Development of a novel carbon fixation pathway enabled by protein engineering

Amanda Lee Smith

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Mary E. Lidstrom, Chair

Daniel T. Schwartz

David Baker

Renata Bura

Program Authorized to Offer Degree:
Chemical Engineering

University of Washington

Abstract

Development of a novel carbon fixation pathway enabled by protein engineering

Amanda Lee Smith

Chair of the Supervisory Committee:
Frank Jungers Chair of Engineering Mary E. Lidstrom
Chemical Engineering

Metabolic engineering in conjunction with computational protein design has the ability to expand the range of possible metabolic pathways. As a demonstration of this utility, a novel one-carbon assimilation pathway, the Formolase pathway, was implemented in *Escherichia coli* using a computationally designed enzyme. This pathway was designed to address a need for more sustainable feedstocks for biochemical production. With the documented ability to produce formate from carbon dioxide, this efficient, linear pathway allows for carbon sequestration and utilization through transformation to chemicals or fuels in industrial hosts that do not naturally fix carbon. The development of formolase, a carboligase, through computational protein design enables this pathway by joining 3-1C molecules into 1-3C molecule that accesses glycolysis via phosphorylation. This glycolytic flux can become any number of biochemical products using well-characterized production pathways. In this work, in vitro function of the formolase pathway is presented and efforts to improve pathway function are detailed.

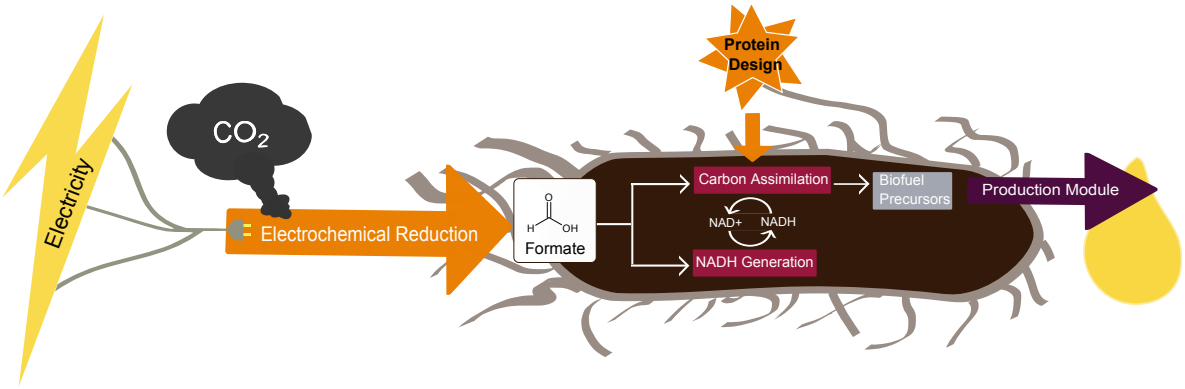


Figure 1: Formolase pathway context and overview. A scheme to convert electrical energy and carbon dioxide into biochemicals. Reducing power in the form of electricity converts carbon dioxide into formate, a molecule that can serve as a cellular energy and carbon source. An engineered microorganism fixes the formate into metabolic intermediates that can serve as biochemical or biofuel precursors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

2/3-PG: 2/3-phosphoglycerate

ACS: acetyl-CoA synthase, E.C. 6.2.1.1, ec = *E. coli*

ACDH: acetaldehyde dehydrogenase, E.C. 1.2.1.10, ec = *E. coli* & lm = *L. monocytogenes*

ADP: adenosine diphosphate

ATP: adenosine triphosphate

AMP: adenosine monophosphate

BAL: benzaldehyde lyase, E.C. 4.1.2.38

COA: coenzyme A

DHAK: dihydroxyacetone kinase, E.C. 2.7.1.29, y = *S. cerevisiae*

DHAP: dihydroxyacetone phosphate

DTT: dithiothreitol

EV: empty vector

FALD: formaldehyde

FDH: formate dehydrogenase, E.C. 1.2.1.2, cm = *C. methylitica*

FLS: formolase

MDH: methanol dehydrogenase

NAD: nicotinamide adenine dinucleotide

NADH: nicotinamide adenine dinucleotide, reduced

PEP: phosphoenolpyruvate

TPP: thiamine pyrophosphate

ACKNOWLEDGMENTS

The author wishes to thank her committee members, especially her mentor, Mary Lidstrom, and members of the Lidstrom Lab throughout her tenure at the University of Washington for their unending support, interesting conversations and generous contributions of time and expertise. Many thanks to members of the CO_2 ARPA-E project team (Janet Matsen, Adam Wargacki, Sean Poust, Jacob Bale, Austin Moon, Curt Fischer and Jason Kelly) and my mentors/co-users of the mass spec (Song Yang, Yanfen Fu, Martin Sadilek, Dale Whittington, Clancey Wolf). Much appreciation for the encouragement, support and discussions about life from fellow grad student Bo Hu. The author also owes a great amount of gratitude to the many wonderful teachers throughout her life: Dr. Ka-Yiu San, Dr. Irene Martinez and Dr. Ann Saterbak, her undergraduate mentors, as well as, Mrs. Crochetiere, Mrs. Brockway, Dr. Chen, Mrs. Applegate and Mrs. Gladden, her middle school and high school science teachers, who inspired her to explore.

DEDICATION

to my husband, Rob, for providing balance in my life, and to my parents, for their boundless support and gift of countless opportunities.

Chapter 1

INTRODUCTION

1.1 Rationale

Metabolic engineering describes the "directed improvement of product formation or cellular properties through the modification of specific biochemical reactions or the introduction of new genes with the use of recombinant DNA technology" [125]. Most of these improvements involve expression of exogenous genes, over-expression of native genes and/or removal of native DNA to knockout competing reactions or change regulation[61, 4, 57]. They commonly result in an increase of a desired product through modification of terminal metabolic pathways or perhaps use of a new substrate by a proximal pathway[149]. As such, metabolic engineers have been limited to introducing and optimizing activities that have been found in other organisms. The advent of protein engineering has begun to change this restriction and expand the accessible metabolic pathway design space[147]. Protein engineering has brought us the ability to use biocatalysis for reactions not found in nature [107, 58, 121]. Incorporating these new enzymes into pathways allows the rewiring of metabolic networks for new functions to address complex problems in sustainable energy and chemical production[144].

Over 80% of renewable energy production in the US results in electricity or heat [135]. Yet, there is a large need for liquid fuels. Liquid fuels are the main source of transportation fuel comprising 71% of liquid fuel usage in 2011, a figure that is only projected to drop by 1% by the year 2040[134]. Liquid biofuels have a great advantage over other renewable energy sources because unlike electricity, liquid fuel as an energy carrier can be easily stored and

transported with the current infrastructure [33]. Electricity production from solar and wind power is subject to fluctuations due to the nature of the resource. Since electrical storage is a challenge and often electric capacity often exists at off-peak hours, engineered fuel-producing microorganisms could be used to harness this excess electricity and convert it into chemical bond energy via cellular energy carriers like NADH [95]. Once the carbon is assimilated into central metabolism, final products could encompass a wide array of electrofuels or electrochemicals (Figure 1.1)[149, 144, 57]. Recently, Li et al. demonstrated the conversion of CO_2 and electricity to 3-methyl-1-butanol with *Ralstonia eutropha*[74].

Using renewable energy in the form of electricity and other ubiquitous chemicals could also reduce dependence on water and land resources, which today's biomass-derived biofuels require. Plants only convert about 4.5-6% of incident solar photons into biomass making the annual photon-to-fuiel conversion efficiency 0.18% for U.S. corn ethanol; microbial electrofuels have the potential to do better[33]. An engineered fuel-producing microorganism using electricity to power the production of a liquid biofuel needs a carbon source as a substrate. To avoid the burdens of a biomass feedstock, a prevalent, low cost chemical is ideal. At over 6 billion metric tons, CO_2 is the most abundant of greenhouse gases, as emissions accounted for approximately 80% of greenhouse gases released by the US in 2007[133]. As such, there is an increasing emphasis on techniques for not only carbon capture/sequestration but also CO_2 utilization. CO_2 has been unattractive as a feedstock for biocatalysis due to inefficiencies in metabolic pathways for CO_2 fixation, but now there is an increased impetus for its use.

Our goal is to create a strain with a corresponding pathway that can use CO_2 and electricity as feedstocks for chemical production. To create a synthetic carbon fixation pathway with the potential for industrial use, we focused on the following characteristics: O_2-tolerance, to maximize the range of energy metabolic schemes possible; efficiency, to maximize productivity; pathway linearity with a minimum number of enzymes, to simplify manipulations;

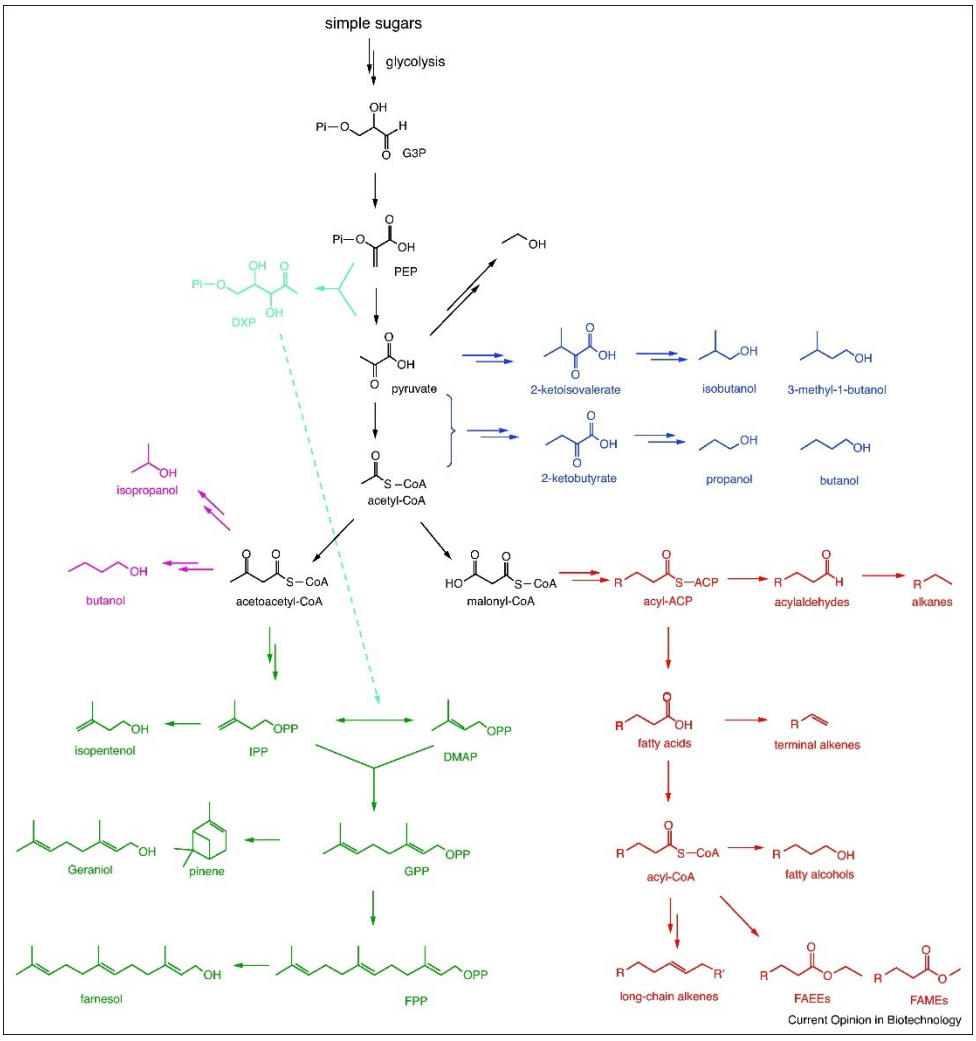


Figure 1.1: A variety of engineered metabolic pathways that branch from central metabolism. These have been engineered for production of advanced biofuels, including isoprenoids, fatty acid esters, alkanes and alcohols. Reproduced with permission from Zhang et al.[149].

and a model host commonly used for biotechnology applications, to facilitate transition to commercialization. This pathway should funnel carbon flux into central metabolic intermediates that serve as precursors to many desirable chemicals. In this way, the pathway can be thought of as an upstream module that can be paired with any number of downstream production modules as described in Figure 1.1.

The crux of this pathway is the ability to bring electricity and CO_2 into cells in accessible forms. Although some microorganisms can utilize electrical current directly, engineering them has numerous challenges[76]. A number of compounds for transferring electrons to cells have been described, however, most do not meet the criteria for our system. The ideal mediator would be plentiful, stable, non-toxic and highly water soluble with the ability to naturally cross cellular membranes in both the reduced and oxidized form. Based on these considerations, the mediator formic acid was a clear choice[10]. The electrochemical reduction of CO_2 into formic acid is well established, although technical challenges for improving efficiency remain[54, 104, 2, 91, 143]. Many microorganisms have pathways involving the use or generation of formate giving rise to the possibility that formate could act as a carbon source as well as an energy source. With formate, we can circumvent the challenges of transferring energy and CO_2 into the cell and instead focus on the transport, partitioning and carboligation of formate itself.

This project is done in collaboration with the Baker group in UW Biochemistry, which has protein engineering expertise. The carbon fixation pathway we describe will serve as the upstream module necessary for electric-powered conversion of formate to acetyl-CoA, where the downstream module contains the pathway required for the transformation of acetyl-CoA to a desired chemical or fuel. Other project partners are Ginkgo Bioworks and the Keasling Lab at UC-Berkeley. The former is investigating alternatives to our pathway, while the latter is developing a downstream production module involving polyketide synthases to build hydrocarbon molecules. We envision this first strain with our pathway as not only a

key element for this project, but also a proof of concept that will be the basis for a library of strains that can transform formate, ultimately CO_2, and electricity into a variety of value-added compounds.

1.2 Other Carbon Fixation Pathways

1.2.1 Native Carbon Fixation Pathways

Several naturally occurring carbon fixation pathways have been identified. These assimilate C1 compounds into the carbon backbone of a larger molecule or utilize C1 compounds either directly or bound to a carrier molecule such as tetrahydrofolate. The most prevalent overall and a member of the former group, the reductive pentose phosphate/Calvin Benson Bassham cycle (rPP) is found in photosynthetic organisms and autotrophic bacteria[14]. Flux through this pathway suffers due to the inefficient carbon dioxide-fixing enzyme Rubisco, which has a low turnover rate and is responsible for losses of over 25% of fixed carbon[103]. Improvements to the enzyme are expected to be difficult and have marginal returns[11]. Four other pathways are known that share two conserved metabolites, acetyl-CoA and succinyl-CoA. In each pathway, acetyl-CoA is carboxylated to succinyl-CoA, then the acetyl-CoA is regenerated by different routes. The reductive TCA (rTCA) cycle and the 3-hydroxypropionate bicycle (HPBC) cycle are found in bacteria that carry out anoxygenic light harvesting in addition to sulfur utilization[12, 22]. Unlike the rTCA cycle, HPBC can function aerobically. The 3-hydroxypropionate/4-hydroxybutyrate cycle (HPHB) and the dicarboxylate/4-hydroxybutyrate cycle (DCHB), which is a combination of the rTCA and the HPHB cycles, are found in Archaea that use H_2 with O_2 and sulfur, respectively, as energy sources[17, 52]. These cycles are aerobic and anaerobic, respectively, and have the largest numbers of reactions at 17 and 14[22].

In the group of C1-unit consumers, the Wood-Ljungdahl/reductive acetyl-CoA pathway (rAcCoA), has the least number of required reactions and is found in strictly anaerobic

bacteria that use H_2 as an energy source[102]. Found in both acetogens and methanogens, it is limited by the O_2 sensitivity of its enzymes and operates near the thermodynamic limit thus requiring a large quantity of protein to support carbon fixation pathway flux[12, 10]. Although not used for autotrophic growth, the glycine synthase (GlcS) pathway and the glyoxylate (GloS) synthetase pathway can potentially also fix carbon[12]. The rAcCoA (acetogens), the GlcS and the GloS pathways use formate created by reduction of CO_2 as an intermediate[12].

Considering formate or other C1 molecules as starting points adds a number of other carbon fixation pathways, including the Serine cycle, the reductive acetyl-CoA pathway and formate oxidation coupled to the rPP[10, 29]. Ribulose mono-phosphate cycle (RuMP; methylotrophic bacteria) and xylulose 5-phosphate pathway (methylotrophic yeast) are other potential routes[60]. RuMP cycle function has been demonstrated in *E. coli*, although only labeling of in vivo metabolites was achieved[90].

Comparisons of these pathways have recently been published. Boyle et al. modeled each pathway in a chosen model organism with thermodynamics and stoichiometry and compared their efficiencies[22]. They found that the chemotrophic systems were always more efficient without the consideration of H_2 production, but with the addition of those costs, the photoautotrophic systems surpass them. The rTCA cycle was identified as slightly more efficient than the rPP cycle, however, the challenges of cell cultivation in an anoxygenic, phototrophic system may outweigh out weight the advantages over the rPP, as well as the fact that oxygenic respiration can produce more energy than fermentative respiration.

Bar-Even et al. also examined natural pathways in addition to considering a new synthetic one. Their pathway analysis hinged on four criteria: pathway specific activity, energetic cost, thermodynamic feasibility and topology. Based on these criteria, an ideal carbon fixation pathway would have high pathway specific activity, an energetic cost that balances a need for a commitment energy input but ultimately has lower ATP and reducing equivalent needs and

a high thermodynamic driving force at each step. Thus, the topology of the pathway would be linear with a minimal number of enzymes and mainly bioorthogonal to existing pathways. This prevents a need for substrate regeneration and simplifies intermediate partitioning issues. Other considerations for such a pathway include function under both anaerobic and aerobic conditions and non-toxic intermediates.

1.2.2 Synthetic Carbon Fixation Pathways

While efforts to move native carbon fixation pathways from one organism to another have continued, synthetic pathways have been explored. Bar-Even et al. proposed a reductive glycine pathway, which is six-step, mainly linear and requires only two non-native enzymes for an *E. coli* host[10]. However, it faces the challenges of potential kinetic limitations and whether the *E. coli* glycine cleavage system can run in the reverse direction. Glycine toxicity may also be an issue[39]. Bogorad et al. created a synthetic pathway called the methanol condensation cycle (MCC)[20]. This pathway combines non-oxidative glycolysis (NOG) with the RuMP cycle. NOG uses a combination of transketolase reactions to prevent the loss of a carbon in the creation of acetyl-CoA from glucose greatly increasing the carbon efficiency of the pathway[21]. Normally, a carbon is lost as CO_2 in the conversion of pyruvate to acetyl-CoA by pyruvate dehydrogenase. As a result of the branch points in the MCC, in vitro chemical production was sensitive to certain enzyme levels, which required adjustment for positive pathway function results. Ethanol and butanol production were achieved in a cell-free system, but no results of in vivo function were reported[20].

1.3 Strain Choice Escherichia coli

Before creating the pathway, we needed to choose a host organism, as this choice dictated what enzymes would be available natively and what type of interference would be expected from native pathways. Since formate will be used a carrier, natural utilization is a consider-

ation, especially since it will need to be transported into the cells. Despite the existence of many formate-utilizing organisms, most are not as well-studied or lack the genetic manipulation tools that exist for industrial model organisms like *Escherichia coli* or *Saccharomyces cerevesiae*[46]. Industrial-scale processes with these model organisms have been realized, creating the advantage of a known quantity for future scale-up with these hosts. In addition, they support high growth rates and carbon fluxes, which many of the natural formate-utilizers do not. This is important as minimum doubling time for autotrophic growth using a simple formate dehydrogenase is predicted to be four hours[10].

As these model organisms do not have pathways for C1 metabolism, attempts to install exogenous carbon fixation pathways into them have had only moderate success. Mattozzi et al. expressed the 3HP cycle in *E. coli* and confirmed function of four pathway modules through reduction of toxicity and amelioration of auxotrophy. However, they were unable to detect growth on formate[82]. Bogorad et al. expressed the methanol condensation cycle, but was only able show function in a cell-free system converting ^{13}C methanol to labeled ethanol and n-butanol[20]. Müller et al. implemented the RuMP pathway with an NAD-linked methanol dehydrogenase in *E. coli*. They were able to demonstrate the in vivo incorporation of ^{13}C methanol into metabolic intermediates in *E. coli*, but did not report growth on methanol [90]. The inherent complexity or unfavorable thermodynamic driving force are potential issues. However, if we are able to design a pathway to avoid these problems, using a model organisms provides a plethora of benefits, including well-developed genetic and computational tools and a large number of well-characterized terminal production pathways. Thus, we have chosen to construct our pathway in *E. coli*.

1.3.1 Native formate metabolism

Formate is produced in *E. coli* during anaerobic and mixed acid fermentation by pyruvate formate lysase (PflB) as it converts pyruvate to acetyl-CoA (Figure 1.2). A formate transporter,

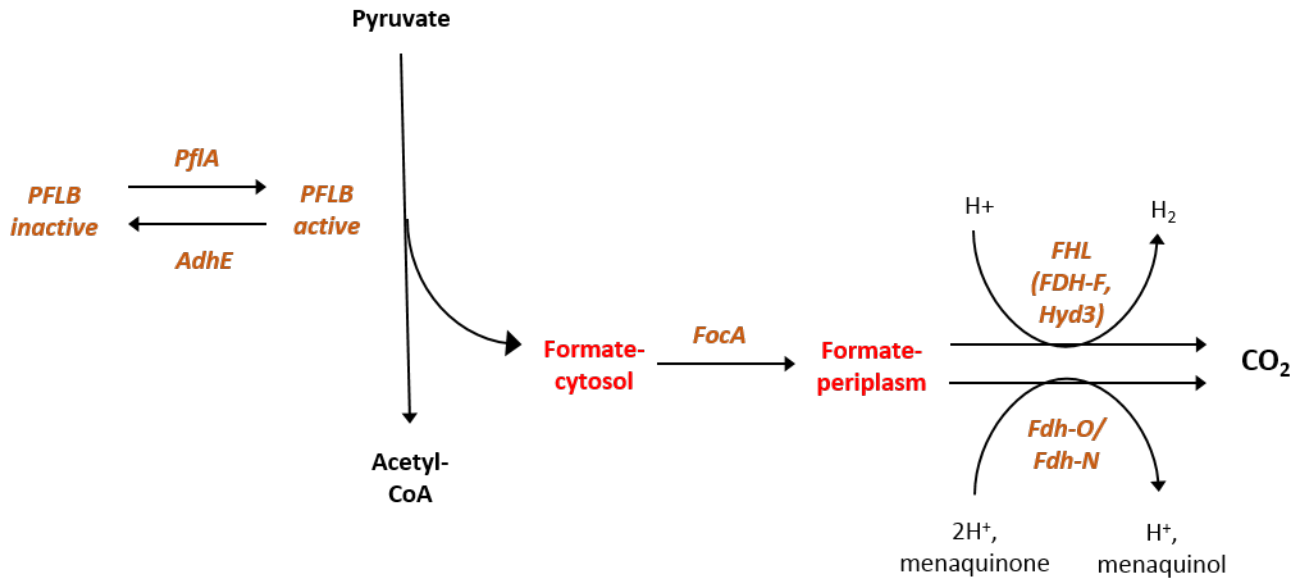


Figure 1.2: Formate metabolism in *E. coli*.

FocA, which is transcribed as part of the *pfl* operon, exports formate from the cytoplasm to the periplasm to prevent acidification of the cytoplasm[127]. Membrane-associated formate dehydrogenases (FDHs) in the periplasm oxidize the formate to carbon dioxide, transferring the electrons to H_2 in the case of FDH-H, part of the formate hydrogen lyase complex, or to a menaquinone in the case of FDH-O or FDH-N[112]. Formate is also involved in the formation of folate molecules and is produced by the glutathione-dependent formaldehyde detoxification pathway[44].

The formation of formate is highly regulated by controlling both transcription and post-translational activity. Transcription of the pfl operon is controlled by Fnr and ArcA, which have weak inductive and suppressive effects, respectively[116]. PflA, a pyruvate formate lyase activating enzyme on the same operon, activates PflB. PflB can also be deactivated by AdhE. Active PflB converts pyruvate to acetyl CoA, thus transferring one-third of the

carbon flux through pyruvate to formate[116]. FocA, part of the pfl transcript, is an integral membrane protein belonging to the formate/nitrite transporter family of membrane transport proteins. The structure of FocA has been determined to be a pentamer channel which allows passive movement of formate[138]. This is important because with the low pKa for formate of 3.77, its buildup in the cytoplasm could lead to acidification. The directionality of the channel can change also. When the medium pH drops below 6.8, FocA can facilitate the uptake of formate. FocB, a putative transporter, may also serve this function[6]. It has a 51% sequence identity with FocA[5]. Protein threading of FocB onto the RSCB database produced a structure of FocB modeled on FocA as the best match and this structure is significantly similar to the FocA structure[62, 146]. However, *focB* is part of the *hyf* operon, which is not expressed under normal conditions in wild type cells[6]. The transcription of this operon, which is predicted to encode a fourth hydrogenase due to its similarity to *hyc*, was only induced during anaerobiosis in the presence of CRP and heterologously expressed HyfR, an analog of formate transcriptional activator, FhlA[115]. FocB is proposed to supply formate to this hydrogenase complex[6].

Once transferred to the periplasm, formate is converted to CO_2 by three isozymes of formate dehydrogenase. FDH-H is part of the formate hydrogen lyase (FHL) complex, which catalyzes the reaction of formate and H^+ to CO_2 and H_2. *fhl* gene expression is controlled by FhlA. Under anaerobic conditions, formate binds to FhlA activating the protein, which in turn activates the transcription of σ54-dependent promoters, including the promoter for the FHL complex[73, 114, 83]. The other two FDH isozymes, FDH-N and FDH-O, are similar and donate electrons to menaquinones upon oxidation of formate to CO_2[111]. The operon *fdnGHI* encoding FDH-N is activated under anaerobiosis in the presence of nitrate and the electrons transferred to the menaquinone pool ultimately are conveyed to nitrate in the cytoplasm[137]. Constitutively expressed FDH-O is thought to be active in the transition from aerobic to anaerobic growth[1]. Each of these enzymes must to be taken into account

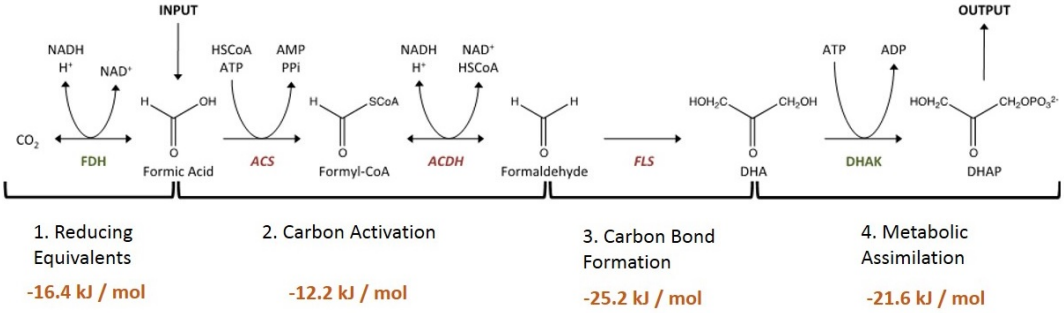


Figure 1.3: The formolase pathway. Enzymes: FDH-formate dehydrogenase, ACS-acetyl CoA synthetase, ACDH-acylating acetaldehyde dehydrogenase, FLS-formolase, DHAK-dihydroxyacetone kinase. Intermediates: DHA-dihydroxyacetone, DHAP-dihydroxyacetone phosphate. Gibb's Free Energy for each step is given[96].

as we design a host strain for the formolase pathway.

1.4 The Formolase Pathway

The pathway we have implemented in *E. coli* has been termed the Formolase pathway. This pathway activates formate to formaldehyde, uses the engineered enzyme formolase (FLS) to convert formaldehyde to dihydroxyacetone (DHA), and transforms DHA to dihydroxyacetone phosphate (DHAP), a central metabolic intermediate. NADH to drive the pathway is generated via formate dehydrogenase (FDH). Figure 1.3 displays the formolase pathway broken down into four steps, excluding formate uptake.

1.5 Pathway Steps

1.5.1 Formate Uptake

Formate enters the periplasm by diffusion potentially through outer membrane proteins, OmpC and/or PhoE. Once in the periplasm, one of the putative formate transporters, FocA

or FocB, will transfer the formate into the cytoplasm, where it is partitioned between formation of reducing equivalents and carbon activation.

1.5.2 Reducing Equivalent Formation

As none of the native *E. coli* FDHs are NAD-linked, a heterologous version must be added to supply the pathway with reducing equivalents. We acquired a FDH gene from our collaborators at Ginkgo Bioworks. Originating from *Candida methylitica*, a methylotrophic yeast, this FDH prefers NADH to NADPH and has been expressed in *E. coli* [18, 3].

1.5.3 Carbon Activation

Native NAD-linked enzymes that convert formaldehyde into formate are irreversible at physiological pH, which is unsurprising given that the Gibbs free energy of the reaction is estimated over 40 kcal/mol[96]. Conversion of formate into formaldehyde thus requires activation of the formate molecule. Enzymes that function on C2 analogs of these compounds, acetate and acetaldehyde, are known and may be promiscuous enough to catalyze the analogous C1 reactions. Acetyl-CoA synthetase (ACS) converts acetate into acetyl-CoA consuming two high energy phosphate bonds, converting ATP to AMP in the process[23, 123]. *E. coli* cells excrete acetate during growth on glucose. As the cells transition to stationary phase, they induce *acs* in the process[70]. The cells can then uptake the excreted acetate and using ACS, activate it to acetyl-CoA, which enters the TCA cycle[23, 70]. Acylating acetaldehyde dehydrogenase (ACDH) catalyzes the reversible reaction of acetyl-CoA to acetaldehyde consuming an NADH[31]. This enzyme is normally involved in the degradation of aromatic compounds. One concern with its over-expression is the potential production of acetaldehyde from the normal acetyl-CoA pools.

1.5.4 Carbon Bond Formation

Carbon bond formation is accomplished with the novel enzyme, Formolase (FLS). The Baker group at University of Washington created FLS using a combination of rational design and random mutagenesis. Designed to catalyze the formose reaction, FLS was generated by first modeling alterations in the active site of benzaldehyde lyase,which catalyzes an analogous reaction (Figure 1.4), then identifying mutations giving function specifically on formaldehyde[120] . Rational design was employed *in silico* using their ROSETTA computational protein design program, which has previously contributed to successful novel enzyme creation[106, 105]. The transition state of the enzyme with formaldehyde as a substrate was used as an input to ROSETTA. Different amino acid positions within a specific proximity to the active site were chosen as targets. Other amino acids were simulated at these sites looking for a minimization of Gibb's free energy and a better fit for the formaldehyde molecule.

Through over four rounds of a simulation, build and test cycle, a 26-fold increase in catalytic proficiency over BAL was achieved with four mutations: A28I, A394G, G419N and A480W[120]. Computationally-guided site-directed mutagenesis identified one beneficial mutation, L90T, and error-prone PCR found two more, W89R and R188H, giving the final FLS with seven mutations and a 100-fold increase in activity for the formose reaction. These mutations effect a variety of changes, including transition state stabilization, packing interactions and unknown functions. This represents a specificity switch of greater than 10-million fold between BAL and FLS (Table 1.5.4)[120]. The Stoddard group at the Fred Hutchinson Cancer Research Center has produced a structure of the FLS (PDB 4QQ8) with X-ray crystallography that matches well with the predicted structure[120].

In addition to its extremely low catalytic efficiency, protein solubility problems are apparent with large quantities of white precipitate occurring in crude cell lysates expressing high levels of FLS at optimum growth temperatures. For this reason, strains with FLS require expression at 18 °C. The desired product for the FLS reaction is the three-carbon

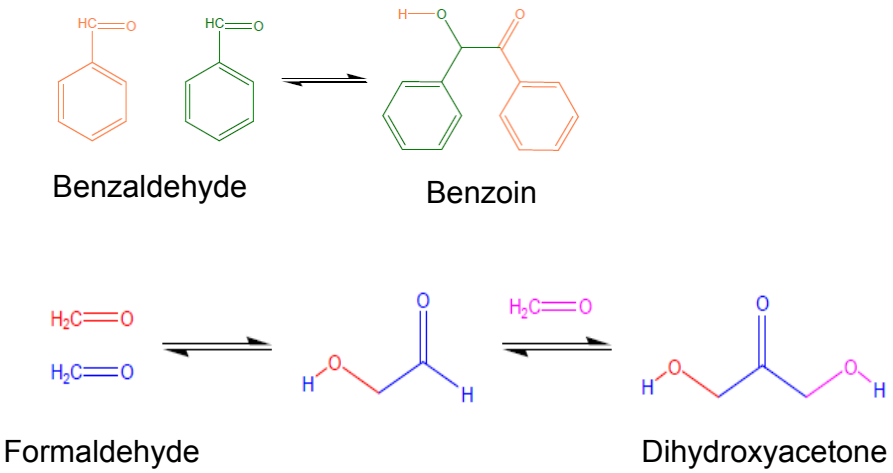
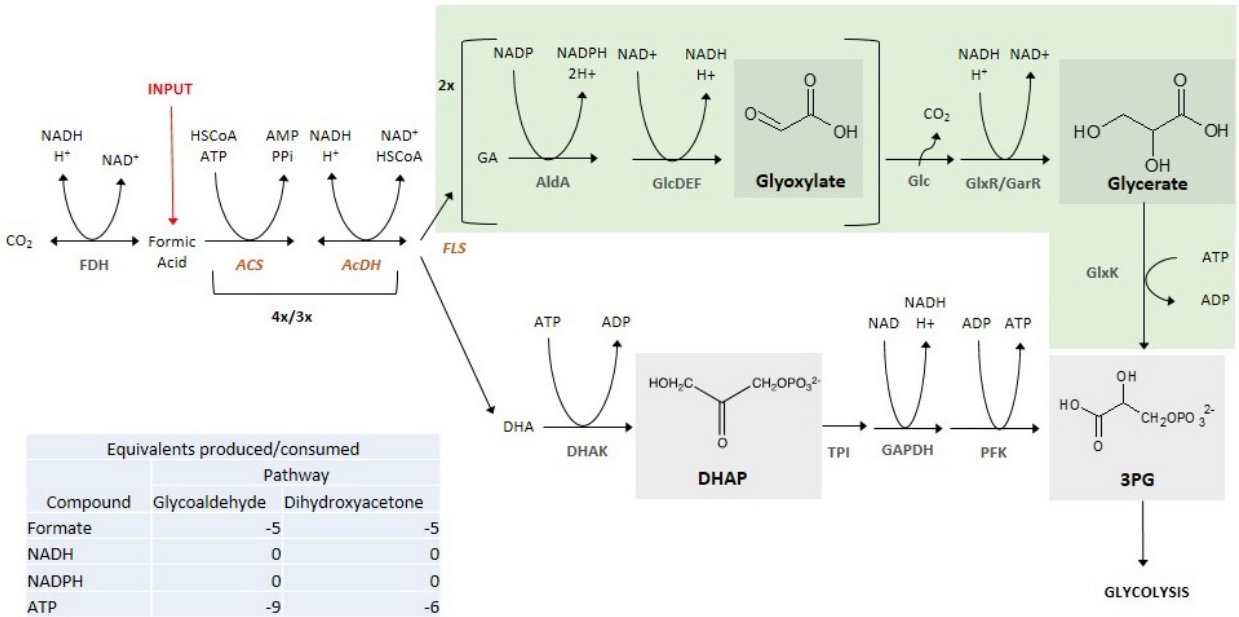


Figure 1.4: Reactions of formolase precursor, benzaldehyde lyase, and formolase. Top. The natural benzaldehyde lyase reaction, which converts benzaldehyde to benzoin. Bottom. The reactions catalyzed by FLS. Two formaldehyde molecules are ligated to form glycolaldehyde (GA), which then reacts with a third formaldehyde molecule to form dihydroxyacetone(DHA)[120].



Equivalents produced/consumed		
	Pathway	
Compound	Glycoaldehyde	Dihydroxyacetone
Formate	-5	-5
NADH	0	0
NADPH	0	0
ATP	-9	-6
HSCoA	0	0
CO2	2	2
3PG	1	1

Figure 1.5: An alternate route to 3-phosphoglycerate via glycolaldehyde shown in green. The normal Formolase pathway is below for reference. The inset table gives a comparison of the pathways in terms of efficiency. Net values are given.

	Catalytic Efficiency, $M^{-1} \cdot s^{-1}$	
	Benzoin Reaction	Formose Reaction
BAL	$1.8 \times 10^3 \pm 1.1 \times 10^2$	$5.0 \times 10^{-2} \pm 4.0 \times 10^{-3}$
FLS	n.d.a	4.7± 0.1

Table 1.1: Kinetic constants of BAL and FLS from Siegel et al.[120]. The catalytic efficiency, (kcat/Km in units of $M^{-1}s^{-1}$) for both FLS and BAL for both the benzoin and formose reaction obtained from linear fits to substrate versus velocity profile. All fits had at least four independently measured rates with an R^2 greater than 0.9. n.d.a. no detectable activity[120].

DHA molecule, however, the two-carbon glycolaldehyde (GA) molecule is also a product and the four-carbon unit made of two GA molecules could also occur. Figure 1.5 presents an alternate pathway for GA consumption that consumes three more ATP than the formolase pathway. This potential pathway joins central metabolism at 3-phosphoglycerate (3PG).

1.5.5 Metabolic Assimilation

The native *E. coli* dihydroxyacetone kinase (DHAK) uses the conversion of PEP to pyruvate to power the phosphorylation of DHA to form DHAP[49]. Induced by DHA, the *dhaKLM* operon encodes three subunits that are involved in its regulation by interactions with the transcriptional activator DhaR. The activity of DhaR is controlled by mutually exclusive binding by DhaK, DHA absent, and DhaL::ADP, DHA present[9]. Since the native version is directly connected to the PTS-system for sugar transport, we chose an exogenous enzyme from *Saccharomyces cerevisiae* that transfers the phosphate from ATP instead of PEP[56].

1.6 Comparison to other CO_2/formate pathways

The formolase pathway is predicted to support a biomass yield higher than any other carbon fixation pathway (6.5 gCDW/mole-Formate) except the reductive TCA cycle (6.7 gCDW/mole-Formate)[10, 120] (Figure 1.6). These values are based on flux balance analysis with the exclusion of ATP maintenance energy. Unlike the rTCA cycle which requires

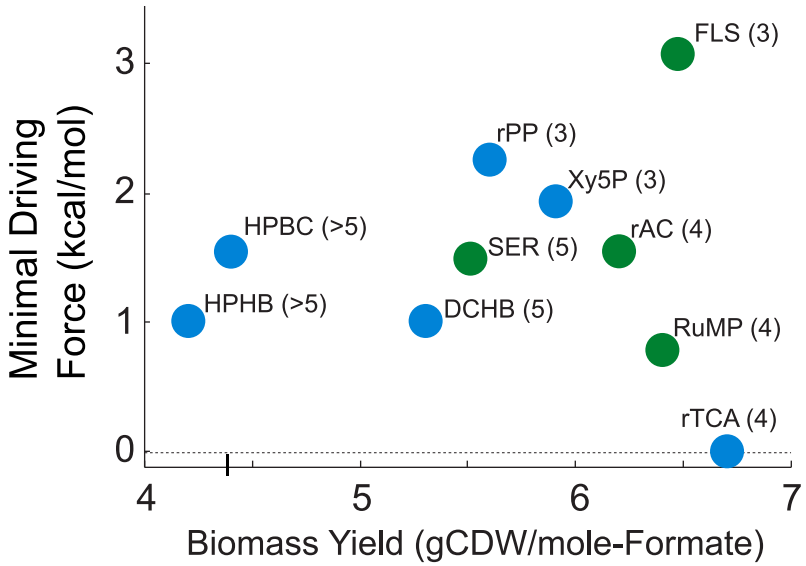


Figure 1.6: Thermodynamics and Carbon Utilization Efficiencies of C1 Pathways reproduced from Siegel et al.[120]. Biomass yield, in gram cellular dry weight (gCDW) per mole of formate consumed, was calculated using flux balance analysis and the core metabolic model of *E. coli* supplemented with pathway enzymes and without considering ATP maintenance[10]. The Min-max Driving Force (MDF) is the lowest value of $-\delta_r G$ in the pathway (i.e. the reaction(s) with the smallest chemical driving force force)[97]. Green dots correspond to pathways that can directly assimilate formate, while blue dots mark carbon fixation pathways that can indirectly assimilate formate by its initial oxidation. Numbers in parentheses indicate the number of foreign enzymes that need to be expressed in *E. coli* to establish an active pathway. Pathway abbreviations are as follows: HPHB, 3-hydroxypropionate-4-hydroxybutyrate cycle; HPBC, 3-hydroxypropionate bicycle; DCHB, dicarboxylate-4-hydroxypropionate cycle; SER, serine cycle; rPP, reductive pentose phosphate cycle/Calvin-Benson-Bassham cycle; Xy5P, xylulose 5-phosphate cycle/dihydroxyacetone cycle); rAC, reductive acetyl-CoA pathway; RuMP, ribulose 4-phosphate cycle, FLS, formolase pathway and rTCA, reductive TCA cycle.

O_2-sensitive enzymes and is thermodynamically constrained, the formolase pathway can operate under aerobic conditions and is thermodynamically favorable. It has a chemical driving force 1.2 kcal/mol greater than any of the other pathways[10]. Under physiological conditions, each step of the formolase pathway dissipates at least 3.6 kcal/mol impeding backward flux[96].

1.7 Summary

Applying protein design to metabolic engineering has enabled the creation of a novel carbon fixation pathway in *E. coli*. Linear, efficient and oxygen-tolerant, unlike native carbon fixation pathways, this formolase pathway assimilates formate units into central metabolism. It could be implemented in biochemical-producing microbes to shift them from sugar utilization to carbon dioxide utilization via an electrochemical reduction of carbon dioxide to formate, thus enabling both the creation of valuable chemicals and the sequestration of a greenhouse gas.

Chapter 2

THE FORMOLASE PATHWAY

This chapter describes the construction of two rounds of pathway DNA constructs, individual component testing and the evaluation of the pathway as a whole.

2.1 Initial Pathway Constructs

Pathway genes were prepared for use with the MIT Standard Registry of Parts Biobricks by removal of specific cut sites[89]. This system uses a set of four cut sites, two on either side of the gene, such that cutting at one cut site on each side followed by ligation results in a construct without a cut site between pieces and the regeneration of the four cut sites [67, 117, 18]. This enables easy movement of DNA pieces into the constructs. Using a set of similar plasmid backbones with a range of copy numbers and promoter-strengths added flexibility to modulate expression levels. Codon usage was also optimized for expression in *E. coli* employing the algorithms developed by DNA 2.0[142]. The initial constructs are shown in Figure 2.1. In addition to the four cut sites that flank all Biobrick parts, the *fls* gene has two additional cut sites flanking the gene, which allows the *fls* gene to be easily exchanged with more recent versions of the gene as improvements are achieved. Ribosome binding sites (RBS) and promoters are also included as part of the vector backbones[89]. Construct B has the inducible *lac* promoter, where a constitutive promoter, $_p$J23100, chosen to have a similar expression level to pLac, was used for Construct A.

Construct A, FLS-DHAK, was cloned into vector pSB1A3, a high copy number plasmid with ampicillin resistance [89]. Therefore, the second plasmid must have a different origin of replication in order to be compatible as two plasmids with replicons in the same incom-

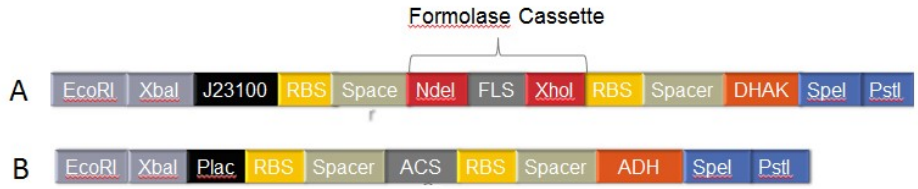


Figure 2.1: Initial gene constructs synthesized by DNA 2.0. Construct A contains the FLS and DHAK genes and was inserted into pSB1A3. Construct B contains genes ACS and ACDH and goes into pSB3K3.

patibility group cannot replicate stably in the same host. The second plasmid, pSB3K3, for construct B, has a compatible but lower copy number origin and a kanamycin resistance gene[89]. The combination of these constructs represents the core of the formolase pathway, the Carbon Activation, Carbon Assimilation and Metabolic Assimilation sections. Another set of Biobrick vectors with the combination of a formate transporter and a formate dehydrogenase and an analogous set with the addition of DHAK were also made. These serve in concert with the core pathway modules. A list of these and the other main constructs used is located in Table 2.1. Some variations on the constructs listed were also made, but differ mainly in the promoter used and are specified when discussed.

2.2 Initial Strain Engineering

An evolution approach for identifying strains with improved growth on or tolerance of formate and formaldehyde was employed. MG1655 with only pSB1A3 FLS-DHAK or pSB1A3 FLS-DHAK and pSB3K3 ACS-ACDH, was serial passaged with a minimal medium for between 60 to 85 passages, with formaldehyde or formate as a substrate, respectively. Figure 2.2 shows improvements in growth achieved over a number of serial passages for both formate and formaldehyde. In the formaldehyde evolution, no growth was measured from any of the

Construct	**Description**
pET29b+	40 copies/cell, pBR322 derived origin, T7 promoter
pSB1A3	100-300 copies/cell, pUC19 pMB1 derived origin, ApR
pSB3K3	20-30 copies/cell, p15A pMR101 derived origin, Km
pSB4C5	4-5 copies/cell, pSC101 origin, CmR
pSBxXx-EV	empty vector versions of the pSBxXx plasmid with no gene expressed
pSB1A3 FLS-DHAK	Vector pSB1A3 with $_p$J23100 FLS-DHAK (Construct A)
pSB3K3 ACS-ACDH	Vector pSB3K3 with ACS-ACDH (Construct B)
pSB1A3 ACS-ACDH FLS-DHAK	Vector pSB1A3 with T7 ACS-ACDH J23100 FLS-DHAK
pSB3K3 T7 FocA	Vector pSB4C5 with T7 FocA
pSB4C5 FDH	Vector pSB4C5 with cmFDH

Table 2.1: List of main initial constructs used in this study[67, 117, 89].

early passages in 35 hours, however, the formaldehyde 40th passaged strain showed growth by hour ten. For the formate evolution strains, increases in growth rate and decreases in lag time were seen with an increase in the number of serial passages. The strains were then cured of their plasmids by repeated streaking on non-selective LB agar. Growth curves comparing the cured strains to the initial wild type MG1655 strain and to a fresh transform of the starting strain with evolved plasmid(s) are shown in Figure 2.2. The cured evolved strains for both formate and formaldehyde, designated ALA1 and ALA2 respectively, had higher growth rates and shorter lag times than the uncured evolved strains, probably due to removal of the plasmid(s). Both fresh transformations with the evolved plasmids performed no better than the empty vectors suggesting that the majority of alterations were in the strain itself.

Next, knockouts of other formate sinks were considered. Collaborators at Ginkgo Bioworks did initial measurements of formate dehydrogenase activity in *E. coli* under various conditions. In their aerobic test, formate dehydrogenase activity was present, attributable to FDH-O and described by literature[18, 110]. The corresponding gene, *fdoG*, was thus removed from our MG1655 strain to eliminate this activity using P1 phage transduction from

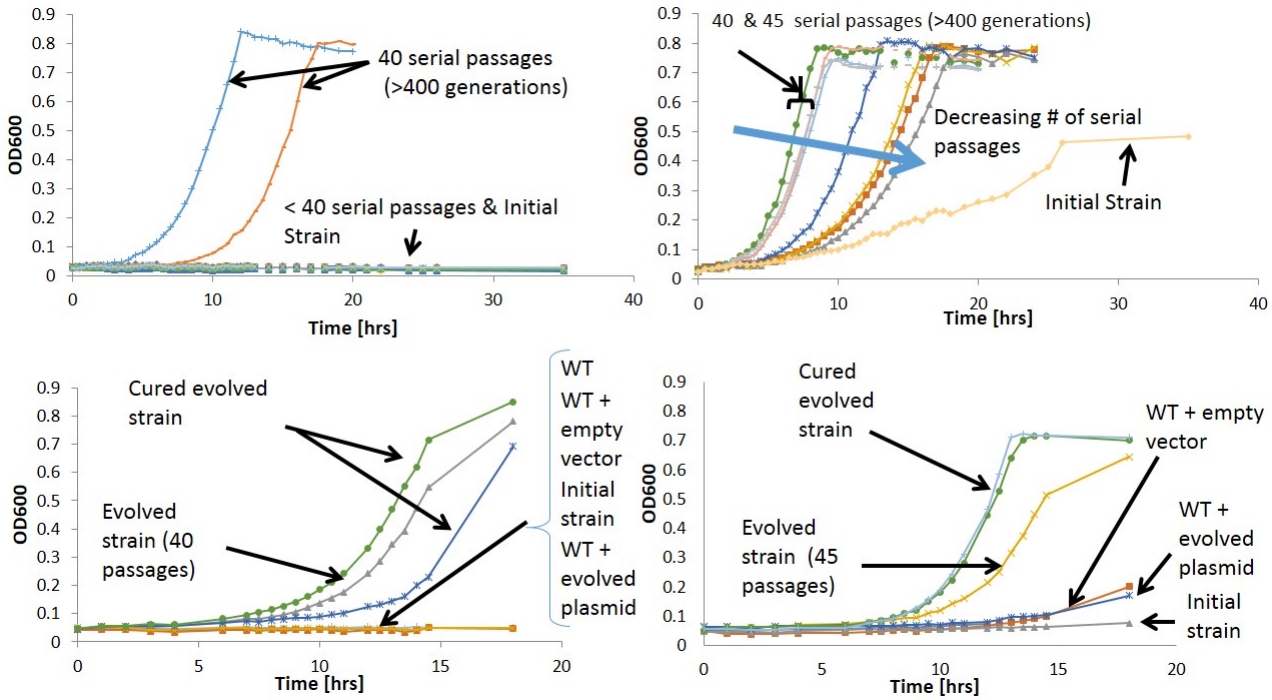


Figure 2.2: Growth curves for serial passaged evolved strains. Growth curves at 37 °C for TOP) WT FLS-DHAK (Formaldehyde) and BOTTOM) WT pSB3K3 ACS-ACDH pSB1A3 FLS-DHAK (Formate). Evolution on LEFT) formaldehyde and RIGHT) formate.

.

a Keio collection strain JW3865-2[8]. Lack of aerobic formate dehydrogenase activity was confirmed in the knockout by measuring a decrease in supernatant formate concentration in strains with and without the knockout.

2.3 Initial Construct Testing

Individual sections of the formolase pathway were assayed for function in vitro. Cell cultures expressing the appropriate proteins were lysed to create extracts that were then treated to remove interfering molecules.

2.3.1 Acetyl-CoA Synthethase (ACS) and Acylating Acetaldehyde Dehydrogenase (ACDH)

First, we checked activity to make sure these enzymes could function on C1 compounds rather than C2 compounds. For ACS, the reaction was coupled to a secondary step where hydroxylamine reacts with the formyl-CoA product to form a hydroxamic acid that absorbs at 520 nm[23]. Although over-expressing ACS only increased specific activity 20% over wild type with formate, the activity on formate was less than 1/3 that of acetate. For ACDH, activity was measured via NADH consumption (absorbance at 340 nm) with formaldehyde as a substrate since formyl-CoA was not available. Significant activity was detected over wild type background. After individual assays, a coupled assay with formate as a substrate ascertained by spectrophotometric measurement of NADH consumption that ACS and ACDH could function together.

In addition, formaldehyde production in cell-free extracts was measured in comparison to a negative control with vector only and to a positive control with a mixture of purified proteins. Formaldehyde measurements were accomplished using the Nash assay, which employs the Hantzech reaction for production of a colorimetric product[92]. Very little production of formaldehyde is measured for any of the crude cell extracts over 20 hours while the purified proteins produced drastically more(Figure 2.3). Purified protein studies with each enzyme

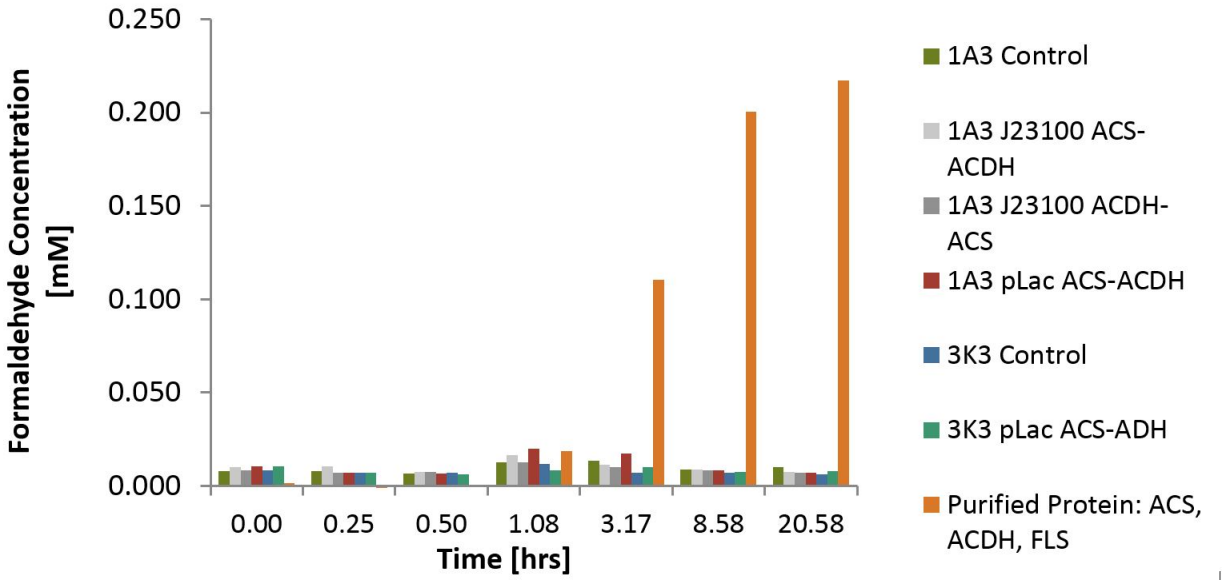


Figure 2.3: Formaldehyde production as measured by the Nash for a variety of crude cell lysates and purified protein. The purified proteins produce drastically greater amounts of formaldehyde. Control designates an empty vector plasmid.

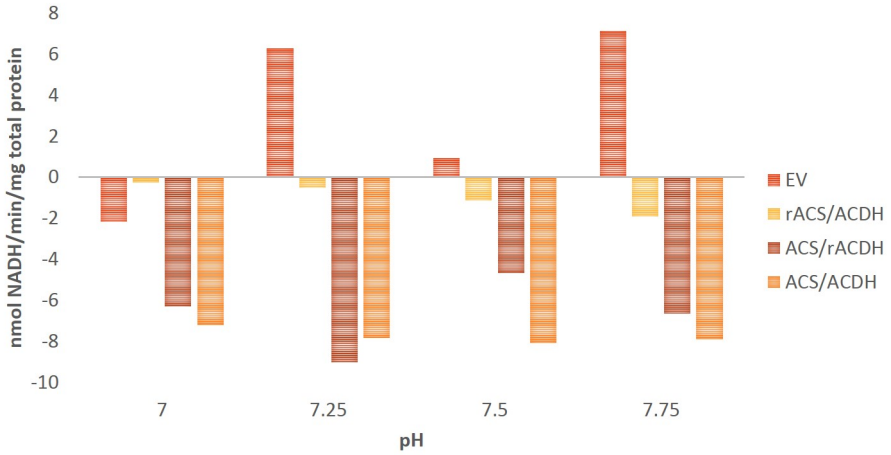


Figure 2.4: NAD-linked assay of the Carbon Activation step with ecACS and ACDH from *L. monocytogenes*. r signifies a reduction in amount of enzyme-expressing lysate to 10% of the original.

amount reduced by 90% also showed that pathway output largely depends on the amount of ACDH and FLS. Another NAD-linked couple enzyme experiment confirmed that ecACDH is that bottleneck of Carbon Activation as decreasing the ecACDH amount reduced the NADH oxidation rate, while decreasing the ACS amount did not[118].This led to the search for an improved ACDH.

Activities of the new ACDH variants were measured by following NADH oxidation. The best of the new variants, lmACDH, has a specific activity of 5.3 $M^{-1}s^{-1}$ on formate compared to 0.050 $M^{-1}s^{-1}$ for ecACDH, over 100-fold improvement[118]. The coupled NADH assay was then used to confirm activity of the new Carbon Activation module. Decreases in lmACDH and ACS only generated dramatic rate reductions when ACS concentration is reduced, identifying ACS as the new bottleneck(Figure 2.4).

2.3.2 Formolase

Measurements of dihydroxyacetone (DHA), the product of the FLS reaction, were made with LC-MS/MS using a PFPP column[145]. With formaldehyde as a substrate and no ATP added, DHA accumulation was detected over the course of 10 hours for the crude cell lysate with FLS-DHAK expressed, but not for the empty vector control (Figure 2.5). With the addition of ATP, there was a lack of DHA and a buildup of DHAP, as expected. Issues with the samples preparation and equipment caused a change in methods such that we could no longer detect DHA via LC-MS/MS, so in further studies other glycolytic compounds were followed instead by adding ATP to the assay mixture [24] (Figure 2.6).

2.3.3 Formaldehyde Tolerance/Formate Growth

As all core pathway enzymes showed activity and a flux balance model of *E. coli* with the pathway added predicted growth, albeit at a low rate, evidence of formaldehyde tolerance/growth or formate growth tied to presence of the pathway was sought[99, 38, 41]. Growth experiments of this nature were completed using a set of four media with all combinations of $-/+$ formate (normally 40 mM) of formaldehyde (varies in the μM range) and $-/+$ glycerol (normally 0.1%, tested up to 0.4%) in a minimal media, Keasling medium (KM) or M9 [93, 64]. Temperatures tested were 37 °C or room temperature ($\sim$22 °C). Cultures were grown in tubes in a shaker (5 mL) or Bioscreen plates in the Bioscreen automated growth curve machine (0.210 mL). In both cases, cultures were kept oxygenated with high shaking, 225 rpm or high shaking setting, respectively. Strains with FLS-DHAK expressed in pSB1A3 did not show increased formaldehyde tolerance in KM minimal media compared to an equivalent strain carrying pSB1A3 empty vector. Formaldehyde concentrations were varied from 0 to 0.6 mM and glycerol concentration was either 10, 20 or 30 mM. No growth was observed in greater than 100 μM formaldehyde. The glycerol concentration seemed to control max OD reached. The presence of the pathway caused decreases in growth rate and

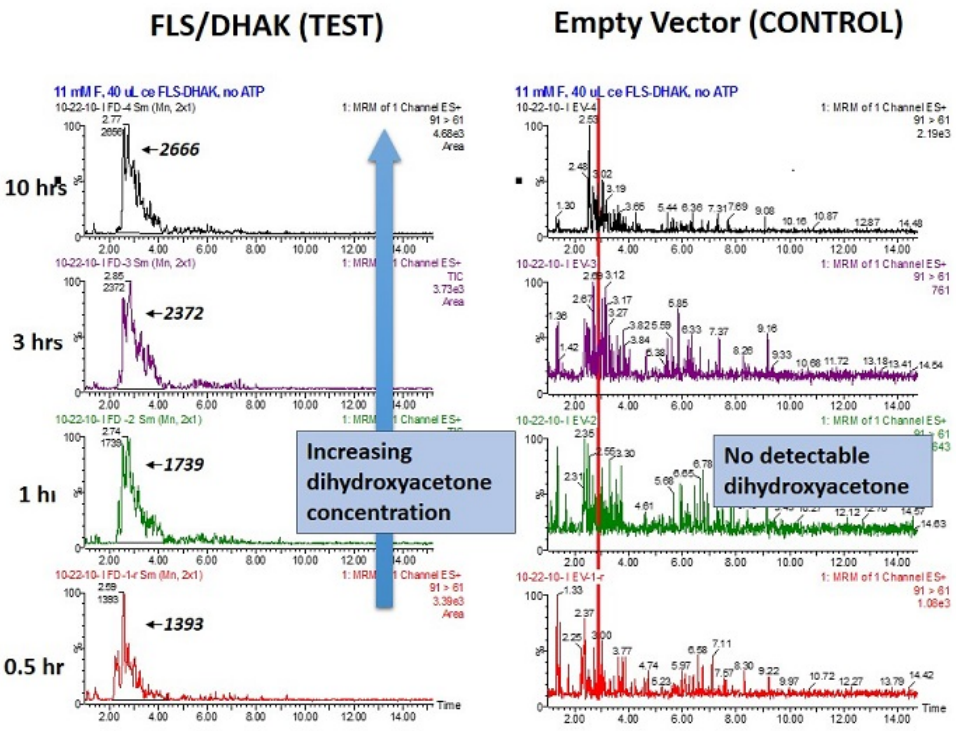


Figure 2.5: Dihydroxyacetone quantification using LC-MS/MS. DHA peak can be seen in the reaction sample taken from the active enzyme, but not from the inactive enzyme.

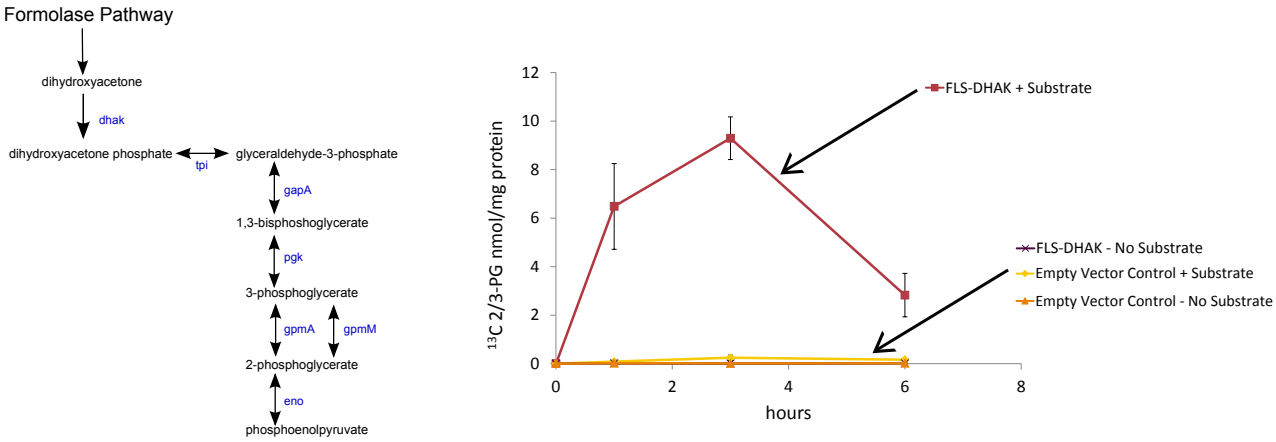


Figure 2.6: Conversion of ^{13}C formaldehyde into glycolytic intermediates. LEFT) Glycolytic metabolism extending from formolase. RIGHT) Accumulation of M+3 labeled 2/3-phosphoglycerate (2/3-PG) from^{13}C formaldehyde.

increases in lag time. Neither combinations with pSB1A3 FLS-DHAK, pSB3K3 ACS-ACDH and pSB4C5 FDH in an MG1655 variant nor pSB1A3 AAFD (ACS, ACDH, FLS, DHAK) with 3K3 $_p$T7 FocA and 4C5 FDH (constitutive promoter) in BL21 AI produced evidence of formate-dependent growth improvement compared to the background strains containing the corresponding empty vectors.

2.4 Formolase pathway function in vitro assessment

Since growth on formate was not detected, a method to confirm that the pathway enzymes could function together was sought. Through enzyme assays with individual pathway sections and the pathway as a whole, in vitro confirmation of the pathway was achieved. The protocol for these assays is shown in Figure 2.7 and detailed in Appendices D and E and consists of four steps: Cell Pellet Preparation, Purified Protein/Cell Extract Preparation, Assay Reaction Setup and Compound Detection. Cell Pellet Preparation consists of the preparation and harvesting of a protein expression culture. The Purified Protein/Cell Extract

Preparation step encompasses the re-suspension of the cell pellets, the lysing of the pellet re-suspensions and purification of the crude lysates. The Assay Reaction setup consists of combining the purified protein or cell extracts with buffer, substrates and cofactors to start the reactions. Time point samples were taken from the assay reactions and analyzed in the Compound Detection step. Compound detection was accomplished using UPLC-MS/MS with an ion pairing method involving tributylamine or with a hydrophilic interaction chromatography (HILIC) method citeBuescher2010. Transitions were setup for pathway intermediates, formyl-CoA, DHAP, 2/3-phosphoglycerate (2/3PG), phosphoenolpyruvate, pyruvate (PYR) and acetyl-CoA. Glycerate and glyoxylate were also monitored for evidence of carbon flux through the alternate GA pathway. Pivalate, norvaline or glutaric acid was added during the sample prep to serve as an internal standard and correct for variations in preparation. Protein concentrations were measured with BCA[122].

2.4.1 Purified Protein

His-tagged proteins expressed from pET29b+ constructs were purified via Cobalt affinity columns. Mixtures of the individually purified proteins combined with the necessary cofactors and substrates converted formate into pathway output DHAP. The DHAP peak was present in the full pathway test sample matching the one in the positive control, but not in any of the negative controls (no formate, no FLS and no ACS or ACDH) [119]. No NADH was added to these reactions; all was generated via the FDH.

2.4.2 Cell Lysate Mixture (Pseudostrain) Data

Since combining purified proteins was successful, another level of complexity was added by including other cellular proteins. This was accomplished by using cell extracts rather than just the tagged proteins from them. Cell extracts from individual protein high expression strains, e.g. BL21 (DE3) pET29b+ FLS (Table 2.1) were mixed together to create "pseu-

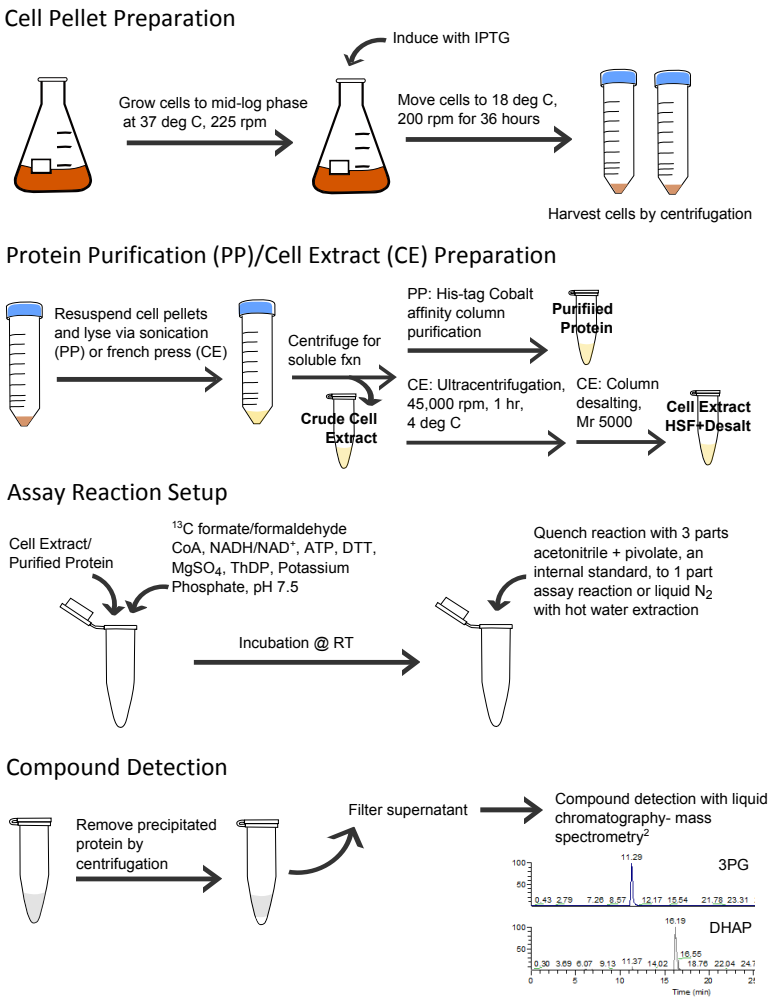


Figure 2.7: Protocol for in vitro pathway assays. These have four main steps and differ slightly depending on whether purified protein or clarified cell lysates are being tested.

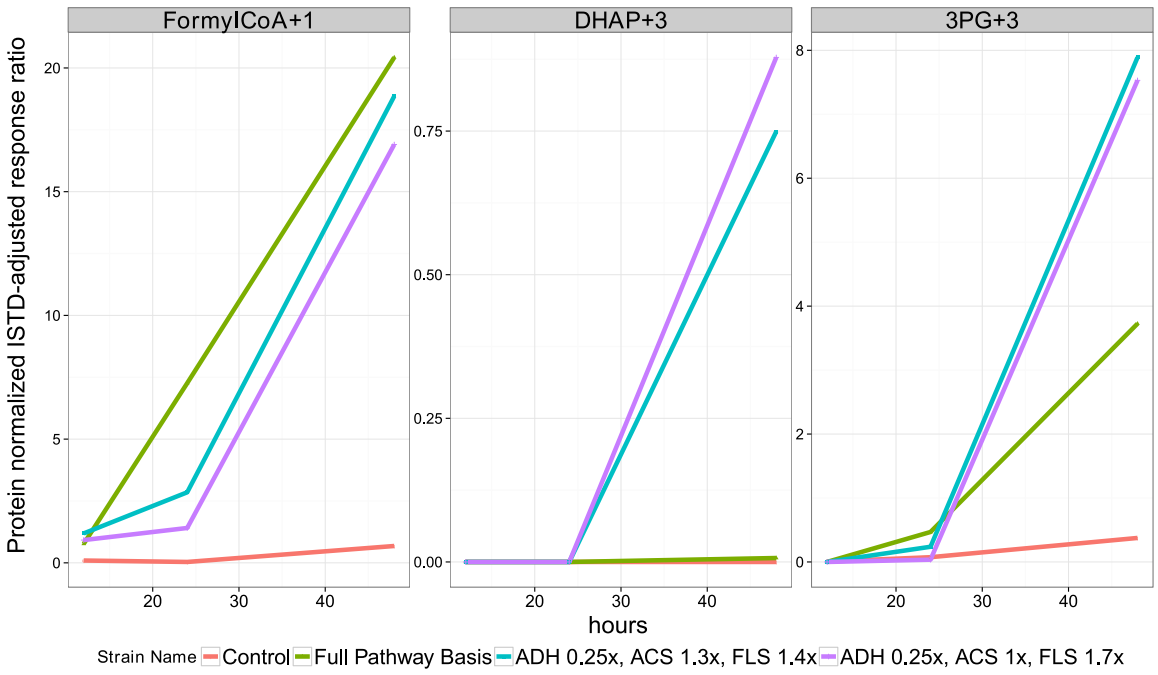


Figure 2.8: A comparison of labeled metabolite levels for one control pseudostrain and three full pathway containing test pseudostrains. The amounts of ACDH, ACS and FLS are given relative to the Full Pathway Basis pseudostrain. FDH and DHAK levels were 1.61% each for all pseudostrains.

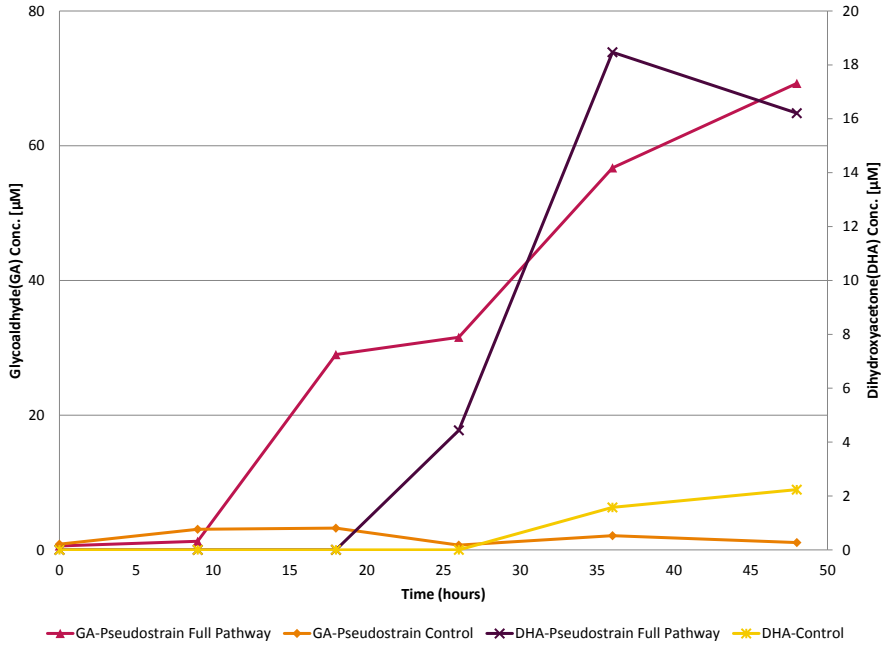


Figure 2.9: Glycolaldehyde and dihydroxyacetone production from the pseudostrains. Measurements done via GC-MS[101].

dostrains." Thus, these mixtures mimicked having a single strain with the entire pathway, but allowed us to trouble-shoot more easily than with strains containing all of the genes together. The crude extracts were combined beginning with a baseline of ratios derived from the purified protein assay and making variations from there to create a number of test pseudostrains. The control is the empty vector crude extract with small percentages of DHAK and FDH corresponding to those in the baseline test strain. The acetonitrile quench followed by ion-pairing LC-MS/MS analysis was used for this set of experiments.

The results (Figure 2.8) show increases in labeled pathway intermediates (FormylCoA+1, DHAP+3, 3PG+3). These amounts increase over time only in the pseudostrains containing the entire pathway (Figure 2.8). The three pseudostrains differ in the proportions of the individual crude extracts. A comparison among the three full pathway pseudostrains shows that an increase in FLS leads to greater conversion into DHAP+3 and more 3PG+3 as a result. GC-MS analysis to measure glycolaldehyde+3 and dihydroxyacetone+3 for pTrcCO_2 3 low FDH (pSB4C5 FDF6) and the pTrc control low FDH (pSB4C5 FDF6) was also completed by Sean Poust in the Keasling group. Increasing trends for both compounds over time were observed only in the full pathway strain (Figure 2.9)[101].

To test whether one of the enzymes might be limiting, pTrc single strains differing only in the expression level of FDH were supplemented with combinations of FDH, ACS and FLS, from the extracts used for the pseudostrain experiments. The strain with the lower FDH levels exhibited the greatest labeled metabolite increases with the addition of FDH alone whereas the higher FDH strain exhibited the most improvement with a mixture of FLS and ACS. However, generally greater increases were seen by the solo FDH addition to the lower FDH expression strain. This suggests that the output is sensitive to FDH level and this should be a focus for tuning expression.

Having the pseudostrain in vitro assay functional allowed us to test variables related to the assay conditions and gene expression levels. FDH dependence was investigated, as

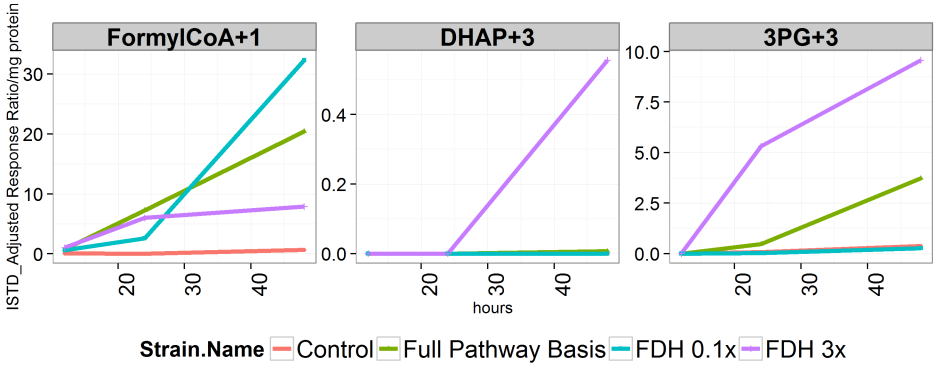


Figure 2.10: Pseudostrain Assay FDH dependence. FDH 0.1x has 0.16% FDH. The Control and Full Pathway basis have 1.61% FDH. FDH 3x has 6.45%.

Bar-Even et al. suggested that the level would be limiting (Figure 2.10). Increasing FDH levels produced more DHAP+3 and 3PG+3 and levels of labeled formyl-CoA+1 decreased as more underwent further conversion.

Originally the assay was run at pH 8 because that was known to be optimal for the FLS. This seemed fine for the other enzymes because four out of five assays for ACDH and ACS used over pH 8 [23, 32, 108]. As seen in Figure 2.11, pH 8 provides superior conversion of ^{13}C formaldehyde as expected, but for ^{13}C formate, the lower pHs provide equal, if better results (Figure 2.11). Thus, reaction pH was lowered to 7.4, more in line with the physiological pH of *E. coli* cytoplasm.

2.5 Higher Expression Constructs

After noting that the large difference in expression between a functional mix of individual clarified cell lysates (CLLs) and the crude cell extract of a full pathway expressing single strain, we investigated a wider array of gene expression options. Gibson cloning was employed by Adam Wargacki in the Baker Lab to create a small library of constructs in the

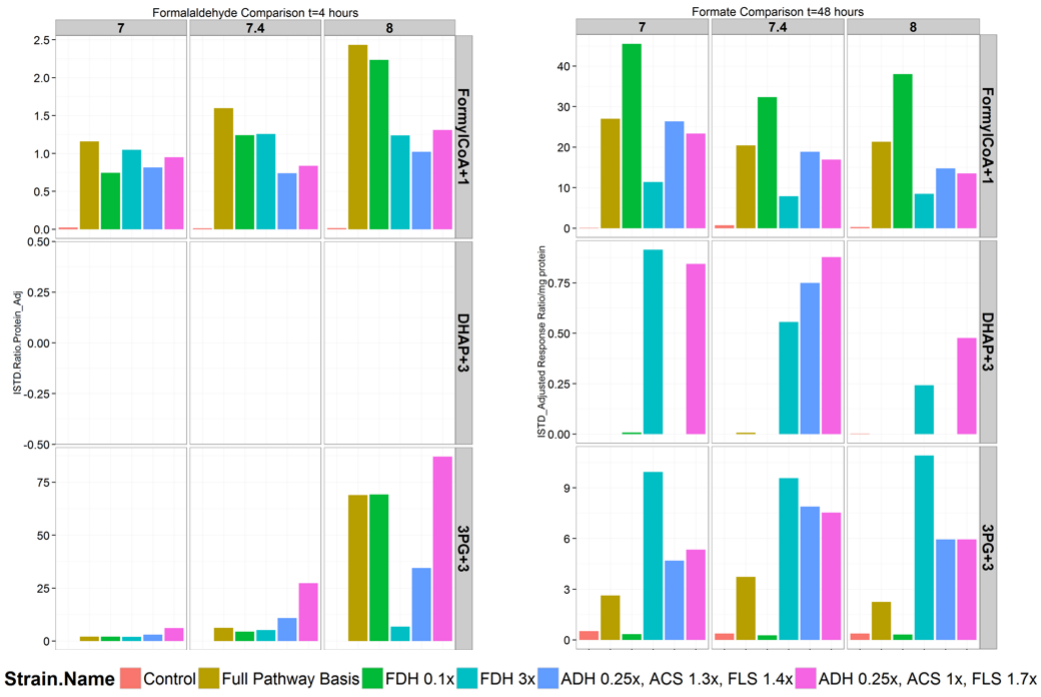


Figure 2.11: pH dependence on labeled metabolite production from ^{13}C formaldehyde and ^{13}C formate. LEFT) Labeled metabolites produced with ^{13}C formaldehyde as a substrate after four hours. RIGHT) Labeled metabolites produced with ^{13}C formate as a substrate after 48 hours. pHs 7, 7.4 and 8 were used in the assay

pTrcHis2 vector backbone (20-40 copies/cell) with the pTrc inducible promoter[120]. Similar constructs were made by Janet Matsen and myself as a T7-controlled combination of Constructs A and B [55, 86, 81]. For the pTrc constructs, ACS, ACDH and FLS were arranged in different orders for the constructs to account for transcriptional attenuation as proximity to the promoter decreases[85]. In these constructs, the ecACDH originally used was replaced with lmACDH. A set of Biobrick vectors with a combination of a formate transporter and a formate dehydrogenase and an analogous set with the addition of DHAK were also made to complement these plasmids (Table 2.2).

Construct	**Description**
pSB1A3 AAFD	pSB1A3 $_p$T7 ACS-ACDH $_p$J23100 FLS-DHAK
pTrc-CO_2-3	pTrc ACS-FLS-ADH
pTrc-CO_2-5	pTrc ADH-ACS-FLS
pTrc-CO_2-6	pTrc ADH-FLS-ACS
pTrc-CO_2-10	pTrc FLS-ADH-ACS
pSB3K3 FF3	pSB3K3 $_p$J23100 FDH $_p$J23114 FocA
pSB3K3 FF4	pSB3K3 $_p$J23100 FDH $_p$J23114 FocB
pSB3K3 FF5	pSB3K3 $_p$J23114 FDH $_p$J23114 FocA
pSB3K3 FF6	pSB3K3 $_p$J23114 FDH $_p$J23114 FocB
pSB4C5 FDFx	Analogs of pSB3K3 FF3-6 with $_p$J23100 DHAK added between the FDH and the FocX in pSB4C5

Table 2.2: Table of pathway higher expression and supporting constructs. pTrc constructs containing ACDH, ACS and FLS were paired with a variety of supporting constructs. All pTrc genes have His-tags. Analogous supporting constructs with FDH and FocA only were also made[120, 55].

2.6 Higher Expression Confirmation

Expression confirmation was carried out by SDS-PAGE. Strong expression bands were observed on SDS-PAGE protein gels for the pSB1A3 T7 ACS-ACH $_p$J23100 FLS-DHAK construct as well as some of the pTrc constructs. Four pTrc-CO_2 plasmids (3, 5, 6, 10) passed initial expression screening. These plasmids differ as to the order of ACS, lmAcDH and FLS

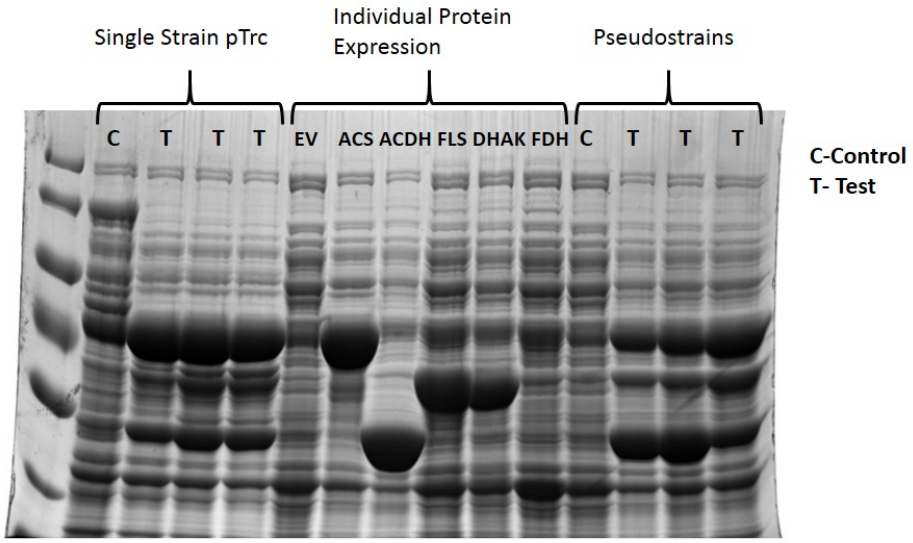


Figure 2.12: SDS-PAGE of individual clarified cell lysates. The first four columns are MG1655 variants with pTrc (control-far left) or CO_2-3 or CO_2-10, which express ACS, ACDH and FLS in different orders. These also contain a secondary plasmid with FDH, DHAK and FocB at lower levels (pSB4C5 FDF4/6). The test single strains from left to right are pTrcCO_2-3 FDF6, pTrcCO_2-10 FDF6 and pTrcCO_2-10 FDF4.The next six columns are individual high expression strain lysates, BL21(DE3)$\star$ pET29b+ empty vector (EV), ACS, lmACDH, FLS, DHAK and cmFDH. The final four are pseudostrain mixtures of the individual high expression lysates. For the pseudostrains, C is a control with FDH and DHAK only, where the Ts are test pseudostrains containing the entire pathway.

(Table 2.2). Further expression testing showed pTrc-CO_2-3 and -10 to be far superior to the other two in band strength of the cloned proteins. In duplicates, pTrc-CO_2-5 only expressed ACS and pTrc-CO_2-6 had faint bands due to far less biomass suggesting a growth defect. Figure 2.12 shows construct expression with pTrcCO_2-3 and pTrcCO_2-10 was compared to negative controls with vector only (ALA2 *fdoG* with pTrc) and to positive controls in which extracts of individually expressed genes on pET29b+ plasmids were mixed (test pseudostrains, varying volumetric ratios of the individual crude extract components). ACS, ACDH and FLS are clearly expressed in both pTrcCO_2-3 and -10. pTrcCO_2-3 was chosen for use in continuing studies.

2.7 Individual gene activity confirmation

2.7.1 Formate Transporter

Sample	Growth Defect with toxic formate analogue? Anaerobic	Aerobic
Expected for FocA/B function	Yes	Yes
Wild Type	Yes	No
Wild Type + Empty Vector	Yes	No
Wild Type + FocA	Yes	Yes
Wild Type + FocB	Yes	No

Table 2.3: Determination of formate transporter function. All cells should have an anerobic growth defect with hypophosphite present, but only have an aerobic growth defect if the expression of the formate transporter results in an active protein.

To ascertain formate transporter function, the assay that Suppmann et al. used to first identify *focA* was employed. Hypophosphite, a toxic analog of formate, causes a growth defect when added to cell cultures only if a formate transporter is active[127]. Normal anaerobic expression of *focA* causes a growth defect with hypophosphite, as should exogeneous expression of a functional formate transporter. The anaerobic growth defect was used a positive

control for this assay. Log-phase, aerobically-grown wild type (WT), *E. coli* MG1655 *fdoG*, were transferred to new media with and without hypophosphite, a formate analogue, in both aerobic and anaerobic environments. *focA* expression gives growth defects in both anaerobic and aerobic conditions while *focB* expression and WT strains do not have an aerobic growth defect. Thus, formate transporter function is only achieved with FocA not FocB and FocA was selected for use in the formolase pathway.

2.7.2 Formate Dehydrogenase

Although the NAD-linked formate dehydrogenase from *C. methylitica* had been expressed in *E. coli*, white precipitate was seen in crude lysates [3]. This precipitate was removed with the remainder of the insoluble fraction and unlysed cells, suggesting solubility issues. Thus, FDH was codon optimized for *E. coli*. Activity of the formate dehydrogenase was assayed by monitoring NADH production as described in Appendix D. The commercially available enzyme appears to be more active than any of the extracts prepped, but the FDH is only a fraction of the total protein amount for the lysate where the commercial preparation is nearly pure(Figure 2.13).

2.7.3 ACS

ecACS is subject to post-transcriptional modification[124]. Changing residue leucine-641 to a proline immunizes the ACS to inactivation by an acetyltransferase. Kinetic studies comparing the L641P mutation to the wild type in *Salmonella* find that the mutant has a Km ten times higher and a turnover rate 60 time lower than the wild type enzyme, however, Janet Matsen consistently measures high activity for the mutant enzyme[81, 27]. The L641P ACS mutant version of pTrcCO_2-3 was made with site-directed mutagenesis and tested with the in vitro assay protocol alongside an unmutated pTrcCO_2-3 (Figure 2.14). The wild type ecACS produced higher fraction of pool labeling and was used for further experiments.

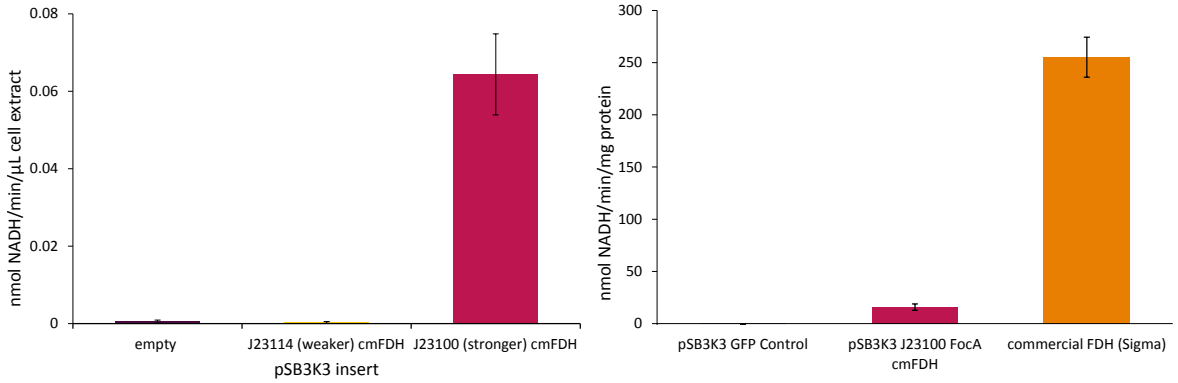


Figure 2.13: cmFDH activity measured via monitoring of A340nm for NADH consumption. LEFT) pSB3K3 - empty (Control), $_p$J23114 cmFDH (Weaker Promoter) and$_p$ J23100 (Stronger promoter) RIGHT) pSB3K3 - GFP (Control) and $_p$J23100 FocA cmFDH (Expressed FDH) compared to FDH from Sigma.

2.8 Single Strain in vitro CCL Assay

Once the assay was functional with the pseudostrains, we repeated the ^{13}C labeling experiment with all pathway components expressed in a single strain. These experiments were done with both pathway construct pathway core - pSB1A3 AAFD (ACS, ACDH, FLS, DHAK) and pTrcCO_2-3 (ACS, ACDH, FLS). The negative controls for these experiments were corresponding strains with empty vectors replacing the pathway core plasmids. Thus, output differences are related to full pathway expression. The best control would actually be the exact core pathway plasmid with an inactive form of FLS. Since pathway activity was not detected in absence of ACS, ACDH or FLS in the purified protein assay, the empty vector is still a reasonable control for this type of assay. Every strain also contained a secondary plasmid, either pSB3K3 FFx (FDH and FocA/B) for the pSB1A3 AAFD pathway core or pSB3K3 $_p$J23100 DHAK for the pTrc-CO_2-3 pathway core. A commercial FDH was added to all strains in the pTrc tests in an amount that approximately balanced the NADH consump-

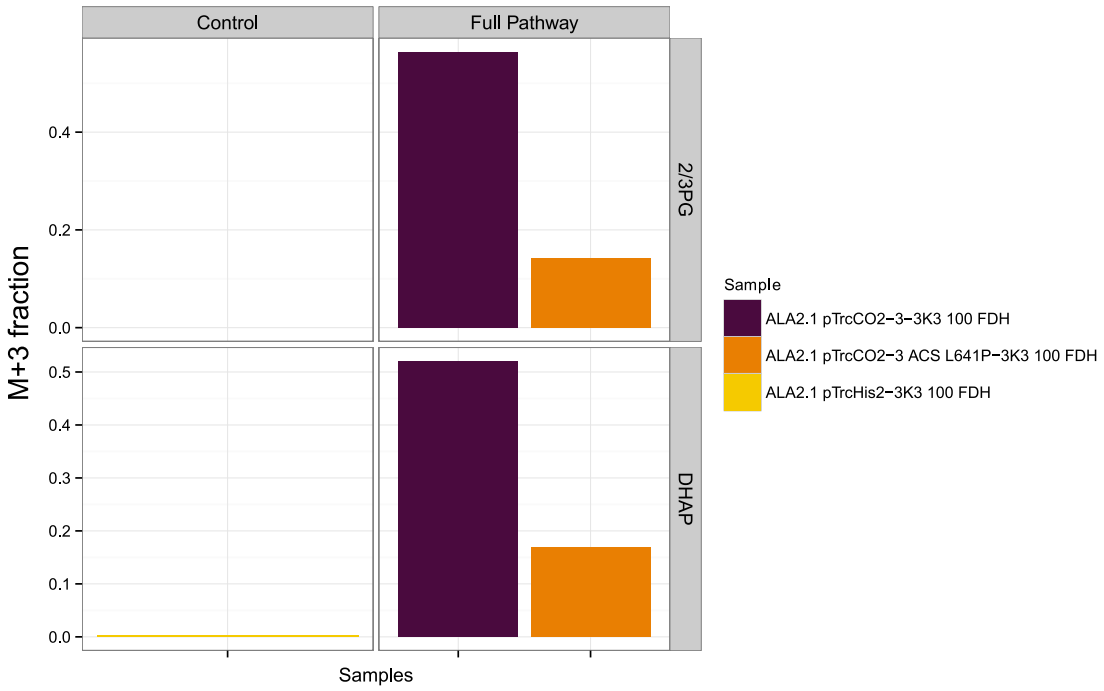


Figure 2.14: Comparison of ACSL641P activity to ACS WT. M+3 2/3-PG production by pTrcHis2, pTrcCO_2-3 with WT ACS and pTrcCO_2 with L641P ACS. All strains contained pSB3K3 $_p$J23100 cmFDH.

tion rate of the test extracts one-to-one. Labeled metabolites from ^{13}C formate were not detected in strains expressing the pathway core on pSB1A3 AAFD, but strains expressing the pathway core on pTrcCO_2-3 produced labeled quantities of metabolites that increased throughout a 24 hour incubation compared to the respective negative controls (Figure 2.15). Based on this data, a doubling time for growth on formate of the test strain was estimated at over three weeks.

The last step to achieving in vitro function with a single strain was to replace the commercial FDH addition with plasmid expression. DHAK expression was also removed as purified protein and in vitro titration studies showed that decreases in its amount do not have a large effect on pathway output and formaldehyde conversion studies in cell lysates have enough native kinase activity to still produce DHAP and other glycolytic intermediates downstream of it. The in vitro experiment was performed on this strain and pathway function remained confirming single strain in vitro formolase pathway function(2.16).

2.9 Formolase Pathway Function in vivo Assessment

With in vitro confirmation of formolase pathway function achieved, proof of in vivo function was sought. Four different methods were used to evaluate pathway activity: in vivo metabolite measurement, in vivo protein incorporation measurement, pathway-dependent biosensor output and growth (2.17). Despite the high doubling time estimate, growth was tested first followed by the others, which should be more sensitive means of in vivo pathway flux detection.

2.9.1 Growth on Formate

Growth experiments were completed using a set of four media with all combinations of $-/+$ formate (normally 40 mM) and $-/+$ glycerol (0.1%) in M9[79]. Temperatures tested were room temperature ($\sim$22 °C) and 18 °C. Cultures (0.210 mL) were grown in Bioscreen

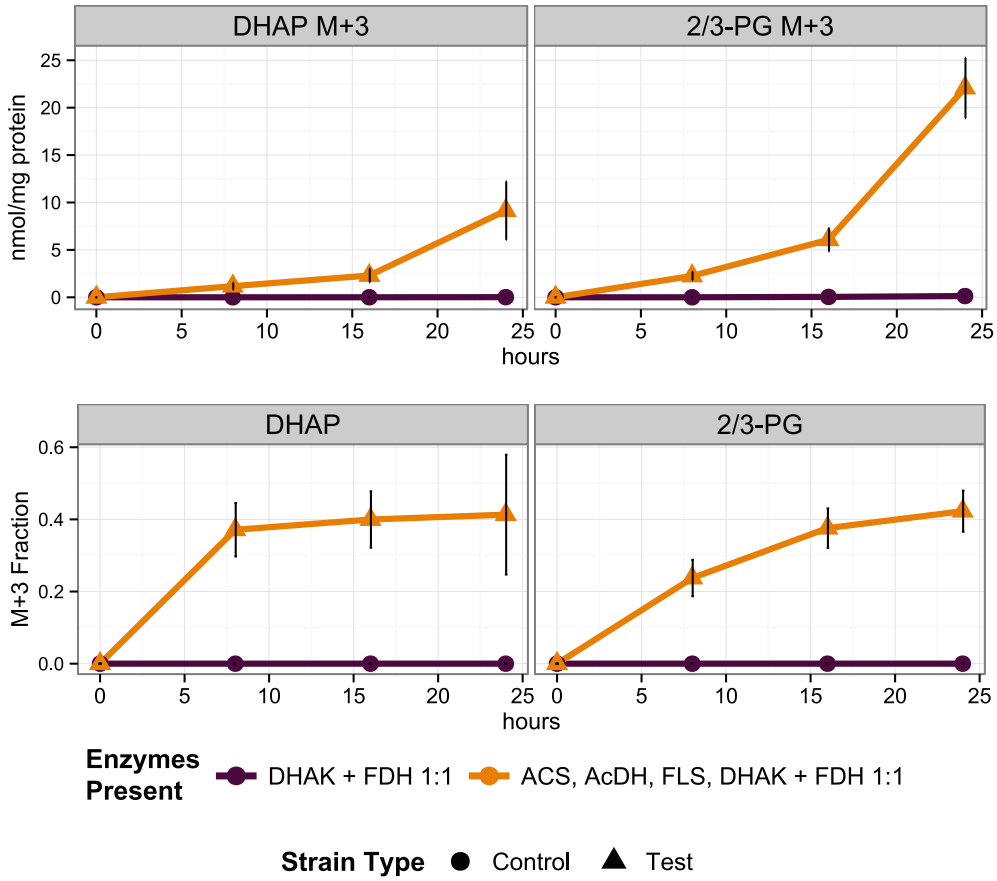


Figure 2.15: in vitro metabolite production from ^{13}C formate. TOP) ^{13}C-labeled (M+3) dihydroxyacetone phosphate (DHAP) and 2/3-phosphoglycerate (2/3-PG) accumulation for clarified cell lysate assays with ^{13}C labeled formate and required co-substrates. Commercial FDH was added to balance NADH oxidation in the test extracts 1:1. No formate controls do not produce labeled intermediates. BOTTOM) ^{13}C-labeled (M+3) DHA and 2/3-PG) accumulation as a fraction of the total pool for each compound for clarified cell extract assays in A. Isotope correction for natural abundance was done with Isocor software[87]. Error bars represent one standard error, n=3.

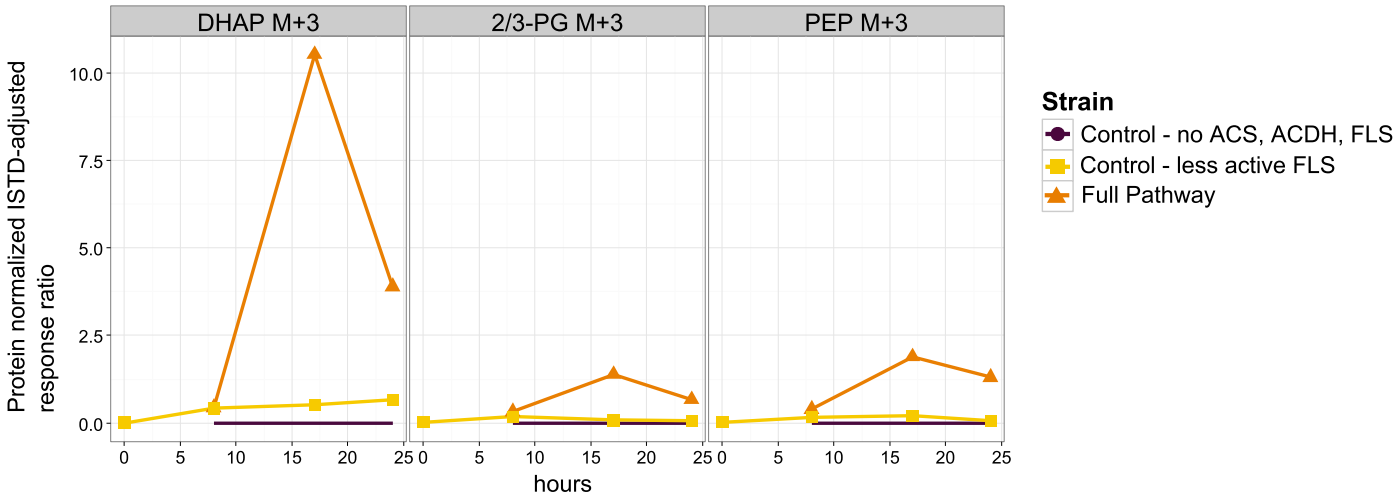


Figure 2.16: in vitro metabolite labeling from ^{13}C formate with FDH, ACS, ACDH and FLS expressed (pTrcCO_2) compared to an pTrc empty vector control (pTrcHis2C) and a mutated FLS control (pTrcCO_2-3 XFLS). All three strains also contain pSB3K3 $_p$J23100 FDH $_p$J23100 FocA.

plates in the Bioscreen automated growth curve machine (Growth Curves Ltd.) . In both cases, cultures were kept oxygenated using the high shaking setting. Growth experiments were carried out as long as 30 days. The cultures tested were expanded sets from the single cell in vitro assay. The pSB1A3 AAFD (ACS, ACDH, FLS, DHAK) pathway core and associated control with all combinations of secondary plasmids pSB3K3 FF3:6 (FDH, FocA/B) and pTrcCO_2-3 and 10 (ACS, ACDH, FLS) pathway cores and associated controls with combination of secondary plasmids pSB4C5 FDF4/6 (FDH, DHAK, FocA/B) were used. Given the four medias described, we looked for improvement in OD during mixed glycerol/formate growth or growth solely on formate as a measure of pathway function. None of the full pathway test strains showed growth beyond its respective control.

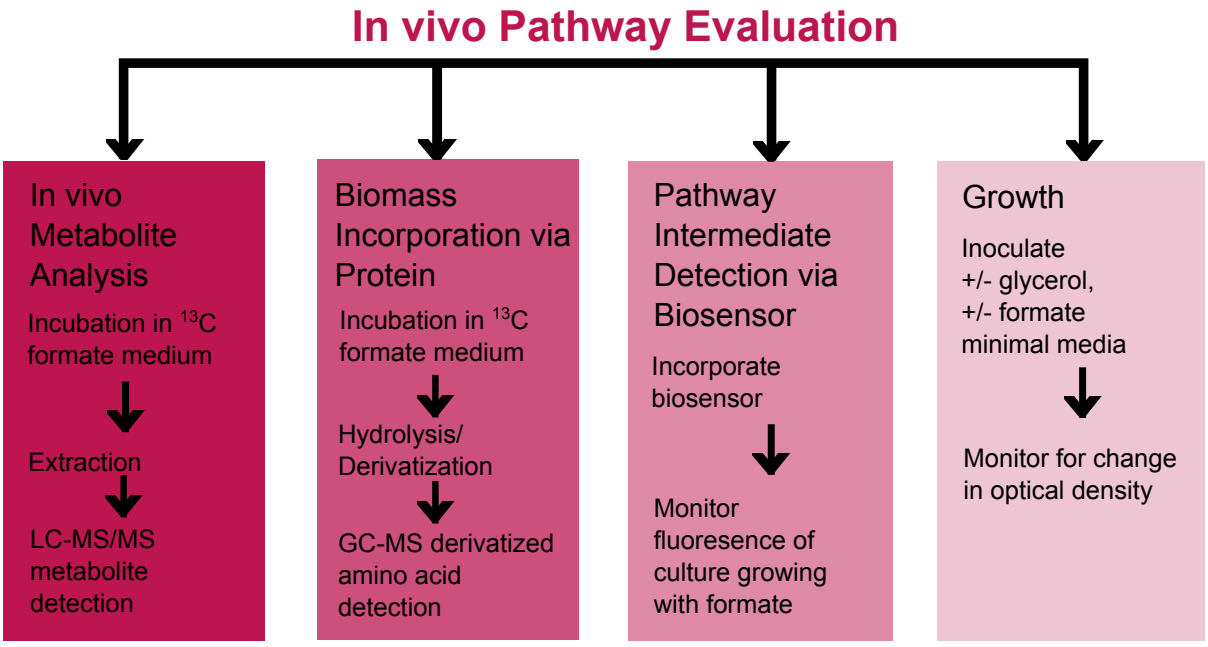


Figure 2.17: Overview of in vivo pathway function analysis methods.

2.9.2 Hydrolyzed Protein Label Incorporation

Since the pathway failed to produce enough biomass to be detected as growth, an alternative approaches were used to test pathway function at lower flux. Next to test ^{13}C formate incorporation into biomass, we investigated the label specifically in protein since it makes up 55% of total dry cell weight[93]. In multiple cases, this experiment was carried out in parallel with the in vitro CCL assay to provide both data sets for the same starting cells. Protein expression was carried out in the same manner as for the in vitro assays described above, and upon completion of the expression period, cells were washed and transferred to 40 mM ^{13}C formate-containing M9 medium. Glycerol was tried as a co-substrate, but did not have an effect. To detect incorporation in protein, the biomass was hydrolyzed to isolate individual amino acids, which were then derivatized with tert-butyldimethylsilyl (TBDMS)[65, 7]. Derivatized samples were analyzed with GC-MS. Data was tabulated for four amino acids serine, alanine, glutamic acid and aspartic acid, and the isotopomer correction was done with IsoCor[87]. These amino acids were chosen because they represent different branch points off central metabolism. Serine, alanine, aspartic acid and glutamic acids are derived from 2/3-PG, pyruvate, oxaloacetate and α-ketoglutarate, respectively. In addition, glutamic acid has an especially large pool because it serves as a precursor of many of the other amino acids[16].

2.9.3 in vivo Metabolite Analysis

Since metabolite pools can turnover quickly, looking for changes in the isotopomer distribution of glycolytic metabolite pools should provide an even more sensitive measure of pathway function. Cells expressing the full formolase pathway were fed ^{13}C labeled formate. Samples were taken from up to 200 hours after inoculation. No M+3 labeling was detected in three independent experiments with different sample prep and mass spectrometry methods (Appendix E). The hot water extraction method was validated using ^{13}C glycerol as a positive

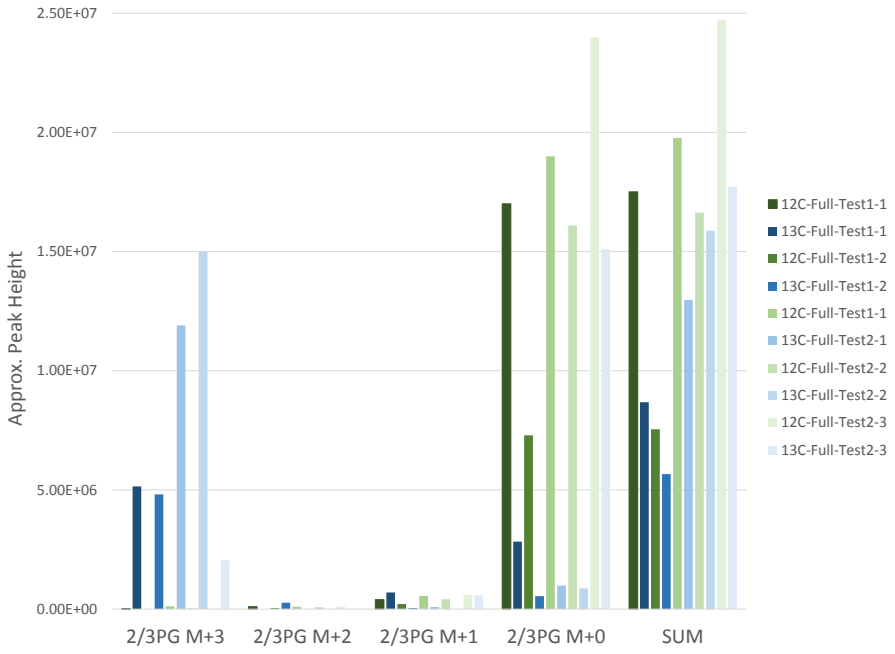


Figure 2.18: in vivo metabolite data for formolase pathway with ^{13}C glycerol as a positive control substrate to confirm the hot water extraction protocol and HILIC detection.

control (Figure 2.18).

2.9.4 Biosensors

Biosensors, which convert the presence of a molecule in the cell to an output that can be non-invasively/indirectly measured, have a variety of uses including straight detection detailed here as well as dynamic regulation described in the next chapter. For the formolase pathway, biosensors could be employed to detect the presence of pathway intermediates and provide an analog output as to the amount of the targeted intermediate. Two useful target compounds

exist, formaldehyde and DHA. Both have natural regulation in *E. coli* that can be co-opted to control their corresponding promoters. Preliminary work on these plasmid-based sensors has been accomplished, but the formaldehyde sensor requires redesign followed by validation and the DHA sensor requires validation. A DHA-biosensor strain was acquired from Bächler et al.[9].

Formaldehyde Biosensor

Regulation by transcriptional activators has been accomplished by building operator sites into hybrid promoters, however, a transcriptional activator has not been identified for formaldehyde. The *frmRAB* operon (formaldehyde detoxification) is induced by formaldehyde and repressed by FrmR[51, 48]. The *frmR* knockout has high levels of frmAB transcript produced, although the binding of FrmR to the formaldehyde promoter has not been confirmed[51]. Previously, the putative promoter region for *frmRAB* was fused with RFP and GFP. Both -187 and -321 regions upstream of *frmR* were tested. However, either did not show sensitivity in the desired range of externally supplied formaldehyde (Figure 2.19. This could be due to the fact that without extra copies of FrmR to repress the operon under normal conditions, the reporter gene is no longer controlled by formaldehyde and is always on. These biosensors should be reworked with a copy of *frmR* in the transcript. An iGEM team from Japan created a formaldehyde sensor with *frmR* included[131]. However, they failed to include a RBS for their reported gene and were unable to measure fluorescence[130].

Dihydroxyacetone Biosensor

A DHA-responsive regulatory system has been described and a DHA-biosensor CBZ strain was obtained from Bächler et al. [9]. This strain(MC4100 *attB::*P_{DhaK} *E. coli-lacZ-bla*) responds to DHA as reported (Figure 2.20) [9]. The formolase pathway was expressed in the CBZ and DHA-induced LacZ activity was sought via M9 formate X-gal plates as well as

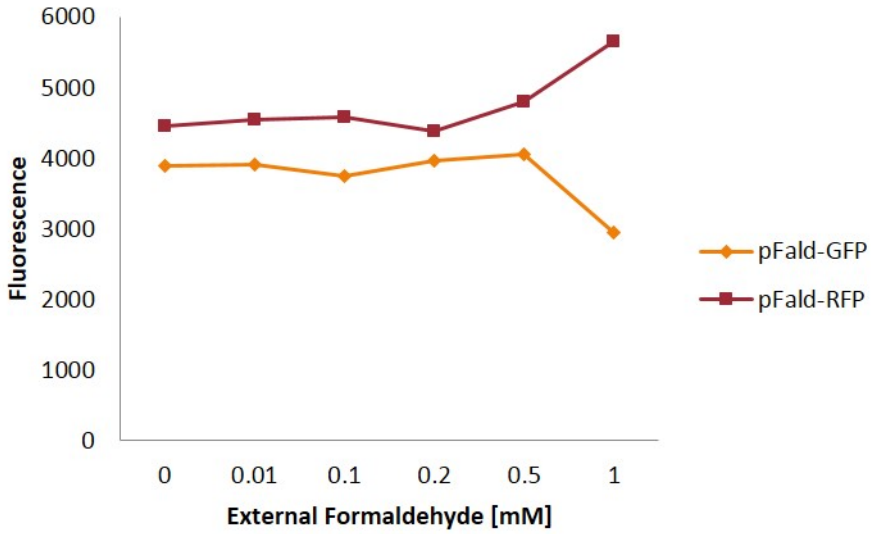


Figure 2.19: Fluorescence data from the promoter fusion of the *frmRAB* promoter with GFP and RFP.

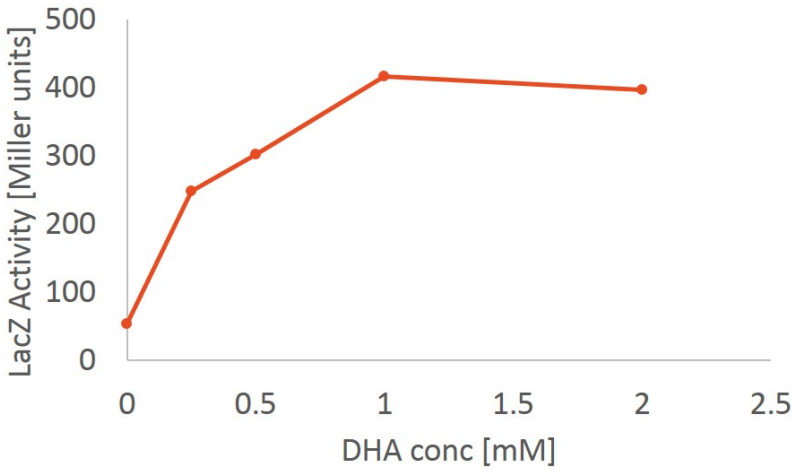


Figure 2.20: CBZ strain response to exogenous DHA measured by LacZ activity.

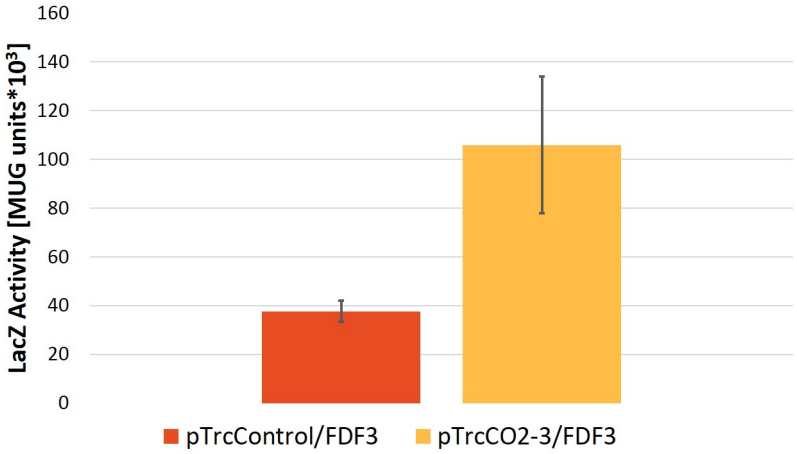


Figure 2.21: DHA-induced LacZ activity in liquid culture by the MUG assay for a CBZ full pathway strain and a control.

M9 formate liquid media with the MUG assay, a fluorescence-based lacZ assay (Appendix E). The liquid media MUG assay results highlight a significant difference in LacZ response between the full pathway and control strains(Figure 2.21).

Plates results also showed a qualitative difference with and without the pathway, however, this difference was not formate dependent (Figure 2.22). Also, the blue color that indicates lacZ activity begins in the anaerobic center of the colony and spreads, suggesting that DHA production is related to anaerobic conditions (Figure 2.22). One hypothesis that would explain this is the fact that *focA* is naturally expressed and formate is naturally produced under mixed acid fermentation in *E. coli*. Perhaps, no formate beyond what is produced in the cells is available to the pathway, suggesting an issue with transporter function.

While the CBZ strain presents an opportunity to test the responsiveness of the system to DHA, having the biosensor on a plasmid is useful for testing with a variety of background strains. While the the preliminary tests completed involving a promoter fusion of GFP/RFP with the *dhaKLM* operon promoter on a plasmid did not show a clear trend with respect

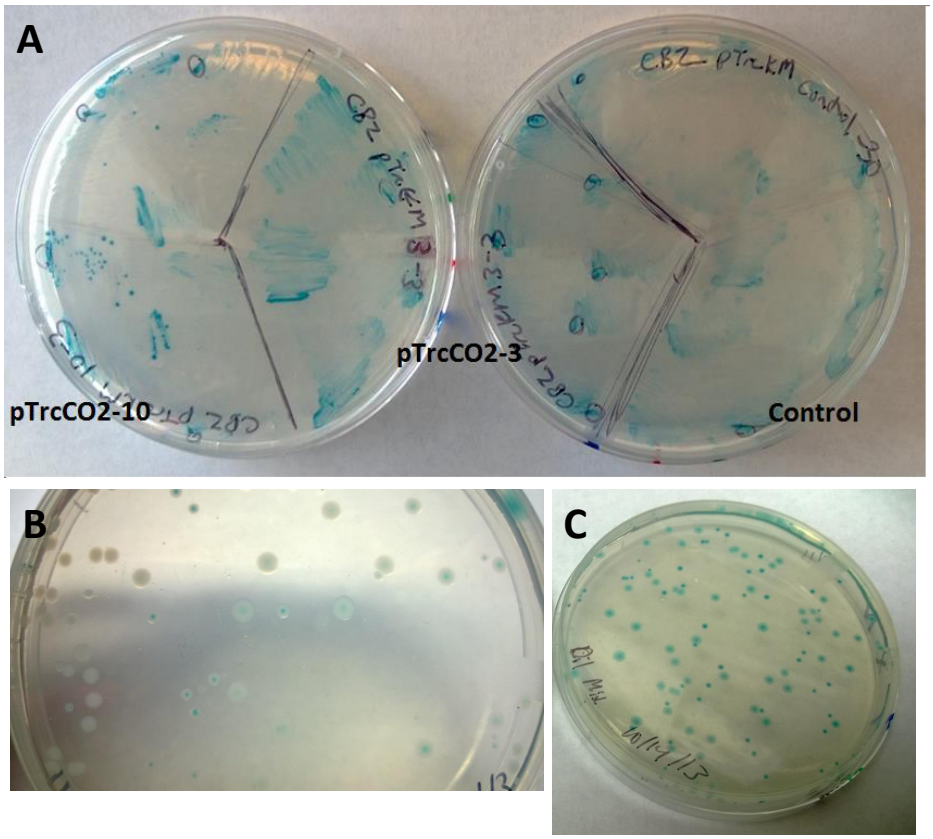


Figure 2.22: DHA-induced LacZ activity on solid media by X-gal hydrolysis. A) Streaks of CBZ strains control (pTrcKM) and tests (pTrcKMCO$_2$-3 and pTrcKMCO$_2$-10) all with pSB4C5 FDF3. pTrcKM is pTrc with a Km cassette substituted for the Amp cassette. B&C) A mix of control and test CBZ cells was plated on M9 with formate and X-gal and incubated at room temperature. The photos in B and C represent 7 days and 11 days after inoculation.

to DHA concentration, however, this seems to be due to growth inhibition by DHA skewing the OD normalized fluorescence. The range of DHA concentrations must be adjusted to determine whether a sensitive range exists. Comparison of the sensor with and without the pathway grown on different carbon substrates will confirm function. If the sensor range is appropriate, it can be coupled with a fluorescent reporter for flow-cytometry-based detection or with an enzyme that can produce a colorimetric product e.g. LacZ.

2.10 Summary

Individual enzyme activity tests confirmed function for FDH, FocA, ACS-ACDH and FLS. The corresponding constructs were combined in evolved strains with a native formate sink removed, then tested for pathway function in vitro and in vivo. in vitro assays confirmed that the pathway is function in clarified cell extracts. A battery of in vivo tests was completed, but the only signs of function are a pathway-dependent, not external formate dependent, difference in Lac-Z activity tied to the P_{DhaK} promoter, and a tiny fraction of in vivo metabolite labeling in a strain with FDH-generation from a co-substrate, mannitol, instead of a NAD-linked FDH.

Chapter 3

IMPROVING THE FORMOLASE PATHWAY: CURRENT AND FUTURE WORK

3.1 Overview

As outlined in the previous sections, the formolase pathway has a large number of issues, including multiple enzymes with poor performance, intermediate toxicity and competition for intermediates. This chapter discusses approaches and results centered around addressing these issues and improving pathway function.

3.2 Challenges to Formolase Pathway Implementation

The problems with the formolase pathway stem from characteristics of the individual enzymes as well as the overall pathway architecture.

1. Low efficiency of enzymes: ACDH and FLS efficiencies as employed are over five orders of magnitude below median enzyme efficiency[13]. Considering the partition between formate dehydrogenase (an exogeneous enzyme, but one that is acting on its native substrate) and the branch controlled by the low efficiencies of ACDH and FLS, the formate may be mostly funneled to CO_2/NADH production leaving little for assimilation. The low efficiencies also limit pathway flux. Estimates of doubling time based on in vitro metabolite measurement are greater than three weeks, making growth detection difficult.

2. Competing pathways/siphoning of intermediates: This issue could cover multiple impacts. For carbon activation, ACDH and ACS are native C2 enzymes. Therefore, they

have much higher activities for acetate and acetyl-CoA than the C1 analogs for which we have re-purposed them. As one of the central metabolites that results from the formolase pathway is acetyl-CoA, the over-expression of these genes may be altering its pool size. Kumari et al. found that acetyl-CoA pools decrease with the induction of *acs* as it likely competes for CoA, acetate and ATP[70]. ACDH could also be converting acetyl-CoA to acetaldehyde, a toxic product, while siphoning acetyl-CoA away from energy and biosynthetic compounds[75]. The formaldehyde produced from ACS and ACDH may not be going to FLS. Glutathione reacts spontaneously to form an adduct with formaldehyde and the product is converted back to formate by native *E. coli* formaldehyde detoxification. Thus, even the small amount of formaldehyde produced is not necessarily available to FLS. In addition, even when DHAP can be made, the methylglyoxal synthase may divert some to toxic methylglyoxal, but these would be small amounts under native MgsA levels[132].

3. Cellular burden: There are numerous reports of longer, more complex pathways expressed in *E. coli*, but the cellular burden of creating large amounts of these enzymes cannot be ignored. By having a growth phase in a richer media, we tried to remove the cellular burden of trying to produce the formolase pathway enzymes with limited carbon flux via the formolase pathway, at least until stationary phase cells expressing the pathway could successfully convert formate.

4. Toxic intermediates: The formolase pathway employs formaldehyde, a toxic compound, as an intermediate. Methylotrophic bacteria naturally have high flux through formaldehyde, so it is possible for microorganisms to deal with such an intermediate[30]. However, *E. coli* has not evolved to do so. Acetaldehyde could also be produced as a consequence of ACDH over-expression. Other microorganisms that have toxic pathway intermediates deal with them by sequestering the pathway inside of a microcompart-

ment preventing the toxic intermediate from interfering with the rest of the cell[28]. A few different microcompartment systems have been discovered and expressed in *E. coli* [109, 128, 100]. Localization of proteins into these microcompartments has also been described[40]. The formolase pathway bears many similarities to the 1,2-propanediol utilization system in the Pdu microcompartment, which encompasses many of the three carbon analogs of its intermediates[34]. Although not discussed further here, encapsulating the formolase pathway in such a way could be beneficial. Other co-localization techniques, including protein scaffolding and fusion proteins, could also serve as ways to improve substrate channeling and reduce toxic intermediate effects without a microcompartment[37, 72, 144].

5. Low flux inducing stress response in *E. coli*: As a fast-growing bacterium (t_d =∼20 min on glucose), *E. coli* induces a variety of stress responses at low growth rates[80].

3.3 Competing Pathway Removal

A number of potentially competing reactions were chosen to generate knockout mutants. The FDH-O, product of gene *fdoG*, consumes formate under aerobic conditions and was removed as a competing reaction. This *fdoG* knockout strain had greater metabolite labeling from ^{13}C formate than the wild type strain. Since formaldehyde is one of the intermediates of the formolase pathway, *E. coli* formaldehyde detoxification was also addressed. Formaldehyde is a toxic compound that causes DNA and protein damage by generating cross-links in addition to creating single strand breaks in DNA and preventing their repair[78, 45, 50]. Formaldehyde spontaneously reacts with glutathione to form an adduct. FrmA, a class III alcohol dehydrogenase, converts this adduct to S-formylglutathione, which undergoes another reaction with FrmB to make formate[48, 53]. The *frmRAB* operon is induced by formaldehyde[51, 47]. Since *frmA* also produces an enzyme that catalyzes a reaction as part of the formaldehyde detoxification pathway, it was also removed. As the first step of

this pathway is a spontaneous reaction with glutathione, *gshA* was knocked out to decrease glutathione levels to address the possibility that the free formaldehyde concentration would increase in turn. With the exception of the *gshA* knockout, gene disruptions, *fdoG* and *frmA*, were moved from existing Keio collection knockout strains to the serial passaged and cured *E. coli* via P1 phage transduction[129, 8]. The *gshA* strain was created using λ red recombination[35]. Enzyme assays confirmed *fdoG* and *frmA* knockouts. The *gshA* knockout was validated by sequencing colony PCR products of the KO region.

3.3.1 in vitro Evaluation of Knockouts

Figure 3.3.1 displays the effect of the knockouts in competing pathways. The data for *fdoG* knockout was collected separately but the isotopologue distribution for the desalted fraction is consistent with a previous experiment incorporating more replicates. The most important changes are those in the soluble fraction because this should be the closest to the actual intracellular environment. Overall, both the single and triple knockouts performed similarly for DSF, but the triple knockout was better for the SF/HSF and will be used henceforth.

Although the *fdoG/frmA* and *fdoG/frmA/gshA* knockout samples labeled a greater portion of their 2/3-PG pools than just the *fdoG* knockout alone, the background for the *fdoG/frmA/gshA* knockout is higher across all the levels of lysate clarification. This result is interesting because Müller et al. reported that the *frmA* knockout reduced their pool labeling[90]. This is exactly what we see here for the -FLS case. Since the *gshA* knockout restores function for the less active formolase, perhaps this is a free formaldehyde availability issue. Without glutathione to form an adduct with, the formaldehyde will be free and available for FLS. However, removing *frmA* should have a similar response by Le Chatelier's principle, as the accumulation of the adduct, S-hydroxymethylglutathione, should leave more free formaldehyde. If this is the blocking issue, then the *frmA* knockout alone should not have a drastically different result than both of the other cases. One hypotheses that could explain

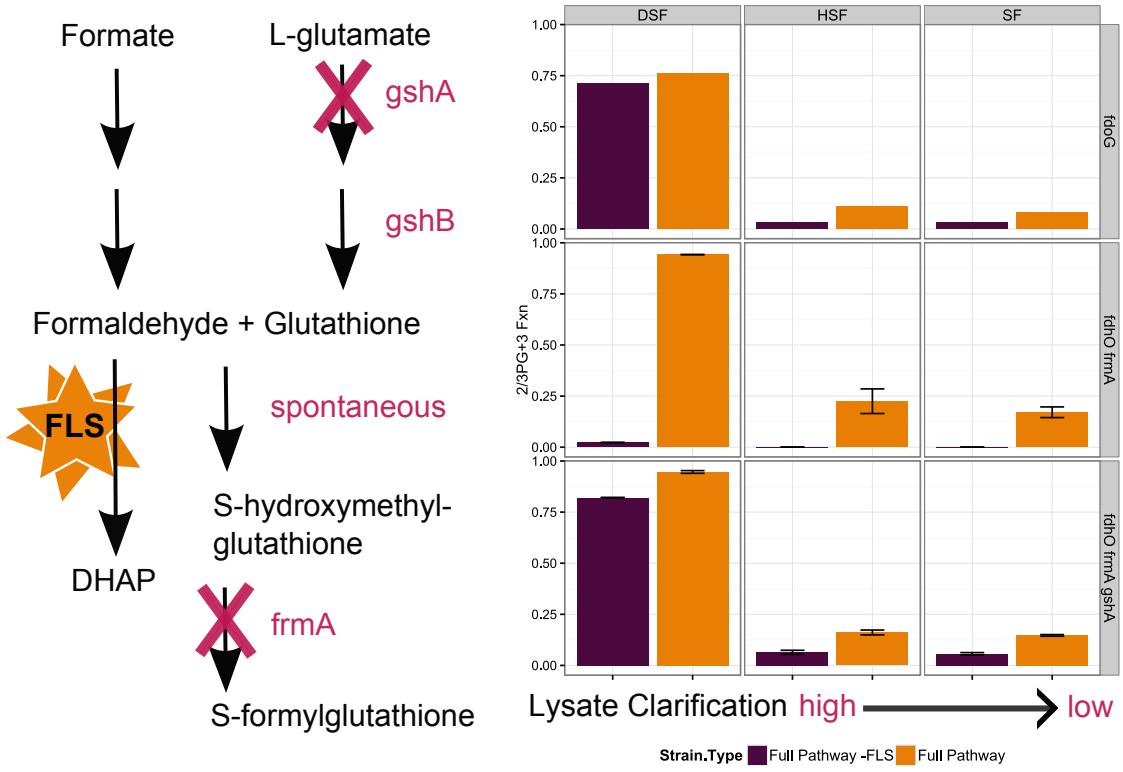


Figure 3.1: Effect on metabolite production from formate by *frmA* and *gshA* knockouts. LEFT) Formaldehyde detoxification pathway in *E. coli*. RIGHT) Three background strains with different combination of knockouts (*fdoG*, *fdoG frmA*, *fdoG frmA gshA*) were used to measure 2/3-PG M+3 pool labeling for in vitro assays with ^{13}C-labeled formaldehyde as a substrate. Full Pathway represents the entire pathway present with a fully active FLS, Full Pathway -FLS represents the entire pathway present with a reduced activity FLS. The horizontal facet represents levels of purification from most (left) to least (right): desalted fraction (DSF), high speed/ultracentrifugation fraction (HSF) and souble fraction (SF). Error bars represent standard error, *fdoG* samples n=1, others n=2. The data for the *fdoG* only knockout, was collected in a separate experiment from the others at a slightly different time point.

this set of results is an accumulation of S-hydroxymethylglutathione. For the *fdoG* knock-out, the formaldehyde detoxification system uses FrmA to convert it to S-/formylglutathione, so no build-up occurs, allowing pathway function. For the *fdoG/frmA/gshA* case, no glutathione is produced. Consequently, the formaldehyde adduct does not form and collect. Again, the formolase pathway is able to function. S-hydroxymethylglutathione is known to be reactive; Lu et al. discovered its involvement in the creation of DNA-glutathione adducts in the presence of formaldehyde[77]. It is not a far stretch to assume that it may be involved in other reactions that would prevent pathway function. The more active FLS must be able to consume enough formaldehyde to keep the adduct concentration below a threshold.

One other notable result, the large change in the 2/3-PG fraction-labeled between the DSF and the HSF, provides insight into why the in vivo tests did not function. The SF is most similar to the cellular environs, so we would expect that in vivo metabolite labeling would be similar to the in vitro labeling of the SF. We should try to increase the signal in the in vitro SF assay, which outputs in a detectable range, in order to have a positive result in vivo. Backing out the levels of lysate clarification here, the gain in signal results from desalting (Figure 3.3.1). Since desalting removes small molecules, we hypothesized that drop-off might be due to the loss of a particular small molecule during this process. Whether that small molecule is a competitive inhibitor, a non-competitive inhibitor or just a part of a competing side reaction that siphons intermediates, etc. is unknown. The desalting flow-through could be collected then fractionated. Adding different fractions to in vitro assay reactions and looking for a loss or decrease in pool labeling would help narrow down which fractions caused inhibition. These fractions can be analyzed by untargeted mass spectrometry to identify candidates for further testing.

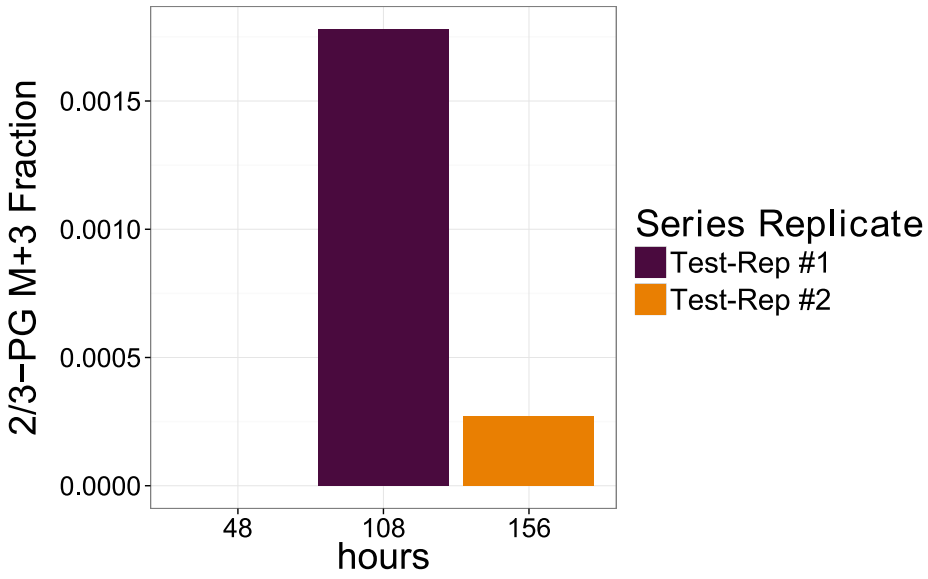


Figure 3.2: Preliminary in vivo metabolite data for formolase pathway without FDH. Strains expressing the full pathway with the exception of the FDH were fed labeled ^{13}C formate with co-substrate mannitol. Metabolite pools were measured after an incubation period. Both replicates were sampled at 48 hours and then sampled again at 108 and 156 hours, respectively.

3.3.2 in vivo evaluation of knockouts

As improvements were seen in with less processing the extracts for the *fdoG, frmA, gshA* knockout strain, in vivo metabolite analysis and growth studies were completed with it. No growth attributable to the formolase pathway was measured and no labeled metabolites were produced over 120 hours from labeled formate. One hypothesis as to why this fails could be the complete oxidation of all the formate to carbon dioxide by the FDH because of the discrepancy in enzyme activity between ACS, which relies on promiscuous activity and FDH, which while exogenous functions on its native substrate. Thus, the formate dehydrogenase was removed from the pathway. Instead, reducing equivalents would be provided by adding a small about of a reduced substrate, e.g. mannitol, preventing quick oxidation of all of the labeled substrate. This approach produced M+3 labeling in the hundredths and thousandths of a fraction for 2/3-PG in four out of ten samples taken from three biological replicates over a period of over 200 hours (Figure 3.2). One replicate without mannitol had tiny amounts of 2/3PG M+3 at the 1 and 48 hour timepoints, but not at 20 hours and 204 hours. The other two samples with labeling were both taken over 100 hours.

3.4 Dynamic Regulation to Prevent Toxic Intermediate Accumulation and Decrease Cellular Burden

Protein levels can be programmed statically using promoters, ribosome binding sites and other DNA sequence variables, however, with the use of regulatory elements such as transcriptional factors and aptazymes, we have the ability to add dynamic regulation to synthetic pathways and allow them to balance themselves. Zhang et al. designed a dynamic sensor-regulator system to control fatty acid ethyl ester production, which functions to repress the downstream sections of the pathway until the upstream section has produced enough of an intermediate to turn-off the repression[148]. They obtained a three-fold increase in yield as a result. Carothers et al. designed, simulated and tested a number of apatazymes, ligand-

dependent ribozymes, which can regulate mRNA half-life depending on ligand presence, to allow dynamic control of protein expression[26]. Since an option exists for a formaldehyde biosensor, it could be used to regulate the formolase pathway. This would allow the cells the dynamically control the production of key pathway enzymes. Production of formaldehyde via ACS-ACDH could trigger FLS expression to ensure that the consumption system is available if the toxic intermediate starts to buildup. The dynamic system could also decrease cellular burden as protein production for the FLS would not be on unless formaldehyde is present.

3.5 Screening and Selection Techniques

Screening and selection techniques provide means of choosing improved constructs. As our main metric of pathway function relies on a time-intensive mass spectrometry method, examination of pathway variants is immensely limited. Therefore, new methods to discern and isolate improvements in the pathway is vital. Screening implies choosing the best of a set from a gradient of output, where selection is more of a Boolean switch, as output consists only of cells that are able to survive in the given conditions. Selections are more powerful because they can be used to evaluate larger numbers of variants. The formolase pathways has several obvious screens and selections available. Challenge with formaldehyde is one selection method. Because formaldehyde is toxic, cells expressing the pathway well should survive better in the presence of formaldehyde. For further challenge, the endogenous formaldehyde system knockouts will be used, so the only method of detoxification is through the formolase pathway. As discussed in Chapter 1, growth on formate is another selection method as *E. coli* cells will not natively grow on it as a carbon substrate. These selection methods were attempted with the current pathway, however, no differences in tolerance or growth attributable to pathway function were seen. As such, we examined other selection and screening techniques that might have a lower bar for a display of formolase pathway

function.

3.5.1 Fluorescence-based Screens

For screening techniques, finding a way to link fluorescence to pathway function will allow enrichment of better strains by flow cytometry. Fluorescence can be tied to pathway function using a redox sensor dye to detect higher levels of respiration with a specific substrate as a sign of cell activity or using a biosensor with GFP or similar as an output connected to production of a pathway intermediate [68, 69]. We have shown that the redox sensor approach works in our *E. coli strain* with glucose as a substrate to trigger cellular respiration, but we have not yet tried formate. For a biosensor strategy, a DHA biosensor described in Chapter 2 could also be used.

3.5.2 Dynamic Sensor-based Selections

Biosensors can be used for selections as well as screens. If functional, the formaldehyde and DHA biosensors could have their outputs be instead tied to host growth rate instead of fluorescence. To accomplish this, pathway function would induce a gene required for growth, e.g. having pathway function allow proline synthesis in a proline auxotroph. Therefore, cells that have higher flux through the added pathway have a growth advantage because of the complementation. Dietrich et al. used tetracycline and transporter, *tetA* with this approach to yield 30% increase in production of butanol[36].

3.5.3 Aerobic growth-based selections with essential gene complementation

Since the formolase pathway cannot support growth on its own, perhaps the pathway flux requirement for growth can be lessened by use of a co-substrate unable to be a sole carbon and energy source for *E.coli* by itself. We can accomplish this by carefully placed knockouts in glycolysis such that the formolase pathway provides carbon for the pentose phosphate

pathway (PPP) and first portion of glycolysis or the latter portion of glycolysis with co-substrates that supply carbon to the other pathways sections beyond the knockout (Figure 3.3). Glycerol enters glycolysis at the same point as the formolase pathway and can be used as a substitute substrate for testing these approaches.

C3 for growth/biosynthesis, C6 for PPP derivatives

As shown in Figure 3.3 top, carbon assimilated via the formolase pathway is funneled into lower glycolysis and the TCA cycle. Knockouts made for this system block assimilation of C6 compounds via glycolysis. Those knockouts considered are essential on glucose but are functional on C3 compounds, like glycerol. Several studies provide information on gene essentiality, although all have different criteria for this designation[59, 8, 141]. Joyce et al. provides data for both glucose and glycerol and lists a set of essential to each conditions as well as both conditions [59]. In addition to knockouts considered based on network structure, this list of glucose-not-glycerol essential genes was a starting point for determining mutations to try. This set of knockouts was cross-referenced across multiple sources, including flux balance simulations described in Appendix C, and pared down to those in Table 3.1 [140]. These strains ordered from the Coli Genome Stock Center were tested on M9 minimal plates with glucose, glycerol, succinate, glucose + glycerol, succinate+glycerol and LB to confirm growth phenotype data (Table 3.2). Based on this data, a subset of knockouts was transferred with P1 phage transduction to ALA2 *fdoG* (Table 3.1 yellow rows). The resulting strains were transformed with pTrcHis2C or pTrcCO_2-3 and pSB3K3 $_p$J23100 FocA cmFDH. M9 minimal tests in solid medium and liquid medium with carbon sources following the plans in Table 3.2 did not confirm any pathway dependent improvements in growth.

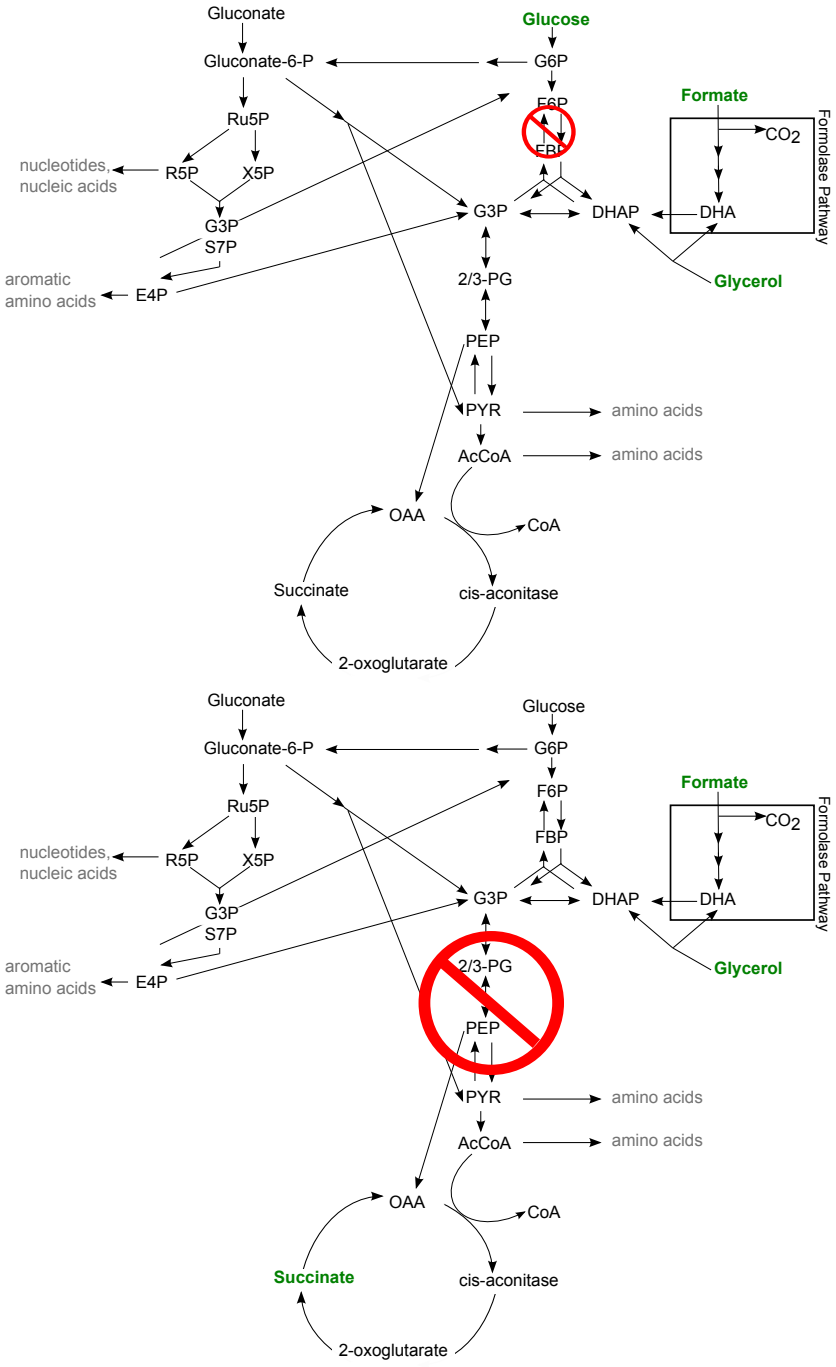


Figure 3.3: Pathway schemes for aerobic growth-based selections. Knockouts considered by inspection are in the regions marked. Green text denotes substrates used for each scenario. TOP) Knockout strains were fed glucose and formate/glycerol. Glucose provides C5/C6 compounds for growth and biosynthesis, but carbon from the formolase pathway or glycerol supplies lower glycolysis and the TCA cycle. BOTTOM) Knockout strains were fed succinate and formate/glycerol. Succinate provides for lower glycolysis and the TCA cycle, where as the formolase pathway supplies the PPP.

C4 for growth/biosynthesis, C3 for PPP derivatives

For this approach, succinate (C4) provides carbon to lower glycolysis and the TCA cycle, while the formolase pathway only provides for C5/C6 compounds in upper glycolysis and the PPP. As the PPP/glycolysis carbon at the glucose-6-phosphate split is reported as 30%/70$, this approach may be the more promising of the two[94]. C4 substrate essential gene data was mostly unavailable, but was simulated as described in Appendix C. Knockout candidates were chosen by manual inspection of the metabolic network, then referenced in the Coli Genome Stock Center to check for confirmed knockout strains (Table 3.3). *gapA* (7563/DS112) and *eno* (5965/DF575) knockouts were ordered from the Coli Genome Stock Center then tested for growth phenotypes on M9 minimal plates with glucose, glycerol, succinate, glucose + glycerol, succinate+glycerol and succinate+glycerol+1x amino acids from Neidhart et al. [93] (Table 3.4). Since strains were not available from the Keio collection, λ red recombination was attempted to make knockouts in the ALA2 *fdoG* background strain, but all colonies from multiple attempts failed colony PCR checks for cassette insertion. Since knocking out these essential genes has proved difficult, future work could transfer the pathway and the *fdoG* knockout to the confirmed *gapA* and *eno* knockout strains.

3.5.4 Anaerobic lactate production-based selection with redox constraints

Many industrial processes are preferably anaerobic because it removes the requirement for oxygen supply, which can be costly to achieve on the large scale. For pathway engineering, anerobic pathways have the advantage of the redox constraints, which provide a forcing function to maintain the balance of reduced and oxidized electron carriers. In this way a production of excess reducing equivalents forces metabolism to use available pathways for dissipation, resulting in reduced terminal products. This approach has been successful at improving ethanol production in macroalgae[139].

In this case, mannitol was used as the reduced substrate since it produces an NADH as it is

oxidized to fructose-6-phosphate. The overabundance of NADH requires the cell to oxidize it back to NAD so energy production can continue via glycolysis. If terminal electron acceptor pathways are eliminated, carbon flux can be directed to one or more reduced products. For this to work with the formolase pathway, two major tasks were required-1) remove the competing anaerobic formate consumption system, *fdhF* and *fdnG*, in addition to *fdoG* and 2) channel carbon to one easily measurable product. Lactate production was chosen since it avoids the need to express the *Zymomonous mobilis* ethanol production pathway. In order to do this, *pflA* and *frdA* were removed (Figure 3.4). These knockouts were completed using P1 phage transduction (Appendix B) from single Keio collection knockout strains sourced by the Coli Genome Stock Center. The reaction balance for this is: $3mannitol + 3formate + 4ADP \Rightarrow 7pyruvate + 4P - Pibonds + 7NADH \Rightarrow 7lactate + 7CO_2 + 4ATP$.

To confirm whether this approach is reasonable, mannitol and glucuronate were used as substrates for anaerobic growth, as glucuronate incorporation requires NADH. The balanced reaction for this is $2mannitol + 1glucuronate + 6ADP \Rightarrow 6lactate + 6CO_2 + 6ATP$. Initial tests showed inhibition of growth probably due to pH drop. Using pH paper, culture pH was determined to drop below 5.5 for the initial trial due to production of acidic fermentation products in the strain containing the pathway pTrcCO_2-3 plasmid. To prevent this, a higher concentration of phosphate, 89 mM approximately the concentration in Terrific broth, was used to buffer the medium for later experiments. A variety of richer media with more mannitol and dilute LB, thiamine or amino acid stock were examined in a second experiment. Of these, M9 + 100 mM mannitol + extra phosphate with and without 4.5 mg/L thiamine were chosen for continuing evaluation. These media with no glucuronate ("noG) controls and glucuronate ("G") were inoculated with ALA2 *fdoG fdnG fdhF pflA frdA* cells and incubated at 18 °C. These cultures were not purged, but after 15 hours, no drop in pH or OD was detected. Samples were collected for HPLC analysis. At that point, purging was done to ensure the switch to fermentation occurred. After an additional 24 hours, another round of

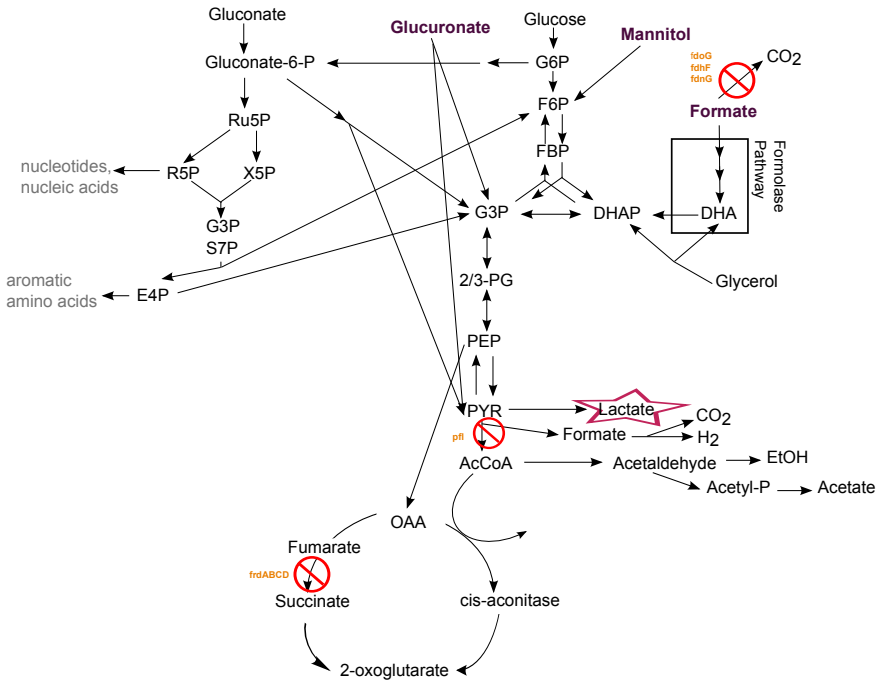


Figure 3.4: *E. coli* anaerobic metabolism with carbon substrates used - mannitol, glucuronate and formate, and knockouts made to funnel carbon to lactate: *fdnG, fdoG, fdhF, pflA, frdA*

.

samples was taken for HPLC analysis after confirmation that OD600 dropped in samples containing glucuronate. The pH did not change. Fermentation product measurements in Figure 3.5 show that glucuronate levels drop between 15 and 29 hours and lactate levels rise as would be expected for this scheme to be valid. Thus, the addition of glucuronate can be replaced with formate and the presence of ACS, ACDH and FLS.

ALA2 *fdoG fdnG fdhF pflA frdA* cells with plasmids pTrcHis2C or pTrcCO_2-3 XFLS or pTrcCO_2-3 were subject to the same analysis except cultures underwent an IPTG-induction and expression period before cell harvest and transfer to mannitol media with and without formate. Samples were collected at 1 day and 5 days. The growth of the formate peak is the main difference between the "no formate" and "formate" samples. The lactate peak does not appear to be larger for the "formate" samples for any of the pTrc plasmids tested. Longer time points or more sensitive analysis may be needed to achieve detection of the lactate.

Despite these issues, this assay is worth pursuing because it eliminates the variables of the formate dehydrogenase since NADH comes from mannitol and formate transporter expression since *focA* is natively expressed under anaerobic conditions. The activity of the transporter was not tested under these assay conditions. If assay conditions did not allow exogenous formate to enter the cells, formate may not have been available to the pathway since *pflA* was knocked out. Once assay conditions confirmed for formate intake are established, this protocol should be re-tested with alterations in other parameters. Increasing the throughput with the HPLC by use of an autosampler would increase the parameter space that could be sampled. If successful increases in lactate production with the FLS pathway and formate are evident, this scheme could be used to screen for mutagenized constructs. Although, it would not allow isolation of one gene at a time. In vivo mutagenesis with a error-prone DNA polymerase I could be used in concert with this method, or the aerobic growth pathway complementation ones, to find improved plasmid versions. This DNA polymerase replicates ColE1 plasmids thus sparing the genome from mutagenesis[25]. As the mutagenesis effects

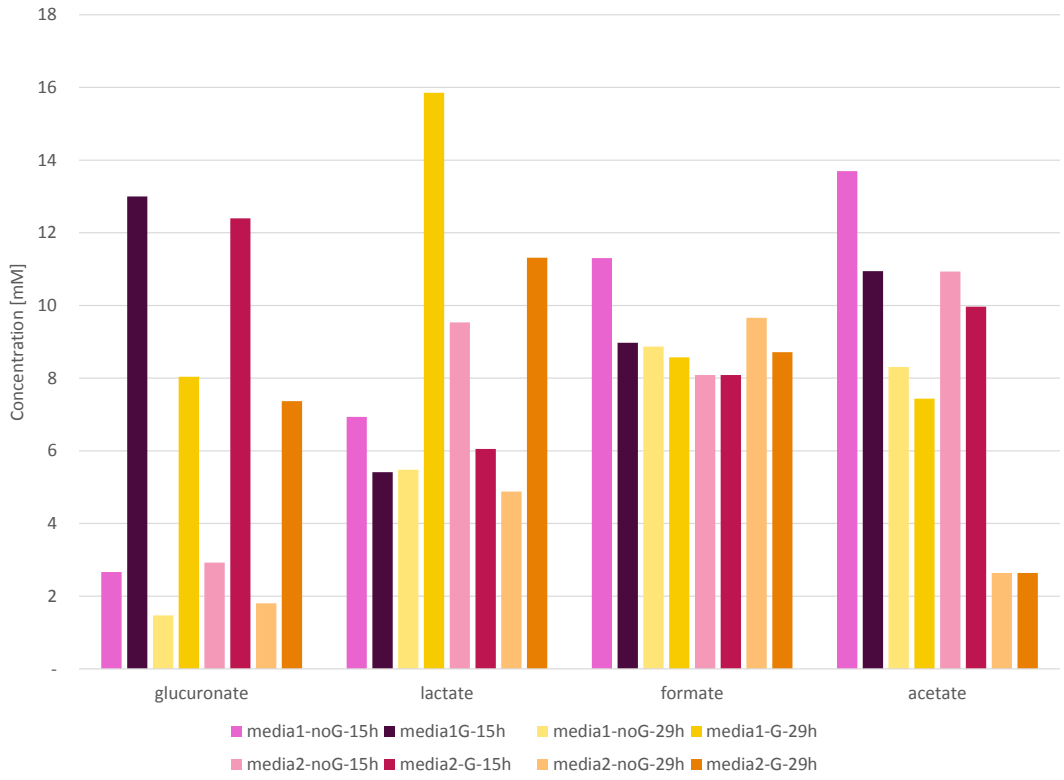


Figure 3.5: Fermentation product measurements by HPLC of anaerobic cultures designed to produce lactate from mannitol only in the presence of glucuronate. noG = no glucuronate, G = glucuronate, media 1 is M9 with 100 mM mannitol and extra phosphate, media 2 is media 1 with added thiamine. Samples were collected at 15 hours and 29 hours.

are strongest with 3kb of the plasmid origin, the current pTrcCO_2-3 would need redesigning to bring the gene targets in this range.

3.6 Methanol Alternative Formolase Pathway

3.6.1 Background

With many diagnosed and perhaps undiagnosed problems in achieving in vivo flux through the formolase pathway, a way to limit these uncertainties is vital. One way to approach this is to isolate sections of the pathway enabling independent assessment and more targeted improvement. Other one-carbon assimilation pathways that have performed relatively well have used methanol as a feedstock. The MCC, a combination of the NOG pathway, the RuMP cycle and an NAD-dependent methanol dehydrogenase, and just the RuMP cycle with an NAD-dependent dehydrogenase, both described previously, have in vitro conversion established and the latter has shown conversion to in vivo metabolites in *E. coli*. These pathways both go through formaldehyde as an intermediate as the formolase pathway does. As such, we can co-opt the production of formaldehyde from methanol eliminating the problematic coupled production of formaldehyde from formate with ACS and ACDH (3.6.1). Since these enzymes are native C2 utilizers, their over-expression may have unintended consequences on acetyl-CoA pools. They also are not efficient on C1 analogs as previously described. Enabling formaldehyde production from a proven methanol dehydrogenase will determine whether FLS can support carbon assimilation on its own.

The redox balance with the methanol dehydrogenase replacement was examined to see if anaerobic conditions would successfully provide an additional constraint. With this pathway, $3Methanol + 4NAD + 1ADP \Rightarrow 1Pyruvate + 4NADH + 1ATP$. Producing ethanol from pyruvate burns two NADH for a single pyruvate causing NADH to accumulate. Therefore, aerobic conditions are appropriate for the methanol formolase pathway.

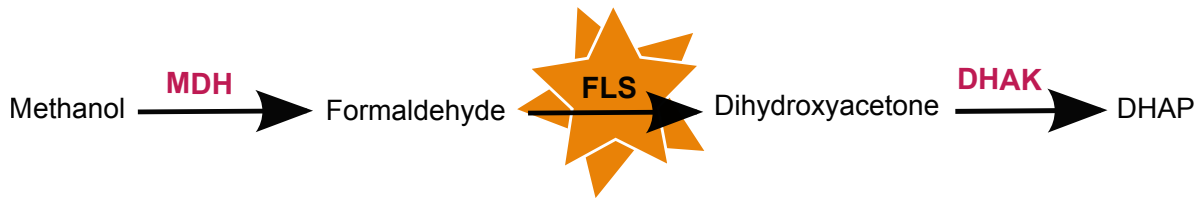


Figure 3.6: Overview of the methanol Formolase pathway. Methanol is converted to formaldehyde with a NAD-linked methanol dehydrogenase. Assimilation via formaldehyde occurs as previously described.

3.6.2 Implementation of the Methanol Formolase Pathway

A set of methanol dehydrogenase-expressing plasmids and the corresponding vector backbone were acquired from the Vorholt lab(Table 3.5) [90]. The pSEVA424 backbone has an RK2 origin and this is compatible with the pTrc plasmid. This set was transformed into MG1655 *fdoG frmA gshA* along with either pTrc BAL, the control, or pTrc FLS, the active form. Cell cultures expressing one plasmid from the pSEVA424 set and either pTrc BAL or FLS, were fed labeled and unlabeled methanol aerobically. Formaldehyde production and OD were monitored with ^{12}C methanol as a substrate, in addition to sample collection for in vivo metabolite analysis with ^{13}C methanol as a substrate.

It might be expected that the presence of an active formate production system, the methanol dehydrogenase, would be toxic without the presence of a formaldehyde consumption system. The background strain lacks the native formaldehyde detoxification pathway leaving the formolase pathway as only formaldehyde consumption system when pTrc FLS is present. Keeping with this idea, controls showed decreased growth when methanol was added. BAL was used as a control because it has minimal activity with formaldehyde as a substrate. Preliminary results show that the presence of FLS instead of BAL may be associated with protection form culture density decline with the addition of methanol, but are not statistically significant since the error bars representing standard deviation overlap. More

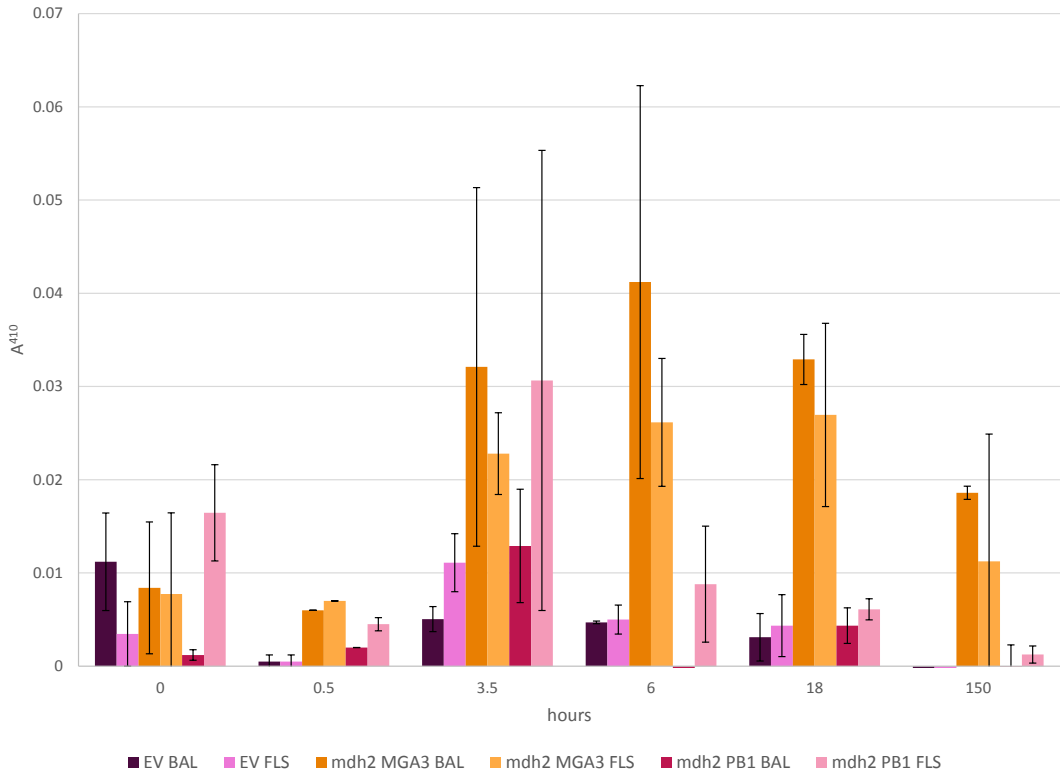


Figure 3.7: Formaldehyde production from methanol formolase pathway and controls. Absorbance at 410 nm reflects formation of Nash assay product. Formaldehyde-less buffer control background subtracted. EV = Empty vector

replicates are needed. Formaldehyde production measurements (3.6.2) in the cell culture medium with the Nash assay showed the formaldehyde production peaks around six hours and the *mdh2* from MGA3 has a higher production level than that from PB1. As a result, in vivo metabolite labeling experiment timepoints should be taken around that time for future experiments of this system instead of the short three to 20 min time points used by Müller et al or the longer timepoints used for the preliminary in vivo study. For this first study, in vivo samples from the pSEVA424 control as well as pSEVA424 *mdh2* PB1 with either pTrc BAL or FLS were analyzed and no metabolite labeling from ^{13}C methanol was detected. Unfortunately, the parallel formaldehyde measurements for PB1 were extremely low. A future experiment with the *mdh2* from MGA3 and shorter timepoints should be completed.

3.7 Summary

There are numerous challenges to overcome in order to achieve formolase pathway function. While this chapter outlines a variety of screening and selection methods, the plethora of"knobs" that must be dialed in may make the goal of growth on formate insurmountable in a single step. The approaches that lend themselves to the refinement of one "knob" should have higher priority as by limiting pathway variables, we may be able to deconvolute the issues surrounding function. Achieving conversion from methanol in vivo would be a first step to confirm FLS utility and allow for improvements via screening. Then, other sections of the pathway can be added and tweaked until it is assembled in its entirety.

x

Potential KOs	Function	LB growth?	Glucose growth?	Glycerol growth?	Source
fbaA	Fructose bisphosphate aldolase	No	No	N.D.	Inspection
fbaB	Fructose bisphosphate aldolase	Yes	Yes	Yes	Inspection
pfkA	6-phosphofructokinase	Yes	Yes (tiny amt.)	Yes	Inspection
pfkB	6-phosphofructokinase-2	Yes	Yes	Yes	Inspection
fsaA	Fructose 6-phosphate aldolase	Yes	Yes	Yes	Inspection
fsaB	Fructose 6-phosphate aldolase	Yes	Yes	Yes	Inspection
aceF	PDH complex	Yes	Yes (tiny amt.)	Yes	Joyce et al.
atpE	ATP synthase	Yes	Indeterminat (tiny amt.)	Yes	Joyce et al.
dnaT	Chromosomal DNA replication	Yes	Mixed	Yes	Joyce et al.
lipA	Lipoate synthase	Yes	Mixed (tiny amt.)	Yes	Joyce et al.
lipB	Lipoyl protein ligase	Yes	Yes (tiny amt.)	Yes	Joyce et al.
lpd	PDH, glycine cleavage, 2-oxoglutarate DH	Yes	No (tiny amt.)	Yes	Joyce et al.
pfkA	Phosphofructokinase	Yes	Yes (tiny amt.)	Yes	Joyce et al.
priA	Restart of stalled replication forks	Yes	Yes	Yes	Joyce et al.
ptsH	Pts systems	Yes	Indeterminat (tiny amt.)	Yes	Joyce et al.
ycaL	Putative heat stock protein	Yes	Yes (tiny amt.)	Yes	Joyce et al.
ilvE	Branched chain aa transferase	Yes	No (tiny amt.)	Yes	Joyce et al.
ubiE	Methyltransferase	Yes	Indeterminat	Yes	Joyce et al.
bioA	Biotin biosynthesis	Yes	No (tiny amt.)	Yes	Joyce et al.

Table 3.1: Knockouts evaluated for use with the "C3 for growth/biosynthesis, C6 for PPP derivatives" system. Growth data collected from EcoCyc, Joyce et al. and Baba et al.[140, 59, 8, 98]. Knockouts in yellow were transferred into ALA2 *fdoG*.

Knock-out	Rich Media	Glucose	Glycerol	Succinate	Succinate + Glycerol	Glucose + Glycerol	Plan of Action
aceF	++	++	++	DS	+	++	Add pathway, look for improvement on succinate+formate
atpE	++	++	+	-	DS (some tiny colonies too)	++	Add pathway, look for improvement on succinate+formate
lipA	++	A few colonies (revertants?)	-	-	+	-	Add pathway, look for improvement on succinate+formate
bioA	++	DS	DS	DS	DS	DS	Add pathway, look for improvement on succinate+formate and glucose+formate
lpd	++	-	+	-	+	Tiny colonies in DS	Add pathway, look for improvement on succinate+formate and glucose+formate
ubiE	++	++	+	-	(faint streak)	++	Add pathway, look for improvement on succinate+formate

Table 3.2: Growth phenotype data for "C3 for growth/biosynthesis, C6 for PPP derivatives" knockout strains. DS = growth only in the dense streak

Potential KOs	Function	LB growth?	Glucose growth?	Glycerol growth?
gapA	Glyceraldehyde-3-Phosphate Dehydrogenase	No	No	N.D.
pgk	Phosphoglycerate kinase	No	No	N.D.
eno	Enolase	No	No	N.D.
pykF	Pyruvate kinase	Yes	Yes	Yes
pykA	Pyruvate kinase	Yes	Yes	Yes
gpmA	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	Yes	Yes (lag)	Yes
gpmM	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	Yes	Yes	Yes
ppc	Phosphoenolpyruvate carboxylase	Yes	No	No

Table 3.3: Knockouts evaluated for use with the "C4 for growth/biosynthesis, C3 for PPP derivatives" system. Growth data collected from EcoCyc, Joyce et al. and Baba et al.[140, 59, 8, 98]. Knockouts in yellow were transferred into ALA2 *fdoG*. N.D. = No Data

Knockouts	Rich media	Glucose	Glycerol	Succinate	Succinate + Glycerol	Glucose + Glycerol	Plans
gapA	++	-	-	+	++	-	Add pathway, look for improvement on succinate+formate
eno	++	-	++	++	++	+	Add pathway, look for improvement on glucose+formate

Table 3.4: Growth phenotype data for "C3 for growth/biosynthesis, C6 for PPP derivatives" knockout strains. DS = growth only in the dense streak

Plasmid	**Gene**
pSEVA424	none
pSEVA424 *mdh2* MGA3	*mdh2* from *B. methanolicus* MGA3
pSEVA424 *mdh2* PB1	*mdh2* from *B. methanolicus* PB1
pTrc BAL	*benzaldehyde lyase* from *Pseudomonas fluorescens biovar I*
pTrc FLS	*fls* design 1 with 4 mutations

Table 3.5: Description of plasmids used for the formolase methanol dehydrogenase pathway.

Chapter 4

CONCLUSION

This work presents a proof-of-concept re-wiring of *E. coli* central metabolism to consume one carbon units using the formolase pathway. This new short, efficient carbon assimilation pathway employs a novel computationally-designed enzyme, formolase, to convert one-carbon units to the larger molecules utilized in central metabolism. The one carbon units, formate molecules, can be derived from carbon dioxide and off-peak electricity, allowing a means of storing electrical energy in the forms of liquid fuels or chemicals.

The formolase pathway requires the expression of five genes. Each gene was expressed individually and as part of a larger construct then tested for activity. Pathway function was determined by first mixing extracts with each gene expressed individually. After receiving a positive result, the entire pathway was expressed in a single strain. Clarified lysate from this strain successfully converted labeled formate to triply labeled three-carbon glycolytic intermediates confirming in vitro conversion by the formolase pathway. In vivo conversion with this strain was examined by feeding labeled formate and measuring labeled metabolites and amino acids from proteins as well as culture density. None of these methods could substantiate pathway function. However, pathway expression in a strain with a biosensor for one pathway intermediate dihydroxyacetone had pathway-dependent differences in sensor output compared to a control strain.

Clearly, pathway activity is extremely low. Based on in vitro results, a doubling time of over a three weeks was estimated. A variety of issues including using promiscuous and designed, i.e. low efficiency, enzymes, having a toxic intermediate, competition/interference from native pathways, etc. could all be preventing higher flux. To make improvements to the

pathway, a range of approaches were explored. Competing pathway knockouts made gains in in vitro incorporation for less processed extracts, but failed to produce in vivo results. Based on discrepancies in enzyme activity for formate at a pathway partition, one branch, FDH, was removed since it is functioning on its native substrate, while the other branch, ACS, relies on promiscuous activity. Relying on a co-substrate for NADH, this produced tiny amounts of labeled metabolites from labeled formate. An aerobic complementation approach where the formolase pathway substitutes for an essential gene, an anaerobic redox balance approach where lactate production requires formolase pathway to function for NADH recycling and a dynamic regulation approach where biosensors for formolase pathway intermediates trigger the expression of an essential enzyme were also investigated. A strategy to isolate a pathway section by producing formaldehyde from methanol is also underway.

Multiple mutations may be required to achieve high enough flux to support growth and to do so these screens and selections will be vital. The ability of these screening and selection methods to improve the formolase pathway depends on being able to achieve consistent evidence of conversion in vivo. It will be much harder to improve a system with no function as opposed to a tiny amount of function. The difficulty is how to get this system to a level of consistent in vivo function with metabolite labeling as a low bar. The formolase is the most important target as it performs the carboligaton. Using the alternate formaldehyde production scheme limits many problematic variables, allowing the focus for improvement to be solely on the formolase. Gains in the formolase can then be translated to the more complex anaerobic system, which adds back in the complications of the ACS-ACDH formaldehyde production, but still avoids formate transport and production of NADH with FDH. Advances at this stage could be transferred to other screening/selection methods requiring the whole pathway in order to finally achieve a final formolase pathway that can support growth on formate.

Realizing our goal of growth on formate from a C1 compound in *E. coli* would be a

milestone for the engineering of central metabolism. Similar projects focused on the implementation of non-native, but known carbon fixation pathways in *E. coli* have at best been able to generate in vivo metabolite labeling[90]. The MCC, another synthetic pathway aiming to fix carbon in *E. coli* only succeeded in the production of ethanol and butanol from methanol in a cell-free system[20]. While the formolase pathway has the benefits of high-driving force, short-length and linearity, the use of computationally designed formolase, which enables it, is one of the major hurdles to success that the other pathways did not need to overcome. Despite this, we were able to achieve metabolite labeling from ^{13}C formate in clarified cell lysates.

Success in the ability of the pathway-dependent screens and selections to improve the formolase as well as the pathway would validate the benefit of placing engineered proteins in the context of a metabolic pathway. Directed evolution, a cycle of mutagenize, screen/select and choose starting points for the next cycle, has long been used to improve de novo enzymes[19]. Often, including the FLS development, screening/selection methods for an engineered enzyme involve an assay that looks directly for a specifically targeted activity[120, 63]. For functional protein use, this may not be the best strategy as natural enzymes are subject to other constraints not tested by this type of method[19]. Directed evolution as part of a metabolic pathway may provide additional power to create more robust enzymes evolved to function in vivo. The screening and selections described here provide the opportunity to test this hypothesis and show the value of combining protein and metabolic engineering tools. If successful, this would show that we can harness the diversity that protein engineering can achieve for function in metabolic pathways.

BIBLIOGRAPHY

[1] H Abaibou, J Pommier, S Benoit, G Giordano, and M A Mandrand-Berthelot. Expression and characterization of the Escherichia coli fdo locus and a possible physiological role for aerobic formate dehydrogenase. *Journal of bacteriology*, 177(24):7141–9, December 1995.

[2] Arun S Agarwal, Yumei Zhai, Davion Hill, and Narasi Sridhar. The electrochemical reduction of carbon dioxide to formate/formic acid: engineering and economic feasibility. *ChemSusChem*, 4(9):1301–10, September 2011.

[3] S.J. Allen and J.J. Holbrook. Isolation, sequence and overexpression of the gene encoding NAD-dependent formate dehydrogenase from the methylotrophic yeast Candida methylica. 1995.

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

[5] S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–10, October 1990.

[6] Simon C Andrews, Ben C Berkst, Joseph Mcclayt, Andrew Amblert, Michael A Quail, Paul Golby, and John R Guest. A 12-cistron Escherichia coli operon (hyf) encoding a putative proton-translocating formate hydrogenlyase system. *Microbiologyy*, 143(1997):3633–3647, 1997.

[7] Maciek R Antoniewicz, Joanne K Kelleher, and Gregory Stephanopoulos. Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. *Analytical chemistry*, 79(19):7554–9, October 2007.

[8] 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, January 2006.

[9] Christoph Bächler, Philipp Schneider, Priska Bähler, Ariel Lustig, and Bernhard Erni. Escherichia coli dihydroxyacetone kinase controls gene expression by binding to transcription factor DhaR. *The EMBO journal*, 24(2):283–93, January 2005.

[10] Arren Bar-Even, Elad Noor, Avi Flamholz, and Ron Milo. Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1827(8-9), October 2012.

[11] Arren Bar-Even, Elad Noor, Nathan E Lewis, and Ron Milo. Design and analysis of synthetic carbon fixation pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 107(19):8889–94, May 2010.

[12] 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, March 2012.

[13] Arren Bar-Even, Elad Noor, Yonatan Savir, Wolfram Liebermeister, Dan Davidi, Dan S Tawfik, and Ron Milo. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50(21):4402–10, May 2011.

[14] J A Bassham, A A Benson, and M Calvin. The path of carbon in photosynthesis. *The Journal of biological chemistry*, 185(2):781–7, August 1950.

[15] S.A. Becker, A.M. Feist, M.L. Mo, G. Hannum, B.Ø. Palsson, and M.J. Herrgard. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols*, 2(3):727–738, 2007.

[16] 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, August 2009.

[17] Ivan A Berg, Daniel Kockelkorn, Wolfgang Buckel, and Georg Fuchs. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science (New York, N.Y.)*, 318(5857):1782–6, December 2007.

[18] Ginkgo Bioworks. Unpublished data. 2009.

[19] Jesse D Bloom and Frances H Arnold. In the light of directed evolution: pathways of adaptive protein evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 106 Suppl:9995–10000, June 2009.

[20] 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*, October 2014.

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

[22] Nanette R. Boyle and John A. Morgan. Computation of metabolic fluxes and efficiencies for biological carbon dioxide fixation. *Metabolic engineering*, 13(2):150–8, March 2011.

[23] T D Brown, M C Jones-Mortimer, and H L Kornberg. The enzymic interconversion of acetate and acetyl-coenzyme A in Escherichia coli. *Journal of general microbiology*, 102(2):327–36, October 1977.

[24] Joerg Martin Buescher, Sofia Moco, Uwe Sauer, and Nicola Zamboni. Ultrahigh performance liquid chromatography-tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites. *Analytical chemistry*, 82(11):4403–12, June 2010.

[25] Manel Camps, Jussi Naukkarinen, Ben P Johnson, and Lawrence a Loeb. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America*, 100:9727–9732, 2003.

[26] James M Carothers, Jonathan A Goler, Darmawi Juminaga, and Jay D Keasling. Model-driven engineering of RNA devices to quantitatively program gene expression. *Science (New York, N.Y.)*, 334(6063):1716–9, December 2011.

[27] Chi Ho Chan, Jane Garrity, Heidi A Crosby, and Jorge C Escalante-Semerena. In Salmonella enterica, the sirtuin-dependent protein acylation/deacylation system (SDPADS) maintains energy homeostasis during growth on low concentrations of acetate. *Molecular microbiology*, 80(1):168–83, April 2011.

[28] Shouqiang Cheng, Yu Liu, Christopher S Crowley, Todd O Yeates, and Thomas A Bobik. Bacterial microcompartments: their properties and paradoxes. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 30(11-12):1084–95, November 2008.

[29] Ludmila Chistoserdova, Marina G Kalyuzhnaya, and Mary E Lidstrom. The expanding world of methylotrophic metabolism. *Annual review of microbiology*, 63:477–99, January 2009.

[30] Rotsaman Chongcharoen, Thomas J Smith, Kenneth P Flint, and Howard Dalton. Adaptation and acclimatization to formaldehyde in methylotrophs capable of high-concentration formaldehyde detoxification. *Microbiology (Reading, England)*, 151(Pt 8):2615–22, August 2005.

[31] David P Clark and John E Cronan. Acetaldehyde Coenzyme A Dehydrogenase of Escherichia coli. *Journal of bacteriology*, 144(1):179–184, 1980.

[32] David P Clark and John E Cronan. Escherichia coli Mutants with Altered Control of Alcohol Dehydrogenase and Nitrate Reductase. *J Bacteriology*, 141(1):179–184, 1980.

[33] Robert J. Conrado, Chad A. Haynes, Brenda E. Haendler, and Eric J. Toone. Electrofuels: a New Paradigm. In James Weifu Lee, editor, *Advanced biofuels and bioproducts*, chapter 38, pages 1037–1064. Springer, New York, NY, 2013.

[34] Christopher S Crowley, Duilio Cascio, Michael R Sawaya, Jeffery S Kopstein, Thomas A Bobik, and Todd O Yeates. Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. *The Journal of biological chemistry*, 285(48):37838–46, November 2010.

[35] Kirill A Datsenko and Barry 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):6640–6645, 2000.

[36] Jeffrey A Dietrich, David L Shis, Azadeh Alikhani, and Jay D Keasling. Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS synthetic biology*, 2(1):47–58, January 2013.

[37] 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, August 2009.

[38] Ali Ebrahim, Joshua A Lerman, Bernhard O Palsson, and Daniel R Hyduke. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC systems biology*, 7(1):74, January 2013.

[39] Marion Eisenhut, Hermann Bauwe, and Martin Hagemann. Glycine accumulation is toxic for the cyanobacterium Synechocystis sp. strain PCC 6803, but can be compensated by supplementation with magnesium ions. *FEMS microbiology letters*, 277(2):232–7, December 2007.

[40] Chenguang Fan, Shouqiang Cheng, Yu Liu, Cristina M Escobar, Christopher S Crowley, Robert E Jefferson, Todd O Yeates, and Thomas A Bobik. Short N-terminal sequences package proteins into bacterial microcompartments. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16):7509–14, April 2010.

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

[42] Daniel G Gibson. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic acids research*, 37(20):6984–90, November 2009.

[43] Ginkgo Bioworks. Standard Assembly of Biobricks.

[44] 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.

[45] R. Grafstrom, A. Fornace, H Autrup, J. Lechner, and C. Harris. Formaldehyde damage to DNA and inhibition of DNA repair in human bronchial cells. *Science*, 220(4593):216–218, April 1983.

[46] Luisa S Gronenberg, Ryan J Marcheschi, and James C Liao. Next generation biofuel engineering in prokaryotes. *Current opinion in chemical biology*, April 2013.

[47] W G Gutheil, E Kasimoglu, and P C Nicholson. Induction of glutathione-dependent formaldehyde dehydrogenase activity in Escherichia coli and Hemophilus influenza. *Biochemical and biophysical research communications*, 238(3):693–6, September 1997.

[48] William G. Gutheil, Barton Holmquist, and Bert L. Vallee. Purification, characterization, and partial sequence of the glutathione-dependent formaldehyde dehydrogenase from Escherichia coli: a class III alcohol dehydrogenase. *Biochemistry*, 31(2):475–81, January 1992.

[49] Regula Gutknecht and Rudolf Beutler. The dihydroxyacetone kinase of Escherichia coli utilizes a phosphoprotein instead of ATP as phosphoryl donor. *The EMBO Journal*, 20(10):2480–2486, 2001.

[50] H D Heck, M Casanova, and T B Starr. Formaldehyde toxicity–new understanding. *Critical reviews in toxicology*, 20(6):397–426, January 1990.

[51] Christopher D Herring and Frederick R Blattner. Global Transcriptional Effects of a Suppressor tRNA and the Inactivation of the Regulator frmR. *Society*, 186(20):6714–6720, 2004.

[52] Harald Huber, Martin Gallenberger, Ulrike Jahn, Eva Eylert, Ivan A Berg, Daniel Kockelkorn, Wolfgang Eisenreich, and Georg Fuchs. A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum Ignicoccus hospitalis. *Proceedings of the National Academy of Sciences of the United States of America*, 105(22):7851–6, June 2008.

[53] B K Hunter, K M Nicholls, and J K Sanders. Formaldehyde metabolism by Escherichia coli. In vivo carbon, deuterium, and two-dimensional NMR observations of multiple detoxifying pathways. *Biochemistry*, 23(3):508–14, January 1984.

[54] B. Innocent, D. Liaigre, D. Pasquier, F. Ropital, J.-M. Léger, and K. B. Kokoh. Electroreduction of carbon dioxide to formate on lead electrode in aqueous medium. *Journal of Applied Electrochemistry*, 39(2):227–232, September 2008.

[55] Invitrogen. pTrcHis2 A, B, and C User's Manual. 2009.

[56] N Itoh, Y Tujibata, and JQ Q Liu. Cloning and overexpression in Escherichia coli of the gene encoding dihydroxyacetone kinase isoenzyme I from Schizosaccharomyces pombe, and its application to dihydroxyacetone phosphate production. *Applied microbiology and biotechnology*, 51(2):193–200, February 1999.

[57] Yu-Sin Jang, Jong Myoung Park, Sol Choi, Yong Jun Choi, Do Young Seung, Jung Hee Cho, and Sang Yup Lee. Engineering of microorganisms for the production of biofuels and perspectives based on systems metabolic engineering approaches. *Biotechnology advances*, 30(5):989–1000, January 2012.

[58] Lin Jiang, Eric A Althoff, Fernando R Clemente, Lindsey Doyle, Daniela Röthlisberger, Alexandre Zanghellini, Jasmine L Gallaher, Jamie L Betker, Fujie Tanaka, Carlos F Barbas, Donald Hilvert, Kendall N Houk, Barry L Stoddard, and David Baker. De novo computational design of retro-aldol enzymes. *Science (New York, N.Y.)*, 319(5868):1387–91, March 2008.

[59] Andrew R Joyce, Jennifer L Reed, Aprilfawn White, Robert Edwards, Andrei Osterman, Tomoya Baba, Hirotada Mori, Scott A Lesely, Bernhard ØPalsson, and Sanjay Agarwalla. Experimental and computational assessment of conditionally essential genes in Escherichia coli. *Journal of bacteriology*, 188(23):8259–71, December 2006.

[60] Nobuo Kato, Hiroya Yurimoto, and Rudolf K Thauer. The physiological role of the ribulose monophosphate pathway in bacteria and archaea. *Bioscience, biotechnology, and biochemistry*, 70(1):10–21, January 2006.

[61] Jay D Keasling. Manufacturing molecules through metabolic engineering. *Science (New York, N.Y.)*, 330(6009):1355–8, December 2010.

[62] Lawrence a Kelley and Michael J E Sternberg. Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols*, 4(3):363–71, January 2009.

[63] Olga Khersonsky, Daniela Röthlisberger, Orly Dym, Shira Albeck, Colin J Jackson, David Baker, and Dan S Tawfik. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *Journal of molecular biology*, 396(4):1025–42, March 2010.

[64] Lance Kizer, Douglas J Pitera, Brian F Pfleger, and Jay D Keasling. Application of functional genomics to pathway optimization for increased isoprenoid production. *Applied and environmental microbiology*, 74(10):3229–41, May 2008.

[65] Maria I Klapa. Proteomic Technologies Ion-Trap Mass Spectrometry Used in Combination with Gas Chromatography for High-Resolution Metabolic Flux Determination. 34(4), 2003.

[66] Roelco J Kleijn, Joerg M Buescher, Ludovic Le Chat, Matthieu Jules, Stephane Aymerich, and Uwe Sauer. Metabolic fluxes during strong carbon catabolite repression by malate in Bacillus subtilis. *The Journal of biological chemistry*, 285(3):1587–96, January 2010.

[67] Thomas Knight, Randall Rettberg, Leon Y Chan, Drew Endy, Reshma P Shetty, and Austin Che. Idempotent Vector Design for the Standard Assembly of Biobricks. 2003.

[68] M C Konopka, T J Strovas, David S Ojala, L Chistoserdova, M E Lidstrom, and M G Kalyuzhnaya. Respiration response imaging for real-time detection of microbial function at the single-cell level. *Applied and environmental microbiology*, 77(1):67–72, January 2011.

[69] Michael C Konopka, Sarah McQuaide, David S Ojala, Marina G Kalyuzhnaya, and Mary E Lidstrom. Single cell methods for methane oxidation analysis. *Methods in enzymology*, 495:149–66, January 2011.

[70] S Kumari, C M Beatty, D F Browning, S J Busby, E J Simel, G Hovel-Miner, and A J Wolfe. Regulation of acetyl coenzyme A synthetase in Escherichia coli. *Journal of bacteriology*, 182(15):4173–9, August 2000.

[71] U K Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259):680–5, August 1970.

[72] Hanson Lee, William C Deloache, and John E Dueber. Spatial organization of enzymes for metabolic engineering. *Metabolic Engineering*, 14(3):242–251, 2012.

[73] S Leonhartsberger, A Ehrenreich, and A Böck. Analysis of the domain structure and the DNA binding site of the transcriptional activator FhlA. *European journal of biochemistry / FEBS*, 267(12):3672–84, June 2000.

[74] Han Li, Paul H. Opgenorth, David G. Wernick, Steve Rogers, T.-Y. Tung-Yun Wu, Wendy Higashide, Peter Malati, Yi-Xin Huo, Kwang Myung Cho, and James C. Liao. Integrated electromicrobial conversion of CO2 to higher alcohols. *Science (New York, N.Y.)*, 335(6076):1596, March 2012.

[75] C S Lieber. Metabolic effects of acetaldehyde. *Biochemical Society transactions*, 16(3):241–7, June 1988.

[76] Derek R Lovley and Kelly P Nevin. Electrobiocommodities: powering microbial production of fuels and commodity chemicals from carbon dioxide with electricity. *Current opinion in biotechnology*, pages 1–6, March 2013.

[77] Kun Lu, Wenjie Ye, Avram Gold, Louise M Ball, and James A Swenberg. Formation of S-[1-(N2-deoxyguanosinyl)methyl]glutathione between glutathione and DNA induced

by formaldehyde. *Journal of the American Chemical Society*, 131(10):3414–5, March 2009.

[78] T H Ma and M M Harris. Review of the genotoxicity of formaldehyde. *Mutation research*, 196(1):37–59, July 1988.

[79] J. Maniatis, T., Fritsch, E.F., Sambrook. *Molecular cloning: a laboratory manual.* CSHL Press, 1982.

[80] Karla Martínez-Gómez, Noemí Flores, Héctor M Castañeda, Gabriel Martínez-Batallar, Georgina Hernández-Chávez, Octavio T Ramírez, Guillermo Gosset, Sergio Encarnación, and Francisco Bolivar. New insights into Escherichia coli metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. *Microbial cell factories*, 11(1):46, January 2012.

[81] Janet Matsen. Unpublished data. 2015.

[82] Matthew d Mattozzi, Marika Ziesack, Mathias J Voges, Pamela A Silver, and Jeffrey C Way. Expression of the sub-pathways of the Chloroflexus aurantiacus 3-hydroxypropionate carbon fixation bicycle in E. coli: Toward horizontal transfer of autotrophic growth. *Metabolic engineering*, 16(null):130–9, March 2013.

[83] J A Maupin and K T Shanmugam. Genetic regulation of formate hydrogenlyase of Escherichia coli: role of the fhlA gene product as a transcriptional activator for a new regulatory gene, fhlB. *Journal of bacteriology*, 172(9):4798–806, September 1990.

[84] MBIOnet. 4x SDS Protein Sample Buffer.

[85] Enrique Merino and Charles Yanofsky. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends in genetics : TIG*, 21(5):260–4, May 2005.

[86] Chuck Merryman and Daniel G Gibson. Methods and applications for assembling large DNA constructs. *Metabolic Engineering*, 14(3):196–204, 2012.

[87] Pierre Millard, Fabien Letisse, Serguei Sokol, and Jean-Charles Portais. IsoCor: correcting MS data in isotope labeling experiments. *Bioinformatics (Oxford, England)*, 28(9):1294–6, May 2012.

[88] Jeffrey Miller. *A short course in bacterial genetics : a laboratory manual and handbook for Escherichia coli and related bacteria*. Cold Spring Harbor Laboratory Press, Plainview N.Y., 1992.

[89] MIT. Registry of Standard Biological Parts [database on the internet], 2003.

[90] 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, January 2015.

[91] S. R. Narayanan, B. Haines, J. Soler, and T. I. Valdez. Electrochemical Conversion of Carbon Dioxide to Formate in Alkaline Polymer Electrolyte Membrane Cells. *Journal of The Electrochemical Society*, 158(2):A167, February 2011.

[92] T Nash. Colorimetric Estimation of formaldehyde concentration by means of the hantzsch reaction. *The Biochemical journal*, 55(3):416–21, October 1953.

[93] Frederick C Neidhardt, Philip L Bloch, and David F Smith. Culture medium for enterobacteria. *Journal of bacteriology*, 119(3):736–47, September 1974.

[94] Frederick C. Neidhardt, John L. Ingraham, and Moselio Schaechter. *Physiology of the Bacterial Cell*. Sinauer Associates, Inc., Sunderland, MA, 1990.

[95] M. Nicholoson. *The Power Makers' Challenge and the Need for Fission Energy*. Springer, London, 2012.

[96] Elad Noor, Arren Bar-Even, Avi Flamholz, and Ron Milo. Equilibrator.

[97] 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):e1003483, February 2014.

[98] Jeffrey D Orth, Tom M Conrad, Jessica Na, Joshua A Lerman, Hojung Nam, Adam M Feist, and Bernhard ØPalsson. A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. *Molecular systems biology*, 7:535, January 2011.

[99] Jeffrey D Orth, Ines Thiele, and Bernhard ØPalsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245–248, 2010.

[100] Joshua B Parsons, Stefanie Frank, David Bhella, Mingzhi Liang, Michael B Prentice, Daniel P Mulvihill, and Martin J Warren. Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Molecular cell*, 38(2):305–15, April 2010.

[101] Sean Poust. Unpublished data. 2013.

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

[103] Christine A Raines. Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: current and future strategies. *Plant physiology*, 155(1):36–42, January 2011.

[104] Torsten Reda, Caroline M Plugge, Nerilie J Abram, and Judy Hirst. Reversible interconversion of carbon dioxide and formate by an electroactive enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, 105(31):10654–8, August 2008.

[105] Florian Richter, Andrew Leaver-Fay, Sagar D Khare, Sinisa Bjelic, and David Baker. De novo enzyme design using Rosetta3. *PloS one*, 6(5):e19230, January 2011.

[106] Carol A Rohl, Charlie E M Strauss, Kira M S Misura, and David Baker. Protein structure prediction using Rosetta. *Methods in enzymology*, 383:66–93, January 2004.

[107] Daniela Röthlisberger, Olga Khersonsky, Andrew M Wollacott, Lin Jiang, Jason DeChancie, Jamie Betker, Jasmine L Gallaher, Eric A Althoff, Alexandre Zanghellini, Orly Dym, Shira Albeck, Kendall N Houk, Dan S Tawfik, and David Baker. Kemp elimination catalysts by computational enzyme design. *Nature*, 453(7192):190–5, May 2008.

[108] Frederick B. Rudolph, Daniel L. Purich, and Fromm. Coenzyme A-linked Aldehyde Dehydrogenase from Escherichia coli. 243(21):5539–5545, 1968.

[109] David F Savage, Bruno Afonso, Anna H Chen, and Pamela A Silver. Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science (New York, N.Y.)*, 327(5970):1258–61, March 2010.

[110] G Sawers, J Heider, E Zehelein, and a Böck. Expression and operon structure of the sel genes of Escherichia coli and identification of a third selenium-containing formate dehydrogenase isoenzyme. *Journal of bacteriology*, 173(16):4983–93, August 1991.

[111] Gary Sawers. The hydrogenases and formate dehydrogenases of Escherichia coli. *Antonie van Leeuwenhoek*, 66(1-3):57–88, 1994.

[112] R G Sawers. Formate and its role in hydrogen production in Escherichia coli. *Biochemical Society transactions*, 33(Pt 1):42–6, February 2005.

[113] Jan Schellenberger, Richard Que, Ronan M T Fleming, Ines Thiele, Jeffrey D Orth, Adam M Feist, Daniel C Zielinski, Aarash Bordbar, Nathan E Lewis, Sorena Rahmanian, Joseph Kang, Daniel R Hyduke, and Bernhard ØPalsson. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0t). *Nature Protocols*, 6(9):1290–1307, August 2011.

[114] Verena Schlensog, Angelika Birkmann, and August Bck. Mutations in trans which affect the anaerobic expression of a formate dehydrogenase (fdhF) structural gene. *Archives of Microbiology*, 152(1):83–89, June 1989.

[115] William T Self, Adnan Hasona, and K T Shanmugam. Expression and Regulation of a Silent Operon , hyf , Coding for Hydrogenase 4 Isoenzyme in Escherichia coli Expression and Regulation of a Silent Operon , hyf , Coding for Hydrogenase 4 Isoenzyme in Escherichia coli . 186(2):580–587, 2004.

[116] Sagit Shalel-Levanon, Ka-Yiu KY San, and George N GN Bennett. Effect of ArcA and FNR on the expression of genes related to the oxygen regulation and the glycolysis pathway in Escherichia coli under microaerobic growth conditions. *Biotechnology and bioengineering*, 92(2):147–59, October 2005.

[117] Reshma P Shetty, Drew Endy, and Thomas F Knight. Engineering BioBrick vectors from BioBrick parts. *Journal of biological engineering*, 2(1):5, January 2008.

[118] Siegel. *Computational Enzyme Design: engineering a novel catalyst, protein therapeutic and carbon fixation pathway*. PhD thesis, 2011.

[119] Justin Siegel. Unpublished data. 2012.

[120] 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*, 112(12):201500545, March 2015.

[121] Justin B Siegel, Alexandre Zanghellini, Helena M Lovick, Gert Kiss, Abigail R Lambert, Jennifer L St Clair, Jasmine L Gallaher, Donald Hilvert, Michael H Gelb, Barry L

Stoddard, Kendall N Houk, Forrest E Michael, and David Baker. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science (New York, N.Y.)*, 329(5989):309–13, July 2010.

[122] P K Smith, R I Krohn, G T Hermanson, A K Mallia, F H Gartner, M D Provenzano, E K Fujimoto, N M Goeke, B J Olson, and D C Klenk. Measurement of protein using bicinchoninic acid. *Analytical biochemistry*, 150(1):76–85, October 1985.

[123] V J Starai and J C Escalante-Semerena. Acetyl-coenzyme A synthetase (AMP forming). *Cellular and molecular life sciences : CMLS*, 61(16):2020–30, August 2004.

[124] Vincent J. Starai, Jeffrey G. Gardner, and Jorge C. Escalante-Semerena. Residue Leu-641 of acetyl-CoA synthetase is critical for the acetylation of residue Lys-609 by the protein acetyltransferase enzyme of Salmonella enterica. *Journal of Biological Chemistry*, 280(28):26200–26205, 2005.

[125] G Stephanopoulos. *Metabolic engineering : principles and methodologies*. Academic Press, San Diego, 1998.

[126] Dan Stoebel and Dan Dykhuizen. Dorman:P1 phage (lysogenic).

[127] B Suppmann and G Sawers. Isolation and characterization of hypophosphite–resistant mutants of Escherichia coli: identification of the FocA protein, encoded by the pfl operon, as a putative formate transporter. *Molecular microbiology*, 11(5):965–82, March 1994.

[128] Shiho Tanaka, Michael R Sawaya, and Todd O Yeates. Structure and mechanisms of a protein-based organelle in Escherichia coli. *Science (New York, N.Y.)*, 327(5961):81–4, January 2010.

[129] L.C. Thomason, N. Costantino, and D.L. Court. E. coli genome manipulation by P1 transduction. In Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, and Kevin Struhl, editors, *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, volume 1, chapter 1: Unit 1. John Wiley & Sons, Inc., July 2007.

[130] TMU-Tokyo iGEM Team. Assay 1, 2012.

[131] TMU-Tokyo iGEM Team. Device 1, 2012.

[132] S Tötemeyer, N a Booth, W W Nichols, B Dunbar, and I R Booth. From famine to feast: the role of methylglyoxal production in Escherichia coli. *Molecular microbiology*, 27(3):553–62, February 1998.

[133] U.S. Energy Information Administration. Emissions of Greenhouse Gases in the United States 2007. Technical report, 2008.

[134] U.S. Energy Information Administration. Annual Energy Outlook 2013 with Projection to 2040. Technical report, 2013.

[135] U.S. Energy Information Administration. Short-Term Energy Outlook (STEO) May 2013. Technical report, May 2013.

[136] Faustino Vidal-Aroca, Michele Giannattasio, Elisa Brunelli, Alessandro Vezzoli, Paolo Plevani, Marco Muzi-Falconi, and Giovanni Bertoni. One-step high-throughput assay for quantitative detection of β-galactosidase activity in intact Gram-negative bacteria, yeast, and mammalian cells. *BioTechniques*, 40(4):433–440, April 2006.

[137] Henian Wang and Robert P Gunsalus. Coordinate regulation of the Escherichia coli formate dehydrogenase fdnGHI and fdhF genes in response to nitrate, nitrite, and formate: roles for NarL and NarP. *Journal of bacteriology*, 185(17):5076–85, September 2003.

[138] Yi Wang, Yongjian Huang, Jiawei Wang, Chao Cheng, Weijiao Huang, Peilong Lu, Ya-Nan Xu, Pengye Wang, Nieng Yan, and Yigong Shi. Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature*, 462(7272):467–72, November 2009.

[139] Adam J. Wargacki, Effendi Leonard, Maung Nyan Win, Drew D. Regitsky, Christine Nicole S. Santos, Peter B. Kim, Susan R. Cooper, Ryan M. Raisner, Asael Herman, Alicia B. Sivitz, Arun Lakshmanaswamy, Yuki Kashiyama, David Baker, and Yasuo Yoshikuni. An engineered microbial platform for direct biofuel production from brown macroalgae. *Science (New York, N.Y.)*, 335(6066):308–13, January 2012.

[140] Daniel Weaver, Socorro Gama-Castro, Luis Muñiz Rascado, Robert Gunsalus, Markus Krummenacker, Carol Fulcher, Deepika Weerasinghe, Anamika Kothari, Ian Paulsen, Verena Weiss, Araceli M. Huerta, César Bonavides-Martinez, Ingrid M. Keseler, Alberto Santos-Zavaleta, Peter D. Karp, Pallavi Subhraveti, Amanda Mackie, Imke Schröder, Martin Peralta-Gil, Suzanne Paley, Julio Collado-Vides, and Aya Kubo. The EcoCyc Database. *EcoSal Plus*, 1(10), December 2013.

[141] Daniel S. Weaver, Ingrid M. Keseler, Amanda Mackie, Ian T. Paulsen, and Peter D. Karp. A genome-scale metabolic flux model of Escherichia coli K12 derived from the EcoCyc database. *BMC Systems Biology*, 8(79), 2014.

[142] Mark Welch, Sridhar Govindarajan, Jon E. Ness, Alan Villalobos, Austin Gurney, Jeremy Minshull, and Claes Gustafsson. Design Parameters of Control Synthetic Gene Expression in Escherichia coli, 2009.

[143] James L. White, Jake T. Herb, Jerry J. Kaczur, Paul W. Majsztrik, and Andrew B. Bocarsly. Photons to formate: Efficient electrochemical solar energy conversion via reduction of carbon dioxide. *Journal of CO2 Utilization*, 7:1–5, September 2014.

[144] Benjamin M Woolston, Steven Edgar, and Gregory Stephanopoulos. Metabolic engineering: past and future. *Annual review of chemical and biomolecular engineering*, 4:259–88, January 2013.

[145] Song Yang, Martin Sadilek, and Mary E Lidstrom. Streamlined pentafluorophenylpropyl column liquid chromatography-tandem quadrupole mass spectrometry and global (13)C-labeled internal standards improve performance for quantitative metabolomics in bacteria. *Journal of chromatography. A*, 1217(47):7401–10, November 2010.

[146] Yuzhen Ye and Adam Godzik. FATCAT: a web server for flexible structure comparison and structure similarity searching. *Nucleic acids research*, 32(Web Server issue):W582–5, July 2004.

[147] Alexandre Zanghellini. De Novo Computational Enzyme Design. *Current opinion in biotechnology*, 29:132–8, October 2014.

[148] 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):354–9, April 2012.

[149] Fuzhong Zhang, Sarah Rodriguez, and Jay D Keasling. Metabolic engineering of microbial pathways for advanced biofuels production. *Current opinion in biotechnology*, 22(6):775–83, December 2011.

Appendix A

MOLECULAR BIOLOGY PROTOCOLS

A.1 DNA Manipulation

A.1.1 Cloning

Codon-optimized versions of *bal*, *fls*, lm*acdh*, cm*fdh* and y*dhak* were obtained from Genescript cloned into pET29b+ vectors (Novagen). ecACDH and ecACS were amplified from genomic material and cloned into the pET29b+ expression vector between *NcoI* and *BamHI*. pTrcCO_2-3 was designed for simultaneous expression of the three critical enzymes in the formolase pathway. EcACS, Des1 and lmACDH were placed under control of a single IPTG inducible trc promoter in the expression vector pTrcHis2C. (Invitrogen) by Gibson cloning[42].

ACS was further amplified from this pET29b+ based template using the following primers:

Trc-ACS	CGATTAAATAAGGAGGAATAAACCATGGGAAGCCAGATCCACAAGCATACCAT
RBS2-R	CATATGACTTTCCTTTATTAGTGGTGGTGGTGGTGGTGCT

FLS was amplified from pET29b+ FLS using the following primers:

RBS2-FLS-F	CACCACTAATAAAGGAAAGTCATATGGCTATGATTACTGGTGGTGAACTGG
RBS3-R	CATTTGACTTTCCTTTACTAGTGGTGGTGGTGGTGGTGCT

lmACDH was amplified from pET29b+ lmACDHusing the following primers:

RBS3-ADH-F	CACCACTAGTAAAGGAAAGTCAAATGAGCCTGGAAGATAAGGACCTGCGTA
Trc2-term-R	CGTAAGCTTCGAATTCCCATATGGTACCTTATTAGTGGTGGTGGTGGTGGTGCT

The expression vector pTrcHis2C (Invitrogen), was subjected to restriction digest by *NcoI*, *BamHI* and calf-intestinal alkaline phosphatase, then gel purified. All PCR products were gel purified (Qiagen). All three gene fragments were assembled into the linearized vector in one step via the method described by Gibson et al.[42]. Ten ng of backbone DNA was used per reaction, along with equi-molar amounts of each insert. NEB 5-alpha cells were transformed with 1 μL of the Gibson product by the heat-shock method. Transformants were selected on LB/Agar carbenicillin plates and sequenced for the correct assemblage of inserts.

An alternate system of expressing ecACS, ecACDH, yDHAK and Des1 was also constructed. Two biscistronic operons were synthesized by DNA2.0 and subcloned by digestion with *EcoRI* and*PstI* (NEB) and ligation with T4 DNA ligase (Promega) into Biobrick vectors to give pSB1A3 FLS (4 mutations)-yDHAK (constitutive promoter) and pSB3K3 ecACS_ecACDH (lac inducible promoter)[43].

y*DHAK* and cm*FDH* were combined in constructs along with formate transporter, *focA*, or putative formate transporter, *focB*. Genes for yDHAK and cmFDH were amplified from pET29b+ yDHAK and pET29b+ cmFDH.

yDHAK XbaI Fwd-new	GCTGGTCTAGAAAAAATGAGCGCTAA
yDHAK BBSuffix Rev-new	GAAAGTATAGGAACTTCACTTCATTTTCTGC
cmFDH29b_XbaIadd_Fwd	GCTGGTCTAGATGAAAATCGTCCTGGTTCTGTATG
cmFDH29b_BBSuffixadd_Rev	AGTTGACTGCAGCGGCCGCTACTAGTATCATTAGG AACCGCCCTTTTTATCGT

focA and *focB* were amplified from *E. coli* BL21 genomic DNA.

These PCR products, created with Phusion Polymerase(Finnzymes/Thermo Scientific), were cloned via the Biobrick system into vector backbone pSB3K3 with high constitutive expression cassette (BBa_K314100) or low constitutive expression cassette (BBa_K314101) in various combinations[117, 89, 43]. pSB3K3 $_p$J23100 DHAK and a set of four constructs, pSB4C5 $_p$J23100/114 cmFDH $_p$J23114 yDHAK $_p$J23114 FocA/B (pSB4C5 FDF3-6), were

F_FocA	GTTTCTTCGAATTCGCGGCCGCTTCTAGAGGTGAAAGCTGACAACCC TTTTGATCTTTTAC
R_FocA	GTTTCTTCCTGCAGCGGCCGCTACTAGTATTAATGGTGGTCGTTTTCA CGCAGG
F_FocB	GTTTCTTCGAATTCGCGGCCGCTTCTAGATGAGAAACAAACTCTCTTT CGACTTGC
R_FocB	GTTTCTTCCTGCAGCGGCCGCTACTAGTATCAGGGTTCCTGACGTAA ATAAATAGCC

created.

A.1.2 Ethanol Precipitation of DNA

Materials

3M Sodium Acetate buffer, pH 5.2 (store at 4 °C) Cold 100% Ethanol (−20 °C) Cold 70% Ethanol in sterile dH_2O (−20 °C) DNA sample 4 °C Microcentrifuge (normal microcentrifuge in cold room works fine). All centrifugations should be on "soft" (no brake) setting.

Procedure

1. Transfer DNA to a container where it fills one fourth the total volume (a 500 μL tube should have no more than 125 μL of DNA solution, for example)

2. Add one tenth volume of Sodium Acetate buffer to equalize ion concentrations (assuming DNA is in TE buffer)

3. Add at least two volumes (2-2.5 of post-salt addition volume) of cold 100% ethanol; mix; let stand in −20 °C freezer for at least one hour

4. Centrifuge sample for 15 minutes at highest speed in a 4 °C microcentrifuge

5. Remove as much supernatant as possible with a 1 mL micropipet; recentrifuge, then remove the rest with a 200 μL pipet

6. Add 200 μL- 1mL of cold 70% ethanol; mix; centrifuge for 5 minutes in a 4 °C centrifuge

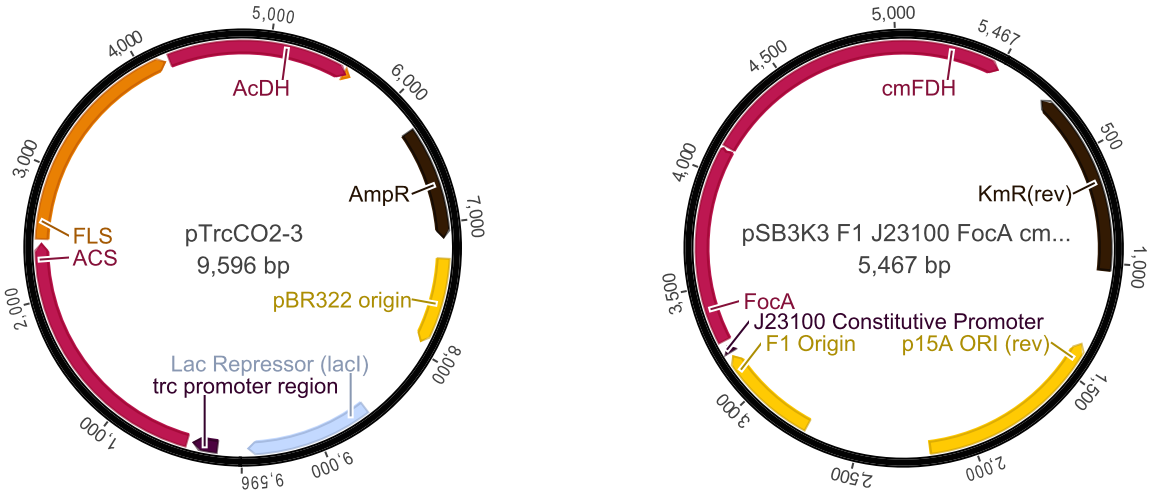


Figure A.1: Higher Expression Plasmid Maps. LEFT) pTrcCO_2-3 expressing ACS, FLS and AcDH on a pTrcHis2 backbone, lac-inducible (Invitrogen). RIGHT) pSB3K3 F1 $_p$J23100 FocA cmFDH, medium copy number, strong constitutve expression. Versions of pSB3K3 F1 $_p$J23100 or $_p$J23114, a weaker promoter, with DHAK, FocA or cmFDH only were also used.

7. Remove supernatant with a 200 μL pipet; evaporate remaining ethanol in a 37 °C water bath or vacuum dry pellet.

8. Resuspend pellet in desired volume of water or TE buffer.

A.1.3 Colony PCR

OneTaq 2x Master Mix (NEB) was used for all colony PCRs. The initial denaturing time was extended to two minutes. Large colonies were resuspended in 20 μ of sterile MilliQ water then 2-4 μ were added for a 10 μL reaction or a colony was picked with a pipet tip, mixed into a PCR then used to start a liquid or plate culture.

A.2 Coding sequences of the pathway genes

cmFDH

```
   1 ATGAAAATCG TCCTGGTTCT GTATGATGCG GGCAAGCACG CGGCGGATGA AGAAAAACTG
  61 TACGGCTGCA CCGAAAATAA GCTGGGCATT GCCAACTGGC TGAAAGATCA GGGCCATGAA
 121 CTGATCACCA CGAGTGATAA GGAAGGTGAA ACCTCCGAAC TGGACAAACA TATTCCGGAT
 181 GCTGACATTA TCATTACCAC GCCGTTTCAC CCGGCGTATA TCACGAAGGA ACGCCTGGAT
 241 AAAGCCAAGA ATCTGAAATC AGTGGTTGTC GCAGGCGTTG GTTCGGACCA TATCGATCTG
 301 GACTACATTA ACCAGACCGG CAAAAAGATT AGCGTCCTGG AAGTGACGGG TAGCAATGTG
 361 GTTTCTGTGG CTGAACACGT CGTGATGACC ATGCTGGTTC TGGTCCGTAA CTTTGTCCCG
 421 GCCCATGAAC AAATCATTAA TCACGATTGG GAAGTGGCGG CCATCGCTAA GGATGCGTAT
 481 GACATTGAAG GCAAAACCAT CGCCACGATT GGCGCAGGTC GTATCGGTTA CCGCGTTCTG
 541 GAACGTCTGC TGCCGTTCAA CCCGAAAGAA CTGCTGTATT ACGATTATCA GGCCCTGCCG
 601 AAAGAAGCAG AAGAAAAAGT GGGCGCGCGT CGCGTTGAAA ATATCGAAGA ACTGGTGGCC
 661 CAAGCAGACA TTGTGACCGT TAACGCTCCG CTGCATGCGG GCACGAAAGG TCTGATCAAC
 721 AAGGAACTGC TGAGTAAGTT CAAGAAGGGT GCATGGCTGG TTAACACCGC CCGCGGTGCA
 781 ATCTGCGTTG CAGAAGATGT CGCAGCTGCG CTGGAATCAG GCCAGCTGCG TGGTTATGGC
 841 GGTGACGTCT GGTTCCCGCA ACCGGCACCG AAAGATCATC CGTGGCGTGA CATGCGCAAC
 901 AAATATGGCG CTGGTAATGC GATGACCCCG CACTACAGCG GCACCACGCT GGATGCTCAG
 961 ACCCGCTATG CGGAAGGCAC GAAAAACATT CTGGAAAGCT TTTTCACCGG TAAATTCGAT
1021 TACCGTCCGC AAGACATCAT TCTGCTGAAT GGCGAATATG TGACGAAAGC GTACGGTAAA
1081 CACGATAAAA AGGGCGGTTC CCTCGAG
```

ecACS

```
   1 ATGGGAAGCC AGATCCACAA GCATACCATT CCTGCCAACA TCGCAGACCG TTGCCTGATA
  61 AACCCTCAGC AGTACGAGGC GATGTATCAA CAATCTATTA ACGTACCTGA TACCTTCTGG
 121 GGCGAACAGG GAAAAATTCT TGACTGGATC AAACCTTACC AGAAGGTGAA AAACACCTCC
 181 TTTGCCCCCG GTAATGTGTC CATTAAATGG TACGAGGACG GCACGCTGAA TCTGGCGGCA
 241 AACTGCCTTG ACCGCCATCT GCAAGAAAAC GGCGATCGTA CCGCCATCAT CTGGGAAGGC
 301 GACGACGCCA GCCAGAGCAA ACATATCAGC TATAAAGAGC TGCACCGCGA CGTCTGCCGC
 361 TTCGCCAATA CCCTGCTCGA GCTGGGCATT AAAAAAGGTG ATGTGGTGGC GATTTATATG
 421 CCGATGGTGC CGGAAGCCGC GGTTGCGATG CTGGCCTGCG CCCGCATTGG CGCGGTGCAT
 481 TCGGTGATTT TCGGCGGCTT CTCGCCGGAA GCCGTTGCCG GGCGCATTAT TGATTCCAAC
 541 TCACGACTGG TGATCACTTC CGACGAAGGT GTGCGTGCCG GGCGCAGTAT TCCGCTGAAG
 601 AAAAACGTTG ATGACGCGCT GAAAAACCCG AACGTCACCA GCGTAGAGCA TGTGGTGGTA
 661 CTGAAGCGTA CTGGCGGGAA AATTGACTGG CAGGAAGGGC GCGACCTGTG GTGGCACGAC
 721 CTGGTTGAGC AAGCGAGCGA TCAGCACCAG GCGGAAGAGA TGAACGCCGA AGATCCGCTG
 781 TTTATTCTCT ACACCTCCGG TTCTACCGGT AAGCCAAAAG GTGTGCTGCA TACTACCGGC
 841 GGTTATCTGG TGTACGCGGC GCTGACCTTT AAATATGTCT TTGATTATCA TCCGGGTGAT
 901 ATCTACTGGT GCACCGCCGA TGTGGGCTGG GTGACCGGAC ACAGTTACTT GCTGTACGGC
 961 CCGCTGGCCT GCGGTGCGAC CACGCTGATG TTTGAAGGCG TACCCAACTG GCCGACGCCT
1021 GCCCGTATGG CGCAGGTGGT GGACAAGCAT CAGGTCAATA TTCTCTATAC CGCACCCACG
1081 GCGATCCGCG CGCTGATGGC GGAAGGCGAT AAAGCGATCG AAGGCACCGA CCGTTCGTCG
1141 CTGCGCATTC TCGGTTCCGT GGGCGAGCCA ATTAACCCGG AAGCGTGGGA GTGGTACTGG
1201 AAAAAAATCG GCAACGAGAA ATGTCCGGTG GTCGATACCT GGTGGCAGAC CGAAACCGGC
1261 GGTTTCATGA TCACCCCGCT GCCTGGCGCT ACCGAGCTGA AAGCCGGTTC GGCAACACGT
1321 CCGTTCTTCG GCGTGCAACC GGCGCTGGTC GATAACGAAG GTAACCCGCT GGAGGGGGCC
1381 ACCGAAGGTA GCCTGGTAAT CACCGACTCC TGGCCGGGTC AGGCGCGTAC GCTGTTTGGC
1441 GATCACGAAC GTTTTGAACA GACCTACTTC TCCACCTTCA AAAATATGTA TTTCAGCGGC
1501 GACGGCGCGC GTCGCGATGA AGATGGCTAT TACTGGATAA CCGGGCGTGT GGACGACGTG
1561 CTGAACGTCT CCGGTCACCG TCTGGGGACG GCAGAGATTG AGTCGGCGCT GGTGGCGCAT
1621 CCGAAGATTG CCGAAGCCGC CGTAGTAGGT ATTCCGCACA ATATTAAAGG TCAGGCGATC
1681 TACGCCTACG TCACGCTTAA TCACGGGGAG GAACCGTCAC CAGAACTGTA CGCAGAAGTC
1741 CGCAACTGGG TGCGTAAAGA GATTGGCCCG CTGGCGACGC CAGACGTGCT GCACTGGACC
1801 GACTCCCTGC CTAAAACCCG CTCCGGCAAA ATTATGCGCC GTATTCTGCG CAAAATTGCG
1861 GCGGGCGATA CCAGCAACCT GGGCGATACC TCGACGCTTG CCGATCCTGG CGTAGTCGAG
1921 AAGCTGCTTG AAGAGAAGCA GGCTATCGCG ATGCCATCGG GGGATCCGAA TTCGAGCTCC
1981 GTCGACAAGC TTGCGGCCGC A
```

lmACDH

```
   1 ATGAGCCTGG AAGATAAGGA CCTGCGTAGC ATCCAAGAAG TGCGTAACCT GATCGAATCT
  61 GCGAATAAGG CCCAAAAAGA ACTGGCGGCG ATGTCACAGC AACAGATTGA TACCATCGTG
 121 AAAGCGATTG CCGACGCAGG CTATGGTGCG CGTGAAAAAC TGGCTAAGAT GGCGCACGAA
 181 GAAACGGGCT TTGGTATTTG GCAGGATAAA GTTATCAAGA ACGTCTTCGC CTCGAAGCAT
 241 GTCTACAACT ACATCAAGGA TATGAAGACC ATCGGTATGC TGAAAGAAGA CAACGAAAAG
 301 AAAGTTATGG AAGTCGCAGT GCCGCTGGGC GTGGTTGCTG GTCTGATTCC GTCAACCAAT
 361 CCGACCTCGA CGGTGATCTA CAAAACGCTG ATTTCAATCA AGGCGGGCAA CAGTATCGTG
 421 TTTAGCCCGC ACCCGAATGC CCTGAAAGCA ATTCTGGAAA CCGTCCGCAT TATCTCAGAA
 481 GCGGCCGAAA AAGCAGGCTG CCCGAAGGGT GCTATTTCGT GTATGACCGT TCCGACGATC
 541 CAAGGCACCG ATCAGCTGAT GAAACATAAG GACACCGCTG TCATTCTGGC AACGGGCGGT
 601 TCTGCGATGG TGAAAGCAGC TTATAGCTCT GGCACCCCGG CAATTGGTGT GGGTCCGGGC
 661 AACGGTCCGG CCTTTATTGA ACGTAGTGCG AATATCCCGC GTGCGGTTAA ACACATCCTG
 721 GATTCCAAGA CCTTCGACAA CGGTACGATT TGCGCCAGCG AACAGTCTGT CGTGGTTGAA
 781 CGTGTCAATA AAGAAGCTGT GATCGCGGAA TTTCGCAAGC AAGGCGCACA CTTCCTGAGT
 841 GATGCTGAAG CGGTGCAGCT GGGCAAATTC ATTCTGCGTC CGAACGGTAG CATGAATCCG
 901 GCGATTGTGG GCAAAAGCGT GCAACATATC GCAAACCTGG CAGGTCTGAC CGTGCCGGCC
 961 GATGCACGTG TTCTGATTGC GGAAGAAACG AAAGTTGGCG CCAAGATCCC GTATAGTCGC
1021 GAAAAACTGG CCCCGATTCT GGCATTTTAC ACCGCTGAAA CGTGGCAGGA AGCATGCGAA
1081 CTGAGCATGG ATATTCTGTA CCATGAAGGC GCTGGTCACA CCCTGATTAT CCATAGCGAA
1141 GACAAAGAAA TTATCCGTGA ATTTGCACTG AAAAAGCCGG TTTCTCGCCT GCTGGTCAAC
1201 ACGCCGGGCG CGCTGGGCGG CATTGGTGCC ACCACGAATC TGGTTCCGGC ACTGACGCTG
1261 GGCTGTGGTG CTGTCGGCGG TAGTTCCTCA TCGGATAATA TCGGTCCGGA AAACCTGTTT
1321 AATATTCGTC GCATCGCCAC CGGCGTGCTG GAACTGGAAG ACATTCGCGA AGGCGGTAGC
```

FLS

```
   1 ATGGCTATGA TTACTGGTGG TGAACTGGTT GTTCGTACCC TGATTAAAGC TGGCGTAGAA
  61 CATCTGTTTG GCCTGCATGG CATTCATATT GACACCATTT TTCAGGCTTG CCTGGACCAC
 121 GACGTCCCAA TCATTGATAC TCGCCACGAA GCGGCGGCAG GCCACGCTGC GGAAGGTTAT
 181 GCCCGCGCGG GCGCTAAACT GGGTGTTGCC CTGGTGACCG CTGGCGGTGG CTTTACCAAT
 241 GCCGTTACGC CGATCGCGAA CGCTCGGACC GATCGCACTC CGGTTCTGTT CCTGACCGGT
 301 TCTGGTGCTC TTCGTGATGA CGAAACCAAC ACCCTGCAGG CCGGTATTGA TCAGGTGGCC
 361 ATGGCGGCCC CGATCACGAA ATGGGCTCAT CGTGTTATGG CAACTGAACA CATCCCGCGT
 421 CTGGTTATGC AGGCCATTCG TGCCGCTCTG AGCGCCCCAC GTGGCCCGGT GCTGCTGGAT
 481 CTGCCATGGG ACATCCTGAT GAACCAAATC GATGAAGATT CCGTTATCAT CCCAGACCTG
 541 GTGCTGTCTG CTCACGGTGC CCATCCAGAC CCGGCTGACC TGGACCAGGC TCTGGCACTG
 601 CTGCGTAAAG CCGAACGCCC AGTTATCGTA CTGGGCTCCG AGGCGTCCCG CACCGCACGC
 661 AAGACCGCAC TGAGCGCATT CGTAGCGGCG ACCGGTGTAC CGGTTTTCGC TGACTATGAA
 721 GGCCTGTCCA TGCTGAGCGG CCTGCCGGAC GCTATGCGTG GCGGCCTGGT GCAGAACCTG
 781 TACTCCTTTG CAAAAGCTGA TGCAGCTCCG GACCTGGTAC TGATGCTGGG TGCTCGTTTC
 841 GGTCTGAACA CCGGTCATGG TTCCGGTCAA CTGATCCCGC ATTCTGCTCA GGTGATCCAG
 901 GTGGATCCAG ACGCGTGTGA ACTGGGTCGC CTGCAAGGCA TCGCGCTGGG TATCGTGGCT
 961 GATGTAGGTG GCACCATTGA AGCGCTGGCT CAGGCGACCG CACAGGACGC CGCGTGGCCG
1021 GACCGCGGCG ACTGGTGCGC CAAGGTAACT GACCTGGCCC AGGAGCGTTA CGCTTCCATC
1081 GCGGCTAAAT CCAGCTCTGA ACATGCGCTG CACCCGTTCC ACGCTTCTCA GGTTATCGCG
1141 AAACACGTGG ACGCAGGCGT GACCGTCGTT GCGGATGGTG GCCTGACTTA TCTGTGGCTG
1201 TCCGAAGTTA TGTCTCGTGT CAAACCAGGC GGCTTCCTGT GCCACGGCTA TCTGAACAGC
1261 ATGGGTGTAG GCTTCGGTAC TGCCCTGGGT GCGCAGGTTG CGGATCTGGA GGCAGGTCGT
1321 CGTACCATCC TGGTGACCGG CGACGGCTCT GTTGGTTATT CCATTGGCGA ATTCGACACC
1381 CTGGTACGCA AACAGCTGCC GCTGATTGTA ATTATCATGA ACAACCAGTC TTGGGGCTGG
1441 ACCCTGCACT TTCAGCAGCT GGCCGTTGGT CCTAACCGTG TCACCGGCAC CCGCCTGGAA
1501 AATGGTTCCT ATCACGGCGT TGCTGCGGCA TTCGGTGCTG ATGGTTACCA CGTCGACTCT
1561 GTCGAGAGCT TCAGCGCCGC TCTGGCTCAG GCACTGGCAC ACAACCGCCC GGCATGCATC
1621 AACGTTGCTG TGGCCCTGGA CCCGATCCCG CCGGAGGAAC TGATCCTGAT TGGCATGGAC
1681 CCGTTTGCGG GCTCCACGGA GAATCTGTAT TTCCAATCCG GCGCG
```

yDHAK

```
   1 ATGTCCGCTA AATCGTTTGA AGTCACAGAT CCAGTCAATT CAAGTCTCAA AGGGTTTGCC
  61 CTTGCTAACC CCTCCATTAC GCTGGTCCCT GAAGAAAAAA TTCTCTTCAG AAAGACCGAT
 121 TCCGACAAGA TCGCATTAAT TTCTGGTGGT GGTAGTGGAC ATGAACCTAC ACACGCCGGT
 181 TTCATTGGTA AGGGTATGTT GAGTGGCGCC GTGGTTGGCG AAATTTTTGC ATCCCCTTCA
 241 ACAAAACAGA TTTTAAATGC AATCCGTTTA GTCAATGAAA ATGCGTCTGG CGTTTTATTG
 301 ATTGTGAAGA ACTACACAGG TGATGTTTTG CATTTTGGTC TGTCCGCTGA GAGAGCAAGA
 361 GCCTTGGGTA TTAACTGCCG CGTTGCTGTC ATAGGTGATG ATGTTGCAGT TGGCAGAGAA
 421 AAGGGTGGTA TGGTTGGTAG AAGAGCATTG GCAGGTACCG TTTTGGTTCA TAAGATTGTA
 481 GGTGCCTTCG CAGAAGAATA TTCTAGTAAG TATGGCTTAG ACGGTACAGC TAAAGTGGCT
 541 AAAATTATCA ACGACAATTT GGTGACCATT GGATCTTCTT TAGACCATTG TAAAGTTCCT
 601 GGCAGGAAAT TCGAAAGTGA ATTAAACGAA AAACAAATGG AATTGGGTAT GGGTATTCAT
 661 AACGAACCTG GTGTGAAAGT TTTAGACCCT ATTCCTTCTA CCGAAGACTT GATCTCCAAG
 721 TATATGCTAC CAAAACTATT GGATCCAAAC GATAAGGATA GAGCTTTTGT AAAGTTTGAT
 781 GAAGATGATG AAGTTGTCTT GTTAGTTAAC AATCTCGGCG GTGTTTCTAA TTTTGTTATT
 841 AGTTCTATCA CTTCCAAAAC TACGGATTTC TTAAAGGAAA ATTACAACAT AACCCCGGTT
 901 CAAACAATTG CTGGCACATT GATGACCTCC TTCAATGGTA ATGGGTTCAG TATCACATTA
 961 CTAAACGCCA CTAAGGCTAC AAAGGCTTTG CAATCTGATT TTGAGGAGAT CAAATCAGTA
1021 CTAGACTTGT TGAACGCATT TACGAACGCA CCGGGCTGGC CAATTGCAGA TTTTGAAAAG
1081 ACTTCTGCCC CATCTGTTAA CGATGACTTG TTACATAATG AAGTAACAGC AAAGGCCGTC
1141 GGTACCTATG ACTTTGACAA GTTTGCTGAG TGGATGAAGA GTGGTGCTGA ACAAGTTATC
1201 AAGAGCGAAC CGCACATTAC GGAACTAGAC AATCAAGTTG GTGATGGTGA TTGTGGTTAC
1261 ACTTTAGTGG CAGGAGTTAA AGGCATCACC GAAAACCTTG ACAAGCTGTC GAAGGACTCA
1321 TTATCTCAGG CGGTTGCCCA AATTTCAGAT TTCATTGAAG GCTCAATGGG AGGTACTTCT
1381 GGTGGTTTAT ATTCTATTCT TTTGTCGGGT TTTTCACACG GATTAATTCA GGTTTGTAAA
1441 TCAAAGGATG AACCCGTCAC TAAGGAAATT GTGGCTAAGT CACTCGGAAT TGCATTGGAT
1501 ACTTTATACA AATATACAAA GGCAAGGAAG GGATCATCCA CCATGATTGA TGCTTTAGAA
1561 CCATTCGTTA AAGAATTTAC TGCATCTAAG GATTTCAATA AGGCGGTAAA AGCTGCAGAG
1621 GAAGGTGCTA AATCCACTGC TACATTCGAG GCCAAATTTG GCAGAGCTTC GTATGTCGGC
1681 GATTCATCTC AAGTAGAAGA TCCTGGTGCA GTAGGCCTAT GTGAGTTTTT GAAGGGGGTT
1741 CAAAGCGCCT TG
```

Appendix B

CELL MANIPULATIONS

B.1 Competent Cells

B.1.1 CCMB Chemical Competent Cells

Supplies

- Plate of *E. Coli* colonies
- SOB-Mg growth medium
- "-" is "minus", indicating that there is no added Mg
- Sterilized 500 mL Erlenmeyer flasks
- 50 mL screw-cap polypropylene tubes
- Freezer tubes (1.5 mL)
- SOC medium for recovery after heat shock (LB/TB is fine instead)

Method

1. Pick several colonies (or a single colony for an overnight) off a freshly streaked plate into ∼1 mL SOB-Mg growth medium
2. Grow cells overnight or several hours in media with appropriate antibiotics if available.
3. Use more inoculum in the next step if cultures weren't grown overnight.
4. Inoculate 50 mL SOB-Mg growth medium with this culture. Use 500 μL of stationary phase culture for 50 mL SOB-Mg medium.
5. Incubate at 275 rpm, 37 °C until OD600 is about 0.3, which corresponds to ∼5*10^7cells/mL. Higher OD isn't usually a problem for routine work.

6. Collect in sterile 50 mL polypropylene centrifuge tube(s) and chill on ice for 10 minutes.
7. Pellet the cells at 750 - 1,000xg (2500 rpm) for 14 min at 4 °C. Decant the supernatant and invert tubes to remove excess culture medium.
8. Disperse cells in ∼1/3 volume of CCMB by gentle vortexing or rapping of the centrifuge tube.
9. Incubate on ice for 20 minutes.
10. Centrifuge at 2500 rpm for 10 min at 4 °C.
11. Resuspend cells in CCMB at 1/12 the original culture volume.
12. Make aliquots in eppendorf tubes, ideally on ice.
13. Flash freeze with liquid nitrogen.
14. Store at −80 °C to preserve them for many months.

Recipes

SOB-Mg (1L)

Ingredient	**Amount**
Bacto Tryptone	20g
Bacto Yeast Extract	5g
1M NaCl	10mL
1M KCl	2.5mL

- Add water to make 1L
- Autoclave

CCMB (1L)

- Prepare a 1M solution of potassium acetate, pH 7.0 using KOH.
- Filter through a 0.2 μM membrane and store frozen.

Ingredient	Amount	Final Concentration
Potassium Acetate, 1M, pH 7	10 mL	10 mM
Glycerol	100g	10% (w/v)
CaCl2*2H2O	11.8g	80mM
MnCl2*4H2O	4g	20mM
MgCl2*6H2O	2.5mL	10mM

- Prepare a solution of 10% potassium acetate, 10% glycerol.
- Add salts, allowing each to enter solution before adding the next.
- Adjust pH to 6.4 with 0.1 M HCl. Do not adjust pH upward with base.
- Filter through a 0.2 μM filter & store at 4 °C.

B.1.2 Chemical Transformation

1. Thaw frozen (−80 °C) competent cells on ice.
2. Add DNA typically 0.5-1 μ of a plasmid prep, 10 μL if ligated plasmid.
3. Flick to mix.
4. Incubate 42 °C for 45 sec - 1 min.
5. Incubate on ice for 2 min.
6. Add 1 mL LB, can use less depending on the amount of cells.
7. Incubate at 37 °C for 45 min - 1 hr in eppendorf tubes. Can tape tubes to a rack or put in a flask.
8. Pellet cells by centrifugation.
9. Keep the ∼200 μL of volume and resuspend the cells.
10. Plate 100 μL cells on LB (+antibiotic(s)) agar plate in a gradient over the plate so there is a section of cells that should have single colonies.

B.1.3 Electrocompetent Cells

1. Pick a single colony for an overnight off a freshly streaked plate into ∼1 mL LB medium.

2. Grow cells overnight in 2 mL of LB + appropriate antibiotics.

3. Inoculate 7 mL in a green cap tube or 50 mL growth medium with this culture. Use a 1:100 dilution for inoculation of stationary culture e.g. 500 μL of stationary phase culture for 50 mL LB medium. Grow at 37 °C with shaking (225 rpm) until mid-log phase is reached ∼0.6 OD600 measured with the Nanodrop or the Tubespec.

4. Cool sterile milliQ and sterile 20% (v/v) glycerol to 4 °C.

5. Harvest cell cultures at mid-log, transfer to plastic tubes for centrifugation and put on ice.

6. Centrifuge cells 10 min or 20 min at 5,000rpm, 4 °C for 7 mL or 50 mL, respectively.

7. Decant supernatant and resuspend pellet in ∼1/3 of the original volume of sterile, cold water.

8. Centrifuge to collect pellets.

9. Repeat washing pellet with large volumes 1/3 to 1/5 of the original volumes twice more.

10. Reduce wash volume to ∼1 mL of 7 mL and ∼2 mL for 50 mL, transfer to eppendorf tubes and perform two more washes. Centrifugation can be done in the smaller rotor, 2-3 min at 14,000rpm, 4 °C. Pipet instead of decanting. Cells should easily be resuspended.

11. Resuspend to final volumes of ∼320 μL for 7 mL and 1 mL for 50 mL using 50% sterile cold water and 50% sterile cole 20% (v/v) glycerol.

12. Flash freeze with liquid Nitrogen or use immediately.

B.1.4 Electroporation

1. Defrost cells on ice.
2. Place a clean, sterile 1 mm gap cuvette on ice.
3. Add DNA (0.5 μL of plasmid, more of linear DNA, ligation or Gibson reactions).
4. Flick to mix.
5. Transfer DNA and cells to cuvette between the plates.
6. Return to ice.
7. Turn on electroporation machine. Settings: 1.2-1.8 kV, low range, 200, capacitance = 25 μF.
8. Remove cuvette from ice and wipe plates to remove liquid.
9. Insert into slide unit making sure plates on the cuvette contact the prongs on the unit.
10. Press two buttons until a beep to shock.
11. Immediately add SOC medium, 0.5 - 1 mL. Pipet up and down to collect cells and transfer to an eppendorf tube.
12. Finish recovery in the same manner as a chemical transformation.

B.2 Strain Engineering

B.2.1 λ Red Recombination

This protocol is adapted from Wanner et al., Baba et al. and openwetware[8, 35].

1. Order primers as given in Baba et al. [8]. The supplement gives H1 and H2 regions which are appended to P1 and P2, respectively, to give the final primers.
2. PCR primers
3. Transform pKD46 into strain to be manipulated

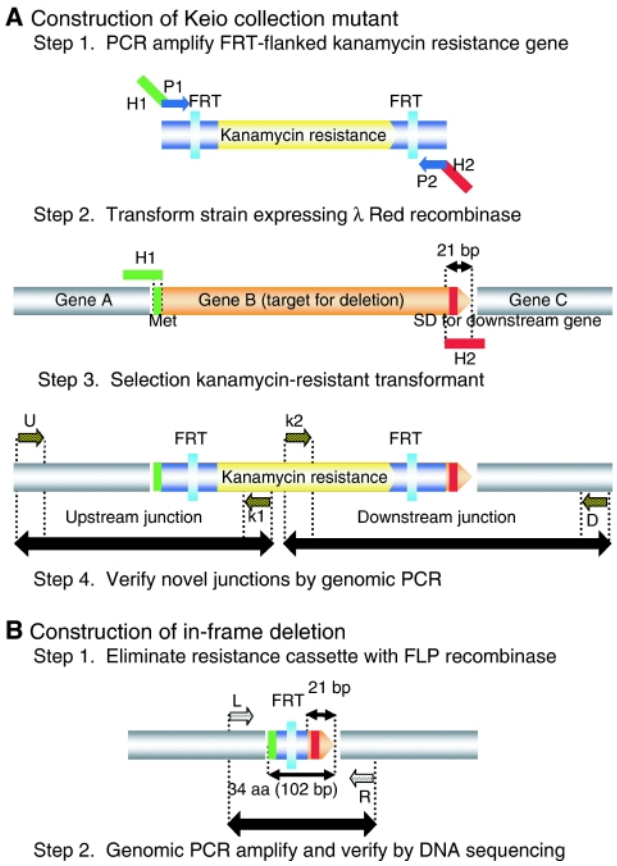


Figure B.1: Overall scheme of lambda red combination mediated knockouts. Reprinted under the CC BY 4.0 license.[8]

4. Make electrocompetent cells of strain to be manipulated wwith pKD46 - These must be supercompetent to uptake linear DNA.

5. Transform PCR product into strain to be manipulated with pKD46 electrocompetent cells and plate on Km (or other antibiotic depending on the antibiotic cassette amplified in the PCR step.

6. Restreak colonies to remove pKD46

7. Confirm presence of antibiotic cassette by PCR and sequencing using primers for the surrounding region to ascertain the cassette's location.

8. Make competent cell of the kanamycin resistant strain.

9. Transform by electroporation with pCP20 (or other plasmid combining the Flippase genes) and plate on Amp/Cb at 30 °C.

10. Restreak colonies on Cb and grow overnight at 30 °C.

11. Restreak a few colonies on LB and grow overnight at 42 °C. Enclose plate in the bag so as not to become desiccated.

12. Repeat.

13. Replica plate colonies on Cb, Km and LB plates. Grow overnight at 42 °C.

14. Pick colonies cured of Km and Cb resistance.

B.2.2 P1 phage transduction

Adapted from P1 Phage Lysogen - Making Phage on OpenWetWare[126].

P1 Phage Propagation

Making Phage Lysates Day 1

1. Streak out lysogenic strain on LB agar

2. Grow overnight at 30 °C.

Day 2

1. Grow lysogenic strains overnight in LB at 30 °C.

Day 3

1. Subculture 200 μL of overnight culture into 15 mL of LB in a 250 mL flask

2. Grow bacteria at 30 °C to an OD600 of $\sim$0.4

3. Move flasks to a $37 - 40$ °C shaking incubator (shaking water bath is best). (Phage will become lytic and lyse the cells)

4. After 30 min, monitor OD600 values every 20 minutes. When value stops increasing (or drops), the cells have started to lyse. Add several drops of chloroform (use glass pipette), swirl, and leave them on the bench for 10 minutes.

5. Pour the liquid into a 15 mL tube, balance all tubes, and spin in the centrifuge at 5,000 rpm for 20 minutes.

6. Pour the supernatant into a fresh 15 mL glass tube and place in the fridge.

P1 Phage Transduction

P1 Phage transduction protocol was adapated from Thomason et al. [129]. The plate lysate creation procedure was used for all infections, except the initial phage propagation. Colonies receiving the Km antibiotic marker were restreaked once on LB agar with Km and citrate then at least once on LB agar with Km before use to ensure no phage carryover. The Km antibiotic marker was removed as described in the λ red recombinase protocol.

Appendix C

COMPUTATIONAL METHODS

C.1 Flux Balance Analysis

Flux balance analysis using constraints-based reconstruction and analysis (COBRA) was performed for a metabolic network incoporating the formolase pathway[99, 15, 113, 38]. This technique employs a metabolic network as a set of constraints then optimizes an objective function, most commonly growth, to predict fluxes through each reaction at a steady state. The formolase pathway was added to the *E. coli* SBML model iAF1260 using COBRApy then biomass accumulation/growth rates were simulated using the following Python code[41, 38].

```
# -*- coding: utf-8 -*-
"""
Created on Tue Jul 15 12:20:43 2014

@author: Amanda
"""
#import modules
import csv
import numpy
import math
import random
import time
import os
import cobra.io
from cobra import Model, Reaction, Metabolite

directory = os.getcwd()
print "working directory is" + directory
```

```
# Load a cobra model from anywhere you like
# model = cobra.io.read_sbml_model('e-coli-core.xml')

#Sets solver
solver = 'glpk' #Change to 'gurobi' or 'cplex' if you have that solver
↪  installed instead.

#Specifies input and output sbml files
sbml_file = directory + '/' + 'iAF1260.xml'
sbml_out_file = directory + '/' + 'iAF1260_FLS.out.xml'
sbml_level = 2
sbml_version = 1 #Writing version 4 is not completely supported.

# Load a cobra model from anywhere you like
cobra_model = cobra.io.read_sbml_model(sbml_file)

cobra_model.optimize(solver=solver)  # 'gurobi' or 'cplex' are also
↪  supported
print cobra_model.solution.f

## Prints a bunch of information about the model
# Iterate through the the objects in the model
'''
for x in cobra_model.reactions:
    print(x.reaction)
for x in cobra_model.metabolites:
    print('%s has formula %s' % (x,x.formula))
for x in cobra_model.genes:
    print('gene %s is associated with reactions:' % x)
    for y in x.reactions:
        print('\t%s' % y)
'''

# Edits the cobra model
cobra_model1 = cobra_model

# Eliminates glucose influx
cobra_model1.reactions.EX_glc_LPAREN_e_RPAREN_.lower_bound = 0.0
```

```
# Formate uptake reaction already exists
# EX_for_LPAREN_e_RPAREN_
# for_e --> (no gene associations)
# gene b0904 (FocA) is associated with reactions:
#        FORt2pp
#        FORtppi

# Edits the lower bound for the formate uptake reaction to represent a
↪  higher external [formate]
cobra_model.reactions.EX_for_LPAREN_e_RPAREN_.lower_bound = -10.

################################################################################
# Adds the NAD-linked formate dehydrogenase reaction
# b3894 - FDH-O alpha subunit
# FDH4pp, FDH5pp
#>>> cobra_model.reactions.FDH4pp.reaction
#'for_p + 2 h_c + q8_c --> h_p + q8h2_c + co2_c'
#>>> cobra_model.reactions.FDH5pp.reaction
#'for_p + 2 h_c + mqn8_c --> h_p + co2_c + mql8_c'

reaction = Reaction('FDH_cm')
reaction.name = 'FDH from CM NAD-linked'
reaction.subsystem = 'Formolase pathway'
reaction.lower_bound = 0.  # This is the default
reaction.upper_bound = 1000.  # This is the default
reaction.reversibility = True  # This is the default
reaction.objective_coefficient = 0.  # This is the default

for_c = cobra_model1.metabolites.get_by_id('for_c')
co2_c = cobra_model1.metabolites.get_by_id('co2_c')
nad_c = cobra_model1.metabolites.get_by_id('nad_c')
nadh_c = cobra_model1.metabolites.get_by_id('nadh_c')
#h2o_c = cobra_model1.metabolites.get_by_id('h2o_c')
#h_c = cobra_model1.metabolites.get_by_id('h_c')

#add the metabolites to the reaction:
reaction.add_metabolites({for_c: -1.0, nad_c: -1.0, co2_c: 1.0, nadh_c:
↪  1.0})
```

```
#Print the reaction to make sure it worked:
print reaction.reaction

print '%i reactions in original model' % len(cobra_model.reactions)
#Add the reaction to the model
cobra_model1.add_reaction(reaction)
print '%i reaction in updated model' % len(cobra_model1.reactions)

################################################################################
# Adds a secondary reaction for ACS
# acs gene b4069
reaction1 = Reaction('FCS')
reaction1.name = 'ACS on formate'
reaction1.subsystem = 'Formolase pathway'
reaction1.lower_bound = 0.  # This is the default
reaction1.upper_bound = 1000.  # This is the default
reaction1.reversibility = 0.  # This is the default
reaction1.objective_coefficient = False  # This is the default
reaction1.add_gene_reaction_rule = 'b4069'

coa_c = cobra_model1.metabolites.get_by_id('coa_c')
atp_c = cobra_model1.metabolites.get_by_id('atp_c')
amp_c = cobra_model1.metabolites.get_by_id('amp_c')
ppi_c = cobra_model1.metabolites.get_by_id('ppi_c')

fcoa_c = Metabolite('fcoa_c', formula='C22H32N7O17P3S',
    name='Formyl-CoA Synthase', compartment='c')

# Add the metabolites to the reaction:
reaction1.add_metabolites({fcoa_c: 1.0, coa_c: -1.0, atp_c: -1.0, amp_c:
 ↪  1.0, for_c: -1.0, ppi_c: 1.0})

# Print the reaction to make sure it worked:
print reaction1.reaction

print '%i reactions in the original model' % len(cobra_model.reactions)
# Add the reaction to the model
cobra_model1.add_reaction(reaction1)
print '%i reactions in the updated model' % len(cobra_model1.reactions)
```

```
##############################################################################
# Adds the ACDH reaction on formate
reaction2 = Reaction('FCDH')
reaction2.name = 'ACDH from LM on formate'
reaction2.subsystem = 'Formolase pathway'
reaction2.lower_bound = 0.  # This is the default
reaction2.upper_bound = 1000.  # This is the default
reaction2.reversibility = True  # This is the default
reaction2.objective_coefficient = 0.  # This is the default

fald_c = cobra_model1.metabolites.get_by_id('fald_c')
h_c = cobra_model1.metabolites.get_by_id('h_c')

# Add the metabolites to the reaction:
reaction2.add_metabolites({fald_c: 1.0, nadh_c: -1.0, h_c: -1.0 , nad_c:
 ↪  1.0, fcoa_c: -1.0, coa_c: 1.0})

# Print the reaction to make sure it worked:
print reaction2.reaction

print '%i reactions in original model' % len(cobra_model.reactions)
# Add the reaction to the model
cobra_model1.add_reaction(reaction2)
print '%i reaction in updated model' % len(cobra_model1.reactions)

##############################################################################
# Adds the formolase reaction
reaction3 = Reaction('FLS')
reaction3.name = 'FLS Formose Reaction'
reaction3.subsystem = 'Formolase pathway'
reaction3.lower_bound = 0.  # This is the default
reaction3.upper_bound = 1000.  # This is the default
reaction3.reversibility = True  # This is the default
reaction3.objective_coefficient = 0.  # This is the default

dhap_c = cobra_model.metabolites.get_by_id('dhap_c')

# Add the metabolites to the reaction:
```

```
reaction3.add_metabolites({dhap_c: 1.0, fald_c: -3.0})

# Print the reaction to make sure it worked:
print reaction3.reaction

print '%i reactions in the original model' % len(cobra_model.reactions)
# Add the reaction to the model
cobra_model1.add_reaction(reaction3)
print '%i reactions in the updated model' % len(cobra_model1.reactions)

################################################################################
# Exclude the following attributes so the final json file will be smaller
EXCLUDE = ['notes', 'charge', 'compartment', 'annotation',
           'subsystem', 'reaction', 'objective_coefficient',
           'notes', 'genes', 'description']

# generate a json file to load into Escher
cobra.io.save_json_model(cobra_model, 'iAF1260.json',
↪  exclude_attributes=EXCLUDE)

# generate a json file to load into Escher
cobra.io.save_json_model(cobra_model1, 'iAF1260_FLS.json',
↪  exclude_attributes=EXCLUDE)

cobra.io.write_sbml_model(cobra_model1, sbml_out_file)
```

Growth rates of single gene deletion knockouts were compared on 10 mM glucose, 14 mM malate, 20 mM glycerol, 50 mM formate using COBRApy and the following code[38]. Results for the single gene deletions are given in Table C.1. The fact that this model predicts growth on glucose is concerning for known glucose-essential genes, however, this approach is useful to provide predictions especially for C4 growth, where data is limited.

```
# -*- coding: utf-8 -*-
"""
Created on Tue Jul 15 11:21:20 2014
```

```
@author: Amanda
"""
#import modules
import csv
import numpy
import math
import random
import time
import os
from cobra.io.sbml import create_cobra_model_from_sbml_file
from cobra.io.sbml import write_cobra_model_to_sbml_file
from cobra import Model, Reaction, Metabolite
from cobra.test import create_test_model, salmonella_pickle, ecoli_pickle
from cobra.flux_analysis import single_deletion
import pandas

#----------------------
# Sets the directory
#----------------------

directory = os.getcwd()
print "working directory is" + directory

#----------------------
# SCRIPT BODY
#----------------------

#Sets solver
solver = 'glpk' #Change to 'gurobi' or 'cplex' if you have that solver
↪  installed instead.

#Specifies input and output sbml files
sbml_file_FLS = directory + '/' + 'iAF1260_FLS.xml'
sbml_out_file_FLS = directory + '/' + 'iAF1260_FLS.out.xml'
sbml_level = 2
sbml_version = 1 #Writing version 4 is not completely supported.

sbml_file = directory + '/' + 'iAF1260.xml'
sbml_out_file = directory + '/' + 'iAF1260.out.xml'
```

```
cobra_model_FLS = cobra.io.read_sbml_model(sbml_file_FLS)
cobra_model = cobra.io.read_sbml_model(sbml_file)

#Removes glucose input
cobra_model_FLS.reactions.EX_glc_LPAREN_e_RPAREN_.lower_bound = 0.0
cobra_model.reactions.EX_glc_LPAREN_e_RPAREN_.lower_bound = 0.0


formate_conc_Start = range(-10,-50,-10)
for x in formate_conc_Start:
    cobra_model.reactions.EX_for_LPAREN_e_RPAREN_.lower_bound = x
    cobra_model.optimize(solver=solver)  # 'gurobi' or 'cplex' are also
    ↪ supported
    print ('%.2f growth rate on %.2f mM formate' %(cobra_model.solution.f,
    ↪ x))


for x in formate_conc_Start:
    cobra_model_FLS.reactions.EX_for_LPAREN_e_RPAREN_.lower_bound = x
    cobra_model_FLS.optimize(solver=solver)  # 'gurobi' or 'cplex' are also
    ↪ supported
    print ('%.2f growth rate on %.2f mM formate'
    ↪ %(cobra_model_FLS.solution.f, x))

# Perform all single gene deletions on a model
# b0115/aceF, b3737/atpE, b0116/lpd, b3916/pfkA, b3956/ppc, b2415/ptsH,
↪ b3770/ilvE, b3833/ubiE, b0774/bioA, b1779/gapA, b2779/eno
GR_for, Opt_for = cobra.flux_analysis.single_deletion(cobra_model,
↪ element_list=['b0115', 'b3737','b0116', 'b3916', 'b3956', 'b2415',
↪ 'b3770', 'b3833', 'b0774', 'b1779', 'b2779' ], method='fba',
↪ element_type='gene', solver='glpk')
GR_FLS_for, Opt_FLS_for =
↪ cobra.flux_analysis.single_deletion(cobra_model_FLS,
↪ element_list=['b0115', 'b3737','b0116', 'b3916', 'b3956', 'b2415',
↪ 'b3770', 'b3833', 'b0774', 'b1779', 'b2779' ], method='fba',
↪ element_type='gene', solver='glpk')
```

```
#Returns both models to glucose input
cobra_model_FLS.reactions.EX_for_LPAREN_e_RPAREN_.lower_bound = 0.0
cobra_model.reactions.EX_for_LPAREN_e_RPAREN_.lower_bound = -0.0
cobra_model_FLS.reactions.EX_glc_LPAREN_e_RPAREN_.lower_bound = -10.0
cobra_model.reactions.EX_glc_LPAREN_e_RPAREN_.lower_bound = -10.0

# Perform all single gene deletions on a model
# b0115/aceF, b3737/atpE, b0116/lpd, b3916/pfkA, b3956/ppc, b2415/ptsH,
 ↪  b3770/ilvE, b3833/ubiE, b0774/bioA, b1779/gapA, b2779/eno
GR_glu, Opt_glu = cobra.flux_analysis.single_deletion(cobra_model,
 ↪  element_list=['b0115', 'b3737','b0116', 'b3916', 'b3956', 'b2415',
 ↪  'b3770', 'b3833', 'b0774', 'b1779', 'b2779' ], method='fba',
 ↪  element_type='gene', solver='glpk')
GR_FLS_glu, Opt_FLS_glu =
 ↪  cobra.flux_analysis.single_deletion(cobra_model_FLS,
 ↪  element_list=['b0115', 'b3737','b0116', 'b3916', 'b3956', 'b2415',
 ↪  'b3770', 'b3833', 'b0774', 'b1779', 'b2779' ], method='fba',
 ↪  element_type='gene', solver='glpk')

#Returns both models to glycerol
cobra_model_FLS.reactions.EX_glc_LPAREN_e_RPAREN_.lower_bound = 0.0
cobra_model.reactions.EX_glc_LPAREN_e_RPAREN_.lower_bound = -0.0
cobra_model_FLS.reactions.EX_glyc_LPAREN_e_RPAREN_.lower_bound = -20.0
cobra_model.reactions.EX_glyc_LPAREN_e_RPAREN_.lower_bound = -20.0

# Perform all single gene deletions on a model
# b0115/aceF, b3737/atpE, b0116/lpd, b3916/pfkA, b3956/ppc, b2415/ptsH,
 ↪  b3770/ilvE, b3833/ubiE, b0774/bioA, b1779/gapA, b2779/eno
GR_glyc, Opt_glyc = cobra.flux_analysis.single_deletion(cobra_model,
 ↪  element_list=['b0115', 'b3737','b0116', 'b3916', 'b3956', 'b2415',
 ↪  'b3770', 'b3833', 'b0774', 'b1779', 'b2779' ], method='fba',
 ↪  element_type='gene', solver='glpk')
GR_FLS_glyc, Opt_FLS_glyc =
 ↪  cobra.flux_analysis.single_deletion(cobra_model_FLS,
 ↪  element_list=['b0115', 'b3737','b0116', 'b3916', 'b3956', 'b2415',
 ↪  'b3770', 'b3833', 'b0774', 'b1779', 'b2779' ], method='fba',
 ↪  element_type='gene', solver='glpk')
```

```
#Returns both models to glycerol
cobra_model_FLS.reactions.EX_glyc_LPAREN_e_RPAREN_.lower_bound = 0.0
cobra_model.reactions.EX_glyc_LPAREN_e_RPAREN_.lower_bound = -0.0
cobra_model_FLS.reactions.EX_mal_D_LPAREN_e_RPAREN_.lower_bound = -14.0
cobra_model.reactions.EX_mal_D_LPAREN_e_RPAREN_.lower_bound = -14.0

# Perform all single gene deletions on a model
# b0115/aceF, b3737/atpE, b0116/lpd, b3916/pfkA, b3956/ppc, b2415/ptsH,
 ↪  b3770/ilvE, b3833/ubiE, b0774/bioA, b1779/gapA, b2779/eno
GR_mal_D, Opt_mal_D = cobra.flux_analysis.single_deletion(cobra_model,
 ↪  element_list=['b0115', 'b3737','b0116', 'b3916', 'b3956', 'b2415',
 ↪  'b3770', 'b3833', 'b0774', 'b1779', 'b2779' ], method='fba',
 ↪  element_type='gene', solver='glpk')
GR_FLS_mal_D, Opt_FLS_mal_D =
 ↪  cobra.flux_analysis.single_deletion(cobra_model_FLS,
 ↪  element_list=['b0115', 'b3737','b0116', 'b3916', 'b3956', 'b2415',
 ↪  'b3770', 'b3833', 'b0774', 'b1779', 'b2779' ], method='fba',
 ↪  element_type='gene', solver='glpk')
```

	lpd	*ubiH*	*ppc*	*aceF*	*bioA*	*ptsH*	*atpE*	*ilvE*	*pfkA*	*eno*	*gapA*
	b0116	b3833	b3856	b0115	b0774	b2415	b3737	b3770	b3916	b2779	b1779
Glucose											
iAF1260	0.825	0.886	0.880	0.852	0.886	0.863	0.318	0.000	0.886	0.712	0.609
iAF1260 + FLS	0.877	0.886	0.880	0.886	0.886	0.863	0.318	0.000	0.886	0.712	0.609
Formate											
iAF1260	0.040	0.040	0.000	0.040	0.040	0.040	0.000	0.000	0.040	0.000	0.000
iAF1260 + FLS	0.240	0.240	0.238	0.240	0.240	0.240	0.000	0.000	0.240	0.223	0.219
Glycerol											
iAF1260	0.890	0.939	0.897	0.897	0.939	0.939	0.318	0.000	0.939	0.739	0.714
iAF1260 + FLS	0.939	0.939	0.927	0.939	0.939	0.939	0.318	0.000	0.939	0.742	0.717
Malate											
iAF1260	0.529	0.565	0.565	0.547	0.565	0.565	0.053	0.000	0.565	0.554	0.000
iAF1260 + FLS	0.560	0.578	0.576	0.578	0.578	0.578	0.056	0.000	0.578	0.574	0.571

Table C.1: COBRApy growth rate simulations for single gene knockouts considered for use in growth selections. Growth rate units are [hr^{-1}]. Genes in red represent those labeled as essential by Weaver et al.[141].

Appendix D

INDIVIDUAL STEP ASSAYS

D.1 Cell Preparation

E. coli cells were transformed with appropriate constructs. Overnight LB cultures of each strain were inoculated with single colonies from a LB plate containing the necessary antibiotics. The overnight cultures were used to inoculate 50 mL TB with the same antibiotics. Growth at 37 °C, 225 rpm was conducted to mid-log phase (OD600 ∼0.6) when cells were induced using 0.5 mM IPTG final concentration when necessary and transferred to 18 °C (225 rpm). After 24 hours at 18 °C, samples of 30 μL were taken to test protein expression via SDS-PAGE, then cell pellets were harvested by centrifugation (Thermo Scientific), 5,000 rpm and 20 min at 4 °C.

D.2 Formate Transporter

Presence of a growth defect on a toxic analogue of formate, hypophosphite, is used to determine function of the formate transporter. Anaerobically, wild-type cells should natively express the transporter causing a growth defect. If the same growth defect is seen aerobically, the formate transporter is concluded to be functional. Cell pellets were washed with 0.9% NaCl. After washing, pellets were resuspended in M9 minimal media, pH 6.5 with 0 and 80 mM sodium hypophosphite and 2 mM potassium nitrate to a equal density across the set of cultures, as close to an approximate 1 OD600. Resuspended cultures were transferred to anaerobic vials. Aerobic cultures were transferred to green cap tubes instead or capped

with autoclaved foil. Anaerobic control tubes were purged for 2 min with nitrogen gas using a filtering needle. Positive pressure was accomplished by removing the outlet filter for 15 second before cutting off the outlet flow. Optical density was monitored with a tube spec starting at time zero.

D.3 Formate Dehydrogenase

Beginning with cell extracts as described in the Pathway Function Assay setup, formate dehydrogenase activity was assessed using NADH production from formate measured at A340 nm. Assay reactions contained 150 mM postassium phosphate buffer, pH 7.5, 0.1 mM DTT, 2 mM ATP, 1 mM NADH, 50 mM formate, 0.2 mM TPP and 2 mM $MgSO_4$.

D.4 ACS and ACDH

D.4.1 NAD-linked

A coupled NAD-linked assay confirms ACS-ACDH function. Cell extracts, ultra-centrifuged and desalted, were combined with 150 mM potassium phosphate buffer, pH 7.5, 0.2 mM CoA, 1 mM NADH, 2 mM ATP, 0.1 mM DTT, 2 mM $MgSO_4$, 0.2 mM TPP and 50 mM formate/acetate. NAD consumption was measured via change in absorbance at 340 nm. The NAD-linked coupled assay was used to fine tune assay conditions for the in vitro pathway functon assay as it can be done at higher throughput.

D.4.2 Nash assay for formaldehyde production

Nash reagent was prepared by dissolving 7.5 g of ammonium acetate, 0.150 mL of acetic acid, and 0.100 mL of 2,4-pentanedione/acetyl acetone in water to make 50.0 mL of reagent solution. 150 μL of cell culture was centrifuged at high speed for 2 min. 125 μL of supernatant

was removed and combined with 125 μL Nash reagent. Samples were briefly vortexed and heated at 65 °C for 10 min. 200 μL was transferred to a 96 well plate and absorbance was measured at 410 nm.

D.5 FLS

Cell extracts with FLS are tested as described in the Pathway Function Assay setup with ^{13}C formaldehyde as a substrate. Enrichment for M+3 product detection confirms function.

D.6 SDS-PAGE

Acrylamide gel (7.5%) with a Tris-Glycine buffer system were used for SDS-PAGE with the Bio-Rad Mini-Protean system [71]. Pageruler Prestained Protein Ladder (Thermo Scientific) was used as a protein standard. Three part protein sample, either lysate or cell culture, was added to one part 4x protein sample buffer. 4x sample buffer made or 4x Laemmli sample buffer (Bio-Rad): 40% Glycerol, 240 mM Tris/HCl pH 6.8, 8% SDS, 0.04% bromophenol blue, 5% β-mercaptoethanol (added right before use)[84]. Samples were boiled for 6 min at 100 °C. GelCode Blue Safe Protein Stain (Pierce Item #1860957, Item# 24594) or a Coommassie blue stain were used to stain the gels.

Appendix E

PATHWAY FUNCTION ASSAYS

E.1 Cell Preparation

ALA2 *fdoG* cells were made chemically competent via CCMB and transformed with appropriate constructs (Appendix B). Overnight LB cultures of each strain were inoculated with single colonies from a LB plate containing the necessary antibiotics. The overnight cultures were used to inoculate 50 mL TB with 100 mg/ml carbenicillin and 100 mg/ml Kan final concentration. Growth at 37 °C, 225 rpm was conducted to mid-log phase (OD600 ∼0.6) when cells were induced using 0.5 mM IPTG final concentration and transferred to 18 °C (225 rpm). After 24 hours at 18 °C, samples of 30 μL were taken to test protein expression via SDS-PAGE, then cell pellets were harvested by centrifugation (Thermo Scientific), 5,000 rpm and 20 min at 4 °C, flash frozen with liquid nitrogen and stored at −80 °C until use or transferred to new media for labelling experiments.

E.2 in vitro Extract Preparation

Cell pellets were resuspended in 3 mL protein buffer (10 mM Potassium phosphate, pH 8.0, 1 mM $MgSO_4$, 0.1 mM TPP and 0.025 mM DTT). Cell lysates were prepared by two passages through a French Press (SIM-Aminco, Spectronic Instruments) with mini-cell (1000 psig). Ultracentrifugation (Beckman) was conducted at 4 °C and 45,000 rpm (Beckman MLS 50) for 1 hour. High-speed extracts (2.5 mL each) were desalted with PD-10 columns (GE Healthcare) eluting in 3.5 mL of the protein buffer to get a clarified cell lysate. NADH

oxidation rates of the clarified cell lysates (CCLs) were measured using a variant of the coupled ACS-ACDH activity assay with assay mixture described in the Assay reaction section below, with 1 mM NADH replacing the NAD. NAD reduction rates by a 10 mg/mL solution of FDH from *Candida boidinii* were also measured in the same assay mixture with the NAD instead of the NADH and no HSCoA using that same method as the ACDH assay. The volume of FDH solution to add per volume of CCL was determined by calculating the average NADH oxidation rate of the full pathway CCLs, then finding amount of FDH needed to balance it out in a 1:1 ratio. This volume ratio was applied to all CCLs to give CCL + FDH 1:1 mixtures, which served as the final CCL preparations.

E.3 in vitro Assay Reaction

The following components were combined in an assay mixture with the given final concentrations, vortexed to mix and incubated for various lengths of time: 150 mM Potassium phosphate, 0.2 mM TPP, 2 mM $MgSO_4$, 11 mM ATP, 0.5 mM NAD, 0.2 mM HSCoA, 50 mM ^{13}C sodium formate (Cambridge Isotope Labs), 0.1 mM DTT and 50% (v/v) CCL + FDH 1:1 preparations. Every eight hours, additional ATP, ^{13}C formate and DTT were added to boost concentrations by 2.5 mM, 10 mM and 10 μM, respectively. At each time point (eight hour intervals), 150 μL of reaction was flash frozen in liquid nitrogen and stored at $-80\,^{\circ}$C.

E.4 Organic Extraction for in vitro Assay Reactions - Acetonitrile Quench

Cell extract assay reaction samples were quenched 3:1 (v/v) with acetonitrile mixed with pivalate 50 μM final concentration. Samples were centrifuged 14,000 rpm for 10 min at 4 °C to remove the precipitated protein. Supernatants was transferred to a Millipore hydrophilic

or phobic filter plate. Samples were shaken with a plate shaker for 5min at 1000 rpm, then centrifuge at 1,500xg, 4 °C for 2 min to filter samples. Plates were covered with plate seals and placed in autosampler at 10 °C for analysis with the ion pairing method.

E.5 Organic Extraction for in vitro Assay Reactions - Fast Centrifugation

The fast centrifugation method from Kleijn et al. was adapted for use with cell extracts[66]. Timepoint samples were thawed on ice. Hot buffered ethanol (Decon Labs) was added to timepoint samples to a final concentration of 60% (v/v). Internal standard, glutaric acid, 7.5 nmol, was added immediately following addition of the hot buffered ethanol. Samples were then heated for 1 min at 78 °C with occasional mixing, placed on ice and centrifuged at 14,000 rpm for 10 min at 4 °C. Supernatant was transferred to new Eppendorf tubes and dried to completion in a Labconco CentriVap with cold trap connected to a Savant Gel pump (GP110). Samples were stored at −80 °C until within a day of LC-MS/MS analysis, when they were re-suspended in 100 μL of Millipore water and vortexed to mix.

E.6 in vivo Labeling

E.6.1 from Formate

Cell Preparation protocol from above was followed. The cells were harvested by centrifugation and resuspended in M9 minimal media with ^{13}C formate with or without a co-substrate. These cultures were returned to 18 °C, 225 rpm shaker until harvesting for in vivo metabolite or hydrolyzed protein analysis.

E.6.2 from Methanol

Adapted from Müller et al. [90]. Cell pellets were prepared as described above except the media used for expression was a M9 with glucose (0.4%). After 24 hours of expression, the pellet was harvested by centrifugation then washed twice with sterile water. After the second wash, the supernatant was pipetted off to retain the cells. The cell pellet was resuspended in a volume of standard M9 (no carbon) with appropriate antibiotics equal to the starting cell culture volume, an approximate OD600 of 1, was added to a final concentration of 0.2 M. Nash assay (Appendix C) was used to measure formaldehyde production. OD at 600 nm was measured using the nanodrop (Thermo Scientific). At least three technical replicates were measured for OD. For in vivo samples, 10 mL of each eculture were harvested and prepped with the hot water extraction method for HILIC sugar phosphate detection.

E.7 Metabolite Analysis- Sample Preparation

E.7.1 Hot water extraction for in vitro metabolite analysis

Day 1: Hot Water Extraction

1. Preheat hot water baths to 100 °C (∼1 hr before using). Add additional DI water so it will cover the most of the tubes.
2. Put the clear plastic tube rack with holes for tube holders in the water bath.
3. Boil filtered sterile water in the microwave (3 min).
4. Put boiling water in boiling water bath to keep hot while doing the next step
5. Add 0.45 mL (3x sample volume of 0.15 mL) boiling water to each sample, cap tightly and vortex to mix.
6. Put tubes in boiling water for 1 min. Make sure the water level in the bath is above the liquid level in the tubes.
7. Put the tubes on ice for 10 min.
8. Pre-cool centrifuge.
9. Vortex tubes for 20 sec each.

10. Centrifuge 5 min, 14,000 rpm, 4 °C
11. Pipet supernatant into fresh tubes. Clarfied lysate samples will have little to no pellet.
12. Flash freeze clean supernatants with liquid Nitrogen. Pause Point: Samples can be stored at −80 °C.
13. Open tubes and cover with parafilm. Puncture holes in parafilm. Put tubes in foam or cardboard holder.
14. Lyophilize until dry (∼overnight).

Day 2: Reconstitute #1

1. Reconstitute in 100 μL sterile, filtered ddH_2O.
2. Centrifuge tubes, max speed, 4 °C, for x 10 minutes
3. Load supernatant into spin column filter (Costar 8161 Spin-X Centrifuge Tube Filter, 0.22μm cellulose acetate in a 2.0 mL polypropylene tube. non-sterile, 100/case)
4. Centrifuge spin column filters for 2 min 14,000 rpm
5. Pipet sample into vial insert in MS vials with split lids
6. Sample is ready for LC-injection.

E.7.2 Hot water extraction for in vivo metabolite analysis

Day 1: Harvest Cells

1. For growing cells, 1-2 doublings suggested, 0.6∼0.8 OD600, harvested 25 mL. For my barely utilizing cells that have undergone 24 hours of expression at 18 °C, harvested 10 mL. Harvest volumes may need adjustment for culture density to prevent clogging the filter.
2. Label 50 mL falcon tubes
3. Setup Filtration Equipment: Filtration setup with vacuum flask (see pictures below), Filters (Nylaflow, Nylon Membrane Filter 0.2 μm, 90 mm, P/N 66603 [Pall Life Sciences]), flat-tipped tweezers and Dewar with liquid Nitrogen
4. Put a new filter in the setup
5. Start vacuum
6. Pour ∼25 mL sterile filtered water through the filter

Figure E.1: Filtration setup.

7. Pipet samples onto the center of the filter - culture sample volumes described above. (It should take ∼30 sec to filter each sample.)
8. Once sample is filtered. Expose filter and fold into quarters with the tweezers.
9. Put folded filter into a clean, labeled 50 mL Falcon tube
10. Close tube and flash freeze in liquid Nitrogen.
11. Repeat for each sample. Pause Point: Samples can be stored at −80 °C at this point.
12. Replace caps on frozen tubes with ones that have holes punched in them.
13. Lyophilize for 8 hrs or overnight (this can be cut shorter if needed, but we do longer just to be safe) [Labconco FreeZone 4.5 −105 °C], vacuum -.008 mbar, temperature −108 °C

Day 2: Hot Water Extraction

1. Preheat hot water baths to 100 °C (∼ 1 hr before using). Add additional DI water so it will cover the most of the tubes.
2. Put a 9-50 ml tube rack in each water bath
3. Boil filtered sterile water in the microwave (3 min).
4. Put boiling water in boiling water bath to keep hot while doing the next step
5. Add 25 mL boiling water to each sample, cap tightly and vortex to mix.
6. Put tubes in boiling water for 20 min. Make sure the water level in the bath is above the liquid level in the tubes.
7. Put the tubes on ice for 40 min. Use a deep ice bucket.
8. Pre-cool centrifuge.
9. Vortex tubes for 1 min each.
10. Remove the filter from each tube. Try to keep all liquid in the tube. Use the pipette to remove pockets that are stuck in the folds.
11. Centrifuge 20 min, 5,000 rpm, 4 °C
12. Decant into fresh tubes.
13. The remaining tubes containing protein and cell debris were centrifuged again, at 4 °C, 5000 rpm for 20 minutes, and supernatants were removed carefully into the clean tubes.
14. Flash freeze clean supernatants with liquid Nitrogen. Pause Point: Samples can be stored at −80 °C.

Figure E.2: Removing filter with cells.

15. Replace caps with caps with holes in the lids.
16. Lyophilize for $\sim$ 24-36 hours until dry.

Day 3: Reconstitute #1

1. Replace caps with ones removed before lyophilization.
2. Reconstitute sample in 1 mL sterile, filtered ddH_2O
3. Vortex to mix and short spin to collect liquid
4. Transfer to 1.7 mL Eppendorf tubes
5. Flash freeze with liquid Nitrogen
6. Open tubes and cover with parafilm. Puncture holes in parafilm. Put tubes in foam holder
7. Lyophilize tubes for until dry ($\sim$ overnight)

Day 4: Reconstitute #2

1. Reconstitute in 100 μL sterile, filtered ddH_2O.
2. Centrifuge tubes, max speed, 4 °C, for x 10 minutes
3. Load supernatant into spin column filter (Costar 8161 Spin-X Centrifuge Tube Filter, 0.22 μm cellulose acetate in a 2.0 mL polypropylene tube. non-sterile, 100/case)
4. Centrifuge spin column filters for 2 min 14,000 rpm , 4 °C
5. Pipet sample into vial insert in MS vials with split lids
6. Sample is ready for LC-injection!

E.7.3 Sample Prep for Amino Acid Analysis from Protein

Cell Prep

1. Grow 1.5 mL of OD550 0.8 - 1.5 ($\sim$0.3 - 0.8 mg) of cell culture
2. Swirl 30-40 seconds in ethanol dry ice bath or liquid N_2 (Don't freeze sample with liquid Nitrogen)
3. Centrifuge 5 min at 13,500 g, 4 °C
4. Discard supernatant

5. Wash pellet with 0.9% 1 mL of sodium chloride at 4 °C, pellet for 5 min at 13,500 x g, 4 °C

6. Repeat previous step. Optional: store pellet at −40 °C to −80 °C

Hydrolysis of Amino Acids

1. Turn on the heating block and equilibrate it to 105 °C - 110 °C Note: Equilibrate the temperature with the hood at the level you intend to leave it at overnight. If you equilibrate it with the hood open, then close it, the temperature will rise due to reduced convection. Don't exceed 110 °C, this can be bad.

2. All steps in this section should be done in the fume hood across from the GC:

3. Suspend cell pellets in 1 mL of 6N HCl.

4. Transfer resuspended cells into 2 mL GC-MS autosampler vials.

5. Seal tubes with screw caps to prevent evaporation of HCl.

6. Bake the well-sealed 2 mL tubes for 24 hours in a heating block set to 105 °C - 110 °C (no hotter than 110 °C) Note: you can pause at this point by storing the samples at −20 °C for 1-2 days

7. Dry the hydrolysate at 95 °C with constant air flow (or N_2) gas flow in the fume hood until the sample is completely dry.

8. Use the "nitrogen tree" to dry the samples: Turn heat on high. Set pressure regulator on tank to 2-4 psi. Flow gauge should read 8 L/min for two samples. Clean capillary tips with ethanol, unscrew white plastic, move metal shaft down (may need to wipe with ethanol to allow this), insert capillary tip into glass sample vial (close to liquid but not touching), screw plastic threading back to lock the metal shaft in place. Leave heating block on. Move sample vial up hourly as liquid evaporates.

9. Bake the dried samples at 105 °C for another 10 minutes to ensure no moisture is left. Note: you can pause at this point by storing the samples at −20 °C. Use new septum for cap if you do.

Derivatization of Amino Acids with TBDMS

1. Remove Ash

   (a) Add 100 μL nanopure water to reconstitute the dried sample (if using 1 mL of OD600 = 0.6 culture)

   (b) Add the volume of water to the vial with the dried sample, pipet to mix, transfer entire volume to a eppendorf tube and vortex. (This is enough for technical replicates.)

   (c) Centrifuge at 13,500 x g for another 1 - 2 minutes to remove ash and collect supernatant. Use a filter centrifuge tube if sampling a large fraction of this volume. If using a very small volume ($\sim$10 μL) you can pellet without a filter and sample from the top of the supernatant. Centrifuge longer (10 min) if using this approach.

2. Add Internal Standard and Dry

   (a) Transfer aliquot into clean GC-MS vial with insert. Use 20 μL (can halve).

   (b) Add 40 μL of internal standard mix to 20 μL of sample. Internal standard is $^{13}C^{15}N$ serine: has mass shift M+4. The internal standard mixture contains $^{13}C^{15}N$ Serine, Alanine, Glycine and Glutamine, equal parts. Ten μL each compound, so 40 μL of mix total is added.

   (c) Repeat drying step with nitrogen tree as done above or speed-vac. If using speed-vac 35 °C for 1-2 hours until dry. Hold vials in 15 mL tubes. Program 9. Check after 1 hr. Appox. Every 30 min after. Put vials in empty 15 ml centrifuge tubes. Balance rotor. Close lid. Press start. Stay with it to make sure it's actually going. Pump will be really noisy at first then will quiet.

   (d) Check speed vac every 20 min to make sure it's still running.

   (e) Optional freeze −20 °C to pause

   (f) Turn on heater, 65 °C.

3. Derivatize

   (a) Prepare pyridine:

      i. Add one layer of molecular sieve to scintillation vial.

      ii. Add 1 mL pyridine per sample.

      iii. Let sit for 5 min. No agitation.

   (b) For each sample:

i. Add 20 μL of molecular-sieve treated pyridine to the dried sample using a syringe(would dissolve pipette tip). The solvent may turn slightly brown,

ii. Add another 20 μL of Trifluoromethanesulfonic acid tert-butyldimethylsilyl ester (TBDMS) and seal well. Use the same blue screw-cap but replace the septum.

iii. Rinse syringe in leftover pyridine, 10 times.

(c) Repeat for all samples,

(d) Incubate for 1 hour at 65 °C). Turn heater off after use.

E.8 Metabolite Analysis

Mass spec signals for glycolytic intermediates were integrated for all isotopomers using LCQuan (Thermo Scientific), Xcalibur(Thermo Scientific) or MassLynx (Waters) software.

E.8.1 GC-MS for Amino Acid Detection

- GC-MS: Agilent 5975 GCMS
- Column: Agilent 19091S-433: 93.92873, HP-5MS 5% Phenyl Methyl Silox, 325 °C: 30 m x 250 μm x 0.25 μm
- Carrier gas: He
- Injection: Splitless injection
- Needle wash: Hexanes

Instructions for Use

1. Check that there is enough Carrier gas. Want at least 40-50 psi on the He tank.
2. Check that there is enough hexanes for needle washing (Two vials should be in the A&B positions on the injection table). Refill if near the min line.
3. Open 579C/Enhanced to control machine. There is an icon on the desktop that may help the GC-MS talk to the computer if they aren't already.

4. Load method (name \ŌRGACD-TBDMS.M) by clicking on folder. Check that the settings are right by comparing to an old sample method file. To access an old sample file, find a sample folder such as D:\Users\Amanda_Janet \2012_11_19-1st_13Cformate_test\5+48.D. Open the file acqmeth.txt

5. Click on thermometer and GC coil icon. Go through the tabs and confirm the settings.

    - ALS: Nothing
    - Valves: Do not set anything
    - Inlets: Septum purge flow: wash out that volume
    - Columns: flow is constant at 1 mL/min. Since temperature increases and He becomes more viscous at higher temps, the pressure ramps up, too. It is held high at the end to bake off anything that stuck.
    - Aux heater: don't want matter to condense before entering mass spec
    - Events: Counts number of injections. Helps you decide when to replace septum.
    - Solvent delay: most of the sample is solvent. Let it run through.
    - Gain Factor: Set to 1.00
    - 1.4 scans/second
    - Time window: how big the white window in the back looks
    - Zones: shows temps in mass spec. Don't want condensation
    - HiVac gauge: sensor broke over weekend.
    - Turbo Speed $\bar{1}00$ is desired.
    - HE output: 40 psi, 700 psi, worry at 2000 psi

6. File –¿ load method: chose default.M (prevents others from accidentally changing our method.)

7. Load samples into autosampler rack.

8. Create file for sample run: will include the vial position and other useful info. Note: you can have it call your method from here. This will reduce the probability that other people will edit your method.

9. Open an existing file (optional)

10. Update your sample names (no slashes!), vial position number, and copy the sample names into the data file names column.

11. Make sure the method for all rows is set to ours.

12. Make a new folder for where you want to store data (button at top left).

13. Set method path at top of the sequence list.

14. Make sure you have an air blank set as the 1st sample with an empty GC/MS vial.

15. Can set the sample type to blank or sample because not using their software.

16. Check that there is enough memory by doing a sequence verification (square button: check mark)

17. Start by pressing run.

Materials and Reagents

- TBDMS (Sigma #: 375934)
- Serine internal standard ($^{13}C^{15}N$ serine. mass shift $= M + 4$.)
- EMD Molecular Sieve Type 3A 8-12 (MX1583D-1)
- Pyridine Sigma ACS Reagent (360570-500mL)
- Agilent Screw Caps with Red PTFE/white silicone Septa 100 pk (5182-723)
- Agilent Extra Septa Red PTFE/white silicone 500 pk (5182-730)
- Agilent Vial Glass Small Volume Inserts (5183-2085)
- Agilent Glass Vial 2 mL with write on spot (5182-0715)
- Costar spinX Centrifuge tube filter, 0.22 μm cellulose acetate in 2.0 mL polypropylene tube sterile, RNAse, DNAse free. 24/pack, 96/case, 8160.

E.8.2 GC-MS for Dihydroxyacetone and Glycolaldehyde Detection

From Sean Poust

1. Add 20 μL 200 mM O-(2,3,4,5,6-Pentafluorobenzyl)hydroxylamine hydrochloride (PFBOA; It takes a few minutes of vortexing to dissolve.) to lysate sample, react for 1 hr at room temperature.

2. Add 20 μL 3M H_2SO_4, check pH is below 5 with pH paper.

3. Centrifuge sample for 2 min at max rpm to remove insoluble material, transfer 140 μL of supernatant to a new tube .

4. Add 80 μL ethyl acetate and vortex.

5. Allow phase separation for 5 minutes.

6. Extract 20 μL ethyl acetate (organic top layer) into GC vial inserts.

7. Add 5 μL N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) under the fume hood.

8. React at room temperature for 5 minutes.

Follow GC-MS setup instructions as above, but use the method described below.

- Program
  - Start 50 °C, hold 1 min
  - Ramp 15 °C per min to 150 °C
  - Ramp 30 °C per min to 300 °C hold 1 min
  - Total 13.67 min
- Injection: 1 μL injection (splitless)
- Inlet temperature: 220 °C
- Flow: 1 mL/min
- Solvent Delay: 8 min
- Scan: 60-500 m/z

- Columns:
  - DB-5 column (Berkeley)
  - Agilent 19091S-433: 93.92873, HP-5MS 5% Phenyl Methyl Silox, 325 °C: 30 m x 250 μm x 0.25 μm (UW)
- Ionization: EI

E.8.3 LC-MS/MS Ion-pairing method

LC-MS/MS analysis was adapted from Buescher et al. [24].

- Column: Water Acquity UPLC HSS T3 1.8 μm 2.1 x 150 mm column
- Mass Spec: Thermo Scientific TSQ Quantum Access
- Mobile Phase A: 10 mM tributylamine, 15 mM acetic acid, 5% (v/v) methanol
- Mobile Phase B: 100% isopropanol

Time [min]	**%A**	**uL/min**
0	100	400
5	100	400
10	98	400
11	91	350
16	91	250
18	75	250
19	50	150
25	50	150
26	100	150
32	100	400
37	100	400

Table E.1: Mobile phase gradients for the ion pairing LC-MS/MS metabolite detection method.

- Scan Width (m/z): 0.01
- Scan Time (s): 0.1

- Mode: Negative mode
- Centroid
- Collision Gas Pressure: 1.5 mTorr
- Spray Voltage: 4000V
- Capillary Temp °C: 350
- Capillary Offset (V): -35
- Sheath Gas Pressure (arbitrary units): 40
- Aux Gas Pressure (arbitrary units): 20

Autosampler vial inserts were loaded with 35 μL of each sample. An extra minute of equilibration time was added to the end of each run. Transitions used for SRM are below. Adjustments to the parent and product ion masses were done to monitor the M+x channels for each targeted compound, where x is the number of carbons in the compound backbone. Internal standard curves made with ^{12}C compounds were used for quantification. Isotope correction for natural abundance was done using the Isocor software[87].

Compound	**Parent Ion**	**Product Ion**	**Tube Lens**	**Collision Energy**
2/3-Phosphoglyceric acid	185.013	79.204	36	134
Dihydroxyacetone phosphate	169.004	97.120	12	46
Glutaric Acid	131.084	87.247	15	146

Table E.2: Mass spec transitions for compounds detected via the ion pairing LC-MS/MS method

Formyl-CoA detection was also performed with this method adapting the given parameters for acetyl-CoA for formyl-CoA, specifically changing the parent ion mass to reflect one

-CH_2 unit while retaining the product ion mass, which represents the CoA group and SRM transition settings. No formyl-CoA standard was available so quantification was unattainable.

E.8.4 Hydrophilic Interaction Liquid Chromatography (HILIC) LC-MS/MS Sugar Phosphate Method

This protocol is underdevelopment by Yanfen Fu.

- Machine: Waters, Xevor QQQ
- Column: Waters 1.50460.0001 Lot # HX543881 SeQuant ZIC-pHILIC, 5 μM, polymeric Sorbent Lot No. P130114, PEEK 150 x 2.1 mm, metal-free HPLC column
- Mobile phase A: 20 mM ammonium bicarbonate in H_20
- Mobile phase B: 100% ACN
- Gradient:

Time [min]	**Flow Rate [mL/min]**	**%A**	**%B**
0	0.150	15.0	85.0
0.50	0.150	15.0	85.0
20.00	0.150	80.0	20.0
21.00	0.150	90.0	10.0
23.00	0.150	90.0	10.0
23.50	0.150	15.0	85.0
30.00	0.150	15.0	85.0

Table E.3: Mobile phase gradients for HILIC LC-MS/MS sugar phosphate detection method

- Column Storage Buffer: Acetonitrile/Ammonium Acetate 5 mM, pH 6.8; 80%/20%

E.9 Growth on Formate

Growth experiments were completed using a set of four M9 minimal media with all combinations of −/+ formate (normally 40 mM) and −/+ glycerol (0.1%) (21). Temperatures tested were room temperature (∼ 22 °C) and 18 °C. Cultures were grown in Bioscreen plates (0.210 mL total volume) in the Bioscreen automated growth curve machine (Growth Curves USA) or glass tubes (5 mL total volume). In both cases, cultures were kept oxygenated using the high shaking setting. Growth experiments were carried out for up to 15 days. The cultures tested included ALA2 *fdoG* with pTrcCO_2-3 (ecACS, lmACDH, FLS) pathway core and associated controls with a combination of secondary plasmids pSB4C5 FDF3-6 (cmFDH, yDHAK, FocA/B). Room temperature growth on M9 minimal medium agar plates with a range of formate concentrations from 0-50 mM was also tested.

E.10 Anaerobic pathway evaluation

E.10.1 Anaerobic Culture Prep

ALA2 *fdoG fdnG fdhF pflA frdA* was used for the mannitol and glucuronate tests. Plasmids pTrcHis2C or pTrcCO_2-3 XFLS, a less active FLS, or pTrcCO_2-3 were added for mannitol and formate tests. Cells were prepared as described in cell preparation in a 50 mL volume for the glucuronate approach confirmation or in smaller volumes for other cultures. For mannitol + formate or glucuronate tests, the cell pellets from approximately 10 mL per final tube were suspended in medium volume (final volume 7-10 mL) without formate or glucuronate (Table E.4. 20 mM glucuronate or 15 mM formate was added separately. Cultures were incubated at 18 °C after harvesting. Optical density at 600 nm was measured using the Nanodrop (Thermo Scientific) for inoculation cultures and tube spectrophotometer (Spectronic 20D)

for cultures in glass culture tubes with and without the lip for sealing. pH was measured with pH strips pH 5-10. To make a tube anaerobic, rubber stopper were crimped into place with aluminum seals. The rubber stopper was cleaned with ethanol and pierced with two sterile needles. A nitrogen line in was attached to one needle for 2 minutes, then the outlet was removed for 15 sec to give positive pressure inside the tube. A 1 mL syringe with 23G needle was used to sample capped tubes for HPLC. The needle was replaced with a syringe filter (0.22 μm) and the sample was filtered into an Eppendorf tube.

Media	Format [mM]	Mannit [mM]	Glucur [mM]	total volume [mL]	Extra phosphate [mL]	5x M9 Salts	Thiami [1 mg/mL	1M MgSO4 [mL]	1M CaCl2 [mL]	0.5 M mannitol [mL]	1 M Sodium Formate [mL]	0.94 M Sod Glu curo [mL
M9 + 100 mM mannitol (#1)	0	100	0	50	3.05	10	0	0.1	0.005	10	0	0
M9 + 100 mM mannitol +Thiamine (#2)	0	100	0	50	3.05	10	0.225	0.1	0.005	10	0	0
Glucuronate to add			20	10								0.21
Formate to add	15			10							0.150	

Table E.4: Media for anerobic pathway studies. pH adjusted to 7.5. Formate or glucuronate was added later in the amounts given. Appropriate antibiotics were also used.

E.10.2 HPLC for Fermentation Product Detection

- Instrument: Shimadzu with UV detector. Manual injection.

- Column: Phenomenex Rezex ROA-Organic Acid H+ (8%), 00H-0138-K0, 300 x 7.8 mm, Ion Exclusion
- Guard Column: Phenomenex Rezex ROA-Organic Acid H+ (8%) LC Guard Column, 03B-0138-K0, 50 x 7.8 mm
- Temperature: 65 °C
- Mobile Phase: 0.005N H_2SO_4, filter 0.22μm
- Degassing method: Bubbled helium, at least 30 min before use.
- Method: Isocratic, 36 min
- Detection: UV, 210 nm

E.11 Biosensors

E.11.1 Fluorescence detection

A microplate fluorimeter (Tecan) was used for fluorescence measurements. RFP: 560 nm (ex) / 590 nm (em) GFP: 485 nm (ex) / 540 nm (em);

E.11.2 MUG: A fluoresence-based high throughput LacZ assay

Protocol (Adapted from Vida-Aroca et al. [136])

1. 20 μL of overnight culture transferred to a 96 well microplate (black/clear optilux flat bottom, BD falcon) with 80 μL Z-buffer[88].
2. Read cell density at 595 or 600 nm.
3. Add 25 μL 1 mg/mL 4-methyl-lumbelligeryl β-D-galactopyranoside (MUG) in DMSO to each well.
4. Incubate room temp for 15 minutes.
5. Stop the reaction with 30 μL 1 M Na_2CO_3
6. Amount of β-galactosidase activity quantified in microplate fluorimeter (Tecan) using the assay on cell-free culture medium sample as a blank.

7. Arbitrary MUG units calculated as $F_{340/35nm/460nm}/(t \times A_{600})$

    - F: Sampled fluorescence at the end of the reaction
    - t: Time of reaction in minutes
    - A_{600}: Absorbance of the cell suspension

E.11.3 X-gal formate plates for CBZ formate growth experiment

Goal: check for slow growth on formate → look for pathway dependent differences

- Strains: CBZ pTrcKM/pTrcKMCO_2-3 XFLS/pTrcKMCO_2-3 ACS L641P/pTrcCO_2-3, all with pSB3K3 $_p$J23100 FocA 10721 FDH
- 200 mL M9 minimal media agar.
- Pour 2 batches M9 minimal plates, one with formate, one without formate.
    - Microwaved agar (3 g agar with 133 mL MillQ water autoclaved 30 min at 121 °C /200 mL M9 agar)
    - + Formate: Added 0.1 mL Cb (50 mg/ml), 200 μL Km, 100μL Cm (34 mg/ml), 0.4 mL 1M $MgSO_4$, 0.020 mL 1 M $CaCl_2$, 1 mL 1 M glucose (5 mM), 2 mL 1 M formate (10 mM), 0.4 mL (20 mg/ml) X-gal in DMSO, IPTG to 0.5 mM, 40 mL 5x M9 stock and sterile water to 200 mL
    - - Formate: Added 0.1 mL Cb (50 mg/ml) , 200 μL Km, 100 μL Cm (34 mg/ml), 0.4 mL 1 M $MgSO_4$, 0.020 mL 1 M $CaCl_2$, 1 mL 1 M glucose (5 mM), 0.4 mL (20 mg/ml) X-gal in DMSO, IPTG to 0.5 mM, 40 mL 5x M9 stock and sterile water to 200 mL