izumi Orita, Naoki Sakamoto, Nobuo Kato, Hiroya Yurimoto, and Yasuyoshi Sakai. Bifunctional enzyme fusion of 3-hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase. *Applied microbiology and biotechnology*, 76(2):439–45, aug 2007.

[113] Hal Alper and Gregory Stephanopoulos. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nature reviews. Microbiology*, 7(10):715–23, oct 2009.

[114] Michael J Smanski, Swapnil Bhatia, Dehua Zhao, YongJin Park, Lauren B A Woodruff, Georgia Giannoukos, Dawn Ciulla, Michele Busby, Johnathan Calderon, Robert Nicol, D Benjamin Gordon, Douglas Densmore, and Christopher A Voigt. Functional optimization of gene clusters by combinatorial design and assembly. *Nature Biotechnology*, 32(12):1241–1249, 2014.

[115] Peng Hu, Sagar Chakraborty, Amit Kumar, Benjamin M Woolston, Hongjuan Liu, David Emerson, and Gregory Stephanopoulos. Integrated Bioprocess for Conversion of Gaseous Substrates to Liquids. *Proceedings of the National Academy of Sciences*, 113(14):3773–3778, 2016.

[116] Peng Hu, Hamid Rismani-yazdi, and Gregory Stephanopoulos. Anaerobic CO2 Fixation by the Acetogenic Bacterium Moorella thermoacetica. *AIChE Journal*, 00(00):1–8, 2013.

[117] Stephen W Ragsdale and Elizabeth Pierce. Acetogenesis and the Wood-Ljungdahl pathway of CO(2) fixation. *Biochimica et biophysica acta*, 1784(12):1873–98, dec 2008.

[118] Benjamin M Woolston, Steven Edgar, and Gregory Stephanopoulos. Metabolic Engineering: Past and Future. *Annual review of chemical and biomolecular engineering*, (March):259–288, mar 2013.

[119] Andrew D Frock and Robert M Kelly. Extreme thermophiles: moving beyond single-enzyme biocatalysis. *Current Opinion in Chemical Engineering*, pages 1–10, jul 2012.

[120] Trond Erik Vee Aune and Finn Lillelund Aachmann. Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. *Applied microbiology and biotechnology*, 85(5):1301–13, feb 2010.

[121] Lu Lin, Houhui Song, Yuetong Ji, Zhili He, Yunting Pu, Jizhong Zhou, and Jian Xu. Ultrasound-Mediated DNA Transformation in Thermophilic Gram-Positive Anaerobes. *PLoS ONE*, 5(9):10, 2010.

[122] Harold R Azencott, Gary F Peter, and Mark R Prausnitz. Influence of the Cell Wall on Intracellular Delivery to Algal Cells by Electroporation and Sonication. *Ultrasound Med Biol.*, 33(11):1805–1817, 2007.

[123] D C Bartoletti, G I Harrison, and J C Weaver. The number of molecules taken up by electroporated cells: quantitative determination. *FEBS letters*, 256(1-2):4–10, oct 1989.

[124] P J Canatella, J F Karr, J a Petros, and M R Prausnitz. Quantitative study of electroporation-mediated molecular uptake and cell viability. *Biophysical journal*, 80(2):755–64, feb 2001.

[125] E a Gift and J C Weaver. Simultaneous quantitative determination of electroporative molecular uptake and subsequent cell survival using gel microdrops and flow cytometry. *Cytometry*, 39(4):243–9, apr 2000.

[126] M R Prausnitz, C D Milano, J a Gimm, R Langer, and J C Weaver. Quantitative study of molecular transport due to electroporation: uptake of bovine serum albumin by erythrocyte ghosts. *Biophysical journal*, 66(5):1522–30, may 1994.

[127] D. M. Fishman and G. D. Patterson. Light scattering studies of supercoiled and nicked DNA. *Biopolymers*, 38(4):535–552, 1996.

[128] M B James and T D Giorgio. Nuclear-associated plasmid, but not cell-associated plasmid, is correlated with transgene expression in cultured mammalian cells. *Molecular therapy : the journal of the American Society of Gene Therapy*, 1(4):339–46, apr 2000.

[129] Shyam K Sharan, Lynn C Thomason, Sergey G Kuznetsov, and Donald L Court. Recombineering: a homologous recombination-based method of genetic engineering. *Nature protocols*, 4(2):206–23, jan 2009.

[130] Harris H Wang, Farren J Isaacs, Peter a Carr, Zachary Z Sun, George Xu, Craig R Forest, and George M Church. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460(7257):894–8, aug 2009.

[131] Bryan Swingle, Eric Markel, Nina Costantino, Mikhail G Bubunenko, Samuel Cartinhour, and Donald L Court. Oligonucleotide recombination in Gram-negative bacteria. *Molecular microbiology*, 75(1):138–48, jan 2010.

[132] Simanti Datta, Nina Costantino, Xiaomei Zhou, and Donald L Court. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proceedings of the National Academy of Sciences of the United States of America*, 105(5):1626–31, mar 2008.

[133] Mark P Taylor, Carlos D Esteban, and David J Leak. Development of a versatile shuttle vector for gene expression in Geobacillus spp. *Plasmid*, 60(1):45–52, jul 2008.

[134] William Dower, Jeff Miller, and Charles Ragsdale. High efficiency transformation of E. coli by high voltaage electroporation. *Nucleic acids research*, 16(13):6127–6145, 1988.

[135] R K Sizemore, J J Caldwell, and a S Kendrick. Alternate gram staining technique using a fluorescent lectin. *Applied and environmental microbiology*, 56(7):2245–7, jul 1990.

[136] Joshua a Mosberg, Christopher J Gregg, Marc J Lajoie, Harris H Wang, and George M Church. Improving lambda red genome engineering in Escherichia coli via rational removal of endogenous nucleases. *PloS one*, 7(9):e44638, jan 2012.

247

[137] R.J. Roberts, T.R. Gingeras, and J.P. Milazzo. A computer assisted method for the determination of restriction enzyme recdgnition sites. *Nucleic acids research*, 5(11):4105–4127, 1978.

[138] John Eid, Adrian Fehr, Jeremy Gray, Khai Luong, John Lyle, Geoff Otto, Paul Peluso, David Rank, Primo Baybayan, Brad Bettman, Arkadiusz Bibillo, Keith Bjornson, Bidhan Chaudhuri, Frederick Christians, Ronald Cicero, Sonya Clark, Ravindra Dalal, John Dixon, Mathieu Foquet, Alfred Gaertner, Paul Hardenbol, Cheryl Heiner, Kevin Hester, David Holden, Gregory Kearns, Xiangxu Kong, Ronald Kuse, Yves Lacroix, Steven Lin, Paul Lundquist, Congcong Ma, Patrick Marks, Mark Maxham, Devon Murphy, Insil Park, Thang Pham, Michael Phillips, Joy Roy, Robert Sebra, Gene Shen, Jon Sorenson, Austin Tomaney, Kevin Travers, Mark Trulson, John Vieceli, Jeffrey Wegener, Dawn Wu, Alicia Yang, Denis Zaccarin, Peter Zhao, Frank Zhong, Jonas Korlach, and Stephen Turner. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science*, 323(January):133–138, 2009.

[139] Benjamin a Flusberg, Dale R Webster, Jessica H Lee, Kevin J Travers, Eric C Olivares, Tyson a Clark, Jonas Korlach, and Stephen W Turner. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature methods*, 7(6):461–5, jun 2010.

[140] Iain a Murray, Tyson a Clark, Richard D Morgan, Matthew Boitano, Brian P Anton, Khai Luong, Alexey Fomenkov, Stephen W Turner, Jonas Korlach, and Richard J Roberts. The methylomes of six bacteria. *Nucleic acids research*, 40(22):11450–62, dec 2012.

[141] John P. Donahue, Dawn A. Israel, Richard M. Peek, Martin J. Blaser, and Geraldine G. Miller. Overcoming the restriction barrier to plasmid transformation of Helicobacter pylori. *Molecular Microbiology*, 37(5):1066–1074, 2000.

[142] Kazumasa Yasui, Yasunobu Kano, Kaori Tanaka, Kunitomo Watanabe, Mariko Shimizu-Kadota, Hirofumi Yoshikawa, and Tohru Suzuki. Improvement of bacterial transformation efficiency using plasmid artificial modification. *Nucleic acids research*, 37(1):e3, jan 2009.

[143] Tohru Suzuki and Kazumasa Yasui. Plasmid Artificial Modification: A Novel Method for Efficient DNA Transfer into Bacteria. *Plasmid*, 765, 2011.

[144] LM Luz-maria M Guzman, Dominique Belin, Michael J Carson, and J Beckwith. Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P_{BAD} Promoter. *Journal of Bacteriology*, 177(14):4121–4130, 1995.

[145] John T Heap, Muhammad Ehsaan, Clare M Cooksley, Yen-Kuan Ng, Stephen T Cartman, Klaus Winzer, and Nigel P Minton. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic acids research*, 40(8):e59, apr 2012.

[146] V Mai and J Wiegel. Advances in development of a genetic system for Thermoanaerobacterium spp.: expression of genes encoding hydrolytic enzymes, development of a second shuttle vector, and integration of genes into the chromosome. *Applied and environmental microbiology*, 66(11):4817–21, nov 2000.

[147] Harish Nagarajan, Merve Sahin, Juan Nogales, Haythem Latif, Derek R Lovley, Ali Ebrahim, and Karsten Zengler. Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of Clostridium ljungdahlii. *Microbial cell factories*, 12:118, jan 2013.

[148] Jin Chen, Jose a Gomez, Kai Höffner, Paul I Barton, and Michael a Henson. Metabolic modeling of synthesis gas fermentation in bubble column reactors. *Biotechnology for biofuels*, 8:89, 2015.

[149] He Huang, Changsheng Chai, Ning Li, Peter Rowe, Nigel Peter Minton, Sheng Yang, Weihong Jiang, and Yang Gu. CRISPR/Cas9-based efficient genome editing in Clostridium ljungdahlii, an autotrophic gas-fermenting bacterium. *ACS Synthetic Biology*, page acssynbio.6b00044, 2016.

[150] Shuning Wang, Haiyan Huang, Haiyan Huang Kahnt, Alexander P. Mueller, Michael Köpke, and Rudolf K. Thauer. NADP-Specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in Clostridium autoethanogenum grown on CO. *Journal of Bacteriology*, 195(19):4373–4386, 2013.

[151] Fungmin Liew, Anne M. Henstra, Michael KÖğpke, Klaus Winzer, Sean D. Simpson, and Nigel P. Minton. Metabolic Engineering of Clostridium autoethanogenum for Selective Alcohol Production. *Metabolic Engineering*, 40(January):104–114, 2017.

[152] Fungmin Liew, Anne M. Henstra, Klaus Winzer, Michael Kopke, Sean D. Simpson, and Nigel P. Minton. Insights into CO2 Fixation Pathway of Clostridium autoethanogenum by Targeted Mutagenesis. *mBio*, 7(3):1–10, 2016.

[153] Irene M Brockman and Kristala L J Prather. Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnology J*, 10(9):1360–1369, 2015.

[154] Lei S Qi, Matthew H Larson, Luke A Gilbert, Jennifer A Doudna, Jonathan S Weissman, Adam P Arkin, and Wendell A Lim. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–1183, 2013.

[155] Brady F. Cress, O. Duhan Toparlak, Sanjay Guleria, Matthew Lebovich, Jessica T. Stieglitz, Jacob A Englaender, J Andrew Jones, Robert J Linhardt, and Mattheos A G Koffas. CRISPathBrick: Modular Combinatorial Assembly of Type II-A CRISPR Arrays for dCas9-Mediated Multiplex Transcriptional Repression in E. coli. *ACS Synthetic Biology*, 4(9):987–1000, 2015.

[156] Sara Cleto, Jaide V K Jensen, Volker F. Wendisch, and Timothy K. Lu. Corynebacterium glutamicum Metabolic Engineering with CRISPR Interference (CRISPRi). *ACS Synthetic Biology*, 5(5):375–385, 2016.

[157] Matthew Deaner and Hal S. Alper. Systematic Testing of Enzyme Perturbation Sensitivities via Graded dCas9 Modulation in Saccharomyces cerevisiae. *Metabolic Engineering*, 40(January):14–22, 2017.

[158] John T Heap, Oliver J Pennington, Stephen T Cartman, and Nigel P Minton. A modular system for Clostridium shuttle plasmids. *Journal of microbiological methods*, 78(1):79–85, jul 2009.

[159] Andrea H Hartman, Hualan Liu, and Stephen B Melville. Construction and characterization of a lactose-inducible promoter system for controlled gene expression in Clostridium perfringens. *Applied and environmental microbiology*, 77(2):471–8, 2011.

[160] Taek Lee, Rachel A Krupa, Fuzhong Zhang, Meghdad Hajimorad, William J Holtz, Nilu Prasad, Sung Lee, and Jay D Keasling. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *Journal of Biological Engineering*, 5(1):12, 2011.

[161] Tao Xu, Yongchao Li, Zhou Shi, Christopher L. Hemme, Yuan Li, Yonghua Zhu, Joy D. Van Nostrand, Zhili He, and Jizhong Zhou. Efficient genome editing in Clostridium cellulolyticum via CRISPR-Cas9 nickase. *Applied and Environmental Microbiology*, 81(13):AEM.00873–15, 2015.

[162] Hongjun Dong, Wenwen Tao, Yanping Zhang, and Yin Li. Development of an anhydrotetracycline-inducible gene expression system for solvent-producing Clostridium acetobutylicum: A useful tool for strain engineering. *Metabolic Engineering*, 14(1):59–67, jan 2012.

[163] Harish Nagarajan, Merve Sahin, Juan Nogales, Haythem Latif, Derek R Lovley, Ali Ebrahim, and Karsten Zengler. Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of Clostridium ljungdahlii. *Microbial cell factories*, 12:118, jan 2013.

[164] Johanna Mock, Yanning Zheng, Alexander P. Mueller, San Ly, Loan Tran, Simon Segovia, Shilpa Nagaraju, Michael Köpke, Peter Dürre, and Rudolf K. Thauer. Energy conservation associated with ethanol formation from H $_2$ and CO $_2$ in *Clostridium autoethanogenum* involving electron bifurcation. *Journal of Bacteriology*, 197(July):JB.00399–15, 2015.

[165] Ralph Bertram and Wolfgang Hillen. The application of Tet repressor in prokaryotic gene regulation and expression. *Microbial Biotechnology*, 1(1):2–16, 2008.

[166] Shilpa Nagaraju, Naomi Kathleen Davies, David Jeffrey Fraser Walker, Michael Köpke, and Séan Dennis Simpson. Genome editing of Clostridium autoethanogenum using CRISPR/Cas9. *Biotechnology for Biofuels*, 9(1):219, 2016.

[167] Hsien-Chung Tseng, Collin H Martin, David R Nielsen, and Kristala L Jones Prather. Metabolic Engineering of Escherichia coli for Enhanced Production of (R)-and (S)-3-Hydroxybutyrate. *Applied and Environmental Microbiology*, 75(10):3137–3145, 2009.

[168] Tessa E F Quax, Nico J. Claassens, Dieter Söll, and John van der Oost. Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular Cell*, 59(2):149–161, 2015.

[169] Markus Fuhrmann, Amparo Hausherr, Lars Ferbitz, Thomas Schödl, and Peter Hegemann. Monitoring dynamic expression of nuclear genes in Chlamydomonas reinhardtii by using a synthetic luciferase reporter gene. *Plant Molecular Biology*, 55:869–881, 2004.

[170] James Bristow, Gareth Butland, and Adam P Arkin. Sequencing Randomly Bar-Coded Transposons. 6(6):1–15, 2015.

[171] Katrin Knittel and Antje Boetius. Anaerobic oxidation of methane: progress with an unknown process. *Annual review of microbiology*, 63:311–34, jan 2009.

[172] Mohamed F Haroon, Shihu Hu, Ying Shi, Michael Imelfort, Jurg Keller, Philip Hugenholtz, Zhiguo Yuan, and Gene W Tyson. Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature*, pages 2–7, jul 2013.

[173] ARPA-E. ARPA-E Announces $66 Million For Transformational Energy Technologies, 2013.

[174] Deepak Dugar and Gregory Stephanopoulos. Relative potential of biosynthetic pathways for biofuels and bio-based products. *Nature biotechnology*, 29(12):1074–8, dec 2011.

[175] Alan G Fast and Eleftherios T Papoutsakis. Stoichiometric and energetic analyses of non-photosynthetic CO2-fixation pathways to support synthetic biology strategies for production of fuels and chemicals. *Current Opinion in Chemical Engineering*, pages 1–16, sep 2012.

[176] Igor W Bogorad, Tzu-Shyang Lin, and James C Liao. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature*, 502(7473):693–7, oct 2013.

[177] Igor W Bogorad, Chang-Ting Chen, Matthew K Theisen, Tung-Yun Wu, Alicia R Schlenz, Albert T Lam, and James C Liao. Building carbon-carbon bonds

using a biocatalytic methanol condensation cycle. *Proceedings of the National Academy of Sciences of the United States of America*, oct 2014.

[178] C Anthony. The prediction of growth yields in methylotrophs. *Journal of General Microbiology*, pages 91–104, 1978.

[179] Shawn E Mcglynn, Grayson L Chadwick, Christopher P Kempes, and Victoria J Orphan. Single cell activity reveals direct electron transfer in methanotrophic consortia. 2015.

[180] Feng-Ping Wang, Yu Zhang, Ying Chen, Ying He, Ji Qi, Kai-Uwe Hinrichs, Xin-Xu Zhang, Xiang Xiao, and Nico Boon. Methanotrophic archaea possessing diverging methane-oxidizing and electron-transporting pathways. *The ISME journal*, 8(5):1069–78, may 2014.

[181] T Ide, S Bäumer, and U Deppenmeier. Energy conservation by the H_2:heterodisulfide oxidoreductase from Methanosarcina mazei Göl: identification of two proton-translocating segments. *Journal of Bacteriology*, 181(13):4076–80, 1999.

[182] Anke Meyerdierks, Michael Kube, Ivaylo Kostadinov, Hanno Teeling, Frank Oliver Glöckner, Richard Reinhardt, and Rudolf Amann. Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. *Environmental microbiology*, 12(2):422–39, feb 2010.

[183] Gunter Wegener, Viola Krukenberg, Dietmar Riedel, Halina E. Tegetmeyer, and Antje Boetius. Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. 2015.

[184] Katharina F Ettwig, Margaret K Butler, Denis Le Paslier, Eric Pelletier, Sophie Mangenot, Marcel M M Kuypers, Frank Schreiber, Bas E Dutilh, Johannes Zedelius, Dirk de Beer, Jolein Gloerich, Hans J C T Wessels, Theo van Alen, Francisca Luesken, Ming L Wu, Katinka T van de Pas-Schoonen, Huub J M Op den Camp, Eva M Janssen-Megens, Kees-Jan Francoijs, Henk Stunnenberg, Jean Weissenbach, Mike S M Jetten, and Marc Strous. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature*, 464(7288):543–8, mar 2010.

[185] Emily J Beal, Christopher H House, and Victoria J Orphan. Manganese- and iron-dependent marine methane oxidation. *Science (New York, N.Y.)*, 325(5937):184–7, jul 2009.

[186] Valerie W. C. Soo, Michael J. McAnulty, Arti Tripathi, Fayin Zhu, Limin Zhang, Emmanuel Hatzakis, Philip B. Smith, Saumya Agrawal, Hadi Nazem-Bokaee, Saratram Gopalakrishnan, Howard M. Salis, James G. Ferry, Costas D. Maranas, Andrew D. Patterson, and Thomas K. Wood. Reversing methanogenesis to capture methane for liquid biofuel precursors. *Microbial Cell Factories*, 15(1):11, 2016.

[187] Hadi Nazem-Bokaee, Saratram Gopalakrishnan, James G Ferry, Thomas K Wood, and Costas D Maranas. Assessing methanotrophy and carbon fixation for biofuel production by Methanosarcina acetivorans. *Microbial cell factories*, 15(1):10, 2016.

[188] Thomas J Mueller, Matthew J Grisewood, Hadi Nazem-Bokaee, Saratram Gopalakrishnan, James G Ferry, Thomas K Wood, and Costas D Maranas. Methane oxidation by anaerobic archaea for conversion to liquid fuels. *Journal of industrial microbiology & biotechnology*, 42(3):391–401, mar 2015.

[189] Zhen Yan, Mingyu Wang, and James G Ferry. A ferredoxin- and F420H2-dependent, electron- bifurcating, heterodisulfide reductase with homologs in the domains Bacteria and Archaea. *mBio*, 8(1):1–15, 2017.

[190] Tina Treude, Victoria Orphan, Katrin Knittel, Armin Gieseke, Christopher H. House, and Antje Boetius. Consumption of methane and CO2 by methanotrophic microbial mats from gas seeps of the anoxic Black Sea. *Applied and Environmental Microbiology*, 73(7):2271–2283, 2007.

[191] Katja Nauhaus, Melanie Albrecht, Marcus Elvert, Antje Boetius, and Friedrich Widdel. In vitro cell growth of marine archaeal-bacterial consortia during anaerobic oxidation of methane with sulfate. *Environmental microbiology*, 9(1):187–196, jan 2007.

[192] Rudolf K Thauer. Anaerobic oxidation of methane with sulfate: on the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO2. *Current opinion in microbiology*, 14(3):292–9, jun 2011.

[193] Conrado Moreno-vivián, Purificación Cabello, Rafael Blasco, Francisco Castillo, N Cabello, Manuel Marti, and Conrado Moreno-vivia. Prokaryotic Nitrate Reduction : Molecular Properties and Functional Distinction among Bacterial Nitrate Reductases MINIREVIEW Prokaryotic Nitrate Reduction : Molecular Properties and Functional Distinction among Bacterial Nitrate Reductases. 1999.

[194] G Unden and J Bongaerts. Alternative respiratory pathways of Escherichia coli: energetics and transcriptional regulation in response to electron acceptors. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1320(3):217–234, 1997.

[195] J W Moir and N J Wood. Nitrate and nitrite transport in bacteria. *Cellular and molecular life sciences : CMLS*, 58(2):215–224, 2001.

[196] Bernhard Jaun and Rudolf K Thauer. *Methyl-Coenzyme M Reductase and its Nickel Corphin Coenzyme F430 in Methanogenic Archaea*, volume 356. 2007.

[197] Divya Prakash, Yonnie Wu, Sang-Jin Suh, and Evert C Duin. Elucidating the process of activation of methyl-coenzyme m reductase. *Journal of bacteriology*, 196(13):2491–8, jul 2014.

[198] Rudolf K Thauer and Seigo Shima. Methane as fuel for anaerobic microorganisms. *Annals of the New York Academy of Sciences*, 1125:158–70, mar 2008.

[199] R Mahadevan, D R Bond, J E Butler, V Coppi, B O Palsson, C H Schilling, and D R Lovley. Characterization of Metabolism in the Fe (III) -Reducing Organism Geobacter sulfurreducens by Constraint-Based Modeling Characterization of Metabolism in the Fe (III) -Reducing Organism Geobacter sulfurreducens by Constraint-Based Modeling âĂă. *Applied and Environmental Microbiology*, 72(2):1558–1568, 2006.

[200] Silvan Scheller, Meike Goenrich, Reinhard Boecher, Rudolf K Thauer, and Bernhard Jaun. The key nickel enzyme of methanogenesis catalyses the anaerobic oxidation of methane. *Nature*, 465(7298):606–8, jun 2010.

[201] Bo Hu and Mary E Lidstrom. Metabolic engineering of Methylobacterium extorquens AM1 for 1-butanol production. *Biotechnology for Biofuels*, 7(1):156, 2014.

[202] William B Whitaker, Nicholas R Sandoval, Robert K Bennett, Alan G Fast, and Eleftherios T Papoutsakis. Synthetic methylotrophy: Engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization, 2015.

[203] Jonas E N Müller, Fabian Meyer, Boris Litsanov, Patrick Kiefer, Eva Potthoff, Stéphanie Heux, Wim J Quax, Volker F Wendisch, and Trygve Brautaset. Engineering Escherichia coli for methanol conversion. *Metabolic Engineering*, 28:190–201, 2015.

[204] W Brain Whitaker, J Andrew Jones, Kyle Bennett, Jacqueline Gonzalez, Victoria R Vernacchio, Shannon M Collins, Michael A Palmer, Samuel Schmidt, Maciek R. Antoniewicz, Mattheos A. Koffas, and Eleftherios T. Papoutsakis. Engineering the Biological Conversion of Methanol to Specialty Chemicals in Escherichia coli. *Metabolic Engineering*, (October):1–11, 2016.

[205] Sabrina Witthoff, Katja Schmitz, Sebastian Niedenführ, Katharina Nöh, Stephan Noack, Michael Bott, and Jan Marienhagen. Metabolic Engineering of Corynebacterium glutamicum for Methanol Metabolism. *Applied and Environmental Microbiology*, 81(6):2215–2225, 2015.

[206] Tung Yun Wu, Chang Ting Chen, Jessica Tse Jin Liu, Igor W Bogorad, Robert Damoiseaux, and James C Liao. Characterization and evolution of an activator-independent methanol dehydrogenase from Cupriavidus necator N-1. *Applied Microbiology and Biotechnology*, 100(11):4969–4983, 2016.

[207] J Vincent Price, Long Chen, W Brian Whitaker, Eleftherios Papoutsakis, and Wilfred Chen. Scaffoldless engineered enzyme assembly for enhanced methanol utilization. *Proceedings of the National Academy of Sciences of the United States of America*, page 201601797, 2016.

[208] Hsien-Chung Tseng, Collin H Martin, David R Nielsen, and Kristala L Jones Prather. Metabolic engineering of Escherichia coli for enhanced production of (R)- and (S)-3-hydroxybutyrate. *Applied and environmental microbiology*, 75(10):3137–45, may 2009.

[209] Yu Jiang, Biao Chen, Chunlan Duan, Bingbing Sun, Junjie Yang, and Sheng Yang. Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System. *Applied and Environmental Microbiology*, 81(7):2506–2514, 2015.

[210] Tung Yun Wu, Chang Ting Chen, Jessica Tse Jin Liu, Igor W Bogorad, Robert Damoiseaux, and James C Liao. Characterization and evolution of an activator-independent methanol dehydrogenase from Cupriavidus necator N-1. *Applied Microbiology and Biotechnology*, 100(11):4969–4983, 2016.

[211] K A Datsenko and B L Wanner. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12):6640–5, jun 2000.

[212] Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill a Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology*, 2:2006.0008, jan 2006.

[213] Adam M Feist, Christopher S Henry, Jennifer L Reed, Markus Krummenacker, Andrew R Joyce, Peter D Karp, Linda J Broadbelt, Vassily Hatzimanikatis, and Bernhard Ø Palsson. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular systems biology*, 3(121):121, jan 2007.

[214] Jonas E.N. Müller, Fabian Meyer, Boris Litsanov, Patrick Kiefer, Eva Potthoff, Stéphanie Heux, Wim J. Quax, Volker F. Wendisch, Trygve Brautaset, Jean-Charles Portais, and Julia a. Vorholt. Engineering Escherichia coli for methanol conversion. *Metabolic Engineering*, 28:190–201, jan 2015.

[215] T Nash. The colorimetric estimation of formaldehyde by means of the Hantzsch reaction. *The Biochemical journal*, 55(3):416–421, 1953.

[216] Christian Zwiener, T. Glauner, and F. H. Frimmel. Method optimization for the determination of carbonyl compounds in disinfected water by DNPH derivatization and LC-ESI-MS-MS. *Analytical and Bioanalytical Chemistry*, 372(5-6):615–621, 2002.

[217] Jason R King, Benjamin Michael Woolston, and Gregory Stephanopoulos. Designing a new entry point into isoprenoid metabolism by exploiting fructose-6-phosphate aldolase side-reactivity of Escherichia coli. *ACS Synthetic Biology*, page acssynbio.7b00072, 2017.

[218] Pierre Millard, Fabien Letisse, Serguei Sokol, and Jean-Charles Portais. IsoCor: correcting MS data in isotope labeling experiments. *Bioinformatics*, 28(9):1294–1296, may 2012.

[219] Bryson D Bennett, Elizabeth H Kimball, Melissa Gao, Robin Osterhout, Stephen J Van Dien, and Joshua D Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature chemical biology*, 5(8):593–9, aug 2009.

[220] Harm Kloosterman, Jan W Vrijbloed, and Lubbert Dijkhuizen. Molecular, biochemical, and functional characterization of a Nudix hydrolase protein that stimulates the activity of a nicotinoprotein alcohol dehydrogenase. *The Journal of biological chemistry*, 277(38):34785–92, sep 2002.

[221] Hui Gao, Manish Kumar Tiwari, Yun Chan Kang, and Jung Kul Lee. Characterization of H 2O-forming NADH oxidase from Streptococcus pyogenes and its application in l-rare sugar production. *Bioorganic and Medicinal Chemistry Letters*, 22(5):1931–1935, 2012.

[222] T Ferenci, T Strom, and J R Quayle. Purification and properties of 3-hexulose phosphate synthase and phospho-3-hexuloisomerase from Methylococcus capsulatus. *The Biochemical journal*, 144(3):477–86, dec 1974.

[223] N Arfman, L Bystrykh, N I Govorukhina, and L Dijkhuizen. 3-Hexulose-6-phosphate synthase from thermotolerant methylotroph Bacillus C1. *Methods in enzymology*, 188:391–397, 1990.

[224] Izumi Orita, Akiko Kita, Hiroya Yurimoto, Nobuo Kato, Yasuyoshi Sakai, and Kunio Miki. Crystal structure of 3-hexulose-6-phosphate synthase, a member of the orotidine 5'-monophosphate decarboxylase suprafamily. *Proteins*, 78(16):3488–92, dec 2010.

[225] Howard M Salis, Ethan A Mirsky, and Christopher A Voigt. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotech*, 27(10):946–950, oct 2009.

[226] W Brain Whitaker, J Andrew Jones, Kyle Bennett, Jacqueline Gonzalez, Victoria R Vernacchio, Shannon M Collins, Michael A Palmer, Samuel Schmidt, Maciek R. Antoniewicz, Mattheos A. Koffas, and Eleftherios T. Papoutsakis. Engineering the Biological Conversion of Methanol to Specialty Chemicals in Escherichia coli. *Metabolic Engineering*, 39(October 2016):49–59, 2016.

[227] Anna Kotrbova-Kozak, Pavel Kotrba, Masayuki Inui, Jiri Sajdok, and Hideaki Yukawa. Transcriptionally regulated adhA gene encodes alcohol dehydrogenase required for ethanol and n-propanol utilization in Corynebacterium glutamicum R. *Applied microbiology and biotechnology*, 76(6):1347–56, oct 2007.

[228] Avi Flamholz, Elad Noor, Arren Bar-Even, and Ron Milo. EQuilibrator - The biochemical thermodynamics calculator. *Nucleic Acids Research*, 40(D1):770–775, 2012.

[229] Elad Noor, Arren Bar-Even, Avi Flamholz, Ed Reznik, Wolfram Liebermeister, and Ron Milo. Pathway Thermodynamics Highlights Kinetic Obstacles in Central Metabolism. *PLoS Computational Biology*, 10(2), 2014.

[230] Bas Teusink, Jutta Passarge, Corinne A Reijenga, Eugenia Esgalhado, Coen C. Van Der Weijden, Mike Schepper, Michael C Walsh, Barbara M Bakker, Karel Van Dam, Hans V Westerhoff, and Jacky L Snoep. Can yeast glycolysis be understood terms of vitro kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry*, 267(17):5313–5329, 2000.

[231] Bryson D Bennett, Elizabeth H Kimball, Melissa Gao, Robin Osterhout, Stephen J Van Dien, and Joshua D Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature chemical biology*, 5(8):593–9, aug 2009.

[232] Meike Goenrich, Rudolf K. Thauer, Hiroya Yurimoto, and Nobuo Kato. Formaldehyde activating enzyme (Fae) and hexulose-6-phosphate synthase (Hps) in Methanosarcina barkeri: A possible function in ribose-5-phosphate biosynthesis. *Archives of Microbiology*, 184(1):41–48, 2005.

[233] Hiroya Yurimoto, Nobuo Kato, and Yasuyoshi Sakai. Genomic organization and biochemistry of the ribulose monophosphate pathway and its application in biotechnology. *Applied microbiology and biotechnology*, 84(3):407–16, sep 2009.

[234] Janet L Donahue, Jennifer L Bownas, Walter G Niehaus, and Timothy J Larson. Purification and Characterization of glpX New Enzyme of the Glycerol 3-Phosphate Regulon of Escherichia coli Purification and Characterization of glpX-Encoded Fructose 1 , 6-Bisphosphatase , a New Enzyme of the Glycerol 3-Phosphate Regulon of Escherichia. *Journal of Bacteriology*, 182(19):5624–5627, 2000.

[235] G N Vemuri, E Altman, D P Sangurdekar, A B Khodursky, and M A Eiteman. Overflow Metabolism in Escherichia coli during Steady-State Growth : Transcriptional Regulation and Effect of the Redox Ratio Overflow Metabolism in Escherichia coli during Steady-State Growth : Transcriptional Regulation and Effect of the Redox Ratio âĂă. *Applied and environmental microbiology*, 72(5):3653–3661, 2006.

[236] Yongjin J Zhou, Wei Yang, Lei Wang, Zhiwei Zhu, Sufang Zhang, and Zongbao K Zhao. Engineering NAD+ availability for Escherichia coli whole-cell biocatalysis: a case study for dihydroxyacetone production. *Microbial Cell Factories*, 12(1):103, 2013.

[237] Ian Wheeldon, Shelley D. Minteer, Scott Banta, Scott Calabrese Barton, Plamen Atanassov, and Matthew Sigman. Substrate channelling as an approach to cascade reactions. *Nature Chemistry*, 8(4):299–309, 2016.

[238] John E Dueber, Gabriel C Wu, G Reza Malmirchegini, Tae Seok Moon, Christopher J Petzold, Adeeti V Ullal, Kristala L J Prather, and Jay D Keasling. Synthetic protein scaffolds provide modular control over metabolic flux. *Nature biotechnology*, 27(8):753–9, aug 2009.

[239] W. W. Cleland. The use of isotope effects to determine enzyme mechanisms. *Archives of Biochemistry and Biophysics*, 433(1):2–12, 2005.

[240] IARC. Monographs on the evaluation of carcinogenic risk to humans: Formaldehyde, 2-Butoxyethanol and 1-tert-Butoxypropan-2-ol. *International Agency for Research on Cancer*, 88:1–390, 2006.

[241] Tunga Salthammer. Formaldehyde in the ambient atmosphere: From an indoor pollutant to an outdoor pollutant? *Angewandte Chemie - International Edition*, 52(12):3320–3327, 2013.

[242] James M Cregg, Knut R Madden, Kevin J Barringer, Gregory P Thill, and Cathy A Stillman. Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. *Mol. Cell. Biol.*, 9(3):1316–23, 1989.

[243] Jonas E. N. Müller, Tonje M. B. Heggeset, Volker F. Wendisch, Julia a. Vorholt, and Trygve Brautaset. Methylotrophy in the thermophilic Bacillus methanolicus, basic insights and application for commodity production from methanol. *Applied Microbiology and Biotechnology*, 99(2):535–551, 2015.

[244] Hiroya Yurimoto, Masahide Oku, and Yasuyoshi Sakai. Yeast methylotrophy: Metabolism, gene regulation and peroxisome homeostasis. *International Journal of Microbiology*, 2011, 2011.

[245] Huba Kalasz. Biological Role of Formaldehyde, and Cycles Related to Methylation, Demethylation, and Formaldehyde Production. *Mini-Reviews in Medicinal Chemistry*, 3:175–192, 2003.

[246] Hiroya Yurimoto, Nobuo Kato, and Yasuyoshi Sakai. Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. *Chemical Record*, 5(6):367–375, 2005.

[247] Jie Zhang, Nikolaus Sonnenschein, Thomas Peter Boye Pihl, Kasper Rud Pedersen, Michael K. Jensen, and Jay D. Keasling. Engineering an NADPH/NADP+ redox biosensor in yeast. *ACS Synthetic Biology*, page acssynbio.6b00135, 2016.

[248] Jingwei Zhang, Jesus F. Barajas, Mehmet Burdu, Thomas L Ruegg, Bryton Dias, and Jay D Keasling. Development of a Transcription Factor Based Lactam Biosensor. *ACS Synthetic Biology*, page acssynbio.6b00136, 2016.

[249] Jie Zhang, Michael K. Jensen, and Jay D. Keasling. Development of biosensors and their application in metabolic engineering. *Current Opinion in Chemical Biology*, 28:1–8, 2015.

[250] Claudio F Gonzalez, Michael Proudfoot, Greg Brown, Yurij Korniyenko, Hirotada Mori, Alexei V Savchenko, and Alexander F Yakunin. Molecular basis of formaldehyde detoxification. Characterization of two S-formylglutathione hydrolases from Escherichia coli, FrmB and YeiG. *The Journal of biological chemistry*, 281(20):14514–22, may 2006.

[251] James Ross Law. *Molecular Basis of Bacterial Formaldehyde Sensing*. PhD thesis, University of Manchester, 2012.

[252] Christopher D Herring and Frederick R Blattner. Global transcriptional effects of a suppressor tRNA and the inactivation of the regulator frmR. *Journal of Bacteriology*, 186(20):6714–6720, 2004.

[253] Tewes Tralau, Pierre Lafite, Colin Levy, John P Combe, Nigel S Scrutton, and David Leys. An internal reaction chamber in dimethylglycine oxidase provides efficient protection from exposure to toxic formaldehyde. *The Journal of biological chemistry*, 284(26):17826–34, jun 2009.

[254] Michael S Packer and David R Liu. Methods for the directed evolution of proteins. *Nature reviews. Genetics*, 16(7):379–394, 2015.

[255] Srivatsan Raman, Jameson K. Rogers, Noah D. Taylor, and George M. Church. Evolution-guided optimization of biosynthetic pathways. *Proceedings of the National Academy of Sciences*, 111(50):201409523, 2014.

[256] Fuzhong Zhang, James M Carothers, and Jay D Keasling. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology*, 30(4):354–359, mar 2012.

[257] Apoorv Gupta, Irene M Brockman Reizman, Christopher R Reisch, and Kristala L J Prather. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nature Biotechnology*, 35(3), 2017.

[258] J C Carlson, A H Badran, D A Guggiana-Nilo, and D R Liu. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat Chem Biol*, 10(3):216–222, 2014.

[259] Christine Nicole S Santos, Mattheos Koffas, and Gregory Stephanopoulos. Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metabolic engineering*, 13(4):392–400, jul 2011.

[260] J Andrew Jones, Victoria R Vernacchio, Daniel M Lachance, Matthew Lebovich, Li Fu, Abhijit N Shirke, Victor L Schultz, Brady Cress, Robert J Linhardt,

and Mattheos A G Koffas. ePathOptimize: A Combinatorial Approach for Transcriptional Balancing of Metabolic Pathways. *Scientific reports*, 5:11301, 2015.

[261] Attila Becskei and Luis Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(June):590–593, 2000.

[262] Jeffrey S Iwig and Peter T Chivers. DNA Recognition and Wrapping by Escherichia coli RcnR. *Journal of Molecular Biology*, 393(2):514–526, 2009.

[263] Jean-Denis Pédelacq, Stéphanie Cabantous, Timothy Tran, Thomas C Terwilliger, and Geoffrey S Waldo. Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology*, 24(1):79–88, jan 2006.

[264] N Arfman, L Bystrykh, N I Govorukhina, and L Dijkhuizen. 3-Hexulose-6-phosphate synthase from thermotolerant methylotroph Bacillus C1. *Methods in enzymology*, 188:391–397, 1990.

[265] Andrea M. Ochsner, Jonas E N Müller, Carlos A. Mora, and Julia A. Vorholt. In vitro activation of NAD-dependent alcohol dehydrogenases by Nudix hydrolases is more widespread than assumed. *FEBS Letters*, 588(17):2993–2999, 2014.

[266] G E de Vries, N Arfman, P Terpstra, and L Dijkhuizen. Cloning, expression, and sequence analysis of the Bacillus methanolicus C1 methanol dehydrogenase gene. *Journal of bacteriology*, 174(16):5346–53, aug 1992.

[267] Kevin M Esvelt, Jacob C Carlson, and David R Liu. A system for the continuous directed evolution of biomolecules. *Nature*, 472(7344):499–503, apr 2011.

[268] Shilpa Nagaraju, Naomi Kathleen Davies, David Jeffrey Fraser Walker, Michael Köpke, and Séan Dennis Simpson. Genome editing of Clostridium autoethanogenum using CRISPR/Cas9. *Biotechnology for Biofuels*, 9(1):219, 2016.

[269] Wenyan Jiang, David Bikard, David Cox, Feng Zhang and Luciano A. Marraffini. CRISPR-assisted editing of bacterial genomes. *Nat biotechnol*, 31(3):233–239, 2013.

[270] Inês Chen and David Dubnau. DNA uptake during bacterial transformation. *Nature reviews. Microbiology*, 2(3):241–9, mar 2004.

[271] J Shaw, David A. Hogsett, and Lee R. Lynd. Natural competence in thermoanaerobacter and thermoanaerobacterium species. *Applied and Environmental Microbiology*, 76(14):4713–4719, 2010.

[272] Erik A. Rodriguez, Robert E. Campbell, John Y. Lin, Michael Z. Lin, Atsushi Miyawaki, Amy E. Palmer, Xiaokun Shu, Jin Zhang, and Roger Y. Tsien. The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. *Trends in Biochemical Sciences*, xx(2):1–19, 2016.

[273] Erik A Rodriguez, Geraldine N Tran, Larry A Gross, Jessica L Crisp, Xiaokun Shu, John Y Lin, and Roger Y Tsien. A far-red fluorescent protein evolved from a cyanobacterial phycobiliprotein. *Nature methods*, 13(9):763–769, 2016.

[274] K Burton. The enthalpy change for the reduction of nicotinamide–adenine dinucleotide. *The Biochemical journal*, 143(2):365–368, 1974.

[275] Gabriel Da Silva, Joseph W. Bozzelli, Nadia Sebbar, and Henning Bockhorn. Thermodynamic and Ab initio analysis of the controversial enthalpy of formation of formaldehyde. *ChemPhysChem*, 7(5):1119–1126, 2006.

Appendix A

Supplemental Information

A.1 Calculation of Yields from H_2 and CO_2 in Figure 1-3

Maximum theoretical hydrogen yields for a variety of products were calculated using known stoichiometry for the conversion of acetyl-CoA (Table A.1), and the *C. ljungdahlii* Wood-Ljungdahl pathway stoichiometry in [41]:

$$4H_2 + 2CO_2 + 0.37ATP \rightarrow AcCoA \tag{A.1}$$

Stoichiometric coefficients were balanced such that there was no accumulation of in-

Product	Stoichiometry
Ethanol (via Acetate)	$2H_2 + AcCoA \rightarrow C_2H_6O + 0.5ATP$
3HB (tesB)	$H_2 + 2AcCoA \rightarrow C_4H_8O_3$
3HB (ptb/buk)	$H_2 + 2AcCoA \rightarrow C_4H_8O_3 + ATP$
Butyrate (ptb/buk)	$2H_2 + 2AcCoA \rightarrow C_4H_8O_2 + ATP$
Butanol	$4H_2 + 2AcCoA \rightarrow C_4H_{10}O$
Butanol (AOR)	$4H_2 + 2AcCoA \rightarrow C_4H_{10}O + 0.5ATP$
2,3-BDO	$3H_2 + 2AcCoA + ATP \rightarrow C_4H_{10}O_2$
Lactate	$2H_2 + AcCoA + 0.5ATP \rightarrow C_3H_6O_3$
4HB	$H_2 + 2AcCoA \rightarrow C_4H_8O_3$
1,4-BDO	$3H_2 + 2AcCoA \rightarrow C_3H_6O_3$

Table A.1: Product Formation Stoichiometry

termediates or biomass. Pathways deficient in ATP production were compensated by the concomitant production of acetate. Production of excess ATP was assumed not to diminish the maximal yield. Electron donors were converted to hydrogen equivalents. For the case of ferredoxin-dependent reactions, this results in an additional cost of 0.5 ATP per reduction (e.g. via the reversal of the ECH-catalyzed reaction). The glucose yield was calculated assuming typical heterotrophic production of acetyl-CoA through glycolysis such that each glucose can produce a maximum of two molecules of acetyl-CoA. This allowed direct calculation of the maximum hydrogen cost for which autotrophic production would be competitive on a feedstock cost basis with glucose-based production.

A.2 Codon Optimized *kanR* and *M. thermoacetica* GAPDH Promoter Sequence

ACCATTTGTGTTGAATAGATAGTGTTTGACGGTACAATCTCCGGCAATT
AGCAATATATCATAATAAATCCTGATTGGGTTAGGAATAATATCAAAAG
CCAAGGAGCCTGAAAGCGGTGGGGGTTGACGCTGCTGGAATTTAACCCT
TGCCGTTACAATAAATATAAGGAGGAGTACAGCATGCGGATCGTGAACG
GCCCCATTATCATGACCCGCGAAGAGCGCATGAAAATTGTTCACGAAAT
CAAGGAACGGATTCTGGATAAATACGGTGACGATGTTAAGGCCATCGGC
GTGTACGGCAGCCTCGGACGGCAGACTGACGGCCCGTATTCGGACATTG
AGATGATGTGCGTAATGAGCACCGAGGAGGCGGAGTTCAGCCACGAGT
GGACGACCGGCGAGTGGAAAGTCGAAGTCAACTTTGACAGCGAAGAAAT
CTTGCTGGATTATGCCTCCCAGGTTGAGTCCGACTGGCCCCTCACCCAC
GGCCAGTTCTTTAGCATTCTCCCCATCTACGACAGCGGGGGCTACCTGG
AAAAAGTATATCAGACCGCCAAGTCCGTGGAAGCCCAGACCTTCCACGA
CGCCATCTGTGCCCTCATTGTGGAAGAACTGTTTGAATACGCTGGAAAG
TGGCGGAACATCCGGGTCCAGGGTCCGACGACCTTTTTGCCGAGCCTCA
CCGTCCAGGTGGCTATGGCCGGCGCCATGCTGATCGGCCTGCACCATCG

TATATGCTACACCACCTCGGCCTCCGTCCTGACTGAAGCCGTCAAGCAG
AGTGATCTGCCGAGTGGCTACGATCATCTGTGCCAGTTTGTGATGAGCG
GTCAGTTGAGCGACAGTGAGAAACTGCTGGAAAGCCTCGAGAATTTCTG
GAATGGTATTCAGGAATGGACTGAAAGGCATGGCTATATCGTGGACGTC
TCGAAACGTATCCCCTTTTAG

A.3   GBlock Used for Expression of sgRNA

CGGGTGGGCCTTTCTGCGTTTTGACAAATTTATTTTTTAAAGTTAAAAT
TAAGTTGTTACCTTTTCATTCCCTACAGTTTTAGAGCTAGAAATAGCAA
GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTTTGAAGCTTGGGCCCGAACAAAAACTCATCTCAGAAGAG
GATCTGAATAGCGCCGTCGACTAAATATCTAGAAGCTCGGCACTGGCCG
TCGTT

A.4 Reactions Added to Core Model iAF1260

| Abbreviation	Reaction
Exchange reactions	
EX_meth(e)	meth_e[C_e]<=>
EX_rib(e)	rib_D_e[C_e]<=>
Transport reactions	
METH_up	meth_e[C_e]<=> meth_c[C_c]
RIB_up	rib_D_e[C_e]+atp_c[C_c]<=>rib_D_c[C_c]+adp_c[C_c]
Metabolic reactions	
MeDH	meth_c[C_c]+nad_c[C_c]<=>h_c[C_c]+nadh_c[C_c]+formald_c[C_c]
HPS	formald_c[C_c]+ru5p_D_c[C_c]<=>h6p_c[C_c]
PHI	h6p_c[C_c]<=>f6p_c[C_c]
RIBK	rib_D_c[C_c]+atp_c[C_c]<=>r5p_c[C_c]+adp_c[C_c]

Table A.2: Reactions Added to Core Model iAF1260

A.5 Calculation of Entropy and Enthalpy of MDH-Catalyzed Reaction

Understanding the effect of temperature on the free energy change (ΔG^0) of NAD-dependent methanol oxidation requires values for the entropy (ΔS^0) and enthalpy (ΔH^0) changes of the reaction. To determine these values, here we use Hess's law in combination with the known heats of formation of methanol, formaldehyde, and hydrogen, and the known ΔH^0 of the H_2-dependent reduction of NAD^+, to first calculate ΔH^0 for the MDH reaction. We then calculate ΔS^0 from the estimated ΔG^0 and the relationship

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 \tag{A.2}$$
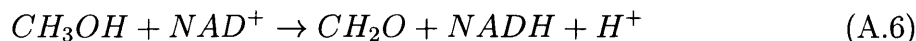
According to [274], the ΔH^0 of the reaction

$$NAD^+ + H_2 \rightarrow NADH + H^+ \tag{A.3}$$
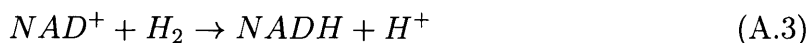
is -29.2 kJ mol^{-1}. To calculate the enthalpy change of the reaction

$$CH_3OH \rightarrow CH_2O + H_2 \tag{A.4}$$

we use the heats of formation shown in Table A.3, and the relationship

$$\Delta H^0 = \sum H_{f_{products}} - \sum H_{f_{reactants}} \tag{A.5}$$

to determine a value of 129.0 kJ mol^{-1}. By Hess's law, the enthalpy change of the reaction

$$CH_3OH + NAD^+ \rightarrow CH_2O + NADH + H^+ \tag{A.6}$$

is 99.8 kJ mol^{-1}. Using this value, the equation A.2, the temperature (T = 298K) and $\Delta G^0 = 34.2$ kJ mol^{-1} we calculate $\Delta S^0 = 0.22$ kJ mol^{-1} K^{-1}.

Species	H_f (kJ mol^{-1})	Reference
CH$_3$OH	-238.0	NIST Webbook
CH$_2$O	-109.0	[275]
H$_2$	0	Definition

Table A.3: Enthalpy of Formation (ΔH_f) of Compounds in Methanol Oxidation

A.6 Reversible Kinetic Model for MDH

MDH was modeled using reversible Michaelis-Menten kinetics for the reaction: $A + B \rightarrow P + Q$

$$v = v_{max}^+ \frac{\frac{[A][B]}{K_A K_B}\left(1 - \frac{\Gamma}{K_{eq}}\right)}{\left(1 + \frac{[A]}{K_A} + \frac{[P]}{K_P}\right)\left(1 + \frac{[B]}{K_B} + \frac{[Q]}{K_Q}\right)} \tag{A.7}$$

A.7 Codon Optimized Genes in Methanol Metabolism

B. methanolicus MGA3 mdh

```
ATGACCACCAACTTTTTTATCCCTCCGGCAAGCGTTATTGGTCGTGGTG
CAGTTAAAGAAGTTGGCACCCGTCTGAAACAAATTGGTGCAAAAAAAGC
ACTGATTGTGACCGATGCATTCTGCATAGCACCGGTCTGAGCGAAGAA
GTTGCAAAAAACATTCGTGAAGCCGGTGTTGATGTTGCCATTTTTCCGA
AAGCACAGCCGGATCCTGCAGATACCCAGGTTCATGAAGGTGTGGATGT
TTTTAAACAAGAGAATTGCGATAGCCTGGTTAGCATTGGTGGTGGTAGC
AGCCATGATACCGCAAAAGCAATTGGTCTGGTTGCAGCAAATGGTGGTC
GTATTAATGATTATCAGGGTGTGAATAGCGTTGAAAAACCGGTTGTTCC
GGTTGTGGCAATTACCACCACAGCAGGCACCGGTAGCGAAACCACCAGT
CTGGCAGTTATTACCGATAGCGCACGTAAAGTGAAAATGCCGGTTATCG
ATGAAAAATCACCCCGACCGTTGCAATTGTTGATCCGGAACTGATGGT
TAAAAAACCGGCAGGTCTGACCATTGCAACCGGTATGGATGCACTGAGC
CATGCAATTGAAGCATATGTTGCAAAAGGTGCAACACCGGTTACCGATG
CCTTTGCAATTCAGGCAATGAAACTGATCAATGAGTATCTGCCGAAAGC
AGTTGCCAATGGTGAAGATATTGAAGCCCGTGAAAAAATGGCCTATGCA
CAGTATATGGCAGGCGTTGCATTTAACAATGGTGGCCTGGGTCTGGTTC
ATAGCATTAGCCATCAGGTTGGTGGTGTGTATAAACTGCAGCATGGTAT
TTGCAATAGCGTGAATATGCCGCATGTTTGTGCCTTTAATCTGATTGCA
AAAACCGAACGCTTTGCCCATATTGCAGAACTGCTGGGTGAAAATGTTG
CGGGTCTGAGCACCGCAGCAGCAGCCGAACGTGCCATTGTTGCACTGGA
ACGTATTAACAAAAGCTTTGGTATTCCGAGCGGTTATGCCGAAATGGGT
GTGAAAGAAGAAGATATCGAGCTGCTGGCAAAAAACGCATATGAAGATG
TTTGTACCCAGAGCAATCCGCGTGTTCCGACCGTTCAGGATATTGCACA
GATTATCAAAAACGCCATGTAA
```

B. methanolicus MGA3 mdh2

ATGACTAACACCCAGTCTGCGTTCTTCATGCCGTCCGTAAACCTGTTTG
GCGCGGGCTCTGTGAACGAAGTTGGTACCCGCCTGGCTGACCTGGGCGT
AAAAAAGGCGCTGCTGGTTACGGACGCTGGCCTGCACGGCCTGGGTCTG
AGCGAAAAATCTCTAGCATCATCCGTGCGGCCGGTGTAGAAGTTTCCA
TCTTCCCGAAAGCGGAACCGAACCCAACCGACAAAAATGTTGCAGAAGG
TCTGGAAGCTTATAACGCGGAGAACTGTGACTCTATTGTGACGCTGGGC
GGTGGCAGCTCTCACGACGCCGGTAAAGCGATCGCTCTGGTTGCGGCTA
ATGGTGGTAAAATCCACGACTATGAAGGCGTGGATGTGAGCAAAGAACC
GATGGTTCCGCTGATTGCTATTAACACGACCGCCGGCACTGGCTCCGAA
CTGACCAAGTTCACGATTATCACTGATACCGAACGTAAAGTTAAAATGG
CTATCGTCGACAAGCATGTGACGCCGACTCTGTCTATCAACGATCCGGA
ACTGATGGTGGGCATGCCACCTAGCCTGACCGCGGCGACTGGCCTGGAC
GCGCTGACCCACGCAATCGAAGCTTATGTCTCCACCGGCGCGACTCCGA
TCACCGATGCCCTGGCTATCCAGGCGATCAAAATCATCAGCAAATACCT
GCCGCGCGCAGTGGCTAACGGTAAAGACATCGAGGCGCGCGAACAGATG
GCGTTCGCACAGAGCCTGGCGGGCATGGCGTTCAACAACGCTGGCCTGG
GCTACGTACACGCTATCGCACATCAGCTGGGCGGTTTCTATAACTTCCC
GCACGGTGTTTGTAACGCTGTCCTGCTGCCGTACGTATGCCGTTTCAAC
CTGATTTCTAAAGTTGAACGCTATGCGGAAATCGCGGCTTTCCTGGGCG
AAAACGTAGATGGTCTGTCCACCTACGACGCTGCGGAAAAAGCGATCAA
GGCAATCGAGCGTATGGCCAAAGATCTGAACATCCCGAAAGGCTTCAAA
GAACTGGGTGCTAAAGAAGAAGACATCGAAACCCTGGCCAAAAATGCCA
TGAAAGACGCGTGTGCCCTGACCAACCCGCGCAAGCCAAAACTGGAGGA
AGTGATCCAGATCATCAAAAACGCCATGTAA

B. methanolicus MGA3 act

ATGGGTAAACTGTTTGAAGAAAAAACCATCAAAACCGAGCAGATTTTTA
GCGGTCGTGTTGTGAAACTGCAGGTTGATGATGTTGAACTGCCGAATGG

271

CCAGACCAGCAAACGTGAAATTGTTCGTCATCCGGGTGCAGTTGCAGTT
ATTGCCATTACCAATGAAAACAAATCGTGATGGTGGAACAGTATCGTA
AACCGCTGGAAAAAGCATTGTTGAAATTCCGGCAGGCAAACTGGAAAA
AGGTGAAGATCCGCGTATTACCGCACTGCGTGAACTGGAAGAAGAAACC
GGTTATGAATGTGAGCAGATGGAATGGCTGATTAGCTTTGCAACCAGTC
CGGGTTTTGCAGATGAAATCATTCATATCTATGTGGCCAAAGGCCTGAG
CAAAAAGAAAATGCAGCAGGTCTGGATGAAGATGAATTTGTGGATCTG
ATTGAACTGACCCTGGATGAGGCACTGCAGTATATCAAAGAACAGCGTA
TCTATGATAGCAAAACCGTTATTGCAGTTCAGTATCTGCAGCTGCAAGA
AGCACTGAAAAACAAATAA

C. glutamicum adhA

ATGACCACTGCTGCACCCCAAGAATTTACTGCTGCTGTTGTTGAAAAAT
TCGGTCATGAGGTGACCGTGAAGGATATTGACCTTCCAAAGCCAGGACC
AAACCAGGCATTGGTGAAGGTACTCACCTCCGGCATCTGCCACACCGAC
CTCCACGCCTTGGAGGGCGATTGGCCAGTAAAGCCGGAACCACCATTCG
TACCAGGACACGAAGGTGTAGGTGAAGTTGTTGAGCTCGGACCAGGTGA
ACACGATGTGAAGGTCGGCGATATTGTCGGCAATGCATGGCTCTGGTCA
GCGTGTGGCACCTGCGAATACTGCATCACCGGCAGGGAAACTCAGTGCA
ACGAAGCTGAGTATGGTGGCTACACCCAAAATGGATCCTTCGGCCAGTA
CATGCTGGTGGATACCCGTTACGCCGCTCGCATCCCAGACGGCGTGGAC
TACCTCGAAGCAGCACCAATTCTGTGTGCAGGCGTGACTGTCTACAAGG
CACTCAAAGTCTCTGAAACCCGCCCGGGCCAATTCATGGTGATCTCCGG
TGTCGGCGGACTTGGCCACATCGCAGTCCAATACGCAGCGGCGATGGGC
ATGCGTGTCATTGCGGTAGATATTGCCGATGACAAGCTGGAACTTGCCC
GTAAGCACGGTGCGGAATTTACCGTGAATGCTCGCAATGAAGATCCAGG
CGAAGCTGTACAGAAGTACACCAACGGTGGCGCACACGGCGTGCTTGTG
ACTGCAGTTCACGAGGCAGCATTCGGCCAGGCGCTGGATATGGCTCGAC
GTGCAGGAACGATTGTGTTCAACGGCCTGCCACCGGGAGAGTTCCCAGC
ATCCGTGTTCAACATCGTATTCAAGGGCCTGACCATCCGTGGATCCCTC

272

GTGGGAACCCGCCAAGACTTGGCCGAAGCGCTCGATTTCTTTGCACGCG

GACTAATCAAGCCAACCGTGAGTGAGTGTTCCCTCGATGAGGTCAATGA

TGTTCTTGACCGCATGCGAAACGGCAAGATTGATGGTCGTGTGGCAATT

CGCTACTAA

G. stearothermophilus mdh

ATGAAAGCAGCGGTTGTCAATGAATTTAAGAAAGCCCTGGAAATCAAAG

AAGTGGAACGCCCGAAACTGGAAGAAGGTGAAGTCCTGGTGAAAATTGA

AGCGTGCGGCGTTTGTCATACCGATCTGCATGCGGCCCACGGTGACTGG

CCGATTAAACCGAAACTGCCGCTGATCCCGGGTCACGAAGGCGTGGGTA

TCGTGGTTGAAGTGGCCAAAGGTGTTAAATCAATTAAAGTCGGCGATCG

TGTGGGTATCCCGTGGCTGTATTCGGCATGCGGCGAATGTGAATACTGC

CTGACCGGTCAGGAAACGCTGTGTCCGCATCAACTGAACGGCGGTTATT

CCGTTGATGGCGGTTATGCAGAATACTGCAAAGCACCGGCTGATTACGT

GGCTAAAATTCCGGATAATCTGGACCCGGTTGAAGTCGCACCGATCCTG

TGTGCTGGCGTCACCACGTATAAAGCACTGAAAGTGAGCGGTGCACGTC

CGGGTGAATGGGTTGCGATTTATGGCATCGGCGGTCTGGGTCACATTGC

CCTGCAGTACGCGAAAGCCATGGGTCTGAACGTCGTGGCAGTGGATATC

AGCGACGAAAAATCTAAACTGGCTAAAGATCTGGGCGCAGACATTGCTA

TCAATGGTCTGAAAGAAGATCCGGTTAAAGCGATTCATGACCAAGTTGG

CGGTGTCCACGCAGCTATCAGCGTGGCCGTTAACAAGAAAGCGTTTGAA

CAGGCCTACCAATCTGTGAAACGTGGCGGTACCCTGGTTGTCGTGGGTC

TGCCGAACGCAGATCTGCCGATTCCGATCTTTGACACCGTTCTGAATGG

TGTCAGTGTGAAAGGCTCCATTGTCGGTACGCGCAAAGATATGCAGGAA

GCACTGGACTTCGCGGCCCGTGGCAAAGTTCGCCCGATTGTCGAAACGG

CGGAACTGGAAGAAATCAATGAAGTGTTTGAACGTATGGAAAAAGGTAA

AATCAACGGTCGTATCGTGCTGAAACTGAAAGAAGATTAA

C. necator mdh2 4.1

ATGACCCACCTGAACATCGCTAATCGCGTCGACAGCTTCTTCATTCCCTG

CGTGACCCTCTTCGGTCCGGGCTGCGTTCGCGAAACGGGCGTGCGCGCC

AGATCACTCGGGGCCAGGAAGGCTCTCATCGTCACGGATGCAGGCTTGC

ACAAGATGGGGCTCTCCGAAGTCGTCGCGGGGCACATTCGCGAAGCCGG

GCTCCAGGCCGTCATCTTTCCGGGTGCCGAGCCCAATCCCACCGACGTT

AACGTTCACGACGGCGTCAAGTTGTTCGAGCGGGAAGAATGCGACTTCA

TCGTTTCGCTCGGCGGCGGCTCATCGCACGACTGCGCGAAAGGCATCGG

CCTCGTTACCGCCGGAGGCGGACATATCCGCGACTACGAAGGCATCGAC

AAATCAACGGTGCCAATGACGCCGCTGATTTCGATCAACACGACCGCTG

GCACTGCTGCGGAAATGACACGCTTTTGCATCATCACTAATTCGAGCAA

TCATGTGAAGATGGTGATCGTCGACTGGCGTTGCACGCCATTAATCGCC

ATCGACGATCCGAGCCTGATGGTCGCGATGCCGCCCGCCTTGACGGCGG

CGACCGGCATGGACGCGTTGACTCACGCCATCGAGGCATACGTTTCCAC

CGCCGCCACGCCAATTACCGATGCCTGTGCGGAGAAGGCGATCGTGCTG

ATCGCCGAATGGCTGCCCAAAGCTGTCGCGAACGGGGACTCGATGGAAG

CACGCGCGGCCATGTGCTACGCCCAATACCTTGCCGGCATGGCCTTCAA

CAACGCATCACTCGGTTACGTGCACGCGATGGCCCATCAACTCGGCGGC

TTCTACAATTTGCCCCACGGCGTGTGCAACGCGATCCTGCTGCCGCACG

TGTCGGAATTCAACCTCATTGCCGCGCCGGAGCGCTACGCGAGAATCGC

CGAACTGCTAGGCGAGAACATTGGGGGCTTGAGCGCGCATGACGCCGCC

AAAGCTGCCGTCTCGGCGATCCGGACCCTTTCCACGTCGATTGGCATTC

CGGCGGGTCTGGCGGGCCTGGGCGTCAAGGCGGACGACCATGAAGTGA

TGGCAAGCAATGCGCAAAAGGATGCTTGCATGCTGACGAATCCGCGCAA

GGCCACGCTGGCGCAAGTCATGGCAATCTTCGCTGCGGCGATGTAA

S. pneumoniae nox

ATGAGCAAAATCGTAGTTGTAGGCGCGAATCACGCAGGCACCGCGTGCA

TCAACACGATGCTGGATAACTTCGGCAACGAAACGAAATCGTGGTATT

TGATCAGAACTCTAACATTAGCTTTCTGGGTTGCGGTATGGCTCTGTGG

ATCGGCGAACAAATTGATGGCGCGGAAGGTCTGTTCTACTCTGACAAAG

AAAAGCTGGAAGCAAAGGGCGCTAAAGTCTACATGAATTCCCCGGTTCT

GAGCATCGACTACGATAACAAGGTCGTCACCGCTGAAGTGGAAGGTAAA

GAGCACAAAGAATCCTATGAAAAACTGATCTTCGCGACCGGCTCCACCC

CGATTCTGCCGCCGATCGAGGGCGTGGAGATTGTTAAAGGTAACCGTGA

ATTCAAAGCTACGCTGGAAAACGTACAGTTTGTAAAACTGTACCAGAAC

GCTGAAGAAGTTATTAACAAACTGAGCGATAAATCTCAGCACCTGGATC

GTATCGCAGTGGTCGGTGGTGGCTACATCGGCGTGGAACTGGCTGAAGC

CTTTGAACGTCTGGGCAAAGAAGTGGTTCTGGTGGACATTGTGGACACC

GTGCTGAACGGCTACTATGATAAAGATTTCACGCAAATGATGGCGAAAA

ATCTGGAGGACCACAACATCCGCCTGGCGCTGGGCCAGACCGTAAAGGC

CATCGAGGGCGACGGCAAAGTTGAACGTCTGATTACCGACAAAGAATCC

TTTGACGTTGACATGGTGATCCTGGCGGTGGGTTTTCGCCCAAACACTG

CGCTGGCTGATGGTAAAATCGAGCTGTTTCGCAATGGCGCGTTTCTGGT

TGATAAGAAACAGGAGACTAGCATCCCAGGTGTGTACGCAGTAGGTGAC

TGCGCTACTGTTTACGATAACGCCCGCAAAGACACCAGCTATATCGCAC

TGGCTTCCAACGCTGTTCGTACCGGTATCGTAGGTGCGTACAACGCGTG

CGGTCACGAACTGGAAGGTATTGGTGTTCAGGGCTCTAACGGTATTTCC

ATCTACGGCCTGCACATGGTTTCCACGGGCCTGACCCTGGAAAAGGCCA

AAGCCGCTGGTTACAACGCCACCGAAACCGGCTTCAATGACCTGCAGAA

ACCGGAGTTTATGAAACACGACAATCATGAAGTGGCCATCAAAATCGTA

TTTGATAAAGATAGCCGCGAAATCCTGGGTGCACAGATGGTGTCTCATG

ACATCGCTATTAGCATGGGTATTCACATGTTCAGCCTGGCTATTCAGGA

ACATGTTACTATCGACAAACTGGCGCTGACCGACCTGTTCTTCCTGCCG

CACTTCAACAAACCGTACAACTACATTACGATGGCCGCACTGACCGCTG

AAAAATAA

A.8 Codon Optimized Genes in 3HB Metabolism

C. necator phaA

CTGCAGAAGGGAGGAAATGAACATGACAGATGTAGTTATAGTGAGCGCA
GCAAGAACAGCAGTAGGAAAATTTGGAGGTAGTTTGGCTAAAATACCTG
CACCAGAATTAGGAGCAGTTGTAATTAAAGCAGCACTAGAAAGAGCAGG
AGTTAAACCAGAACAAGTAAGTGAAGTAATAATGGGACAGGTTTTAACA
GCAGGAAGTGGACAAAATCCAGCAAGACAAGCAGCAATAAAAGCAGGAT
TACCAGCTATGGTACCAGCAATGACAATAAATAAAGTATGCGGATCTGG
TTTAAAGGCAGTAATGTTAGCAGCAAATGCAATTATGGCAGGCGATGCA
GAGATAGTAGTAGCAGGTGGACAAGAAAATATGAGCGCAGCACCACACG
TACTTCCAGGTAGTAGAGATGGATTTAGAATGGGAGATGCAAAGCTAGT
AGATACAATGATAGTAGATGGACTTTGGGATGTATATAATCAATATCAC
ATGGGTATTACAGCAGAAAATGTAGCAAAGAATATGGAATAACAAGGG
AAGCACAAGATGAATTTGCAGTAGGTTCACAGAATAAAGCAGAAGCAGC
TCAAAAAGCAGGAAAATTTGACGAGGAAATAGTACCAGTATTAATACCA
CAAAGAAAAGGCGACCCAGTAGCTTTTAAGACTGATGAATTCGTAAGAC
AAGGCGCAACATTGGATTCTATGAGTGGATTAAAACCAGCATTCGATAA
AGCAGGAACAGTAACAGCAGCAAACGCTTCAGGATTAAATGATGGCGCA
GCAGCAGTAGTAGTTATGAGTGCAGCTAAAGCAAAGAACTAGGACTTA
CACCATTAGCAACAATAAAGAGTTATGCAAATGCAGGAGTAGATCCAAA
AGTTATGGGAATGGGACCAGTACCAGCATCAAAAAGGGCATTATCAAGG
GCAGAATGGACACCACAAGATTTAGATTTAATGGAGATAAACGAAGCTT
TTGCAGCACAGGCACTTGCAGTACATCAACAAATGGGCTGGGACACATC
AAAGGTAAATGTAAACGGCGGAGCAATAGCAATTGGACACCCTATAGGA
GCAAGCGGATGCAGAATATTAGTAACTTTATTACACGAAATGAAAAGAA
GAGATGCTAAAAAGGGCTTAGCAAGTCTATGTATAGGAGGAGGAATGG
GAGTAGCATTAGCAGTAGAAAGAAAATAA

C. necator phaAB

TTTAAAAGGAGGAATATTAAACATGAAAAAGGTATGTGTTATAGGTGCA

GGTACTATGGGTTCAGGAATTGCTCAGGCATTTGCAGCTAAAGGATTTG

AAGTAGTATTAAGAGATATTAAAGATGAATTTGTTGATAGAGGATTAGA

TTTTATCAATAAAAATCTTTCTAAATTAGTTAAAAAAGGAAAGATAGAA

GAAGCTACTAAAGTTGAAATCTTAACTAGAATTTCCGGAACAGTTGACC

TTAATATGGCAGCTGATTGCGATTTAGTTATAGAAGCAGCTGTTGAAAG

AATGGATATTAAAAAGCAGATTTTTGCTGACTTAGACAATATATGCAAG

CCAGAAACAATTCTTGCATCAAATACATCATCACTTTCAATAACAGAAG

TGGCATCAGCAACTAAAGACCTGATAAGGTTATAGGTATGCATTTCTT

TAATCCAGCTCCTGTTATGAAGCTTGTAGAGGTAATAAGAGGAATAGCT

ACATCACAAGAAACTTTTGATGCAGTTAAAGAGACATCTATAGCAATAG

GAAAAGATCCTGTAGAAGTAGCAGAAGCACCAGGATTTGTTGTAAATAG

AATATTAATACCAATGATTAATGAAGCAGTTGGTATATTAGCAGAAGGA

ATAGCTTCAGTAGAAGACATAGATAAAGCTATGAAACTTGGAGCTAATC

ACCCAATGGGACCATTAGAATTAGGTGATTTTATAGGTCTTGATATATG

TCTTGCTATAATGGATGTTTTATACTCAGAAACTGGAGATTCTAAGTAT

AGACCACATACATTACTTAAGAAGTATGTAAGAGCAGGATGGCTTGGAA

GAAAATCAGGAAAAGGTTTCTACGATTATTCAAAATAA

E. coli tesB

AAATTTCAGGAGGTTAGTTAGAATGTCACAAGCATTAAAAAATTTGTTA

ACATTATTAAATTTAGAAAAAATAGAAGAAGGATTATTTAGAGGTCAAA

GTGAAGATTTAGGATTAAGACAAGTATTCGGTGGACAAGTAGTAGGACA

AGCATTATACGCAGCAAAAGAGACAGTTCCAGAAGAAAGATTAGTACAT

TCATTTCATAGTTATTTTTTAAGACCAGGAGATAGTAAGAAACCAATAA

TATACGACGTAGAAACATTAAGAGATGGAAATTCATTTTCAGCAAGAAG

AGTAGCAGCAATTCAAAATGGAAAACCAATTTTTTATATGACAGCATCA

TTTCAAGCTCCAGAAGCAGGTTTTGAACACCAAAAAACTATGCCAAGCG

CACCAGCACCTGACGGATTACCTAGCGAAACACAAATAGCACAATCATT
AGCACACCTTTTACCACCTGTATTAAAAGACAAGTTTATATGCGACAGA
CCATTAGAAGTAAGACCAGTAGAATTCCACAATCCACTTAAAGGACACG
TAGCTGAACCTCACAGACAAGTTTGGATTAGAGCTAATGGAAGTGTACC
AGATGATTTAAGAGTTCACCAATATCTATTAGGCTACGCATCAGACTTA
AATTTTTTACCAGTTGCTTTACAACCACACGGAATAGGATTTTTAGAAC
CAGGAATACAGATAGCTACAATAGATCACAGTATGTGGTTTCACAGACC
TTTTAATCTAAATGAATGGCTTTTATATAGCGTAGAATCAACATCAGCA
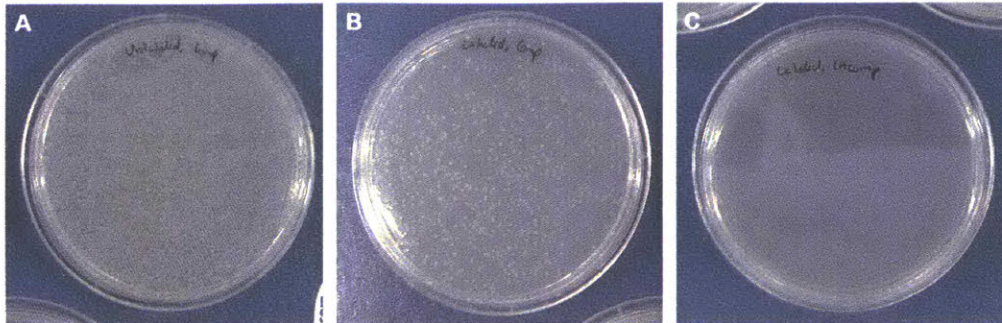TCTAGCGCAAGAGGTTTTGTAAGGGGAGAATTTTATACACAAGATGGCG
TATTAGTTGCAAGTACAGTACAAGAAGGAGTTATGAGGAATCACAATTA
A

Figure A-1: **Verification of Transformation of *E. coli* with Cy5-labeled pUCG18.** Cells were recovered for one hour in SOC following electroporatoin, and plated on LB with 100 μg mL^{-1} ampicillin. A) cells electroporated with unlabeled pUCG18. B) cells electroproated with Cy5-labeled pUCG18. C) cells mixed with labeled pUCG18 but not electroporated. The difference in colony number between A and B is most likely due to a decrease in replication efficiency in the labeled variant, as expected per the manufacturer's guidelines.
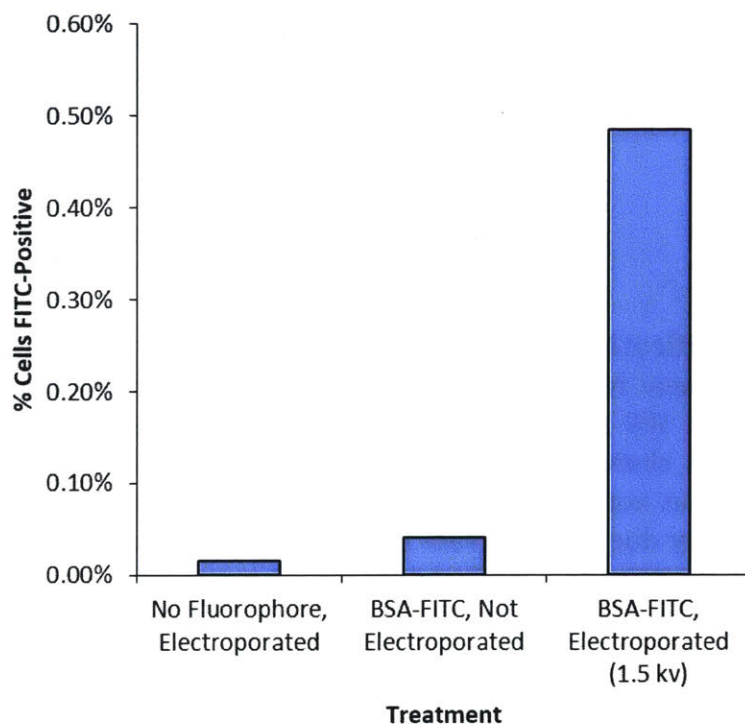
Figure A-2: **Transformation of *Moorella thermoacetica* with fluorescein-labeled plasmid.** Cells of *M. thermoacetica* were transformed exactly as for the Cy5-labeled plasmid, but using a FITC-labeled plasmid, producing similar quantitative results. Due to spectral overlap, PI could not be used to correct for dead cells in the experiment.
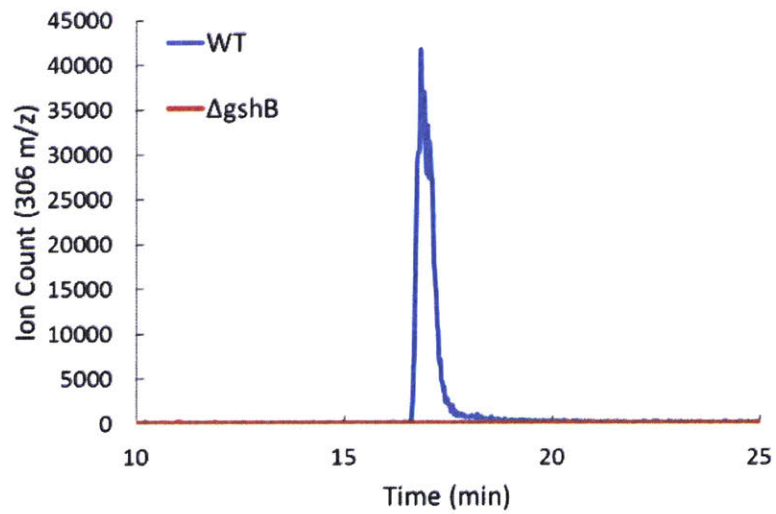
Figure A-3: **LC-MS/MS Confirmation of Glutathione Deficiency in MG1655(DE3) $\Delta gshB$.** Metabolite extracts were analyzed by LC-MS/MS detection of the glutathione fragment (306/143). Blue: WT cells, Red: $\Delta gshB$
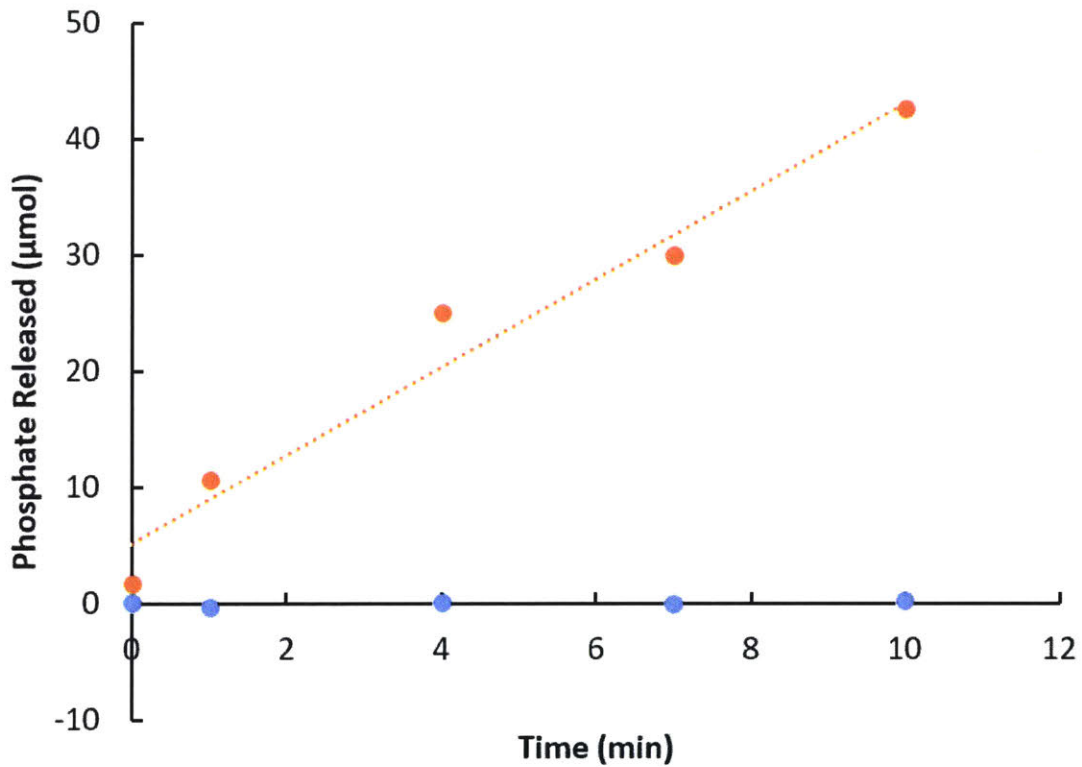
Figure A-4: **Assay of ACT Activity in Crude Lysates of pETMEOH140 and pETMEOH150.** Activity was determined with ADP-Ribose as substrate, and measured by the release of phosphate by Calf Intestinal Phosphatase after generation of the mononucleotide by ACT. Orange: pETMEOH150. Blue: pETMEOH140. Crude extracts were diluted 100-fold before assay.