Improving the Formolase Pathway by In Vitro and In Vivo Enzyme Evolution

Janet Bickford Matsen

General Exam
April 2, 2015

University of Washington
Chemical Engineering
Lidstrom Lab

Abstract

The Formolase Pathway has been invented by the Lidstrom and Baker Labs [1]. It includes three enzymes that each catalyze reactions not seen in nature, and is intended to be the growth module of a biological electrofuel system. Such a system could create transportation fuels or other chemicals from CO_2 and renewable energy. Initially only in vitro evidence of Formolase Pathway function could be demonstrated using purified enzymes; no evidence of in vivo activity could be seen. Subsequent experiments highlighted the need for better enzymes, which should increase flux through the pathway and decrease toxicity to the host. This work uses in vitro and in vivo directed enzyme evolution to improve the Formolase Pathway enzymes with the goal of increasing the commercial potential of the pathway.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Electrosynthesis/electrofuels introduction

Microbial electrosynthesis and electrofuel projects aim to make fuels and chemicals from CO_2 and renewable energy. This approach has the potential for better energy efficiency, resource efficiency, land efficiency, and scalability than available methods [2]. Such electrosynthesis requires a microbial catalyst that can utilize renewably produced electrons to make carbon-carbon bonds for biomass and product formation. One strategy is to deliver electrons to chemolithoautotrophs via compounds such as H_2 [3], H_2S, NH_3 [4], and Fe^{2+} [5], or perhaps even directly from a cathode [6]. This could be coupled with several naturally occurring carbon fixation pathways [2, 7]. These strategies are currently limited by the necessity to use non-model organisms such as *Ralstonia eutropha* (H_2 utilizer), *Acidithiobacillus thiooxidans* (H_2S user), *Nitrosomonas europaea* (NH_3), and *Acidithiobacillus ferrooxidans* (Fe^{2+} utilizer [2]). Formate can be produced electrochemically from CO_2 [8], representing another promising energy carrier that could also serve as a carbon source [9]. While there are some organisms that can naturally grow on formate, they are not amenable to rapid commercialization like *E. coli* or *Saccharomyces cerevisiae*. The Lidstrom and Baker Labs at the University of Washington designed a novel pathway called the Formolase Pathway for use in such model organisms. The Formolase Pathway (Figure 1.1) supplies cells with formate for both carbon and energy and can be used in conjunction with a variety of biofuel production pathways [10].

1.2 The Formolase Pathway

In the Formolase Pathway, cells are provided carbon and reducing power in the form of formate, which can be made from CO_2 by passing current over a lead electrode at low pH [8, 11]. Formate is assimilated using three sequential enzymes (Table 1.1) including a computationally designed enzyme that carboligates three formaldehyde molecules into dihydroxyacetone, a reaction not known in nature. The other two core enzymes catalyze very similar reactions in nature (Table 1.2).

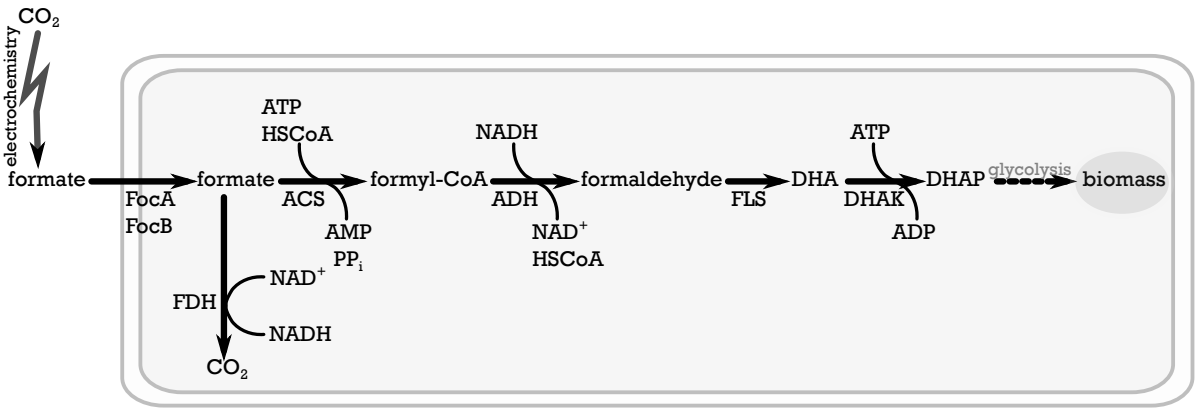

Figure 1.1: The Formolase Pathway. FDH = formate dehydrogenase. ACS = acetyl-CoA synthase. ADH = acetaldehyde dehydrogenase. FLS = formolase. DHAK = dihydroxyacetone kinase. DHA = dihydroxyacetone. DHAP = dihydroxyacetone phosphate.

In *E. coli*, formate diffuses into the periplasm though porins then enters the cytoplasm via a recombinantly expressed transporter, either FocA or FocB [12, 13]. Cytoplasmic formate is partitioned between two branches: oxidation to gain reducing power and assimilation into cellular building blocks. The oxidative branch results in NADH production from a heterologous formate dehydrogenase. Formate assimilation begins with activation to formyl-CoA by acetyl-CoA synthase (ACS), driven by conversion of ATP to AMP + PP_i [14]. Formyl-CoA is further reduced to formaldehyde by acetaldehyde dehydrogenase (ADH), oxidizing one NADH. Formolase (FLS) carboligates three formaldehyde molecules into dihydroxyacetone (DHA). DHA is subsequently converted to the easily assimilated glycolytic intermediate dihydroxyacetone phosphate (DHAP) by dihydroxyacetone kinase.

Table 1.1: Formolase Pathway enzyme abbreviations and sources. Protein sequences are available in Appendix A.1.

Enzyme Name	**Abbreviation**	**Host Organism**	**PDB Code**
formate dehydrogenase	FDH	*Candida methylica*	
acetyl-CoA synthase	ACS	*E. coli*	2P2F
acetaldehyde dehydrogenase	ADH	*Listeria monocytogenes*	3K9D
formolase	FLS	novel	4QPZ, 4QQ8
dihydroxyacetone kinase	DHAK	*Saccharomyces cerevisiae*	

Table 1.2: Repurposed enzymes in the Formolase Pathway

enzyme abbreviation	**native activity**	**repurposed activity**
FDH	$\text{formate} + \text{NAD}^+ \longleftrightarrow \text{CO}_2 + \text{NADH}$	
ACS	$\text{acetate} + \text{CoA} + \text{ATP} \longleftrightarrow \text{acetyl}-\text{CoA} + \text{AMP} + \text{PP}_\text{i}$	$\text{formate} + \text{CoA} + \text{ATP} \longleftrightarrow \text{formyl}-\text{CoA} + \text{AMP} + \text{PP}_\text{i}$
ADH	$\text{acetyl}-\text{CoA} + \text{NAD}^+ \longleftrightarrow \text{acetaldehyde} + \text{NADH}$	$\text{formyl}-\text{CoA} + \text{NAD}^+ \longleftrightarrow \text{formaldehyde} + \text{NADH}$
FLS	benzoin $\longleftrightarrow$ 2 benzaldehyde	3 formaldehyde $\longleftrightarrow$ dihydroxyacetone
DHAK	$\text{DHA} + \text{ATP} \longleftrightarrow \text{DHAP} + \text{ADP} + \text{PP}_\text{i}$	

The enzymes used are summarized Tables 1.1 and 1.2. Formate dehydrogenase from *Candida methylica* [15] is used because the native *E. coli* FDH is not NAD-linked and cannot produce the NADH required by the Formolase pathway. ACS is repurposed from acetate scavenging, and has been demonstrated to work on the single-carbon (C1) version of the native two-carbon (C2) reaction [16]. ADH from *Listeria monocytogenes* is repurposed from acetaldehyde detoxification and was also shown to have C1 activity. FLS was designed computationally by the Baker Lab [16] using benzaldehyde lyase as a template. Mutations causing the specificity to switch from benzaldehyde to formaldehyde were identified from libraries of candidate designs [1]. An NADH-linked dihydroxyacetone kinase from *Saccharomyces cerevisiae* is used to convert dihydroxyacetone to dihydroxyacetone phosphate because the *E. coli* DHAK is dependent on phosphoenolpyruvate [17].

1.3 Challenges in Formolase Pathway implementation

The difficulty of obtaining high flux through new metabolic pathways using chemistry not found in nature is reflected by the absence of published examples. The goal for our pathway is to sustain growth on formate using the Formolase Pathway. Our team's first step toward this goal was to verify enzyme activity using purified enzymes in controlled enzyme assays. Purified enzymes were incubated with ^{13}C labeled formate and the metabolites produced were detected with mass spectrometry [1]. In addition to requiring either purified enzymes or highly clarified (ultracentrifuged and desalted) cell extracts, long reaction times (hours) were necessary to detect trace concentrations of downstream metabolites. Carbon flux calculations suggest that if the flux observed in the assay was sustained in vivo, the pathway could support an organism with a 10+ day doubling time (Matsen, Smith & Lidstrom, unpublished).

So far, growth of *E. coli* has not been demonstrated using the FLS pathway despite testing varied media conditions and gene expression levels. M9-based growth medium with differing formate concentrations was tested with numerous plasmids containing different promoters (pLac, pTrc and T7) and varying gene orders. No growth was seen.

To troubleshoot failed growth experiments, sensitive in vivo ^{13}C labeling experiments were done. Cells harboring the Formolase Pathway in minimal M9 medium

were fed ^{13}C formate. Mass spectrometry was used to look for assimilated ^{13}C in amino acids [18, 19]. No incorporation of ^{13}C formate into living cells was observed in several tests despite obtaining signal when positive controls with ^{13}C glycerol were tested (Smith, Matsen, & Lidstrom, unpublished). These experiments highlighted substantial differences between the reaction conditions in vitro and in vivo, which became the focus of this work.

1.4 Potential issues preventing growth

There are numerous differences between the highly controlled reaction conditions that demonstrated early proof-of-principle activity and the clearly problematic in vivo system. Potential issues are listed below to illustrate our motivation to break down the goal into smaller pieces.

1.4.1 Limiting transport

Extensive enzyme assays show the current versions of our enzymes require high substrate concentrations. This is often the case when enzymes are used to catalyze reactions they did not evolve to perform. Poor formate transport could limit the concentrations of all Formolase Pathway intermediates. The intracellular concentration of formate provided by the heterologously expressed transporter is not known. Formate transport efficiency is being assayed (Smith & Lidstrom, unpublished).

1.4.2 Poor ACS and ADH specificity

ACS and ADH are wild-type enzymes that catalyze 2-carbon (C2) reactions in nature. Both interact unfavorably with metabolites present in *E. coli* cells, especially acetyl-CoA and acetate (Figure 1.2). Preference of native C2 substrates may be resulting in significant flux through undesirable reactions. The cellular pool of acetyl-CoA may be drained as it is converted to acetate and toxic acetaldehyde at significant rates. Additionally, ACS drains the ATP pool [20]. Enzymes may be ”busy” catalyzing undesirable reactions, limiting flux through the intended C1 reactions.

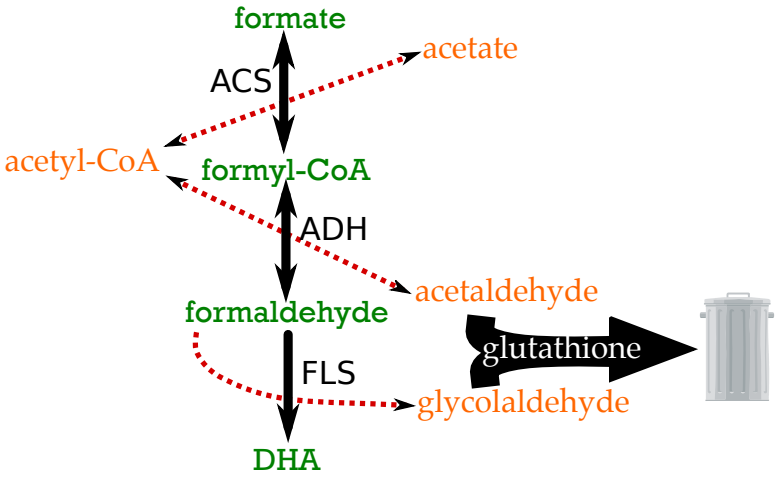

Figure 1.2: Formolase pathway with undesirable side reactions shown by red dashed arrows. Intended metabolites are colored green and undesirable metabolites of side reactions are shown in orange. Aldehydes are detoxified by large intracellular glutathione pools in *E. coli*.

1.4.3 Flux loss to unintended reactions

Formaldehyde reacts with nucleophiles, forming adducts and crosslinking biomolecules including DNA and proteins [21]. Formaldehyde and other aldehydes can also be lost by reacting with glutathione, a thiol-based antioxidant produced by many types of organisms to reduce oxidative stress. In *E. coli*, glutathione is one of highest concentration metabolites [22], representing a potentially large formaldehyde sink.

FLS is known to produce glycolaldehyde, an undesirable C2 product resulting when only two formaldehyde molecules are combined instead of three (Figure 1.2). Collaborators at UC Berkeley are probing what fraction of the product is glycolaldehyde, and whether FLS can produce dihydroxyacetone by combining one glycolaldehyde with one formaldehyde.

1.4.4 Low catalytic efficiencies

Flux through the intended reactions may be too low to support measurable growth. The catalytic efficiencies of ADH and FLS on C1 substrates are approximately 5 [16], whereas natural enzymes usually have catalytic efficiencies (k_{cat}/K_m) between 10^3 and 10^8 $M^{-1}s^{-1}$ [23]. The catalytic efficiency of ACS for formate is known to be poor but cannot be calculated because the rate of ACS is proportional to the formate concentration until the solubility limit is approached (750 mM formate; Matsen, unpublished). These poor efficiencies are caused in part by high K_m values. Consequently, the concentrations of all intermediates of the Formolase Pathway likely need to be very high in order to sustain appreciable flux.

1.4.5 Pathway burden

In order to obtain growth, the pathway must benefit the host more than it burdens it. This means that the beneficial energy and carbon flux the cell receives from the pathway must be be greater than the energetic costs associated with expression of

the enzymes and all of the problems described above. It is difficult to quantify the undesirable fluxes and energetic costs of each issue, and hence to know how far the pathway is from breaking even. Understanding that many of these issues stem from poor enzyme properties motivated new efforts to improve the enzyme kinetics.

1.5 Strategies For Pathway Improvement

The goal of high in vivo flux has been broken down into smaller subgoals that limit the number of issues preventing progress:

1. **Improvement of single enzymes in vitro**, removing the issues of toxicity and most side reactions. Both random and rational approaches will be used. See Chapter 2.

2. **In vivo selection of segments of the FLS pathway**, relaxing the requirement that all Formolase Pathway enzymes need to perform well. Use of a methylotrophic host allows unique opportunities for formaldehyde based selections. See Chapter 3.

Improving portions of the pathway by these means may allow for powerful whole-pathway selections. True growth-based selections should enable rapid improvement in the Formolase Pathway and increase its commercial potential.

Chapter 2

Directed Evolution of Single Enzymes In Vitro

.

Directed enzyme evolution has been used to improve countless enzymes since its establishment in the 1990s [24]. These techniques can improve a range of properties including substrate specificity, thermostability, and solvent tolerance for use in industrial catalysis or in vivo catalysis.

Directed evolution will be used to increase the activity and substrate specificity of the Formolase Pathway enzymes with higher in vivo performance as the goal. Both random and rational approaches will be used. Enzyme improvements will be assessed by enzymatic assays of kinetic rate constants. In vivo performance gains will be assessed via growth assays and ^{13}C labeling studies. As soon as the Formolase Pathway enzymes have been improved enough to support growth, higher throughput selections (see Chapter 3) will be used to accelerate the rate of evolution.

2.1 ACS and ADH require engineering

Three enzymes in the Formolase pathway are used for reactions not known to occur in nature: ACS, ADH, and FLS. Of these three, only formolase has been engineered to prefer the intended substrate over its native substrate [1]. FLS specificity and activity were enhanced by iteratively testing computationally designed libraries [16]. ACS and ADH have not been engineered. They retain their full wild-type preference for C2 metabolites and are thought to limit in vivo performance. If ACS and ADH can be evolved to have higher activity and specificity, in vivo selection strategies for the entire pathway including formolase may be possible (see Section 3.2).

2.2 Four assays are available for ACS and ADH engineering

Several assays available for either ACS or ADH (Figure 2.1, Table 2.1):

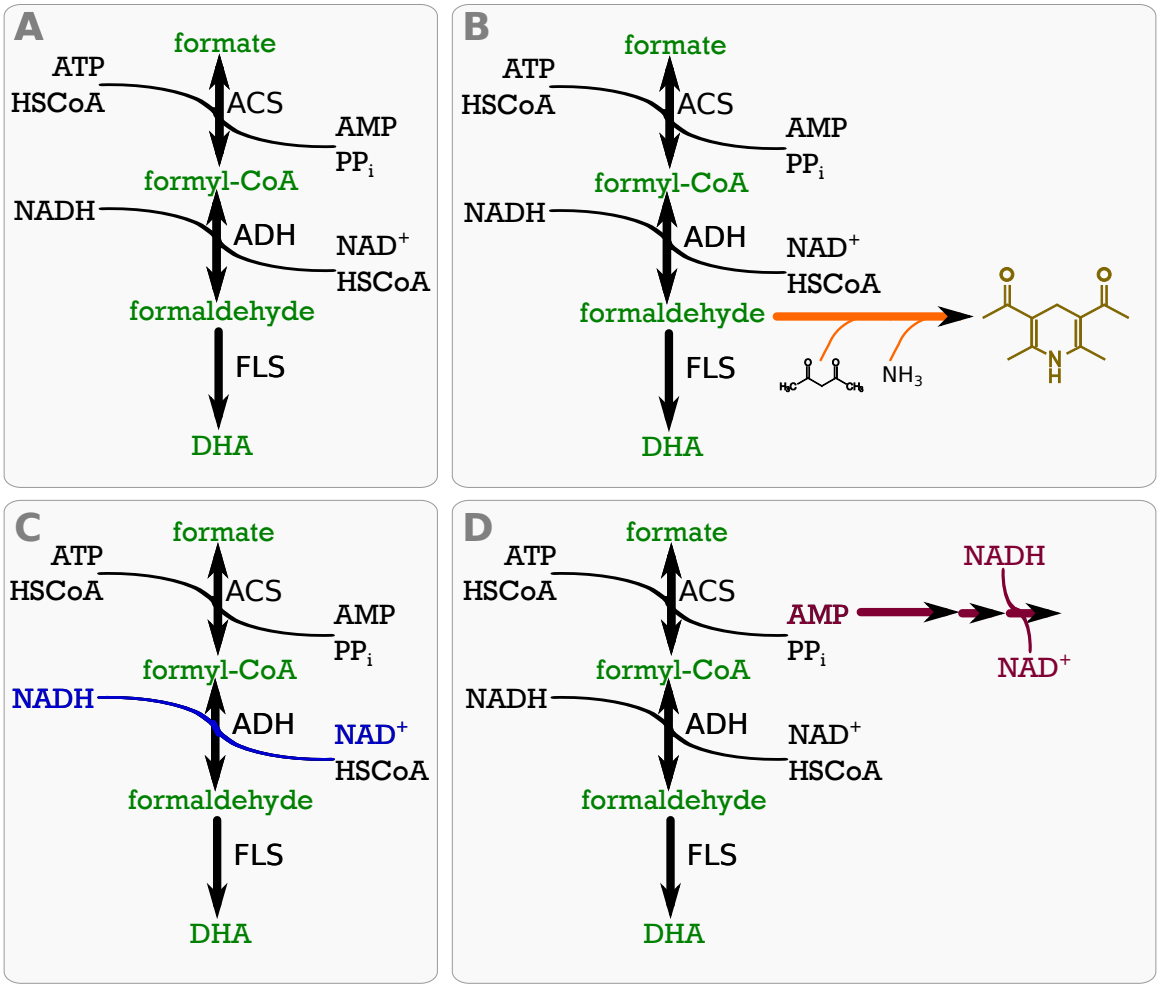

Figure 2.1: The Formolase Pathway with available enzymatic assays for ACS and ADH evolution. Colored lines/arrows highlight the chemistry allowing for rate observations. A: FLS pathway only. B: The Nash assay for formaldehyde. The orange arrow shows production of the yellow pigment diacetyldihydrolutidine. C: NADH/NAD^+ assay of ACS/ADH combined activity. D: ACS assay with 3 coupling enzymes (maroon arrows) that link AMP production to NADH oxidation.

Table 2.1: Enzyme assays validated for ACS and ADH library screening and kinetic characterization.

	Nash assay	**ACS + ADH NADH monitoring**	**AMP monitoring**
works for	ACS or ADH	ACS or ADH	ACS
summary	a yellow pigment is produced from formaldehyde made by ACS + ADH	NADH consumption caused by ACS + ADH activity is monitored by changes in A340 nm	NADH oxidation is coupled to AMP production by 3 enzymes
assay reaction	formaldehyde is covalently incorporated into a yellow pigment	formyl-CoA is reduced to formaldehyde by ADH, causing NADH to be oxidized and A340nm to drop	NADH is oxidized to NAD+ (A340nm drops) as AMP is produced
kinetic or end-point	end-point	kinetic	kinetic
parameter read	yellowness; A412nm	NADH oxidation (A340 nm)	NADH oxidation (A340nm)
sample types	whole cells or purified protein	purified protein only	purified protein only
suitable for	initial library screens	library refinement, kinetic characterization	kinetic characterization and substrate preference assays

Nash assay for formaldehyde

The Nash assay is a colorimetric formaldehyde end-point assay [25] that can be used to screen ACS or ADH variants. A yellow compound is produced when formaldehyde reacts with the Nash reagent mix. Absorbance at 412nm is used to infer the formaldehyde concentration after protein has been precipitated and removed. This assay can be used to assay either ACS or ADH by altering which enzyme is limiting. The Nash assay is well-suited for high throughput screening because it is an endpoint assay using inexpensive materials and measures the product of interest nearly directly.

ACS + ADH dependent NADH oxidation

ACS or ADH can be assayed by monitoring A340 drop as NADH is oxidized in a mixture containing ACS and ADH with appropriate substrates and cofactors [16]. NADH absorbs strongly at 340nm, whereas NAD^+ does not, enabling inference of the NADH concentration and hence the assay rate. Purified enzymes are used in NADH-based assays due to unspecific NADH oxidation activity caused by other enzymes.

AMP production by ACS

Production of AMP can also be inferred by an A340 drop associated with NADH oxidation. Three coupling enzymes convert two molecules of NADH to NAD^+ per

molecule of AMP produced [26, 20, 27]. This assay works equally well regardless of the substrate, enabling analysis of substrate specificity. It is less suitable for high throughput screening because it requires purified enzymes, is kinetic in nature, and uses relatively expensive commercially purified enzymes.

2.3 ACS was chosen as the first engineering target

ACS was chosen over ADH for initial evolution experiments. Improvement of the specificity and C1 activity of ACS was predicted to have the greatest effect on in vivo pathway performance by increasing energy efficiency and reducing toxicity, as follows:

1. ACS is known to be toxic in *E. coli* and *Salmonella*, as demonstrated in studies of its tight transcriptional and post-transcriptional regulation [28, 20]. Depletion of ATP and accumulation of AMP are thought to be the source of toxicity [20]. For every mole of substrate catalyzed, more than 2 moles of ATP are required for catalysis and restoration of AMP back to ATP [20]. Thus the poor substrate specificity of ACS (Subsection 1.4.2) may have a large impact on in vivo efficiency. Reducing the rate at which ACS catalyzes undesirable reactions represents potentially significant energy savings for the cell.

2. The toxicity of ADH is thought to be lower than that of ACS. When ACS is the first pathway gene expressed in polycistronic pTrc His2C vector variants, growth is poorest (Smith and Matsen, unpublished). Since the first enzyme is usually expressed at the highest levels, this result supports the idea that over expression of wild-type ACS is detrimental.

ACS is a also better studied than ADH. The enzymology of the ADH currently in use is so poorly understood that the directed evolution approaches available are limited. ADH 3K9D is a member of the fairly diverse acylating aldehyde dehydrogenase superfamily, but little beyond that is known. Other than a putative catalytic cysteine, nothing about the active site is known. ACS, on the other hand, is well characterized with dozens of papers about its biochemistry, enzymology, and regulation available [14]. The superfamily ACS is a member of catalyzes equivalent reactions with larger substrates [26, 29, 30], suggesting the active site is capable of evolving [31, 32]. Furthermore, a crystal structure with bound substrate and Coenzyme A is available (PDB accession 2P2F [26]) making the active site and key residues clear. Both directed engineering and enzyme evolution approaches are being used.

2.4 In vitro ACS library screening strategy

As noted above, two strategies are proposed to improve pathway enzymes: directed in vitro engineering and enzyme evolution. This section describes progress on the in vitro engineering approach. See Chapter 3 for in vivo selections.

Assays of varying speed and precision will be used to screen ACS libraries. First quick tests using whole cells will be used to identify promising candidates. Then

a more quantitative (but time-consuming) assay will be used to confirm activity enhancements using purified enzymes.

In the whole-cell Nash assay, BL21(DE3) colonies from the mutated libraries are transferred to deep-well 96-well plates and grown in autoinduction medium [33]. Two days later 4uL of cells are assayed in polystyrene 96-well plates with half of the normal cross-sectional area of normal plates. Each 40 μL reaction contains 2μL of 10X BugBuster, 6 mM NADH, 1 mM CoA, 4 mM ATP, 5 mM $MgCl_2$, 700 mM sodium formate, and 7 μM purified ADH. Plates are incubated for 7 minutes at room temperature while formaldehyde is produced, then 40 μL of Nash reagent is added.

A yellow pigment is produced from formaldehyde as plates are incubated at 50^{o}C for 30 minutes. Wells are scored visually from 0 to 3, with 2 corresponding to the average yellowness of control wells using cells expressing the unmutated template plasmid. High-scoring wells are re-tested, and carried into subsequent rounds of testing. If desired, the formaldehyde concentration can be quantified by comparing formaldehyde standard 412 nm absorbances to those of unknown assay mixtures that have been clarified by trichloroacetic acid precipitation. This assay is extremely fast, requires minimal labor, and can identify promising mutants from among the inactive and moderately active variants before doing more rigorous tests using purified enzymes. Mutants that help with expression, solubility, activity, and perhaps specificity/toxicity are expected; all of these are desirable for in vivo use.

Enzyme variants that look promising after replicate Nash assay tests have been performed will be characterized more thoroughly. Candidates will be his-tag purified in 96-well plates and assayed by the ACS + ADH kinetic A340 nm test. Mutants with enhanced specific activity will be purified at flask-scale for more careful validation. All methods used are posted to GitHub's GitPages: http://janetmatsen.github.io/protocols/.

2.5 In vitro ACS engineering progress

ACS engineering began with validation of a mutation reported to increase activity [34]. ACS functions as an acetate scavenging enzyme in nature [34] and is regulated post-translationally by acetylation on Lys609 [35]. Previous studies showed that this regulation can be broken by mutating Lys641 to proline, resulting in a higher activity enzyme [34]. This mutation increases the specific activity of ACS when expressed in *E. coli* BL21(DE3) grown on rich medium (Figure 2.2) and was included in the template for all libraries generated.

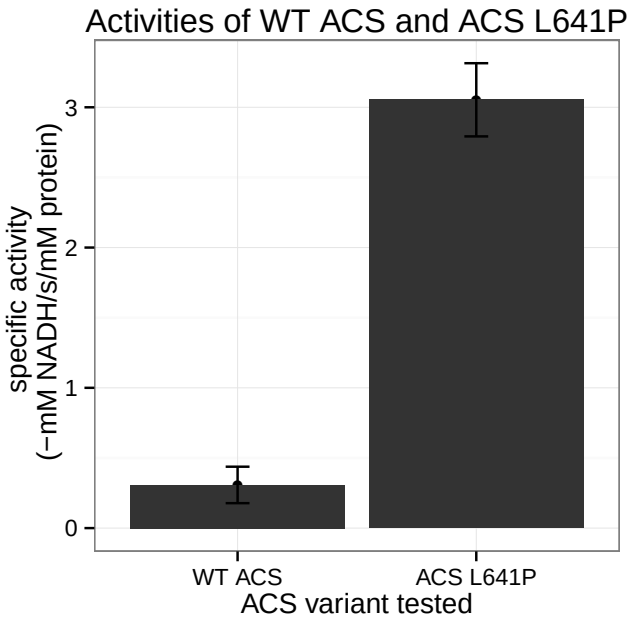

Figure 2.2: ACS L641P has about 10x higher activity than wild-type ACS (Matsen, unpublished) when expressed in *E. coli* BL21(DE3) grown on rich TB medium. In this assay, production of AMP by ACS was monitored by coupling enzymes [26] using triplicate biological replicates. The activity ratio is equivalent when using ADH as the coupling enzyme, and regardless of whether the substrate is formate or acetate.

2.6 Rational mutations tested

Before high throughput screening methods had been developed, rational mutations were tested based on ACS crystal 2P2F. A library of 37 rational point mutations (Table 2.2, Figure 2.3) in the active site was generated via the QuikChange II kit. Sites were chosen based on enzymology, multiple sequence alignments with homologous proteins specific to larger substrates, and distance from acetate in crystal 2P2F.

Table 2.2: Rational mutations tested in ACS L641P background. See Figure 2.3 for location relative to the substrate.

residue	mutation motivation	substitutions tested
V310	close to substrate	I, L, M, F, D
T311	close to substrate	N, Q, V, I, L, M, D, E
V386	close to substrate	I, L, M, R
G387	close to substrate	A, V, S, T, D
N521	close to substrate	Q
Y496	tighten active site by pushing on Q415-T418	F
Y355	pushes on V386, T311	W
T412	pushes on W414 in active site	N, Q
G420	pushes on W414 in active site	A, V, I, L, M, F, Y, W
F260	tighten active site by pushing on V310	Y
L262	tighten active site by pushing on V310	M

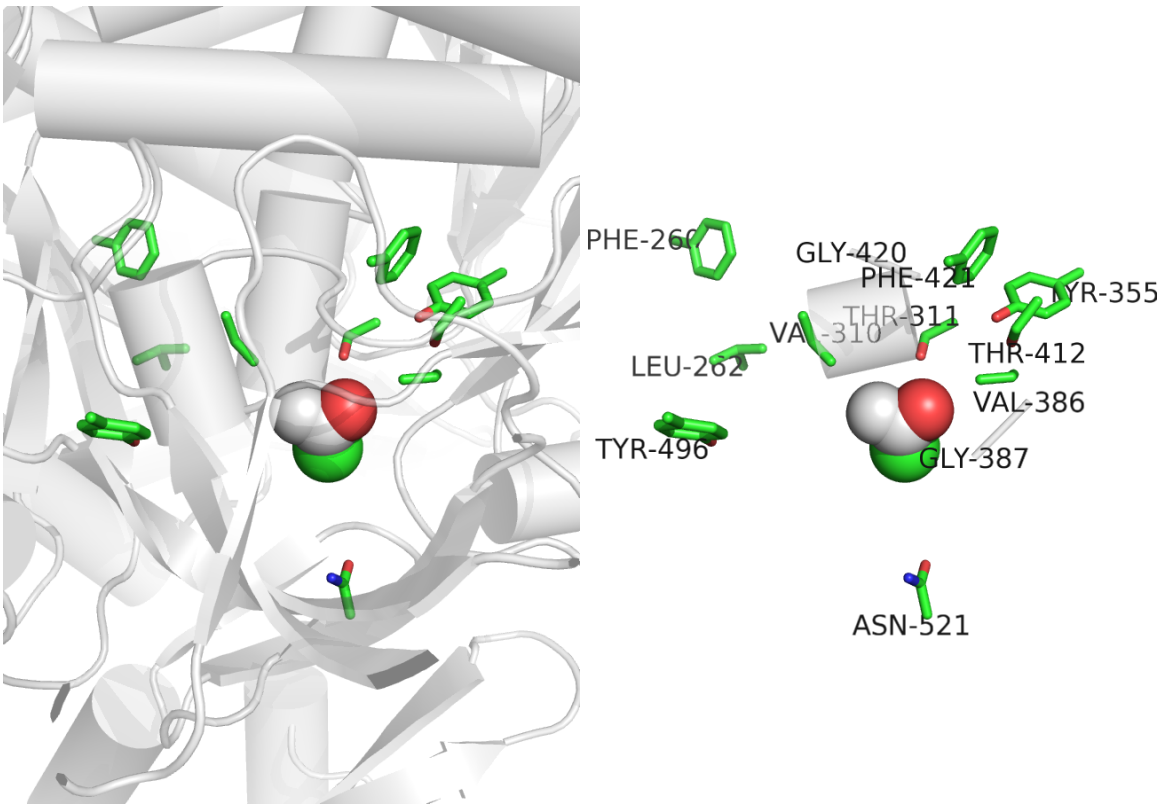

Figure 2.3: Rational mutations tested in ACS shown by their location in crystal 2P2F. Acetate is shown using spheres. The side chains of mutated residues are shown with green carbon atoms. ATP and CoA are not shown for simplicity.

Substitutions were mostly selected with the intention of increasing side chain length and tightening up the space inside the binding pocket with the intention of excluding acetate but not formate. None of the mutations significantly improved formate activity or specificity. Some combinations of apparently neutral mutations involving larger side chains were tested, but also failed to enhance catalysis or

specificity. Despite negative results, these assays served as a testbed for substantial assay method development and made random library screening possible.

2.7 Random library testing

The failure of the rational mutations tested motivated adoption of a random approach. Random mutations using ACS L641P as a template are being tested now using the whole-cell Nash assay. Up to eight 96-well plates of mutants can be tested per week. Mutations are being introduced via the GeneMorph II kit (Agilent Technologies) and assembled with Gibson Assembly [36]. The first libraries tested have mutations that span the entirety of the gene (Table 2.3). Ten plates of such mutants have been assayed and promising candidates were stored for re-testing. A mutation rate that leads to 40-50% inactive mutants is being sought in order to sample sequence space with appropriate depth [37]. Mutations that increase protein expression and stability are expected. Benefits to catalysis are also possible. Subsequent libraries will have mutations only in portions of the gene near the active site (Figures 2.4 and 2.5). Sample focused libraries have been prepared but not screened yet. Often improved enzymes are identified after screening 10^3–10^4 variants from good quality libraries.

Table 2.3: Random libraries created and tested to date

library	**region mutated**	**# of plates tested**	**relative mutation load**	**comments**
GM19	whole gene	1	high	7 mutations/protein on average; nearly all variants are inactive
GM20	whole gene	10	medium	Appropriately mutated: 3 mutations/protein on average. About 26% lack insert. About 32% have insert but are inactivated by mutations.
GM26	residues 284-318	0	low	Nearly all have complete genes
GM27	residues 284-318	0	medium	Nearly all have complete genes
GM28	residues 284-318	0	high	Nearly all have complete genes
GM29	residues 284-417	0	low	Nearly all have complete genes
GM30	residues 284-417	0	medium	Assemblies may have issues

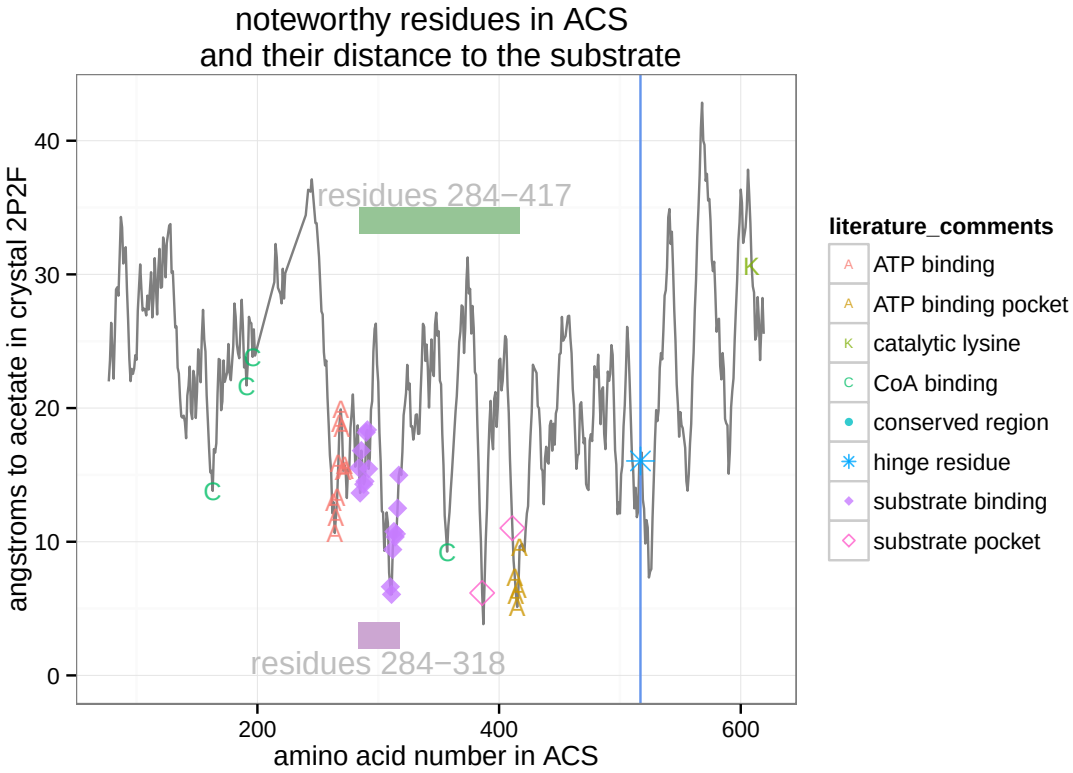

Figure 2.4: Residue distance from the substrate with functional annotations from literature research. Regions targeted for mutagenesis (residues 284-318, residues 284-417) are emphasized by the green and purple boxes. The vertical blue line marks a conserved hinge residue that separates the N-terminal domain from the swinging C-terminal domain.

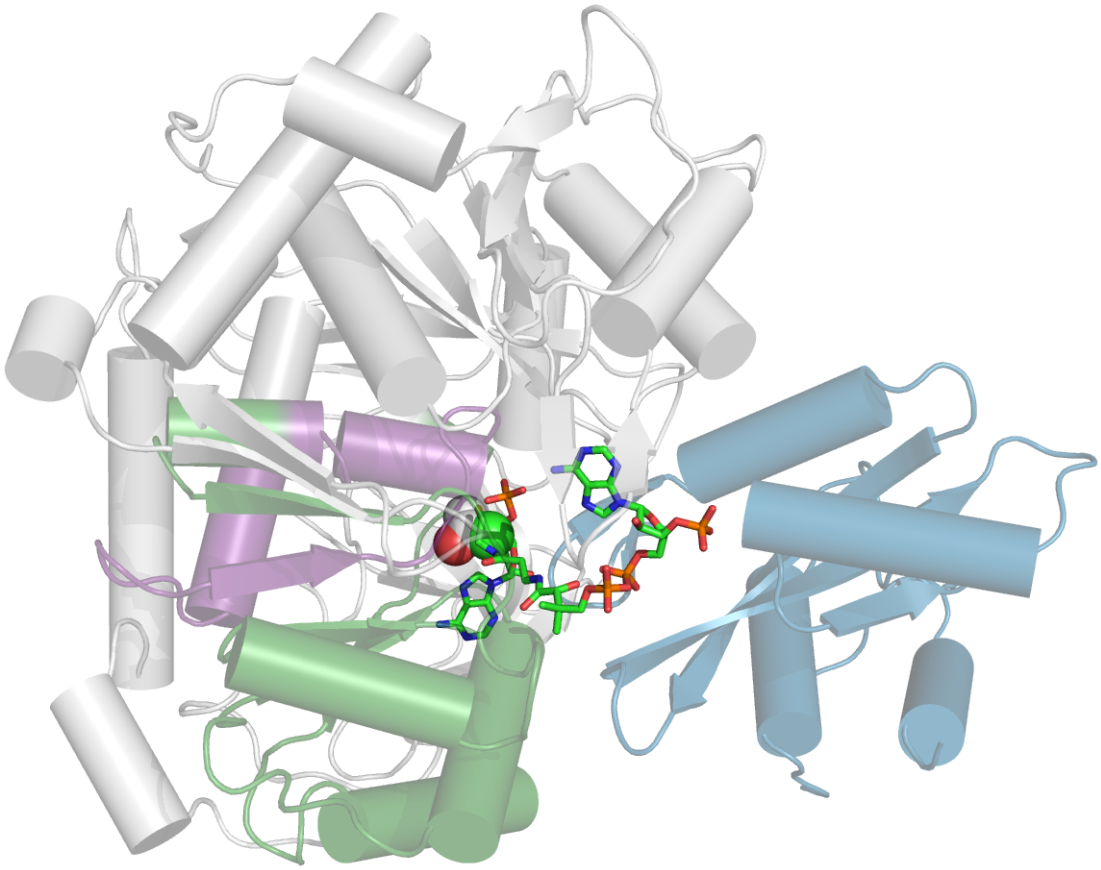

Figure 2.5: Targeted mutagenesis regions for ACS using crystal 2P2F. Purple barrels represent residues 284-318 and green barrels represent residues 318 - 417, which will be targeted with random mutations. The blue barrels represent the independently mobile C-terminal domain containing catalytic Lys609. Acetate is shown as spheres in the center. CoA and AMP are shown as sticks.

Chapter 3

In Vivo Enzyme Evolution

3.1 Motivation

A true selection is the most promising means of improving the efficiency of the Formolase Pathway. Orders of magnitude more enzyme variants could be screened once such a selection scheme is established. Such increased throughput in genetic variants is the most likely way to improve the Formolase pathway enough to have industrial potential. The inability to demonstrate growth using the whole Formolase Pathway motivated the design of selection schemes that target portions of the pathway. Discovery of mutations that improve the pathway enzymes or the cell host may unlock the ability to carry out whole pathway selections.

3.2 Enzyme selections using *Methylovorous glucosetrophus* SIP3-4

Methylovorous glucosetrophus SIP3-4 (referred to as SIP3-4 subsequently) is a bacterium isolated by the Lidstrom Lab that was identified as a promising host for selections of ACS + ADH, FLS, and eventually the whole pathway (Figure 3.1). This bacterium does not naturally grow on formate, but it naturally produces and assimilates formaldehyde, enabling selections for portions of the Formolase pathway that begin at or end with formaldehyde. Furthermore, SIP3-4 is expected to tolerate the Formolase pathway enzymes better than *E. coli* and support higher Formolase Pathway flux for the following reasons:

1. **Reduced loss of formaldehyde to detoxification systems is predicted.** All carbon assimilated by SIP3-4 outside of minor carboxylation reactions is first converted to formaldehyde when SIP3-4 is grown on its preferred substrate, methanol. Such flux through formaldehyde suggests low levels of formaldehyde sinks such as glutathione. Reduced loss of formaldehyde to glutathione sinks and the predicted tolerance of relatively high formaldehyde concentrations suggests the possibility that higher flux through portions of or the entire Formolase Pathway will be seen.

2. **SIP3-4 is predicted to have lower flux through acetyl-CoA**, a metabolite that is likely drained by ACS and perhaps ADH activity. The metabolism of SIP3-4

shows that acetyl-CoA is used only as a building block for some cell components suggesting lower flux through acetyl-CoA than in typical growth of *E. coli*.

Modules of the C1 metabolism in SIP3-4 will be replaced with portions of the Formolase Pathway to allow for selection of improved enzyme variants. Three different selection schemes are envisioned (Figure 3.1): (A) only ACS and ADH, (B) FLS only, or (C) the combination of ACS, ADH, and FLS. For selection scheme A, wild-type SIP3-4 cells can be used. For selections B and C, a strain unable to grow on C1 compounds will be created and used in order to force flux through FLS. This mutant will be deficient in both copies of hexulose-6-phosphate synthase, the formaldehyde assimilating enzyme. Selection A will be tested first while the double-knockout strain is being produced. Appendix A.2 contains a more detailed metabolic map for SIP3-4.

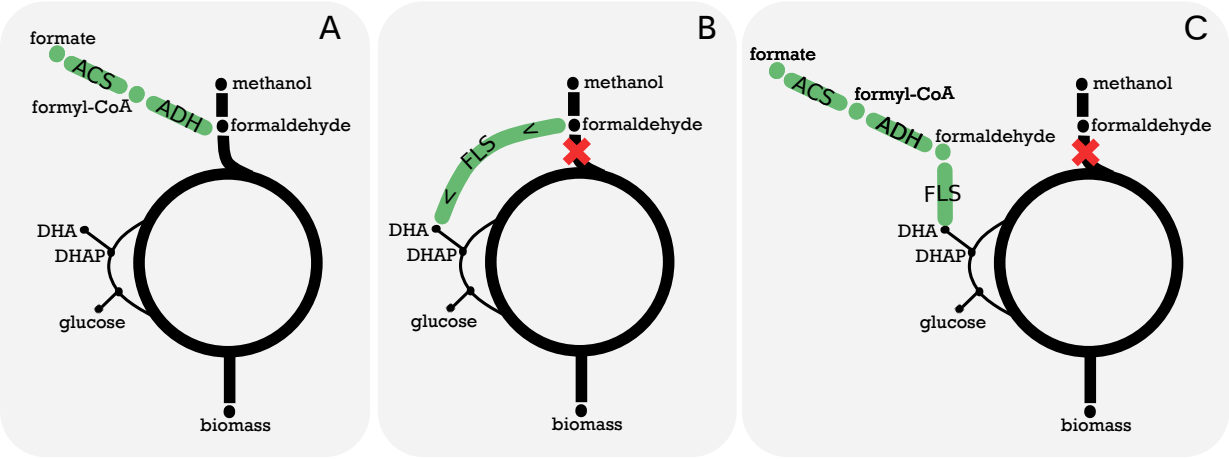

Figure 3.1: Three selection strategies planned for SIP3-4. The enzymes ACS, ADH, and FLS can be added in different combinations. A: add only ACS and ADH to the wild type strain and select for growth on formate. B: add only FLS to a hexulose-6-phosphate synthase knockout (red x) and select for growth on methanol. C: Add the entire pathway to a hexulose-6-phosphate synthase mutant and select for growth on formate. DHA is dihydroxyacetone and DHAP is dihydroxyacetone-phosphate. Cultures can be supplemented with glucose if necessary.

Why was SIP3-4 chosen over other methylotrophs, when all strains have comparably high flux through formaldehyde when growing on C1 compounds? RuMP cycle organisms are preferred because they assimilate formaldehyde and cannot grow on formate. Serine cycle methylotrophs are less desirable for this application because they can assimilate C1 units as either formate or formaldehyde [38], making it difficult to force flux through ACS and ADH.

SIP3-4 is a RuMP cycle methylotroph, which assimilates C1 units at the formaldehyde oxidation level [38] and cannot grow on formate. It is also one of the few RuMP cycle methylotrophs known to grow on an alternative multi-carbon compound, glucose (Figure 3.1). Growth on glucose allows the possibility of making a mutant that can no longer assimilate C1 units and the possibility of FLS selections. Glucose assimilation also allows for supplementation when selecting for ACS and ADH variants. In addition, SIP3-4 has a sequenced genome [39] (NCBI BioProject [40] accession number PRJNA33241), and RNA-seq data is available [41]. The other possible candidate facultative RuMP methylotrophs are very few in number, less well

studied, and could easily be ruled out due to lack of genomic and transcriptomic information.

3.3 SIP3-4 selections: plans and progress

SIP3-4 has never been genetically engineered or used for biotechnology. Progress has been made toward use of SIP3-4 for Formolase Pathway selections including identification of promoters that drive expression, improvement of expression vectors, validation of electroporation, and creation of unmarked knockouts. Progress toward use of SIP3-4 for Formolase Pathway selections has also been made: soluble ACS and ADH expression has been confirmed, ACS activity has been confirmed, and modifications of the strain to allow FLS selections have been made.

3.3.1 Identification of promoters with varying strengths

First promoters of varying strengths were identified. Promoter sequences derived from SIP3-4 transcriptomic data and commonly used *E. coli* sequences were tested. pHps1 and pHps2 are the 177 and 250 bp regions upstream of each copy of hexulose-phosphate synthase, the SIP3-4 formaldehyde assimilating enzyme. RNA-seq data for methanol-grown cells was used to determine the regions for harvest. Promoters were assembled into a trimmed (Matsen, unpublished) version of the broad-host range cloning vector pCM66 [42] in front of the gene for red fluorescent protein. Figure 3.2 shows the strengths of each promoter as measured by fluorescence tests. See Appendix A.4 for promoter sequences used and links to plasmid files.

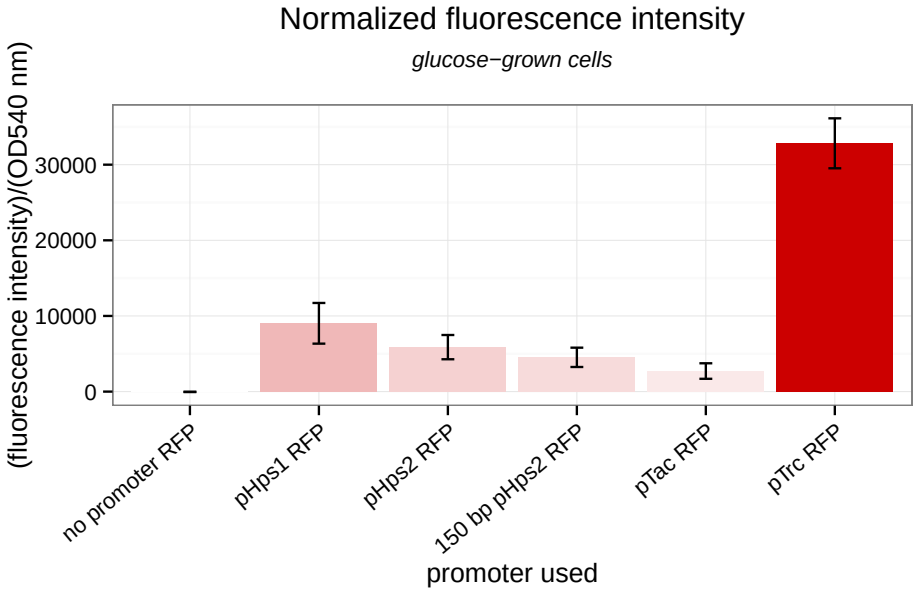

Figure 3.2: Promoter strengths of SIP3-4 strains grown on glucose plates.

3.3.2 Verification of formate uptake

Uptake of formate into the cytosol is necessary for selections involving ACS or ADH. While no formate-specific transporter has been identified, the following evidence suggests formate is transported into the cytoplasm. Addition of formate to SIP3-4 cultures grown on glucose enhances growth (Figure 3.3), presumably by formate dehydrogenase activity providing NADH. The formate dehydrogenase enzymes are thought to exist solely in the cytosol, suggesting formate does cross both the outer and inner membranes.

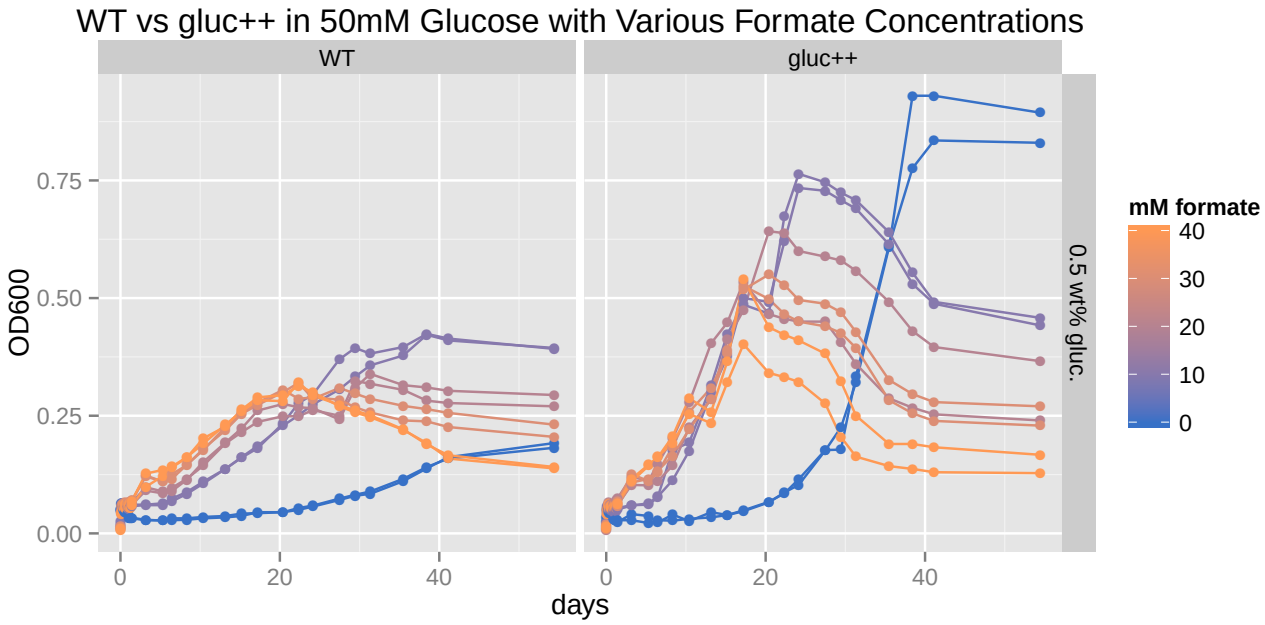

Figure 3.3: Growth curves of two strains of SIP3-4 grown on glucose with varying formate concentrations. The left panel has a wild type strain and the right panel shows data from a strain isolated from a short glucose selection experiment.

3.3.3 Confirmation of enzyme expression

The selection for ACS or ADH variants in wild-type SIP3-4 will be carried out first. Expression was demonstrated using plasmid pJ69 which contains promoters pLac and pHps before ACS, then promoter pHps2 before ADH (Matsen unpublished). Cells with this plasmid were grown in parallel with an empty vector control on methanol plates, lysed with BugBuster (EMD Millipore), and clarified by centrifugation. ACS and ADH are clearly visible in the SDS-PAGE gel (Figure 3.4), indicating soluble protein expression.

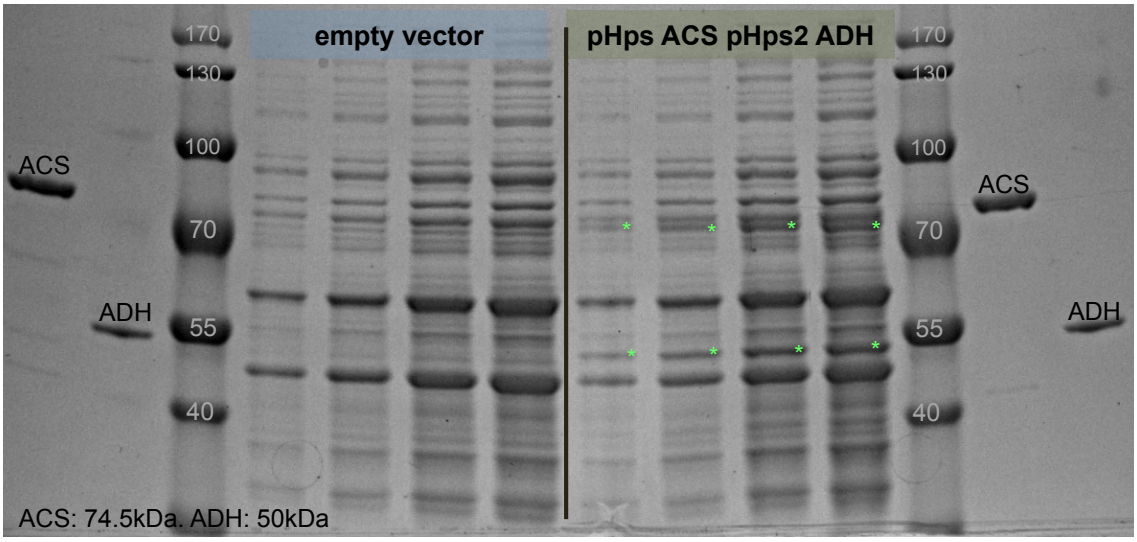

Figure 3.4: SDS-PAGE showing SIP3-4 expresses ACS and ADH. The left set is an empty vector control and the right set is from a strain with an ACS + ADH expression plasmid added. Different volumes of each lysate are added within each set. ACS and ADH are marked with green stars.

3.3.4 Enzyme activity verification

Activity of ACS expressed in SIP3-4 was confirmed using the Nash assay (Figure 3.5). Addition of purified ADH to the extract produced enough formaldehyde to confirm ACS activity.

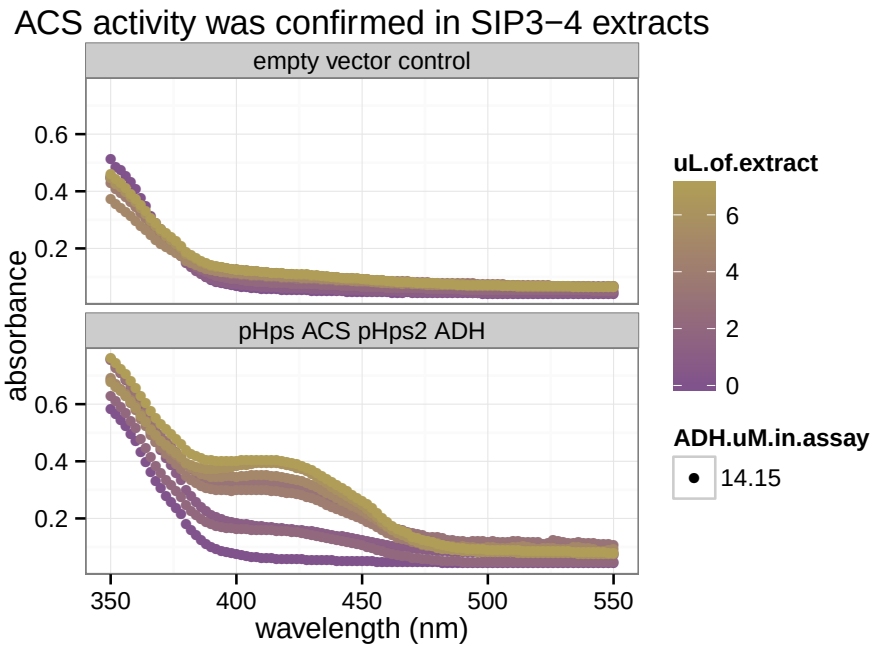

Figure 3.5: Demonstration of ACS activity from from a SIP3-4 extract using the Nash assay. The top panel shows an empty vector control, and the bottom is a strain expressing ACS and ADH. Elevated absorbances in the 400-450 nm range indicate extract-dependent formaldehyde production.

The activity of ADH expressed in SIP3-4 was probed by adding excess purified ACS to SIP3-4 extracts. Though no activity was confirmed, subsequent experiments and calculations showed that the ADH expression was about four times lower than required for the assay conditions tested. Future tests will be done with more concentrated SIP3-4 lysates.

3.3.5 Growth of SIP3-4 on formaldehyde produced by ACS and ADH

No growth has been detected yet, but only one plasmid has been tested (pJ69). In vitro studies suggest ADH is limiting in a SIP3-4 strain expressing this plasmid. New expression plasmids are being constructed for SIP3-4 with altered gene expression levels for both proteins. If none of these new plasmids lead to growth, in vivo ^{13}C assimilation studies will be used to see whether flux into intermediates can be detected.

3.3.6 Preparation of SIP3-4 for formolase selections

Selections involving FLS require a mutant with both copies of hexulose-6-phosphate synthase knocked out. Each copy has been removed in single strains (Matsen, unpublished), and the antibiotic markers were easily removed using *cre* recombinase encoding pCM157 [42]. Attempts to build a double knockout have failed. Slow growth prevented the ability to get single colonies on glucose plates. Glucose plates must be used because a double mutant will be unable to grow on methanol.

By serial passaging SIP3-4 in glucose medium a few times, the doubling time on glucose was improved from 8 days for wild type (Figure 3.3) to about 4 days (Figure 3.3). Further improvement are needed before the double knockout can be made. An attempt to increase the growth rate by expressing a sugar transporter homolog failed (Matsen, unpublished). This test used recombinant expression of an *E. coli* galactose transporter homolog (63% amino acid identity) that was cloned from the SIP3-4 genome. The expression level tested did not lead to noticeably greater growth rates on plates, but higher and lower expression levels could be tested next. Helen Chan, a UW Chemical Engineering undergraduate is working to increase the glucose growth rate by serial passaging of SIP3-4 in glucose medium. Mutagenesis can be accelerated using ethyl methanesulfonate (EMS) [43, 44] or UV mutagenesis [45] if necessary.

3.3.7 Electroporation validation

An electroporation method was developed to easily transform libraries of DNA variants into SIP3-4 and make knockouts. *E. coli* conjugation is typically used to transfer plasmids to methylotrophs, but has disadvantages and requires 2-3 times longer to accomplish than electroporation. Methylotrophic cultures have to be restreaked to purify away the *E. coli*, which linger by consuming SIP3-4 excretions and lysis products. The purification requirement becomes problematic when glucose is used as the substrate for SIP3-4.

Electroporation of SIP3-4 was tested using methods from *E. coli* literature, except cultures are grown on agar plates for ease. Cells were collected off the agar plate, washed once with 10% glycerol, resuspended in fresh 10% glycerol, and electroporated with 1 μL of miniprepped DNA at 1.2 kV in an electroporation cuvette with a 1 cm gap.

Figure 3.6: Plating of a typical single electroporation. The transformation efficiency was greater than 10^4 CFU/ug of DNA in this un-optimized experiment.

3.4 SIP3-4 potential obstacles

3.4.1 Enzymatic loss of formate and formaldehyde

Excessive formate dehydrogenase activity would result in depletion of formate, leaving little for assimilation by the Formolase Pathway. Alternately, flux through the tetrahydromethanopterin (H_4MPT)-linked formaldehyde oxidation pathway may be too high, countering flux through our pathway (Figure 3.7B). In either case, it may be necessary to knock out the linear H_4MPT pathway and force the cells to use the dissimilatory RuMP cycle (Appendix A.3) to gain reducing power [46]. SIP3-4 is predicted to be able to grow without either FDH or the H_4MPT pathway, although glucose supplementation may be necessary.

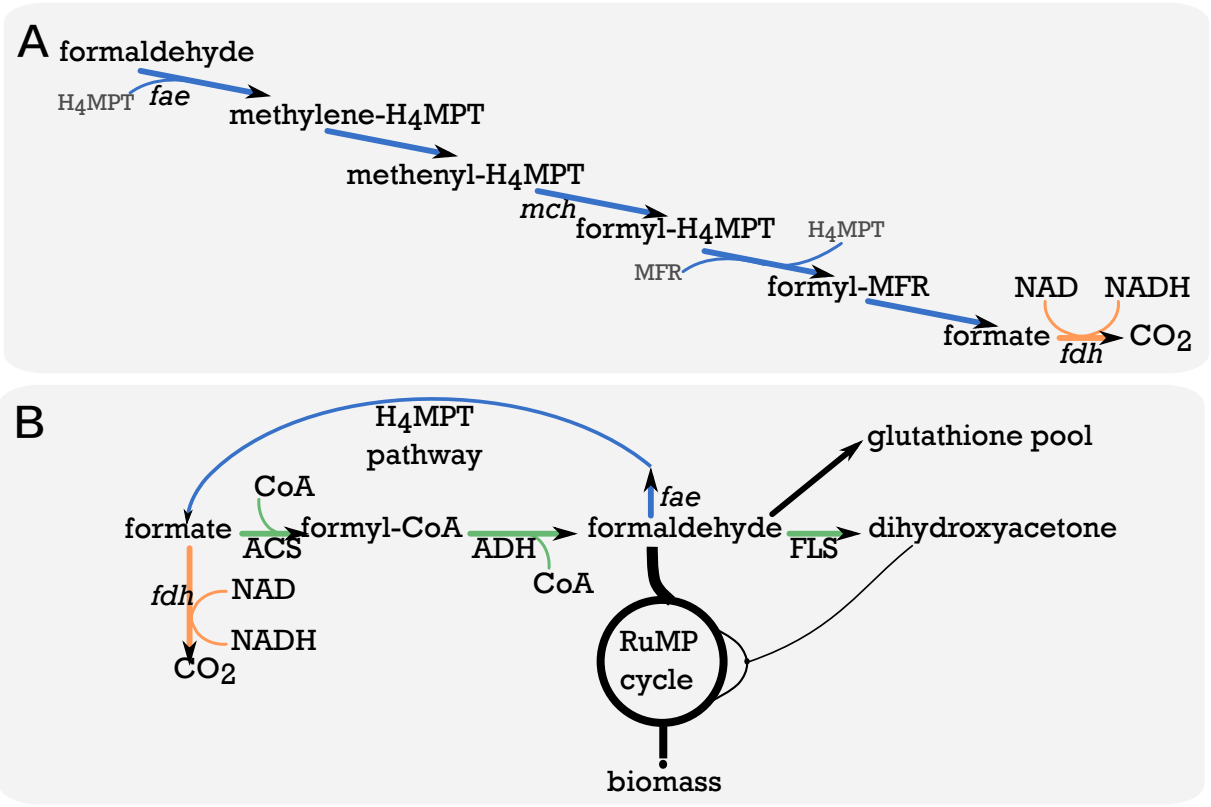

Figure 3.7: (A) The native linear formaldehyde oxidation pathway. (B) Potential for undesirable amounts of flux through H_4MPT pathway or formate dehydrogenase. Enzymes of the Formolase pathway are shown with green reaction arrows. The linear formaldehyde oxidation pathway is depicted with blue arrows, and formate dehydrogenase is shown with an orange arrow. ATP and NADH associated with Formolase Pathway enzymes are omitted for simplicity.

Deletion of the single copy of methenyl H_4MPT cyclohydrolase (*mch*, Figure 3.7A) is the simplest way to knock out the linear formaldehyde oxidation pathway, and has been done previously in a related methylotroph [47]. Two copies of formate dehydrogenase are present; each could be knocked out. Figure 3.7B also depicts the possibility for formaldehyde loss to the glutathione-based detoxification system. Given the known high flux through formaldehyde in this strain, this will likely not be a problem. Nonetheless, knockouts of glutathione synthesis genes could be tested.

3.4.2 Balanced formaldehyde fluxes are required

It is expected that there is some limit on the intracellular formaldehyde concentration that can be tolerated by SIP3-4. Overproduction may overwhelm the formaldehyde homeostasis mechanisms, which are poorly understood in methylotrophs. Consequently, formaldehyde production and consumption rates will likely need to be balanced to some extent. Testing growth under varying levels of gene expression is predicted to be the best way to find conditions with balanced formaldehyde production and consumption rates for all three selection strategies.

3.4.3 No growth despite in vitro enzyme activity

If enzyme activity can be demonstrated by the Nash assay but growth cannot be demonstrated after testing varied expression levels, troubleshooting will be necessary. Varying media composition is the first thing to test. Growth of SIP3-4 in 5mL airtight test tubes has proven useful for testing variants in culture media because cell concentrations can be inferred using a tube spectrophotometer and the media doesn't evaporate over experiments that span weeks.

First it will be important check that formate is supplied at an appropriate rate. The concentration needs to be kept low enough to minimize toxicity but remain high enough to prevent depletion. Formate both benefits growth, and has toxic effects as seen in Figure 3.3. Addition of formate to glucose-grown cultures shortens the lag phase, but the final OD and stability of stationary phase cells are harmed. To maintain low but steady formate concentrations, formate can be added to cultures over time either by addition to culture test tubes (high throughput) or a chemostat (best control). Formate concentrations can be measured by HPLC or enzymatic assay.

If formate availability is validated but pathway-dependent growth is not observed, in vivo flux through the pathway can be assessed via $^{13}C-$labeling of intermediates. The Lidstrom Lab has substantial expertise with observing metabolic flux via ^{13}C labeling studies. Incorporation into either metabolites [48] or protein-derived amino acids [49, 50] can be assessed.

3.4.4 ATP depletion

The toxicity of ACS expression in *E. coli* appears to be a result of imbalances in the ATP and AMP pools, as discussed in Section 2.3. Such an imbalance may be problematic for selections involving ACS in SIP3-4. More ATP can be produced by three means: formate supply, methanol supplementation, and addition of phosphofructokinase. Formate allows for NADH production and thereby ATP production by formate dehydrogenase and the electron transport chain. Limiting methanol could be supplied to selection scheme A (Figure 3.1) with only ACS and ADH. A phosphofructokinase gene can be added to allow an extra mole of ATP to be generated per mole of glucose used.

3.4.5 Unpredicted interference with SIP3-4 metabolism

If for some reason in vivo activity of the Formolase Pathway enzymes cannot be shown, *Methylobacterium extorquens* AM1 can be used as an alternative host. A

formate-tetrahydrofolate ligase (*ftfL*) mutant can be used to prevent growth on formate. If formaldehyde produced by ACS and ADH enhances growth of a *Methylobacterium extorquens* AM1 *ftfL* mutant growing on methylamine, this selection can be used.

Chapter 4

Bacterial Microcompartments

4.1 Encapsulation of the Formolase Pathway in bacterial microcompartments

Many bacteria increase the efficiency of toxic and inefficient metabolic pathways by encapsulating these enzymes in proteinacious shells called bacterial microcompartments (BMCs). These shells are composed of thousands of copies of several types of hexameric and pentameric proteins that self-assemble into 100-200 nm polyhedrals [51, 52]. They are widespread across bacteria, exemplified by 40 genera of bacteria having been found to contain genes that appear to encode microcompartment shells [53]. The three systems that are best understood are listed in Table 4.1.

Table 4.1: The three best studied microcompartment systems

BMC system	**model system**	**metabolic specialty**	**aldehyde intermediate?**
carboxysome	cyanobacteria	overcomes inefficiency of RuBisCO and carbonic anhydrase	no
Pdu	*Salmonella, Citrobacter*	1,2-propanediol	yes
Eut	*E. coli*	ethanolamine utilization	yes

Two of the best studied systems evolved to sequester sequential enzymatic steps with a toxic aldehyde intermediate, which exactly parallels our metabolic system. Encapsulation of the Formolase pathway in such a compartment may elevate flux and reduce cytotoxicity. This could simultaneously increase performance of the Formolase pathway and be a groundbreaking case study of microcompartment use in biotechnology.

The Formolase Pathway will be inserted into the Pdu microcompartment used by *Salmonella* to metabolize 1,2-propanediol, a byproduct of plant cell wall decomposition. Pdu is the best studied microcompartment system with an aldehyde intermediate, and shares some characteristics with the Formolase Pathway (Figure 4.1).

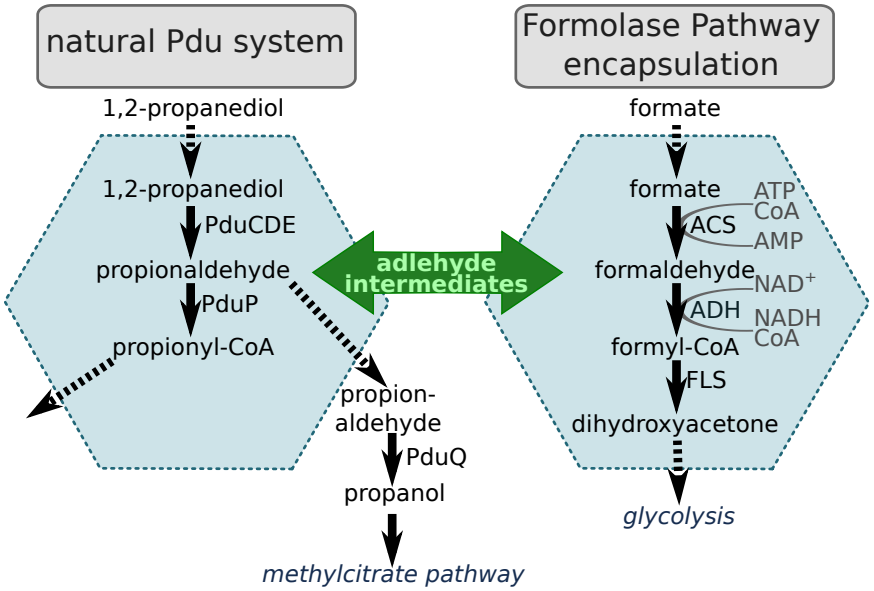

Figure 4.1: Similarities between the natural Pdu System and the proposed system. Both systems include aldehyde intermediates and Coenzyme A activated compounds.

4.1.1 The Pdu compartment is suitable for the Formolase Pathway

Scientists have determined the minimal set of proteins required for recombinant Pdu microcompartment expression [54], and have identified 20-30 amino acid peptides that localize heterologous proteins to the microcompartments [55, 56]. Green fluorescent protein with such an N-terminal peptide sequence appeared to localize inside Pdu BMCs [55, 56]. These peptide sequences will be used to localize FLS and ADH. Encapsulation may enhance pathway performance by elevating the local substrate concentration, reducing formaldehyde toxicity, and reducing the rate of undesirable reactions. ACS may or may not need to be encapsulated.

4.1.2 Progress

The ability to use PduP instead of the ADH 3K9D enzyme used would reduce the number of heterologous enzymes that need to be encapsulated by one. Recombinantly expressed PduP was demonstrated to have activity on formate (Figure 4.2), and can be used instead of ADH in microcompartment research. The enzyme PduP from *Salmonella* was cloned into a high expression inducible pTrc His2C vector and purified. The specific activity and substrate specificity was assayed in parallel with the ADH currently in use.

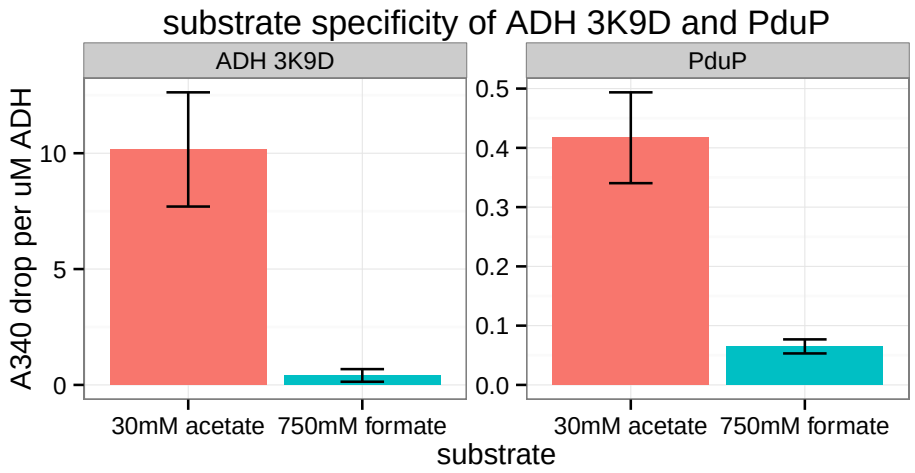

Figure 4.2: Substrate specificities of the ADH currently in use and PduP from the Salmonella enterica Pdu microcompartment shown with different y-scales.

PduP was determined to have better substrate specificity than the ADH currently in use Figure 4.2). Though PduP has about 6x lower activity on formate, this may be an artifact of the His-tag. The tag position and linker length could be optimized. Lower activity of PduP might not be an issue when it is encapsulated inside a microcompartment, where the local substrate concentration is thought to be elevated.

Collaborators at the Tullman-Ercek lab of UC Berkeley are currently working to demonstrate encapsulation of the Formolase pathway enzymes.

Chapter 5

Summary of Recent Accomplishments and Research Plans

5.1 Recent Accomplishments

1. Broke down goal of Formolase Pathway dependent growth into smaller problems
    - Identified ACS as a good target for in vitro engineering.
    - Planned in vivo selections for ACS and ADH that don't require FLS activity.
2. Developed and validated several assays for ACS or ADH library screening and analysis.
    - Can assay enzymes using whole cells, 96-well His-tag purified enzymes, and flask-scale purifications.
    - Two kinetic assays for quantifying enzyme activity and Michaelis-Menten kinetic constants have been validated.
3. Selected ACS L641P as a genetic background to base libraries upon.
    - This mutant is a better starting point for enzyme engineering because it provides more signal per protein.
    - The L641P mutation should allow for the same flux in *E. coli* with lower expression levels and thus enhance pathway performance.
4. Began screening ACS libraries.
    - The whole-cell Nash assay is being used as an initial screen.
    - Up to eight 96-well plates can be screened per week inexpensively.
    - To date, ten plates with random mutations scattered across the whole gene have been screened.
5. Begun development of SIP3-4 for directed evolution experiments
    - Identified 5 promoters of varying strength.
    - Verified formate uptake.

- Validated enzymatic formate concentration quantification.
- Isolated a SIP3-4 strain that grows faster on glucose.
- Knocked out both copies of hexulose-6-phosphate synthase separately.
- Validated electroporation for accelerated genetic work.

6. Begun explorations of bacterial microcompartments
    - Validated use of the native PduP as a substitute for the ADH enzyme currently used.
    - Initiated a collaboration with experts able to test for encapsulation of our enzymes.

5.2 Research Plans and Their Sequence

The ultimate goal is to demonstrate growth of *E. coli* on formate using the Formolase Pathway. Shorter-term goals include evolving individual enzymes in vitro or portions of the pathway in *Methylovorus glucosetrophus* SIP3-4. The work will be carried out approximately in this order.

1. Evolve ACS in vitro.
    - Screen 1000-2000 ACS variants with Helen Chan, an undergraduate researcher. Focus on libraries with randomization restricted to regions close to the active site.
    - Recombine mutations found in promising variants; iterate screening.
2. Test whether improved ACS variants allow for growth in SIP3-4.
3. Demonstrate in vivo flux through ACS and ADH in SIP3-4 (concurrent with ACS evolution).
    - Verify ADH activity in SIP3-4 extracts.
    - Alter ACS and ADH expression levels to increase probability of balanced expression that confers growth.
    - Test for growth using varied expression plasmids in varying media conditions, especially formate and glucose concentrations.
    - Troubleshoot if necessary: check formate depletion, test for ^{13}C incorporation into amino acids.
    - If growth caused by ACS and ADH can not be demonstrated, wait for discovery ACS/ADH enzymes with better catalytic properties from in vitro screens.
    - Test use of PduP as an alternative ADH in vivo
4. Improve ADH in vitro if necessary.
    - Evolve either ADH 3K9D or PduP for enhanced formate activity or substrate specificity.

- Test for higher in vivo performance.

5. Evolve FLS in vivo
    - Obtain a strain that grows better on glucose so a C1 mutant can be made.
    - Knock out the second copy of hexulose-6-phosphate synthase so SIP3-4 can no longer assimilate formate.
    - Test the ability for FLS to complement the double knockout and rescue growth.
    - Select for FLS variants by selecting for cells that can produce dihydroxyacetone from formaldehyde produced by either the native cell metabolism or ACS + ADH.
6. Re-test whether improved pathway enzymes allow *E. coli* to grow on formate.
7. Test use of bacterial microcompartments to increase flux and decrease toxicity.
    - Optimize expression and activity of PduP. Alter the location of His*6 tag and length of the spacer.
    - Verify localization of ACS, ADH, and FLS with collaborators.
    - Verify activity of enzymes modified for encapsulation.
    - Identify whether encapsulation of the pathway enhances in vivo performance.
8. If time allows, evolve the whole Formolase Pathway in chemostat or turbidostat (requires growth).

Chapter 6

Glossary

C1	Single-carbon compounds such as methanol, CO_2, or formate
C2	Two-carbon compounds such as acetate and acetyl-CoA
2P2F	The crystal structure of the acetyl-CoA synthase (ACS) used in the Formolase pathway
3K9D	The crystal structure of the acetaldehyde dehydrogenase (ADH) in the Formolase pathway. Also used to specify which acylating aldehyde dehydrogenase is being specified.
ACS	Abbreviation for the enzyme acetyl-CoA synthase, which converts formate to formyl-CoA
ADH	Abbreviation for the enzyme acetaldehyde dehydrogenase, which converts formyl-CoA to formaldehyde. Unless specified otherwise, it is the *Listeria* enzyme crystallized as Protein Data Bank entry 3K9D.
bacterial microcomparment	(BMC) Widespread bacterial organelles that are made of a protein shell that surrounds and encloses various enzymes
BL21(DE3)	An *E. coli* strain commonly used for protein production
^{13}C formate	Formate with an extra mass unit on the carbon. Useful in mass spectrometry based enzyme assays and detection of in vivo assimilation.
chemolithoautotroph	An autotrophic microorganism that obtains energy by oxidizing inorganic compounds
DHAK	Dihydroxyacetone kinase
dihydroxyacetone kinase	The enzymatic link between the Formolase pathway and glycolysis in *E. coli* metabolism.
electrofuel	An emerging class of carbon-neutral drop-in replacement fuels that are made by storing electrical energy from renewable sources in the chemical bonds of liquid or gas fuels.
electrosynthesis	The synthesis of chemical compounds in an electrochemical cell. This document is focused on microbial electrosynthesis.

ethyl methanesulfonate	(EMS) A chemical mutagen used to intentionally cause mutations in directed evolution experiments.
FocA	A formate transporter
FocB	A formate transporter
FLS	Abbreviation for formolase
formolase	A computationally designed enzyme from the Baker Laboratory at the University of Washington. Converts three formaldehyde molecules into dihydroxyacetone in the Formolase Pathway.
His-tag	(polyhistidine-tag) An amino acid motif in proteins that consists of at least six histidine (His) residues
***Methylobacterium extorquens* AM1**	A model serine cycle methylotroph studied by the Lidstrom lab for decades. This strain could serve as a methylotrophic host for evolving ACS and ADH if the primary SIP3-4 strategy fails.
methylotroph	A diverse group of microorganisms that can use reduced one-carbon compounds, such as methanol or methane, as the carbon source for their growth. This documents discussion is restricted to aerobic gram positive methylotrophs.
***Methylovorus glucosetrophus* SIP3-4**	A RuMP cycle methylotroph that grows on methanol and glucose
microbial electrosynthesis	A form of microbial electrocatalysis in which electrons are supplied to living microorganisms via a cathode in an electrochemical reaction. The microbe uses these electrons to build cell mass and industrially relevant products.
OD	Optical density, a proxy for cell concentration
PDB	Protein data bank (http://www.pdb.org/)
pHps1	A promoter region harvested from the annotated copy of hexulose-6-phosphate in *Methylovorus glucosetrophus* SIP3-4
pHps2	A promoter region harvested from the un-annotated copy of hexulose-6-phosphate in *Methylovorus glucosetrophus* SIP3-4
pTac	Another promoter derived from the *lac* operon of *E. coli*
pTrc	A strong inducible promoter based on the *lac* operation of *E. coli*
RuMP cycle	A methylotrophic metabolic pathway for assimilating methanol. Cellular carbon is assimilated at the oxidation level of formaldehyde.
SIP3-4	Short for *Methylovorus glucosetrophus* SIP3-4, a RuMP cycle methylotroph that can grow on methanol or glucose. Of interest for three in vivo screening strategies for the Formolase Pathway.

serine cycle A methylotrophic pathway for assimilating methanol using formaldehyde as an intermediate; 2 molecules of formaldehyde and 1 molecule of carbon dioxide are utilized in each cycle forming a three-carbon intermediate. http://biocyc.org/META/new-image?object=PWY-1622

Appendix A

Appendix

A.1 Genes used in the Formolase Pathway

gene	sequence
ACS (wild-type)	MSQIHKHTIPANIADRCLINPQQYEAMYQQSINVPDTFWGE QGKILDWIKPYQKVKNTSFAPGNVSIKWYEDGTLNLAANCL DRHLQENGDRTAIIWEGDDASQSKHISYKELHRDVCRFANT LLELGIKKGDVVAIYMPMVPEAAVAMLACARIGAVHSVIFG GFSPEAVAGRIIDSNSRLVITSDEGVRAGRSIPLKKNVDDALK NPNVTSVEHVVVLKRTGGKIDWQEGRDLWWHDLVEQASD QHQAEEMNAEDPLFILYTSGSTGKPKGVLHTTGGYLVYAAL TFKYVFDYHPGDIYWCTADVGWVTGHSYLLYGPLACGATT LMFEGVPNWPTPARMAQVVDKHQVNILYTAPTAIRALMAE GDKAIEGTDRSSLRILGSVGEPINPEAWEWYWKKIGNEKCPV VDTWWQTETGGFMITPLPGATELKAGSATRPFFGVQPALVD NEGNPLEGATEGSLVITDSWPGQARTLFGDHERFEQTYFSTF KNMYFSGDGARRDEDGYYWITGRVDDVLNVSGHRLGTAEI ESALVAHPKIAEAAVVGIPHNIKGQAIYAYVTLNHGEEPSPE LYAEVRNWVRKEIGPLATPDVLHWTDSLPKTRSGKIMRRILR KIAAGDTSNLGDTSTLADPGVVEKLLEEKQAIAMPS
ADH	MSLEDKDLRSIQEVRNLIESANKAQKELAAMSQQQIDTIVKA IADAGYGAREKLAKMAHEETGFGIWQDKVIKNVFASKHVY NYIKDMKTIGMLKEDNEKKVMEVAVPLGVVAGLIPSTNPTS TVIYKTLISIKAGNSIVFSPHPNALKAILETVRIISEAAEKAGCP KGAISCMTVPTIQGTDQLMKHKDTAVILATGGSAMVKAAYS SGTPAIGVGPGNGPAFIERSANIPRAVKHILDSKTFDNGTICA SEQSVVVERVNKEAVIAEFRKQGAHFLSDAEAVQLGKFILRP NGSMNPAIVGKSVQHIANLAGLTVPADARVLIAEETKVGAK IPYSREKLAPILAFYTAETWQEACELSMDILYHEGAGHTLIIH SEDKEIIREFALKKPVSRLLVNTPGALGGIGATTNLVPALTLG CGAVGGSSSSDNIGPENLFNIRRIATGVLELEDIREGGS

FLS MAMITGGELVVRTLIKAGVEHLFGLHGIHIDTIFQACLDHDV
PIIDTRHEAAAGHAAEGYARAGAKLGVALVTAGGGFTNAV
TPIANARTDRTPVLFLTGSGALRDDETNTLQAGIDQVAMAA
PITKWAHRVMATEHIPRLVMQAIRAALSAPRGPVLLDLPWD
ILMNQIDEDSVIIPDLVLSAHGAHPDPADLDQALALLRKAER
PVIVLGSEASRTARKTALSAFVAATGVPVFADYEGLSMLSGL
PDAMRGGLVQNLYSFAKADAAPDLVLMLGARFGLNTGHG
SGQLIPHSAQVIQVDPDACELGRLQGIALGIVADVGGTIEAL
AQATAQDAAWPDRGDWCAKVTDLAQERYASIAAKSSSEH
ALHPFHASQVIAKHVDAGVTVVADGGLTYLWLSEVMSRVK
PGGFLCHGYLNSMGVGFGTALGAQVADLEAGRRTILVTGD
GSVGYSIGEFDTLVRKQLPLIVIIMNNQSWGWTLHFQQLAVG
PNRVTGTRLENGSYHGVAAAFGADGYHVDSVESFSAALAQ
ALAHNRPACINVAVALDPIPPEELILIGMDPFAGSTENLYFQS
GALE

A.2 *Methylovorus glucosetrophus* SIP3-4 metabolism in more detail

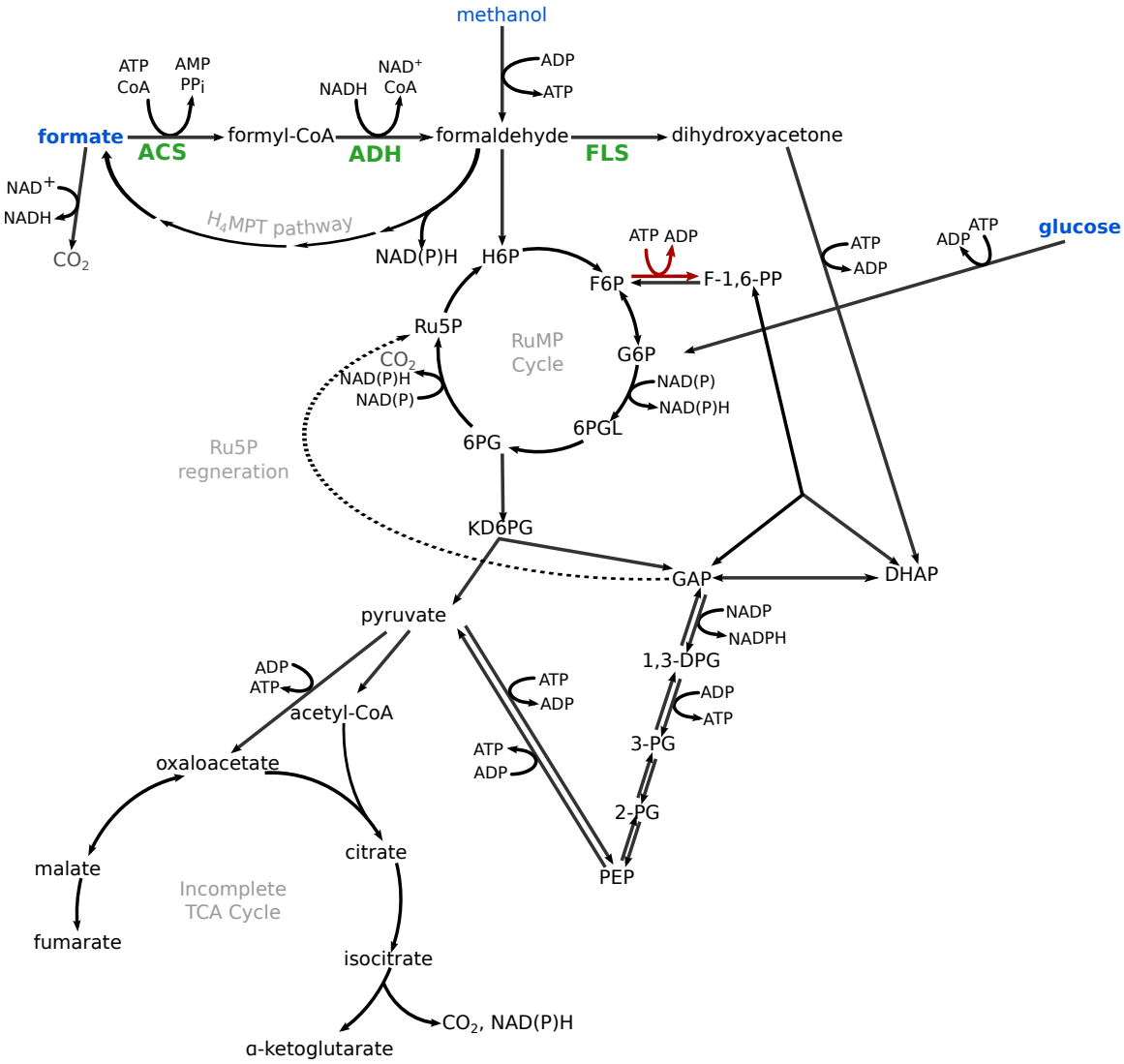

Figure A.1: *Methylovorus glucosetrophus* SIP3-4 metabolism in more detail, with Formolase Pathway Enzymes depicted in green. The arrow in red represents phosphofructokinase, which is missing in the SIP3-4 genome but could be added.

A.3 The dissimilatory RuMP cycle for formaldehyde oxidation

Below is an image of the assimilatory and dissimilatory RuMP cycles [47].

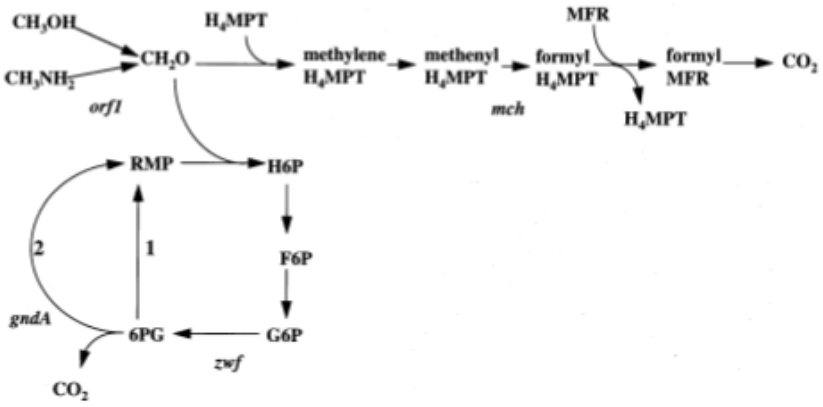

Fig. 1. Major C_1 metabolic pathways in *M. flagellatus* KT. Methanol and methylamine are oxidized by corresponding oxidative systems to produce formaldehyde, which is then either bound by H_4MPT or reacts with ribulose 5-phosphate (RMP). Assimilatory (1; for complete sequence of reactions see Anthony, 1982) and dissimilatory (2) RMP cycles branch at the level of 6-phosphogluconate (6PG). Genes targeted in this study are indicated. H6P, hexulose 6-phosphate, F6P, fructose 6-phosphate, G6P, glucose 6-phosphate, MFR, methanofuran.

Figure A.2: The assimilatory and dissimilatory RuMP cycles for formaldehyde assimilation and oxidation, directly from [47].

A.4 Promoters tested in *Methylovorus glucosetrophus* SIP3-4

There are two copies of hexulose-6-phosphate synthase in the *Methylovorus glucosetrophus* SIP3-4 genome. Promoter sequences from both were harvested. pHps1 indicates DNA upstream of the annotated copy. pHps2 represents the DNA before the start codon of the unannotated copy, which is misannotated as a protein of unknown function translated in the reverse direction. See Figure 3.2 for their strengths when used in SIP3-4 grown on glucose.

promoter	base pairs	sequence
pHps1	177	ACTTAAAAGCGATTATGATGCTGCCTAATGATG CATTTAAGGTGCATCATGGCAAAATATTGCTAA AATTCAACCGATTTACCCTATTCACAAGAAAA AATGTATCATTACGGAGTGAGTAAATCATAAC CCTGTAACTTAAGCGGAGGAAGTATCGTGGCA CAACCATCAGTACAA
pHps2	250	GGTTTCGTGGTAGACCAAAACAAGCTGGAAAA GGCACTGGCTGAAAAGCTGTAATATTGCTGACT TGAATAGCGCTAATAATCAATTTATTAGCGCTA TTCACTTTATTCAAGTAGTACATAAGTGCTATT GGATTTACCTAGCTTAACGTTGACATCAACTTT GATTCCAACTTAAGATAGTTTTTCTAAAAACTA AGCACTTAATTTTTTAAACCAACTGGAGGAAGT ACATCGTGGCATTAACTCAA
pHps2 trimmed	150	CACTTTATTCAAGTAGTACATAAGTGCTATTGG ATTTACCTAGCTTAACGTTGACATCAACTTTGA TTCCAACTTAAGATAGTTTTTCTAAAAACTAAG CACTTAATTTTTTAAACCAACTGGAGGAAGTAC ATCGTGGCATTAACTCAA
pTrc	223	CTGTTGACAATTAATCATCCGGCTCGTATAATG TGTGGAATTGTGAGCGGATAACAATTTCACAC AGGAAACAGCGCCGCTGAGAAAAAGCGAAGC GGCACTGCTCTTTAACAATTTATCAGACAATCT GTGTGGGCACTCGACCGGAATTATCGATTAACT TTATTATTAAAAATTAAAGAGGTATATATTAAT GTATCGATTAAATAAGGAGGAATAAACC
pTac	61	GAGCTGTTGACAATTAATCATCGGCTCGTATAA TGTGTGGTCACACAGGAAACAGAATTCT

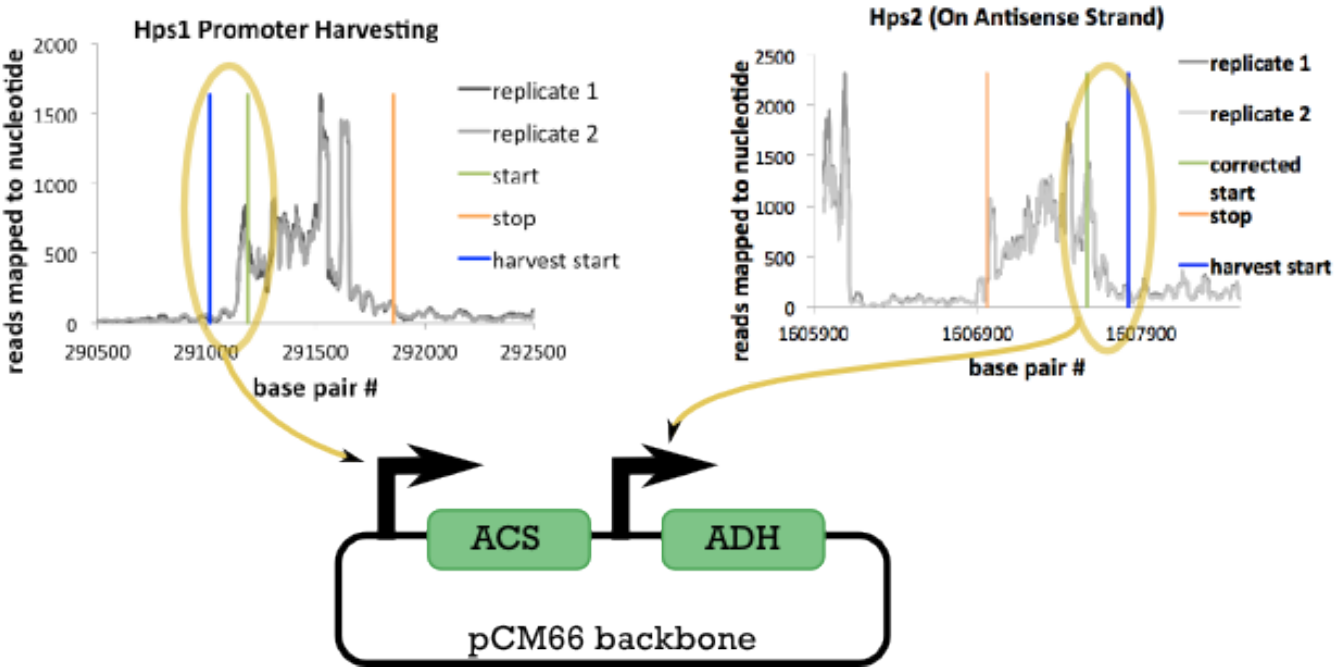

Figure A.3: Harvesting promoter regions for a SIP3-4 ACS ADH expression plasmid pJ69.

Plasmid files that show the promoter in context can be found here:

pHps1: (pJ71)

- https://github.com/JanetMatsen/Plasmids/tree/master/73%20-%20pHps1%20RFP%20pCM66T

pHps2: (pJ66)

- https://github.com/JanetMatsen/Plasmids/tree/master/66%20-%20pHps2%20RFP%20pCM66T

pHps2 trimmed: (pJ77)

- https://github.com/JanetMatsen/Plasmids/tree/master/77%20-%20150%20bp%20pHps2%20RFP%20pCM66T

pTrc: (pJ72)

- https://github.com/JanetMatsen/Plasmids/tree/master/72%20-%20pTrc%20RFP%20pCM66T

pTac: (pJ71)

- https://github.com/JanetMatsen/Plasmids/tree/master/71%20-%20pTac%20RFP%20pCM66T

no promoter control: (pJ70)

- https://github.com/JanetMatsen/Plasmids/tree/master/70%20-%20no%20promoter%20RFP%20pCM66T

RFP amino acid sequence:
MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLP
FAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQD
SSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMR
LKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEG
RHSTGA**

Bibliography

[1] 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*, 2015.

[2] Robert J Conrado, Chad A Haynes, Brenda E Haendler, and Eric J Toone. Electrofuels: A new paradigm for renewable fuels. In *Advanced Biofuels and Bioproducts*, pages 1037–1064. Springer, 2013.

[3] Han Li, Paul H Opgenorth, David G Wernick, Steve Rogers, 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*, 335(6076):1596, Mar 2012.

[4] Korneel Rabaey and René A Rozendal. Microbial electrosynthesis - revisiting the electrical route for microbial production. *Nat Rev Microbiol*, 8(10):706–16, Oct 2010.

[5] Scott Banta and Alan West. Electrofuel production using genetically engineered acidithiobacillus ferrooxidans. In *Meeting Abstracts*, number 50, pages 2290–2290. The Electrochemical Society, 2014.

[6] Miriam A Rosenbaum and Alexander W Henrich. Engineering microbial electrocatalysis for chemical and fuel production. *Curr Opin Biotechnol*, 29:93–8, Oct 2014.

[7] Aaron S Hawkins, Yejun Han, Hong Lian, Andrew J Loder, Angeli L Menon, Ifeyinwa J Iwuchukwu, Matthew Keller, Therese T Leuko, Michael WW Adams, and Robert M Kelly. Extremely thermophilic routes to microbial electrofuels. *Acs Catalysis*, 1(9):1043–1050, 2011.

[8] B Innocent, D Liaigre, D Pasquier, F Ropital, J-M Léger, and KB Kokoh. Electro-reduction of carbon dioxide to formate on lead electrode in aqueous medium. *Journal of Applied Electrochemistry*, 39(2):227–232, 2009.

[9] 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. *Biochim Biophys Acta*, 1827(8-9):1039–47, 2013.

[10] Pamela P Peralta-Yahya, Fuzhong Zhang, Stephen B del Cardayre, and Jay D Keasling. Microbial engineering for the production of advanced biofuels. *Nature*, 488(7411):320–8, Aug 2012.

[11] 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, 2014.

[12] Wei Lü, Juan Du, Nikola J Schwarzer, Elke Gerbig-Smentek, Oliver Einsle, and Susana L A Andrade. The formate channel foca exports the products of mixed-acid fermentation. *Proc Natl Acad Sci U S A*, 109(33):13254–9, Aug 2012.

[13] 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, Nov 2009.

[14] V J Starai and J C Escalante-Semerena. Acetyl-coenzyme a synthetase (amp forming). *Cell Mol Life Sci*, 61(16):2020–30, Aug 2004.

[15] 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. *Gene*, 162(1):99–104, Aug 1995.

[16] Justin Bloomfield Siegel. *Computational Enzyme Design: Engineering a Novel Catalyst, Protein Therapeutic, and Carbon Fixation Pathway*. PhD thesis, University of Washington, 2011.

[17] R Z Jin and E C Lin. An inducible phosphoenolpyruvate: dihydroxyacetone phosphotransferase system in escherichia coli. *J Gen Microbiol*, 130(1):83–8, Jan 1984.

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

[19] Annik Nanchen, Tobias Fuhrer, and Uwe Sauer. Determination of metabolic flux ratios from 13c-experiments and gas chromatography-mass spectrometry data. In *Metabolomics*, pages 177–197. Springer, 2007.

[20] 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. *Mol Microbiol*, 80(1):168–83, Apr 2011.

[21] Richard J Hopkinson, Philippa S Barlow, Christopher J Schofield, and Timothy D W Claridge. Studies on the reaction of glutathione and formaldehyde using nmr. *Org Biomol Chem*, 8(21):4915–20, Nov 2010.

[22] 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. *Nat Chem Biol*, 5(8):593–9, Aug 2009.

[23] 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.

[24] Ryan E Cobb, Ran Chao, and Huimin Zhao. Directed evolution: past, present, and future. *AIChE Journal*, 59(5):1432–1440, 2013.

[25] T NASH. The colorimetric estimation of formaldehyde by means of the hantzsch reaction. *Biochem J*, 55(3):416–21, Oct 1953.

[26] Albert S Reger, Jill M Carney, and Andrew M Gulick. Biochemical and crystallographic analysis of substrate binding and conformational changes in acetyl-coa synthetase. *Biochemistry*, 46(22):6536–46, Jun 2007.

[27] John R Williamson and Barbara E Corkey. Assays of intermediates of the citric acid cycle and related compounds by fluorometric enzyme methods. *Methods in enzymology*, 13:434–513, 1969.

[28] Suman Kumari, Christine M Beatty, Douglas F Browning, Stephen JW Busby, Erica J Simel, Galadriel Hovel-Miner, and Alan J Wolfe. Regulation of acetyl coenzyme a synthetase in escherichia coli. *Journal of bacteriology*, 182(15):4173–4179, 2000.

[29] Stefan Schmelz and James H Naismith. Adenylate-forming enzymes. *Curr Opin Struct Biol*, 19(6):666–71, Dec 2009.

[30] Andrew M Gulick, Xuefeng Lu, and Debra Dunaway-Mariano. Crystal structure of 4-chlorobenzoate:coa ligase/synthetase in the unliganded and aryl substrate-bound states. *Biochemistry*, 43(27):8670–9, Jul 2004.

[31] Andrew Currin, Neil Swainston, Philip J Day, and Douglas B Kell. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem Soc Rev*, Dec 2014.

[32] Kerstin Steiner and Helmut Schwab. Recent advances in rational approaches for enzyme engineering. *Comput Struct Biotechnol J*, 2:e201209010, 2012.

[33] F William Studier. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*, 41(1):207–34, May 2005.

[34] 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. *J Biol Chem*, 280(28):26200–5, Jul 2005.

[35] V J Starai, I Celic, R N Cole, J D Boeke, and J C Escalante-Semerena. Sir2-dependent activation of acetyl-coa synthetase by deacetylation of active lysine. *Science*, 298(5602):2390–2, Dec 2002.

[36] Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, 3rd, and Hamilton O Smith. Enzymatic assembly of dna molecules up to several hundred kilobases. *Nat Methods*, 6(5):343–5, May 2009.

[37] Oriana Salazar and Lianhong Sun. Evaluating a screen and analysis of mutant libraries. *Methods Mol Biol*, 230:85–97, 2003.

[38] C Anthony. *Biochemistry of methylotrophs*. Academic Press Inc., 1982.

[39] Alla Lapidus, Alicia Clum, Kurt Labutti, Marina G Kaluzhnaya, Sujung Lim, David A C Beck, Tijana Glavina Del Rio, Matt Nolan, Konstantinos Mavromatis, Marcel Huntemann, Susan Lucas, Mary E Lidstrom, Natalia Ivanova, and Ludmila Chistoserdova. Genomes of three methylotrophs from a single niche reveal the genetic and metabolic divergence of the methylophilaceae. *J Bacteriol*, 193(15):3757–64, Aug 2011.

[40] Tanya Barrett, Karen Clark, Robert Gevorgyan, Vyacheslav Gorelenkov, Eugene Gribov, Ilene Karsch-Mizrachi, Michael Kimelman, Kim D Pruitt, Sergei Resenchuk, Tatiana Tatusova, Eugene Yaschenko, and James Ostell. Bioproject and biosample databases at ncbi: facilitating capture and organization of metadata. *Nucleic Acids Res*, 40(Database issue):D57–63, Jan 2012.

[41] Alexey Vorobev, David A C Beck, Marina G Kalyuzhnaya, Mary E Lidstrom, and Ludmila Chistoserdova. Comparative transcriptomics in three methylophilaceae species uncover different strategies for environmental adaptation. *PeerJ*, 1:e115, 2013.

[42] Christopher J Marx and Mary E Lidstrom. Broad-host-range cre-lox system for antibiotic marker recycling in gram-negative bacteria. *Biotechniques*, 33(5):1062–7, Nov 2002.

[43] Yu-Ping Lai, Jing Huang, Lin-Fa Wang, Jun Li, and Zi-Rong Wu. A new approach to random mutagenesis in vitro. *Biotechnol Bioeng*, 86(6):622–7, Jun 2004.

[44] G A Sega. A review of the genetic effects of ethyl methanesulfonate. *Mutat Res*, 134(2-3):113–42, 1984.

[45] Jeffrey L Bose. Chemical and uv mutagenesis. *Methods Mol Biol*, Feb 2015.

[46] Ludmila Chistoserdova. Modularity of methylotrophy, revisited. *Environ Microbiol*, 13(10):2603–22, Oct 2011.

[47] L Chistoserdova, L Gomelsky, J A Vorholt, M Gomelsky, Y D Tsygankov, and M E Lidstrom. Analysis of two formaldehyde oxidation pathways in methylobacillus flagellatus kt, a ribulose monophosphate cycle methylotroph. *Microbiology*, 146 (Pt 1):233–8, Jan 2000.

[48] Song Yang, Martin Sadilek, Robert E Synovec, and Mary E Lidstrom. Liquid chromatography-tandem quadrupole mass spectrometry and comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry measurement of targeted metabolites of methylobacterium extorquens am1 grown on two different carbon sources. *J Chromatogr A*, 1216(15):3280–9, Apr 2009.

[49] Annik Nanchen, Tobias Fuhrer, and Uwe Sauer. Determination of metabolic flux ratios from 13c-experiments and gas chromatography-mass spectrometry data: protocol and principles. *Methods Mol Biol*, 358:177–97, 2007.

[50] Nicola Zamboni, Sarah-Maria Fendt, Martin Rühl, and Uwe Sauer. (13)c-based metabolic flux analysis. *Nat Protoc*, 4(6):878–92, 2009.

[51] Todd O Yeates, Michael C Thompson, and Thomas A Bobik. The protein shells of bacterial microcompartment organelles. *Curr Opin Struct Biol*, 21(2):223–31, Apr 2011.

[52] Todd O Yeates, Cheryl A Kerfeld, Sabine Heinhorst, Gordon C Cannon, and Jessup M Shively. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat Rev Microbiol*, 6(9):681–91, Sep 2008.

[53] Shouqiang Cheng, Yu Liu, Christopher S Crowley, Todd O Yeates, and Thomas A Bobik. Bacterial microcompartments: their properties and paradoxes. *Bioessays*, 30(11-12):1084–95, Nov 2008.

[54] 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. *Mol Cell*, 38(2):305–15, Apr 2010.

[55] 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. *Proc Natl Acad Sci U S A*, 107(16):7509–14, Apr 2010.

[56] Chenguang Fan and Thomas A Bobik. The n-terminal region of the medium subunit (pdud) packages adenosylcobalamin-dependent diol dehydratase (pducde) into the pdu microcompartment. *J Bacteriol*, 193(20):5623–8, Oct 2011.