RICE UNIVERSITY

Synthetic Metabolic Pathways for Efficient Utilization of One-Carbon (C1) Compounds

By

Seung Hwan Lee

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Sibani Biswal (Nov 28, 2022 16:45 CST)

Sibani Lisa Biswal

Professor, Chemical and Biomolecular Engineering, Materials Science and NanoEngineering

Ramon Gonzalez (Nov 29, 2022 09:20 EST)

Ramon Gonzalez

Adjunct Professor, Chemical and Biomolecular Engineering

Ross Thyer (Nov 28, 2022 17:54 CST)

Ross Thyer

Assistant Professor, Chemical and Biomolecular Engineering

Laura Segatori (Nov 29, 2022 09:17 CST)

Laura Segatori

Professor, Bioengineering, Chemical and Biomolecular Engineering, and BioSciences

HOUSTON, TEXAS

November 2022

RICE UNIVERSITY

Synthetic Metabolic Pathways for Efficient Utilization of One-Carbon (C1) Compounds

by

Seung Hwan Lee

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Ramon Gonzalez, PhD, Thesis Director
Adjunct Professor, Chemical and Biomolecular Engineering

Sibani Lisa Biswal, PhD, Chair
Professor, Chemical and Biomolecular Engineering, Materials Science and NanoEngineering

Ross Thyer, PhD
Assistant Professor, Chemical and Biomolecular Engineering

Laura Segatori, PhD
Professor, Bioengineering, Chemical and Biomolecular Engineering, and BioSciences

HOUSTON, TEXAS
December 2022

Abstract

Synthetic Metabolic Pathways for Efficient Utilization of One-Carbon (C1) Compounds

by

Seung Hwan (Allen) Lee

One-carbon (C1) compounds derived from waste gases such as carbon dioxide, carbon monoxide and methane start to be recognized as carbon feedstock in the field of metabolic engineering and industrial biotechnology. Numerous enzymes and pathways have been identified and engineered for efficient C1 assimilation into multi-carbon molecules and the list continues to increase at an unprecedented pace with advances in synthetic biology. This thesis aims to provide two synthetic pathways into the list, based on new-to-nature biochemistries, each having unique characteristics and advantages over the preexisting pathways. Specifically, there are four chapters in the thesis: first chapter provides a comprehensive review on C1-utilizing enzymes and metabolic pathways, both natural and synthetic, with comments on cross-platform capabilities and industrial applications and highlighting the pathway dependency to the host metabolism. Second chapter introduces synthetic C1 utilization pathways named Formyl-CoA Elongation (FORCE) pathways. FORCE pathways operate in an orthogonal manner to the host metabolism, exemplified by the abilities to generate products directly from C1 compounds in a growth-decoupled bioconversion in *Escherichia coli*. Also, FORCE pathways' potential to be harnessed in a synthetic C1-trophy

segregating C1 assimilation and native substrate utilization is demonstrated in a two-strain co-culture system. Third chapter discusses approaches to improve FORCE pathway flux by identifying and engineering more efficient variants of the key condensation enzyme, 2-hydroxyacyl-CoA synthase (HACS), which is also identified as a major rate limiting step in the pathways. A variant with more than 10-fold improvement in activity was discovered, which was applied in the pathway to demonstrate significantly improved product titer, rate, and yield. Fourth chapter explores the journey to engineer *E. coli* to utilize a non-native substrate methylsuccinate, a metabolic precursor from oxygen-independent methane activation via fumarate addition. Combination of rational pathway design and adaptive laboratory evolution is used to achieve a strain growing efficiently on methylsuccinate as sole carbon source, which could be used as a selection platform to screen for methane activation enzymes and ultimately as a chassis for synthetic methanotrophy.

Acknowledgements

I would like to first thank Dr. Ramon Gonzalez and funding from NSF, DOE, Rice University and University of South Florida for financial support during my PhD research. It has been dynamic 5.5 years with moving from Houston to Tampa, COVID-19, and marriage but has been a great journey thanks to people around me and good science that has driven me to get to work every day.

Dr. Gonzalez has been great in keeping me motivated with new ideas while providing freedom in pursuing my own independent research. He also trusted me in being involved in many different projects which ultimately made me a better scientist. I would also like to thank him for supporting me to attend different conferences. Watching talks and networking with inspirational people in the field have been instrumental in expanding my views in research and shaping my career post PhD.

I would also like to thank my former and current lab mates for all the good memories. Special thanks to Alex for the fundamental training in molecular biology and more. Also, James for being a great role model in being a good scientist. Thank you Biki and Andy for hosting me with good foods every time I visit Houston and especially Biki for all the great scientific discussion. Thank you Fayin for collaborations and guidance in experiments.

My PhD would not have been possible without the support from Rice CHBE department. I would like to thank Dr. Michael Wong, Barbara Windish, Rice OISS and others who allowed me to continue research as a Rice PhD student while physically away from campus. I would also like to thank Dr. Lisa Biswal for serving as my committee chair.

Most importantly, I want to thank my wife, Seolyeong, and my parents for always supporting me.

Table of contents

List of Figures

List of Tables

Introduction

It is becoming more evident that climate change, primarily caused by anthropogenic carbon emission, is affecting our lives. Over the past few years, more and more concrete scientific data are piling up with regards to the environmental[1], social[2], cultural[3] and economic[4] cost of climate change. Scientists emphasize that it is imperative to transform the current fossil fuel-based infrastructure to carbon-free, sustainable energy and manufacturing to circumvent the imminent disaster. Thanks to rapid deployment of renewable energy, considerable progress has been made in the energy infrastructure and other sectors via electrification, notably light-duty transportation[5]. However, power generation from 100% renewable sources like solar and wind cannot address all carbon emissions because there are sectors that are hard to be replaced by electricity such as heavy-duty transportation and aviation or require carbon as feedstock such as chemical industries. Biomanufacturing, having its basis on sustainable feedstock and enzymatic catalysis taking place under mild temperature and pressure, emerged as a promising alternative to traditional fuel and chemical production. Starting from pharmaceutical, natural product and food additives[6] to commodity chemical building blocks, fuels, textile and materials, industrial biotechnology has shown massive potential in wide range of product profiles[7].

One of the key advantages of biomanufacturing is its flexibility in utilizing different feedstock. Traditional petrochemical industry relies on oil and gas as raw materials and feedstock for chemical production is restrained to several building blocks, such as ethylene, propylene and BTX[8]. Therefore, substituting feedstock to sustainable carbon sources like CO_2 is extremely challenging in the current chemical industry. On the other hand, biomanufacturing leverages biochemical reactions catalyzed by enzymes that can harness myriads of different organic molecules. Although traditional industrial biotechnology relies heavily on industrial organisms

utilizing sugars and other biomass-derived feedstock, recent advances in synthetic biology and metabolic engineering have greatly expanded the feedstock profile of biomanufacturing ranging from industrial off-gas[9] to waste plastics[10] for value-added product synthesis in an industrial scale.

Among sustainable carbon sources that can be harnessed via biomanufacturing, one-carbon (C1) molecules are particularly interesting as they can be efficiently generated from CO_2 via electrochemical or enzymatic reactions. While there are numerous autotrophic organisms that can natively fix CO_2, they need external energy sources like light, hydrogen and/or chemical energy to utilize CO_2 and the efficiency is substantially lower than electrochemical reduction[11]. However, C1 molecules with higher degrees of reduction, such as methane, methanol, formaldehyde and formate generated from CO_2 via electrochemical reduction can be utilized by organisms with higher efficiency than direct CO_2 fixation[12]. One of the key advantages of utilizing reduced C1 compounds is that they do not require external energy sources and associated machineries to convert the sources to cellular energy. For example, C1 assimilation pathways utilizing methanol or formate have been successfully transferred to non C1-trophic *Escherichia coli* (*E. coli*) for synthetic methylotrophy[13, 14] and formatotrophy[15] without the need for external energy source. Model organisms like *E. coli* have the advantages of high cell growth rate and density, as well as plenty of genetic tools and product synthesis pathways demonstrated and available for them.

This thesis aims to introduce synthetic C1 utilization pathways demonstrated in *E. coli* that are designed to outcompete natural pathways in the context of (1) pathway independence from the host metabolism that allows modular control of the pathway and high flux from C1 substrate to product, and (2) kinetically and thermodynamically more efficient for C1 activation and

assimilation. In specific, this thesis consists of four main chapters, starting with a literature review on known C1 utilization enzymes and assimilation pathways. The next two chapters discuss synthetic C1 utilization pathways named Formyl-CoA Elongation (FORCE) pathways that are designed to bypass host metabolism and directly generate multi-carbon products from C1 compounds, attributing to better control of the pathway flux and higher product yields. A major challenge with synthetic pathways building upon new-to-nature chemistries is identifying enzymes that exhibit high kinetic parameters for the desired non-native reactions. Chapter 3 discusses the work on identifying and engineering highly efficient FORCE pathway enzymes via bioprospecting and protein engineering, thereby achieving substantially improved product titer, rate, and yield. Finally, the last chapter discusses a synthetic methylsuccinate metabolism demonstrated in *E. coli*. Methylsuccinate is a key metabolic precursor of oxygen-independent assimilation of methane via fumarate addition. Although still a postulation, this reaction provides substantially more energy efficient yet thermodynamically favorable route for methane utilization than native pathways. The engineered *E. coli* strain growing efficiently on methylsuccinate can serve as a steppingstone to a synthetic methanotrophy, which operates with significantly higher carbon and energy efficiency than native methanotrophs for diverse product synthesis from methane.

Chapter 1. Overview of one-carbon (C1) utilizing enzymes and C1 metabolism

1. C1 utilizing enzymes

One-carbon (C1) compounds encompass gaseous methane, CO_2, and carbon monoxide as well as water-soluble methanol, formaldehyde, and formic acid at different redox states (Fig. 1-1). Nitrogen-containing C1 compounds such as methylamine and nitromethane are not included in the subject of this study. This section focuses on the efficiency and cross-platform/chassis transferability of the C1 utilizing enzymes in the context of biomanufacturing using C1 feedstock. Details of individual enzyme structures and catalytic mechanisms are not in the scope of this study and can be found in other studies and reviews.

1.1. Interconversion among C1 compounds

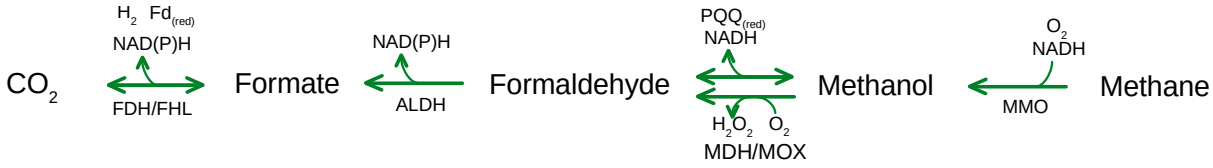


Figure 1-1. Enzymatic conversion among C1 compounds at different redox states. C1 compounds from oxidized (left) to reduced (right) states can be biochemically interconverted by oxidoreductases utilizing various redox cofactors including but not limited to molecular hydrogen (H_2), ferredoxin (Fd), NAD(P)/NAD(P)H, pyrroloquinoline quinone (PQQ) and molecular oxygen (O_2). FDH: formate dehydrogenase; FHL: formate hydrogen-lyase; ALDH: aldehyde dehydrogenase; MDH: methanol dehydrogenase; MOX: methanol oxidase; MMO: methane monooxygenase.

1.1.1. Methane oxidation to methanol

Biological methane oxidation to methanol is catalyzed by methane monooxygenases (MMO) (Fig. 1-1). There are two types of MMO, which are soluble MMO (sMMO) and particulate MMO (pMMO). Although both catalyze the same reaction, their structures, cofactor requirements and mechanisms are very different[16]. pMMO is more abundant in nature and there are many species of methanotrophs that have only pMMO or both pMMO and sMMO[17]. As the name

implies, sMMO is a soluble protein, which makes it relatively easier to purify and analyze than membrane-bound pMMO. On the other hand, pMMO has better access to methane due to its location on the membrane[18]. One of the key limitations of MMO is its requirement of molecular oxygen (O_2) as electron acceptor, leading to loss of reducing equivalent. sMMO requires two electrons in the form of NAD(P)H to cleave O_2 molecule to make radical oxygens, which are very strong electron acceptors[19]. The oxygen radical then rapidly oxidizes methane to form methanol[16]. In turn, the overall reaction becomes highly exothermic necessitating extra cooling demand in industrial scale fermentation[20]. Co-utilization of explosive methane and oxygen could also cause safety concerns in large scale facilities. Heterologous expression of both pMMO[21] and sMMO[22] has been demonstrated although pMMO was only shown *in vitro* after assembly in a biosynthetic scaffold[21]. Functional expression of sMMO was shown *in vivo* in *E. coli* by co-expressing chaperone proteins with methanol oxidation rate of up to 20 mg methanol/gDCW/h[22].

1.1.2. Methanol oxidation to formaldehyde

Enzymes catalyzing methanol oxidation to formaldehyde are largely categorized into three: NAD^+-dependent methanol dehydrogenases (MDH), PQQ (pyrroloquinoline quinone)-dependent MDH, and O_2-dependent methanol oxidases (MOX/AOX)[23] (Fig. 1-1). Methanol oxidation coupled with NAD^+ reduction to NADH is thermodynamically unfavorable in ambient temperature[24, 25], leading to suboptimal kinetics of NAD^+-dependent MDHs. Most native NAD^+-dependent MDHs exhibit low affinity and turnover number toward methanol with K_m in the order of 100 mM and k_{cat} at 0.1-0.3 s^{-1}[23]. Therefore, native methylotrophs harboring NAD^+-dependent MDHs grow on high temperature, where the NAD^+-dependent methanol oxidation becomes more thermodynamically favorable[26], and utilize MDH activator enzymes (ACT)

which improves the catalytic efficiency of NAD^+-dependent MDH through conformational change of NAD^+-MDH complex to enhance the electron transfer[27]. Improvement in kinetic parameters of NAD^+-dependent MDHs by adding ACT was demonstrated *in vitro* but not in non-methylotrophic host *in vivo*[23]. Alternatively, several ACT-independent MDHs were identified and engineered to improve the catalytic efficiencies with the best reported enzyme having catalytic efficiency (k_{cat}/K_m) of 66.8 $M^{-1}s^{-1}$[27, 28]. Despite poor kinetics, NAD^+-dependent MDHs are widely employed in engineering synthetic methylotrophs using model organisms like *E. coli* due to their abilities to utilize universal reducing cofactor, NAD^+/NADH.

PQQ-dependent MDHs are present in the periplasm of gram negative methylotrophs [23]. Unlike NAD^+-dependent MDHs, PQQ-dependent MDHs have higher affinity toward methanol than longer chain alcohols, with K_m as low as 0.02 mM [24]. As a result, PQQ-dependent MDHs exhibit superior catalytic efficiency than NAD^+-dependent MDHs[29]. However, use of PQQ-dependent MDH in heterologous host has not been reported likely due to the requirement of unnatural redox cofactor. Some PQQ-dependent MDHs utilize rare earth metal ions like Lanthanides (Ln^{3+}) as cofactors[29].

Methanol oxidases (alcohol oxidases) are found in methylotrophic yeasts such as *Pichia pastoris* (*Komagataella phaffii*). Unlike MDHs, MOX utilizes molecular oxygen as electron acceptor to produce hydrogen peroxide, which is often coupled with catalase to form water. As highly electronegative oxygen is used as electron acceptor, the reaction is thermodynamically downhill. As a result, MOXs have high catalytic efficiency (k_{cat}/K_m of *P. pastoris* AOX1 = 10,000 $M^{-1}s^{-1}$)[30]. However, the electrons lost by water forming reaction cannot be recovered, meaning methanol oxidation reaction cannot be coupled to downstream reduction to form energy-containing molecules. Nevertheless, MOXs have been expressed in other non-methylotrophic yeasts like

Saccharomyces cerevisiae[31] and *Yarrowia lipolytica*[32] to utilize methanol. Moreover, leveraging its high catalytic efficiency, it was used in a synthetic C1 utilization pathway *in vitro* coupled with downstream formaldehyde utilization enzyme with poor affinity to drive the reaction[33].

1.1.3. Formaldehyde oxidation to formate

While formaldehyde is a key metabolite in several C1 utilization pathways, it is also generated as a byproduct from diverse metabolic activities found in all organisms[34]. Formaldehyde is a highly toxic compound that can cause irreversible DNA-protein crosslinking[13]. To rapidly remove formaldehyde, nature developed various formaldehyde detoxification mechanisms, most notably glutathione-dependent oxidation to formate. Formaldehyde and glutathione spontaneously form S-(hydroxymethyl)-glutathione. S-(hydroxymethyl)-glutathione dehydrogenase is an extremely efficient (k_{cat}/K_m in the order of 10^6 $M^{-1}s^{-1}$) enzyme that catalyzes NAD^+-dependent oxidation of S-(hydroxymethyl)-glutathione to S-formyl-glutathione. This formaldehyde detoxification mechanism is present in many different classes of organisms encompassing mammals, plants and bacteria[35]. Some organisms utilize different forms of thiols instead of glutathione, such mycothiol[36] from actinomycetes and bacillithiol[37] from several gram-positive bacteria. Glutathione-independent formaldehyde dehydrogenases are also identified which are unique to closely related *Pseudomonas putida* and *Pseudomonas aeruginosa*[38] (Fig. 1-1).

1.1.4. Formate oxidation to CO_2

Formate dehydrogenase (FDH) catalyzes reversible formate oxidation reaction to CO_2. Various redox cofactors including NAD^+[39], $NADP^+$[40], ferredoxin/$NAD(P)^+$[41], quinones[42], cytochrome c[43], coenzyme F420[44] and H_2[45] are identified to participate in the electron

transfer coupled to the reaction (Fig. 1-1). Oxidation of formate to CO_2 could be coupled with energy generation in the cells, including proton motive force[42] and reducing equivalents[39, 40, 44]. Under oxygen-limited conditions, formate oxidation coupled with reduction of electron acceptors, such as nitrate[42], sulfate[43] and hydrogen ion[45], was demonstrated. Autotrophic acetogens and methanogens can fix CO_2 using reverse FDH reaction under reductive environment. Various electron donors, including H_2 via electron-bifurcating hydrogenase, ferredoxin and mixture of ferredoxin and NAD(P)H are shown to be utilized for CO_2 reduction reaction in acetogens[46]. In methanogens, on the other hand, CO_2 is initially bound to C1 carrier, methanofuran (MFR) followed by reduction to formyl-MFR mediated by ferredoxin as a redox cofactor[47, 48].

1.2. C1 activation to C1 carriers

While enzymes that catalyze interconversion between C1 compounds at different redox states exist, only formaldehyde and CO_2 are the two metabolic nodes that are directly assimilated into the metabolic pathways for biomass production. Nature developed numerous carboxylases and associated CO_2 fixation pathways presumably due to the electrophilic nature and atmospheric abundance of CO_2. On the other hand, despite the strong reactivity of formaldehyde, there exist only two native metabolic pathways that assimilate C1 at formaldehyde node, D-ribulose 5-phosphate (RuMP) and D-xylulose 5-phosphate (XuMP) cycles. All other native C1 metabolism involves activated C1 via transfer to C1 carrier molecules, including but not limited to tetrahydrofolate (THF, H_4F), coenzyme M (CoM), coenzyme A (CoA), methanofuran (MFR), tetrahydromethanopterin (THMPT, H_4MPT) and tetrahydrosarcinapterin (THSPT, H_4SPT). Activated C1 is likely preferred (1) to avoid toxicity of C1 compounds, especially formaldehyde; and (2) to enhance enzyme catalytic efficiency and specificity toward the desired substrate by

attaching small C1 residue to bulky carrier molecules. As a result, many native and synthetic C1 assimilation pathways use activated C1 molecules, notably THF as a carrier for Wood-Ljungdahl pathway[46], Serine cycle[49] and Reductive Glycine pathway[50], CoM and THMPT/THSPT as carriers for methanogenesis and reverse methanogenesis[51] and CoA as a carrier for Formyl-CoA Elongation (FORCE) pathways[25] (Figure 1-2).

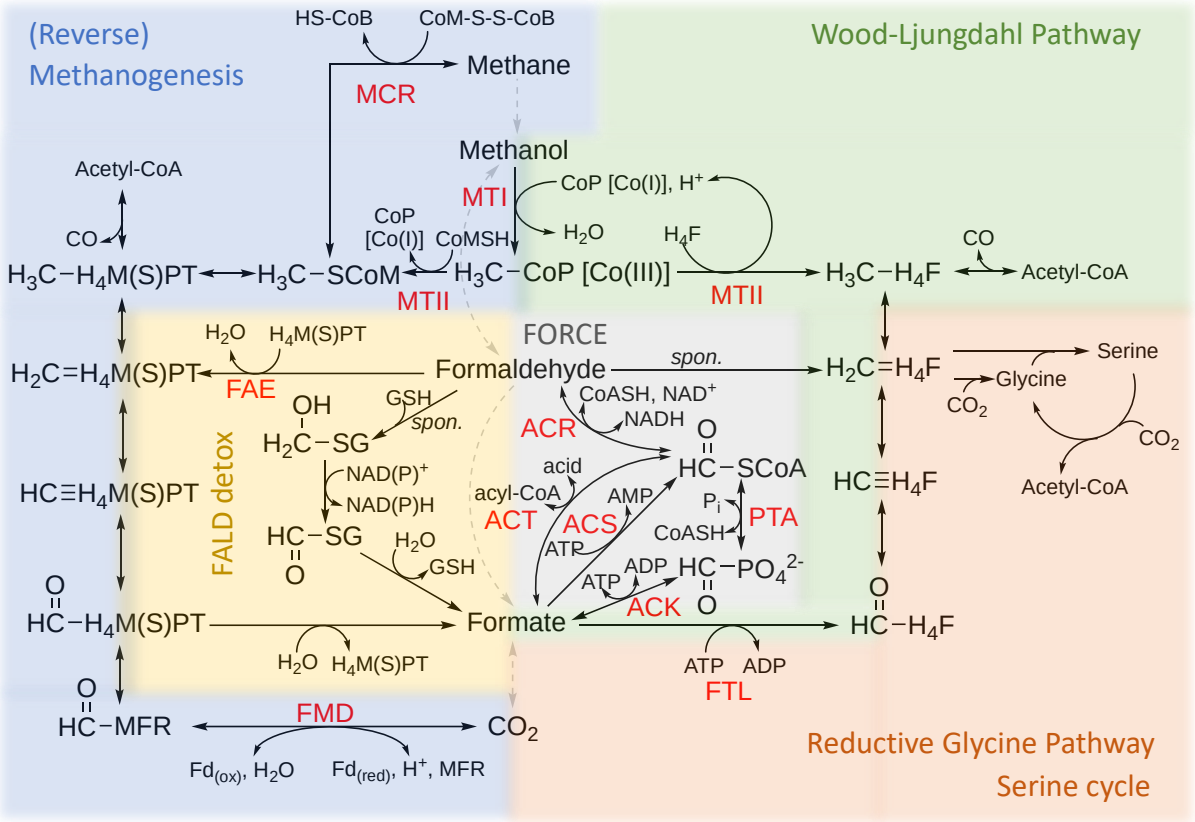


Figure 1-2. C1 metabolism that involves activation of C1 to C1 carrier molecules. Largely, methanogens utilize CoM, H4MPT/H4SPT and MFR as C1 carriers for (reverse) methanogenesis; acetogens utilize THF as C1 carriers for Wood-Ljungdahl Pathway; Reductive Glycine Pathway and Serine cycle also involve either formaldehyde or formate activation to methylene-THF and formyl-THF, respectively; FORCE pathway utilizes formyl-CoA as C1 elongation unit for iterations[25]. Appropriate C1 activation enzymes to C1 carriers are highlighted in red font. Abbreviations: *spon.*: spontaneous reaction; FALD detox: formaldehyde detoxification (oxidation) pathway.

1.2.1. Methane to methyl-CoM

Methyl-coenzyme M reductase (MCR) is an enzyme responsible for the final step of methanogenesis, as well as the initial activation step of methane in the reverse methanogenesis. Anaerobic methane-oxidizing archaea (ANME) capable of anaerobic oxidation of methane (AOM) via reverse methanogenesis were first identified from a consortium containing sulfate-reducing bacteria (SRB)[52]. Because ANME and SRB share interdependent metabolism, isolation of pure ANME has not been successful[48]. However, experimental validation of the methanogenesis in reverse direction has been demonstrated in vitro[53] and in vivo via heterologous expression of MCR[54].

MCR catalyzes reversible conversion from methane and heterodisulfide of coenzyme M and coenzyme B (CoM-S-S-CoB) to methyl-CoM and Coenzyme B (Fig. 1-2). It has been established that electrons required for methyl-CoM reduction to methane come from hydrogen via electron transfer from methanophenazine-dependent hydrogenase to CoM-S-S-CoB reductase, using methanophenazine as electron carrier[48]. In case of AOM, it is hypothesized that electrons from methane oxidation are transferred to SRB, nitrate, or metals via multiheme cytochrome c proteins (MHCs), but the exact mechanism is not yet clear[48]. Coenzyme F_{430} is the key redox cofactor involved in MCR activity and its biosynthesis in *E. coli* was recently demonstrated[55]. A study shows that MCR is highly specific toward CoM and CoB, which are not commonly present in organisms other than methanogens and ANMEs[56]. However, some bacterial species utilizing CoM in alkene metabolism[57] and others harboring genes for CoM biosynthesis were identified[58] which opens possibility for transferring this unique chemistry into more tractable organisms. However, an inherent challenge with methane activation via MCR is the unfavorable thermodynamics with $\Delta G = +30$ kJ/mol, which results in MCR being the rate limiting step during

reverse methanogenesis[53]. As a result, MCR exhibits high K_m ranging from 1.1 to 37 mM (equivalent to 3 MPa) methane, which is the likely condition in the deep marine sediments where ANME grows[48].

1.2.2. Methanol to methyl-CoM/THF

While aerobic methanotrophs and methylotrophs typically utilize methanol via oxidation to formaldehyde followed by assimilation via RuMP or Serine cycle, anaerobes such as methanogenic archaea or acetogenic bacteria utilize methanol via methyl group transfer to C1 carrier such as coenzyme M (CoM, methanogens) and tetrahydrofolate (THF, acetogens) (Fig. 1-2)[59]. A key advantage with this route is that it avoids formation of highly toxic formaldehyde. The methyltransferase system consists of two enzymes: (1) methyltransferase I (MTI) which cleaves the C-O bond in methanol and attaches cobalamin-binding corrinoid protein (CoP) as an intermediate C1 carrier; (2) methyltransferase II (MTII) which transfers methyl group from CoP to THF or CoM. CoP[Co(I)] is highly reactive and readily oxidizes to inactive Co(II) state, which can be reactivated by ATP-consuming activating enzymes (AE) to Co(I)[60]. There is limited information about the kinetics of this mechanism but the growth rate of a methylotrophic acetogen, *Eubacterium limosum*, on methanol as sole carbon source was reported to have a doubling time of 8 hours[61].

1.2.3. Formaldehyde to S-(hydroxymethyl)-glutathione/mycothiol/bacillithiol

Spontaneous (non-enzymatic) formaldehyde activation and transfer to glutathione, mycothiol or bacillithiol as C1 carriers occurs as a part of the formaldehyde detoxification pathway present in all classes of organisms as discussed in Section 1.1.3.

1.2.4. Formaldehyde to methylene-H_4MPT/THF

Condensation of formaldehyde and THF and H_4MPT to methylene-THF and methylene-H_4MPT, respectively, are known to occur spontaneously[62, 63]. However, in aerobic methanotrophs and methylotrophs, where there is a constant supply of formaldehyde from MDH activities (especially from highly efficient PQQ-dependent MDHs), the rate of spontaneous condensation could be insufficient to detoxify formaldehyde[64]. To address this, *Methylobacterium extorquens* AM1 expresses formaldehyde-activating enzyme (FAE), which facilitates the formation of methylene-H_4MPT from formaldehyde and H_4MPT (Fig. 1-2). This enzyme was shown to be essential for growth on methanol[64]. Intriguingly, FAE has no activity with THF even though *M. extorquens* AM1 assimilates C1 at the methylene-THF node instead of methylene-H_4MPT via Serine cycle[64]. This finding supported by other experimental and computational analysis points that formate, not formaldehyde, is the main branch point for methanol assimilation, where formaldehyde is first converted to formate via H_4MPT-dependent formaldehyde detoxification pathway, followed by formate activation to formyl-THF and reduction to methylene-THF to enter the Serine cycle[65]. Thus, approaches to heterologously incorporate C1-THF-based pathways into non-C1 trophic hosts like *E. coli* all designated formate as a node for C1 activation[66, 67] (catalyzed by formyate-THF ligase discussed in Section 1.2.6). Unlike H_4MPT, THF is a commonly available C1 carrier in all organisms and serves as a point of C1 assimilation in several C1 utilization pathways, which makes these pathways good candidates to be transferred into non-native hosts.

1.2.5. Formaldehyde to formyl-CoA

Acyl-CoA reductases (ACRs, or acylating aldehyde dehydrogenases) catalyze reversible oxidation of aldehyde to acyl-CoA typically utilizing $NAD(P)^+$ as electron acceptor. ACRs and

alcohol dehydrogenases (ADHs) catalyze terminal reactions in the fermentative pathway, where short-chain acyl-CoAs, such as acetyl-CoA and butyryl-CoA are reduced to aldehydes and eventually to alcohols. There are both enzymes only with ACR activities and bifunctional enzymes that have both ACR and ADH activities such as *E. coli* AdhE. Among enzymes that only have ACR activities, acyl-CoA reductase from *Listeria monocytogenes* (LmACR) was shown to have activity for formaldehyde oxidation[68] and ACR from *Parageobacillus thermoglucosidasius* (PtACDH) and *Bacillus methanolicus* MGA3 (BmACDH) showed activities for formyl-CoA reduction[69] (Fig. 1-2). As the primary substrate for these enzymes are not likely C1, they have relatively low catalytic efficiency, with the best reported value of 95 $M^{-1}s^{-1}$ with formaldehyde[68].

1.2.6. Formate to formyl-THF

Formate-THF ligase catalyzes reversible activation of formate to formyl-THF with expense of one ATP equivalent (ADP-forming) (Fig. 1-2). Although THF is a common C1 carrier involved in many metabolisms including amino acid and nucleic acid synthesis, FTL in not present in non-C1 trophic organisms like *E. coli* presumably because there is no need to have extra supply of C1-THF from formate. In the C1 assimilation pathways, FTL serves as the C1 activation enzyme for formatotrophy in both aerobic Serine cycle/Reductive Glycine pathway (rGlyP) and anaerobic Wood-Ljungdahl pathway. Heterologous expression of FTL in *E. coli* was demonstrated to enable formate consumption in *E. coli* via Serine cycle[67] and rGlyP[70]. There is limited information available on the kinetics of this enzyme but the K_m of FTL with formate is in the range of 5-20 mM from aerobic methylotrophs and anaerobic acetogens[71].

1.2.7. Formate to formyl-CoA

Formate activation to formyl-CoA can be catalyzed by three different categories of enzymes: an acyl-CoA transferase (ACT) that can transfer CoA from acyl-CoA to formate, an

acyl-CoA synthetase (ACS) that activates formate to formyl-CoA with expense of two ATP equivalents (AMP-forming) and an acetate kinase-phosphoacyltransferase (ACK-PTA) pair which first activates formate to formyl-phosphate with expense of one ATP equivalent (ADP-forming), followed by CoA transfer to replace phosphate[25] (Fig. 1-2). All these routes are demonstrated experimentally. Some promising candidates for ACT are formyl-CoA transferase from *Oxalobacter formigenes* (OfFrc), which was originally identified to use oxalyl-CoA[72] as a donor and later identified to use succinyl-CoA[25] as well, and 4-hydroxybutyrate CoA-transferase from *Clostridium aminobutyricum* (CaAbfT) which was shown to utilize glycolyl-CoA as donor with high affinity (K_m = 12 μM)[73]. Several ACSs were identified to have formate activation activity with K_m ranging from 50 to 150 mM and k_{cat} ranging from 5 to 10 s^{-1}[69]. In case of ACK, ACK from *Clostridium cylindrosporum* (CcAck) was reported to have formate activation activity, but it suffers from high K_m of 400 mM[74]. Although ACT can be an effective route if CoA can be recycled efficiently from downstream acyl-CoA, the activity depends on the availability of CoA donors, which are not as abundant as free CoA, and despite high affinity, low turnover number (0.4 s^{-1} for CaAbfT using glycolyl-CoA as donor) could make this step rate limiting. For ATP-dependent activation, ACK-PTA is more ATP-efficient but less thermodynamically favorable than ACS, leading to weaker driving force. For cell-free implementation of formate activation, an ATP-regenerating module could be required for constant ATP supply, while in vivo, surplus reducing equivalent coupled with oxidative phosphorylation can be an alternative option. Functional expression of ACS in vivo requires either mutation in the ACS residue responsible for acetylation[75] or deletion of native deacetylase gene (*patZ* in *E. coli*)[69] to prevent post-translational repression of the enzyme activity.

1.2.8. CO_2 to formyl-MFR

Formyl-methanofuran dehydrogenase (FMD) is the first step in methanogenesis that catalyzes the reduction of CO_2 and condensation with methanofuran (MFR) to form formyl-MFR (Fig. 1-2). Hydrogen and formate can provide electrons required for the reaction through flavin-based electron bifurcation catalyzed by heterodisulfide reductase (Hdr)[76]. While the reaction is an endergonic process (ΔG = +16 kJ/mol), it is coupled with exergonic methyl-CoM reductase (MCR) reaction (ΔG = - 30 kJ/mol) to compensate the unfavorable thermodynamics[77]. Based on recently solved protein structure of FMD, it was proposed that FMD forms a multi-enzyme complex with Hdr and formate dehydrogenase (FDH) to efficiently channel electrons to the CO_2 reduction reaction[76].

1.3. Enzymes catalyzing C-C coupling reactions using C1 or C1 carrier as a substrate

In the previous section, various C1 utilizing enzymes that can either interconvert between C1 compounds at different redox states or activate C1 compounds to C1 carrier are explored. While interconversion and activation of C1 compounds have important roles in metabolism, such as redox cofactor recycling, electron transfer and formaldehyde detoxification, carbon-carbon (C-C) coupling is an essential step in C1-trophy that assimilates C1 into the cell metabolism for biomass production and/or value-added product synthesis. There are several enzymes that catalyze C-C coupling between two C1 residues, and many more existing and proposed mechanisms that attaches C1 compound or carrier into multi-carbon compounds. Enzymatic CO_2 assimilation catalyzed by carboxylases and associated CO_2 fixation pathways are not covered in this thesis, but comprehensive reviews on this topic can be found elsewhere[78, 79].

1.3.1. Carbon monoxide dehydrogenase (CODH)-acetyl-CoA synthase (ACS) complex

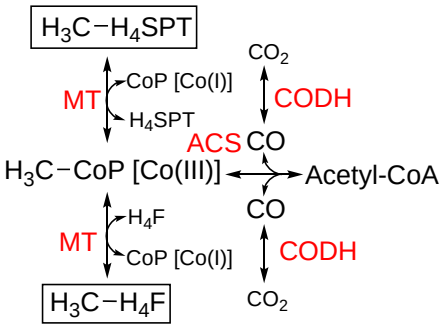


Figure 1-3. Methyl-THF/THSPT and CO2 reaction to form acetyl-CoA catalyzed by CODH-ACS complex in anaerobic acetogens. MT: methyltransferase; ACS: acetyl-CoA synthase; CODH: carbon monoxide dehydrogenase.

Carbon monoxide dehydrogenase-acetyl-CoA synthase (CODH-ACS) complex catalyzing C-C coupling reaction between a C1 moiety, methyl-THF (or methyl-THSPT), and a C1 compound, CO_2 (reduced to CO as an intermediate), is one of the two identified naturally occurring enzymes that catalyze C1-C1 coupling reactions. It is known to be one of the oldest enzymatic complexes[80] present in acetogens and some methanogenic archaea. CODH-ACS catalyzes the key reaction in acetogenic bacteria to assimilate C1 compounds, such as CO_2 and CO to generate biomass[81], whereas in certain methanogens, such as *Methanosarcina Barkeri*, it is primarily employed in a reverse direction to dissimilate acetate into methane and CO_2 via methanogenesis[82]. It is also hypothesized that some methanogens growing on C1 as sole carbon source utilize the same CODH-ACS complex to assimilate C1 for growth[83].

As seen in Figure 1-3. the reaction takes place in three steps: (1) cobalamin-binding corrinoid protein (CoP) transfer to the methyl-THF/THSPT catalyzed by methyltransferase (this step is similar to methyl transfer step during methanol activation described in Section 1.2.2); (2) CO_2 reduction to CO mediated by redox cofactor such as NAD(P)H or ferredoxin; (3) C-C bond formation followed by condensation with free CoA to form acetyl-CoA[84]. As most acetogens

and methanogens are anaerobes, many enzymes and cofactors in their metabolic pathways are also oxygen sensitive including CODH-ACS and associated cofactors like CoP. Moreover, some cofactors are unique to acetogens and methanogens, potentially limiting the transfer of the pathway to heterologous platforms. Nevertheless, if CODH-ACS can be implemented in synthetic C1 utilization pathways with help of advanced protein and platform engineering, countless biosynthetic pathways can be derived from this core reaction as acetyl-CoA is the key metabolic precursor for diverse value-added bioproducts.

1.3.2. Glycine cleavage complex (GCV) and Serine hydroxymethyltransferase (SHMT)

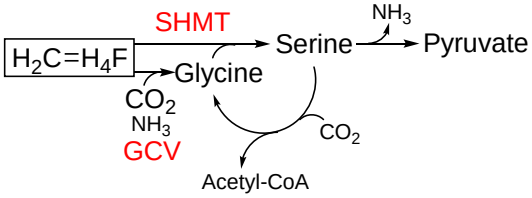


Figure 1-4. Methylene-THF, CO2 and NH3 to form glycine catalyzed by GCV and methylene-THF and glycine to form serine catalyzed by SHMT. GCV: glycine cleavage complex; SHMT: serine hydroxymethyltransferase.

While methyl-THF is the key metabolic node for anaerobic C1 assimilation via CODH-ACS complex, methylene-THF, an oxidation product of methyl-THF, serves as a node for oxygen-tolerant C1 assimilation routes via glycine cleavage complex (GCV) and serine hydroxymethyltransferase (SHMT)[85] (Fig. 1-4). Both enzymes exist in heterotrophic organisms incapable of C1 assimilations, with primary role of interconverting two essential amino acids, glycine and serine. SHMT serves as the entry point for a number of native and engineered C1 assimilation pathways including Serine cycle[49], Modified Serine cycle[67] and Serine threonine cycle[70]. GCV is the key component of the Reductive Glycine pathway (rGlyP) which condenses two C1 units, methylene-THF and CO_2 into C2 amino acid, glycine. GCV, along with CODH-ACS, are the only two naturally occurring C1-C1 coupling reactions but most organisms natively

harboring GCV and SHMT including *E. coli* cannot support growth on C1 sources via rGlyP with native enzymes[85]. However, literature shows that recombinant expression of heterologous formyl-THF ligase (FTL) enables *E. coli* to utilize formate via rGlyP and partial Serine cycle [66, 67, 70]. It was also shown that further pathway optimization of rGlyP[15] and Serine threonine cycle[86] and adaptive laboratory evolution enabled *E. coli* strain to grow on formate as sole carbon source.

1.3.3. 2-hydroxyacyl-CoA synthase (HACS) (carbonyl group + formyl-CoA)

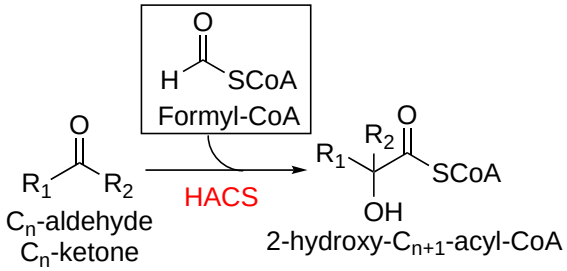


Figure 1-5. Carbonyl compound condensation with formyl-CoA to form 2-hydroxyacyl-CoA catalyzed by HACS. HACS: 2-hydroxyacyl-CoA synthase.

2-hydroxyacyl-CoA synthase (HACS) is a thiamine-diphosphate (TPP) dependent enzyme that catalyzes condensation of carbonyl group with formyl-CoA to form 2-hydroxyacyl-CoA (Fig. 1-5). It was first identified from the α-oxidation of long chain fatty acids in mammalian metabolism, which cleaves a long chain fatty acyl-CoA into an aldehyde and a formyl-CoA. HACS also catalyzes the condensation reaction in reverse direction and has wide substrate specificity including not only aldehydes with varying chain lengths but also ketones such as acetone[68]. A wide substrate range of HACS allows condensation between C1 (formyl-CoA) and a carbonyl compound with varying chain lengths and functional groups (R_1 and R_2 in Fig. 1-5), which can be further converted to other value-added products via α-reduction and termination pathways[25]. A major bottleneck of the HACS-based pathways is poor kinetics of HACS especially with short chain aldehydes and ketones[68]. Oxalyl-CoA decarboxylase, an enzyme from the same TPP-

dependent enzyme family, was successfully engineered to have HACS activity with improved specificity and catalytic efficiency (k_{cat}/K_m = 400 M^{-1} s^{-1}) toward formaldehyde and formyl-CoA condensation[73]. Utilizing an uncommon substrate like formyl-CoA provides advantages that there is no competing endogenous enzyme to deplete this molecule. However, it also poses a challenge that there is limited information on enzymes that produces this molecule efficiently.

1.3.4. Aldolases utilizing formaldehyde as a substrate

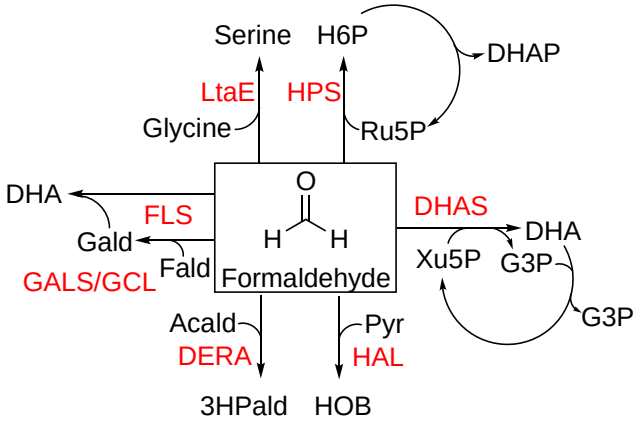


Figure 1-6. Various aldolases utilizing formaldehyde as substrate. LtaE: L-threonine aldolase; HPS: 3-hexulose-6-phosphate synthase; DHAS: dihydroxyacetone synthase; HAL: 4-hydroxy-2-oxohutanoate aldolase; DERA: deoxyribose-5-phosphate aldolase; GALS: glycolaldehyde synthase; GCL: glyoxylate carboligase; FLS: formolase.

There are enzymes that form carbon-carbon bond directly on C1 molecules without the need for C1 carrier. Other than CO_2, formaldehyde is the only molecule that undergoes such reaction presumably because it is a highly reactive compound unlike other C1 molecules. However, its extreme toxicity forces high catalytic efficiency of formaldehyde assimilation enzymes to maintain low formaldehyde concentration. Two native enzymes, 3-hexulose-6-phosphate synthase (HPS) from RuMP cycle and dihydroxyacetone synthase (DHAS) from XuMP cycle, are shown to have the characteristics of low K_m and high k_{cat} to ensure rapid detoxification of formaldehyde generated from methanol oxidation[87, 88] (Table 1-1). However, they utilize 5-carbon D-ribulose

5-phosphate and D-xylulose 5-phosphate as co-substrates, requiring efficient regeneration of these molecules in the downstream pathways (Fig. 1-6).

Several other enzymes are reported to have promiscuous activities with formaldehyde, which are then combined with downstream reactions to make synthetic C1 assimilation pathways. *E. coli* L-threonine aldolase (LtaE) can catalyze reversible serine cleavage reaction to glycine and formaldehyde at near 1/60 catalytic efficiency of the original L-threonine cleavage reaction[89] (Fig. 1-6). Several enzymes including *E. coli* YfaU[90] and KHB[91] were found to catalyze aldol reaction between formaldehyde and pyruvate to form 2-keto-4-hydroxybutyrate (or 4-hydroxy-2-oxohutanoate (HOB)) (Fig. 1-6). However, these HOB aldolases (HAL) suffer from low affinity with formaldehyde, which could limit their use in vivo. Enzymes catalyzing aldol reaction between acetaldehyde and formaldehyde were identified and a variant from thermophilic organism was found to have improved tolerance with formaldehyde[92]. This enzyme also has decent affinity with formaldehyde at K_m of 2.54 mM (Table 1-1).

There have been efforts in protein engineering to make synthetic C1 utilization enzymes. A notable example is formolase (FLS), a computationally designed and engineered enzyme using benzaldehyde lyase (BAL) as a starting reference[93]. FLS catalyzes two consecutive reactions using formaldehyde as the only substrate: two formaldehyde aldol reaction to form glycolaldehyde, and subsequent glycolaldehyde and formaldehyde aldol reaction to form dihydroxyacetone (DHA) (Fig. 1-6). FLS was further engineered using directed evolution from several groups[33, 94, 95] with limited success. One of the major drawbacks of FLS is that it produces mixture of glycolaldehyde and DHA and the ratio changes with varying formaldehyde concentrations[96]. Moreover, DHA, a product of FLS, is a three-carbon molecule which suffers from loss of C1 to form acetyl-CoA, which makes the overall pathway less carbon efficient. To address this, a similar

FLS-like enzyme with higher specificity toward glycolaldehyde production was engineered using benzoylformate decarboxylase (BFD) as a reference[97]. The engineered glycolaldehyde synthase (GALS) also suffers from high K_m with formaldehyde (170 mM), which prevents its use in vivo[97] (Table 1-1). Another engineered version of *E. coli* glyoxylate carboligase showed GALS activity with improved K_m at 18 mM but suffers from low turnover number at 0.1 s^{-1} (Table 1-1). Overall, engineered FLS and GALS provide an efficient route for multi-carbon product synthesis directly from C1 compounds, but catalytic efficiency of these enzymes at current stage is substantially below the level needed to detoxify formaldehyde for cell survival or support growth on C1.

Table 1-1. Summary of kinetic parameters for formaldehyde utilizing aldolases

Name	Organism	Co-substrate	Kinetics with formaldehyde			Ref.
			K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})	
3-hexulose-6-phosphate synthase (HPS)	*Methylomicrobium alcaliphilum* 20Z	D-ribulose 5-phosphate	0.98	45.8	46800	[87]
Dihydroxyacetone synthase (DHAS)	*Candida boidinii*	D-xylulose 5-phosphate	0.43			[88]
L-threonine aldolase (LtaE)	*E. coli*	Glycine				[89]
2-keto-4-hydroxybutyrate aldolase (KHB)	*E. coli*	Pyruvate	500>			[91]
Deoxyribose-5-phosphate aldolase (DERA)	*Thermotoga maritima*	Acetaldehyde	2.54			[92]
Formolase (FLS)	Engineered (*Pseudomonas fluorescens biovar I*)	Formaldehyde, Glycolaldehyde			4.7	[93]
Glycolaldehyde synthase (GALS)	Engineered (*Pseudomonas putida*)	Formaldehyde	170	1.58	9.3	[97]
Glyoxylate carboligase (GCL)	Engineered (*E. coli*)	Formaldehyde	18	0.1	5.2	[98]

1.3.5. Other postulated radical mechanisms to activate C1 molecules

C1 compounds other than formaldehyde and CO_2 are relatively unreactive molecules and high activation energy is required to break the C-H bond to form the C-C bond. In case of methane activation catalyzed by MMO, highly exergonic oxygen reduction reaction is coupled with C-H

activation, thus wasting valuable reducing equivalents. An alternative way to overcome high activation energy is by using radical mechanisms. There are specific types of enzymes, known as glycyl radical enzymes (GRE), which utilize radical mechanisms to activate C1 compounds and other unreactive compounds with high activation energy such as alkanes. GREs are oxygen-sensitive in its active form, but some variants expressed in facultative anaerobes have protective mechanisms to deactivate in the presence of oxygen and reactivate in anoxic environment by activating enzymes[99]. From five different classes of GREs characterized to date[100], formate-lyases and X-succinate synthases are particularly interesting as they have potential to be employed in synthetic C1 assimilation pathways.

i. Pyruvate formate lyase (PFL) and other formate lyases

Formate lyases catalyze the cleavage of 2-ketoacid to acyl-CoA and formate, which include pyruvate formate lyase (PFL) and other 2-ketoacid formate lyases such as *E. coli* TdcE[101]. PFL catalyzes reaction from pyruvate to acetyl-CoA and formate, a key metabolic node that connects glycolysis and TCA cycle in the central metabolism. Although PFL is known to mainly catalyze pyruvate cleavage reaction, it is fully reversible under physiological conditions based on the thermodynamic analysis[85]. Although there is no native C1 assimilation pathways using PFL and other formate lyases, numerous synthetic C1 assimilation pathways are postulated around PFL and other 2-ketoacid formate lyases[85]. PFL-dependent co-assimilation of acetate and formate was demonstrated in engineered *E. coli* in vivo[102], which was also demonstrated in an aerobic culture using synthetic bacterial microcompartment to protect oxygen-sensitive PFL[103]. This pioneering work provides new opportunities in employing not just the PFL-based pathways but also other highly energy- and carbon-efficient C1 utilization pathways involving oxygen-sensitive enzymes in industrial microbes under aerobic fermentation.

ii. Methylsuccinate synthase (MSS)

A new-to-nature oxygen-independent methane activation via fumarate addition catalyzed by methylsuccinate synthase (MSS) is postulated based on catalytic function of GREs known as X-succinate synthases. Known X-succinate synthases utilize toluene and C3+ alkanes as substrates and catalyze C-C coupling reaction with fumarate, generating benzyl/alkylsuccinates. Although alkylsuccinate synthases (ASS) are generally known to activate subterminal carbon of alkanes generating (1-methylalkyl)succinates, a study reports the evidence of terminal carbon activation from sulfate-reducing bacteria utilizing C3 (propane) and C4 (butane) alkanes[104]. This result opens the possibility for utilizing shorter-chain hydrocarbons including ethane and methane although there is no direct evidence of C1 and C2 activation. A potential energy surface (PES)-based thermodynamic analysis also indicates that the overall reaction of methane activation via fumarate addition is thermodynamically favorable[105]. The challenge would be overcoming the bond dissociation energy of methane (440 kJ/mol), which is substantially higher than that of glycine residue (350 kJ/mol)[106, 107]. Nevertheless, sufficient driving force generated by high substrate concentration and/or low product concentration could enhance the reaction as evidenced by terminal carbon activation of propane requiring 410 kJ/mol of energy. A recent publication reports functional expression of full ASS complex in *E. coli*, demonstrating activation of C3-C6 alkanes in vivo via fumarate addition[108]. This report signifies that methane activation via methylsuccinate synthase (MSS) could be employed in industrial chasses as it requires common central metabolite, fumarate as co-substrate and does not require any specific cofactor. The next objective for MSS-based synthetic methanotrophy would be developing a novel methylsuccinate

catabolism pathway as there is no known metabolic pathway that directly utilizes methylsuccinate as substrate or intermediate.

2. C1 metabolism

Numerous review papers discuss different aspects of native and synthetic C1 utilization pathways[24, 85, 109, 110]. There are important factors to be considered when comparing different pathways, which include energy (ATP and reducing equivalent) efficiency, thermodynamics and driving force of the pathway, number of enzymes, catalytic efficiency, oxygen tolerance and specific cofactor requirements of core enzymes, and toxicity of intermediates among others. Here, I would also like to introduce another criterion for assessing pathway orthogonality to the host metabolism, which could be determined by the degree of pathway overlap with the central metabolism and other essential metabolites. While the presence of native enzymes and substrates as intermediates could reduce the number of foreign enzymes that need to be introduced, the enzyme expressions could be affected by host regulatory mechanisms and there could be competing endogenous enzymes diverting the flux away from the engineered metabolic pathways. Autocatalytic cycles, which represent a great portion of both native and synthetic C1 utilization pathways, could bring more challenges in finetuning the flux at the branchpoint[111].

Traditional metabolic engineering focuses on redirecting metabolic flux toward desired product typically from common metabolic precursor like pyruvate and acetyl-CoA. Broad overlap with the central metabolism would not be a problem for native C1-trophs as the same approach can readily be applied for product synthesis by engineering these organisms. However, engineering synthetic C1-troph by introducing non-native C1 utilization pathways could bring in undesired complexities if the pathway is interconnected with highly regulated host metabolism. Often, these challenges are addressed by combinatorial approach of rational design and adaptive laboratory

evolution (ALE), and recent success in engineering synthetic methylotroph shows the sheer amount of work required to engineer a strain to utilize non-native C1 substrates[13, 14]. It is also important to note that incorporating C1 utilization pathway that shares less enzymes and metabolites with the central metabolism, such as Reductive Glycine pathway, requires substantially less rounds of evolution (and number of mutations) to achieve C1-trophic phenotype[15] than pathways heavily interconnected with central metabolism like RuMP cycle[13, 14]. In addition to the benefits and drawbacks of each C1 assimilation pathways, this section also addresses the interconnection of the pathway to the native carbon metabolism as displayed in Figure 1-7 and challenges that are associated with engineering the pathways to achieve synthetic C1-trophy and/or product synthesis through the pathway.

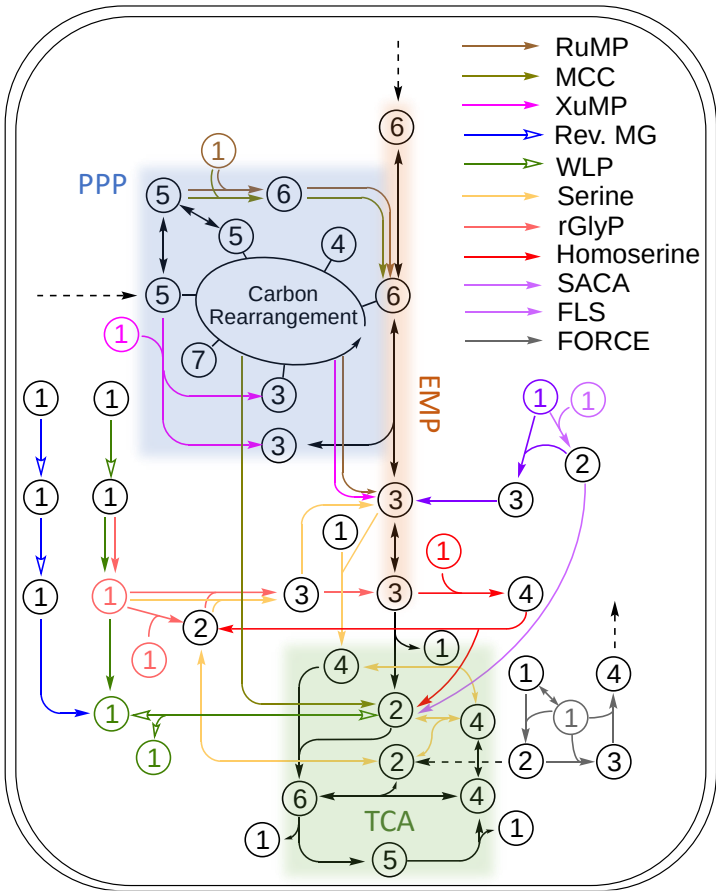


Figure 1-7. Simplified schematic for different native and synthetic C1 utilization pathways and central metabolic pathways: pentose phosphate pathway (PPP), Embden-Meyerhof-Parnas (EMP) pathway (glycolysis), and tricarboxylic acid (TCA) cycle. Numbered circles represent number of carbons in metabolites. Dotted arrow represents entry point for glucose (top) and xylose (left). Hollow arrowhead represents oxygen sensitive reaction or pathway. RuMP: Ribulose Monophosphate cycle; MCC: Methanol Condensation Cycle; XuMP: Xylulose Monophosphate cycle; Rev. MG: reverse methanogenesis; WLP: Wood-Ljungdahl pathway; Serine: Serine cycle (includes Modified Serine cycle and Serine-Threonine cycle); rGlyP: Reductive Glycine pathway; Homoserine: Homoserine cycle; SACA: Synthetic Acetyl-CoA pathway; FLS: Formolase pathway; FORCE: Formyl-CoA Elongation pathway.

2.1. C1 assimilation connected to the pentose phosphate pathway

Pentose phosphate pathway (PPP) plays an important role in the central metabolism for assimilation and dissimilation of carbon molecules that do not belong to hexose (C6), triose (C3) sugars or C2 metabolites like acetyl-CoA and glyoxylate. For example, pentose (C5) sugars such

as ribose and xylose are channeled to the central metabolism via PPP. PPP also has crucial role in producing reducing equivalent NADPH and metabolic precursor for nucleotide synthesis[112]. Because carbon molecules of varying chain lengths from C3 to C7 are involved during the carbon rearrangement in PPP (Fig. 1-7), it makes a good entry point for C1 molecules as well, as demonstrated in naturally occurring C1 assimilation pathways such as RuMP and XuMP cycles. However, the substantial overlap of these C1 assimilation pathways with the highly regulated central metabolism, as well as the complexity of the carbon rearrangement reactions in the PPP makes the engineering of these C1 assimilation pathway challenging.

2.1.1. Ribulose Monophosphate (RuMP) cycle

Type I methanotrophic and methylotrophic bacteria operate via Ribulose Monophosphate (RuMP) cycle to assimilate C1 compounds, such as methane, methanol, and formaldehyde. Fast growing methanotrophs and methylotrophs harboring RuMP cycle were employed to demonstrate product synthesis from methane and methanol. Examples include Type I methanotroph *Methylomicrobium buryatense* 5GB1C engineered to produce lactic acid[113] and C4 carboxylic acids[114] from methane and thermophilic methylotroph *Bacillus methanolicus* producing amino acids from methanol[115]. However, due to challenges in engineering these organisms with limited information and genetic tools, no report on value-added products with industrially relevant titer, yield, and rate has been reported.

There have been great efforts in employing the RuMP cycle in non-methylotrophic model organisms like *E. coli* to construct synthetic methylotroph or methanotroph. In theory, methanol dehydrogenase (MDH), formaldehyde-assimilating 3-hexulose-6-phosphate synthase (HPS) and 6-phospho-3-hexuloisomerase (PHI) are the only three heterologous enzymes needed for incorporating full RuMP cycle into *E. coli* (Fig. 1-7). However, it was found that substantial

metabolic rewiring in the downstream pathway is needed to enable proper balance in the metabolic flux regenerating Ru5P and exiting the pathway for cell growth. Initial approaches for introducing RuMP cycle in *E. coli* relied on providing multi-carbon co-substrates, such as xylose and gluconate, and constructing "methanol auxotrophs" to demonstrate synthetic methylotrophy[116-118]. Full synthetic methylotroph using *E. coli* was eventually accomplished via extensive rational genome engineering and adaptive laboratory evolution (ALE), one strategy[13] using genome integrated constitutive expression of MDH, HPS and PHI under serial dilution approach with decreasing nutrient-rich media, and the other strategy[14] using plasmid-based expression in strong inducible promoter using a continuous chemostat with decreasing secondary carbon source, pyruvate. In both approaches, significant amounts of time and effort were put into evolving the strains to build the phenotype and numerous mutations in the genome were identified from the evolved strain. Interestingly, one of the key common findings from the ALE approach of two independent studies was the importance of high expression of heterologous core pathway enzymes, especially MDH. This signifies the limitation of heterologously expressed NAD^+-dependent MDH in non-methylotroph like *E. coli* in the absence of activator enzymes. Due to these issues, methanol uptake and specific growth rates of these strains growing on methanol as sole carbon source do not reach the level of native methylotrophs[119].

2.1.2. Xylulose Monophosphate (XuMP) cycle

Whereas methylotrophic bacteria employ RuMP to assimilate C1, methylotrophic yeasts harbor Xylulose Monophosphate (XuMP) cycle which has high similarity with the RuMP cycle in that C1 unit (formaldehyde) is assimilated into the pentose phosphate pathway (Fig. 1-7). However, unlike prokaryotes, eukaryotes possess different organelles within the cell, which give them advantages in avoiding formaldehyde toxicity by employing C1 utilization pathways in a separate

compartment from the cytosol, such as peroxisomes in yeast. Moreover, yeast utilizes alcohol oxidases which generate hydrogen peroxide as an intermediate, which could be scavenged inside the peroxisomes as well[120]. Methylotrophic yeasts, such as *Komagataella phaffii* (*Pichia pastoris*) and *Ogataea polymorpha* have been used to produce various products including proteins[121], fatty acids[122], alcohols[123] and alkenes[124] using methanol as the carbon source. However, it is worth noting that even employing native methylotrophs required substantial engineering and adaptive evolution to allow the cells to tolerate high methanol concentration and produce desired products at high titer, yield and rate. Efforts to engineer non-methylotrophic model yeast, *Saccharomyces cerevisiae*, to build synthetic methylotroph had partial success and still required additional nutrients (yeast extract) to support growth on methanol[125].

2.1.3. Methanol Condensation Cycle (MCC)

One of the drawbacks of RuMP and XuMP cycles is that they generate a C3 molecule (glyceraldehyde-3-phosphate) as an intermediate entering the central metabolism, which inevitably undergoes loss of a C1 unit (formate or CO_2) in the downstream glycolysis from pyruvate to acetyl-CoA. A synthetic Methanol Condensation Cycle (MCC) is designed to address this issue. At its core, MCC is a combination of RuMP cycle and phosphoketolase-based nonoxidative glycolysis (NOG) which generates 3 molecules of C2 units (acetyl-phosphate) instead of 2 molecules of C3 units to bypass the carbon loss in central metabolism[126] (Fig. 1-7). A key finding in the study of MCC is the importance of optimum phosphoketolase expression level in maintaining steady state and avoiding kinetic trap. Ensemble modeling for robustness analysis (EMRA) was used to identify the control points, which were then confirmed by cell-free experiments[126]. By employing MCC, they achieved 80% carbon yield of ethanol production from methanol in vitro, which exceeds the theoretical maximum of RuMP cycle followed by

glycolysis[126]. However, a major bottleneck in the pathway was again identified to be the NAD^+-dependent methanol dehydrogenase, due to the limitations discussed in Section 1.1.2, which forced excess amount of commercial alcohol dehydrogenase to be added to drive the reaction[126].

2.2. C1 assimilation involving amino acids as intermediates

A native C1 assimilation pathway, Serine cycle, and its synthetic/engineered variants as well as naturally occurring Reductive Glycine pathway (rGlyP) all involve proteogenic α-amino acids as intermediates including glycine, serine, and threonine, which are indispensable in all organisms. As a result, most of these pathways are closely interconnected to the central metabolism, not just with amino acid metabolism but also with glycolysis and TCA cycle (Figure 1-7). Moreover, except for the rGlyP, all Serine cycle-derivative pathways are cyclic pathways which further complicate engineering approaches as described in cyclic C1 assimilation pathways associated with pentose phosphate pathway in Section 2.1.

2.2.1. Serine cycle

Serine cycle is a native C1 assimilation pathway in certain methylotrophs known as Type II methylotrophs, such as *Methylobacterium extorquens* AM1. Methylene-THF is the metabolic precursor entering the pathway, which could be generated from different C1 compounds as described in Section 1.2. The main route for methylene-THF generation was previously thought to be the spontaneous condensation between formaldehyde and THF but it was later found that formate is the main branch point for the Serine cycle[127]. Formaldehyde generated from PQQ-dependent methanol dehydrogenase is activated to methylene-H_4MPT instead of methylene-THF, catalyzed by formaldehyde activating enzyme (FAE) which is reported to produce methylene-H_4MPT specifically[64]. Following cyclohydrolase and dehydrogenase activities, formyl-H_4MPT is hydrolyzed to formate, which is then reactivated to formyl-THF by formyl-THF ligase (FTL).

This bypass requires extra enzymes and an ATP in the process as opposed to direct condensation of formaldehyde to methylene-THF, but postulated to maintain high flux downstream of formaldehyde preventing toxic effect from formaldehyde accumulation[127].

Methylene-THF enters the Serine cycle via condensation with glycine to form serine, catalyzed by serine hydroxymethyltransferase (SHMT) as discussed in Section 1.3.2. Serine then enters glycolysis via 3-phosphoglycerate, followed by carboxylation at phosphoenolpyruvate (PEP) node to enter TCA cycle via oxaloacetate (Fig. 1-7). The resulting C4 molecule undergoes cleavage to acetyl-CoA and glyoxylate at malyl-CoA, where acetyl-CoA is used for cell growth and product synthesis while glyoxylate is converted to glycine to close the cycle[128]. Because the pathway is closely interconnected with the central metabolism, especially cyclic pathways including TCA cycle and ethylmalonyl-CoA (EMC) cycle, optimizing metabolic flux, and enhancing acetyl-CoA production requires more sophisticated approach than a simple push-pull-block strategy[129].

2.2.2. Modified Serine cycle

Heterologous implementation of Serine cycle in *E. coli* was demonstrated with modifications to avoid the toxicity of intermediate, hydroxypyruvate, and the promiscuity of *E. coli* GhrA on glyoxylate reductase and hydroxypyruvate reductase activities[67]. The Modified Serine cycle bypasses the four-step reactions from serine to PEP via two-step shortcut consisting of deamination to pyruvate followed by phosphorylation to PEP[67]. Moreover, an alternative, more efficient route for glycine production from glyoxylate was investigated[67] as *E. coli* does not have ethylmalonyl-CoA pathway to efficiently regenerate glyoxylate from acetyl-CoA[128]. However, likely due to the pathway flux imbalance, full Modified Serine cycle could not support cell growth nor production of C2 product, such as acetate and ethanol, using methanol as sole carbon source[67].

2.2.3. Serine-threonine cycle

Serine cycle and Modified Serine cycle utilize malate ligase and malyl-CoA lyase to generate two C2 units, acetyl-CoA and glyoxylate from C4 (malate). These two enzymes are not universal in organisms not operating Serine cycle and the resulting glyoxylate amination to glycine could also be challenging as seen from implementing Modified Serine cycle in *E. coli*. Serine-threonine cycle (STC) bypasses these reactions via 7-step reactions, which involves amino acid intermediates including aspartate, homoserine and threonine[70, 85]. A key advantage to this pathway is that all enzymes are universally available in the native amino acid metabolism and glycine can be efficiently regenerated potentially at higher flux than direct amination of glyoxylate. In addition, this pathway has reduced overlap with the central metabolism as oxaloacetate produced from carboxylation of PEP immediately enters the amino acid metabolism via aspartate, instead of entering TCA cycle in forming malate. However, five extra enzymes and an ATP are required in the process which makes the pathway less energy- and enzyme-efficient than the Serine cycle and the Modified Serine cycle. Recently, STC was successfully incorporated in *E. coli* to enable growth on formate as sole carbon source via adaptive laboratory evolution[86], which paves the way for utilizing Serine cycle-derivative pathways in non-methylotrophs for synthetic C1-trophy.

2.2.4. Homoserine cycle

Whereas two synthetic variants of Serine cycle still involve C1 assimilation via SHMT (methylene-THF as C1 node) and PEP carboxylase, identical to the original Serine cycle, synthetic Homoserine cycle utilizes promiscuous aldolase reactions utilizing formaldehyde to construct the cycle similar to the Serine-threonine cycle (STC). Specifically, two promiscuous formaldehyde aldolases: serine aldolase (SAL)[89] and 4-hydroxy-2-oxobutanoate aldolase (HAL)[91] utilizing

glycine and pyruvate as co-substrates, respectively, make synthetic Homoserine pathway[130] (Figure 1-7). SAL catalyzes direct condensation of glycine and formaldehyde to serine bypassing methylene-THF intermediate. HAL enables bypass to homoserine from pyruvate via 4-hydroxy-2-oxobutanoate (HOB) as intermediate, reducing six-step enzymatic reaction from STC to two-step and eliminating two ATP and two NADPH requirements. As a result, unlike STC, this pathway is more energy and enzyme efficient than Serine cycle and Modified Serine cycle for acetyl-CoA production from methanol[130]. Although the pathway relies on promiscuous activities of two aldolases, the identified SAL and HAL variants could generate sufficient flux to synthesize essential amino acids allowing growth of auxotrophic strains. Still, finding or engineering SAL and HAL with better activity and finetuning of interconnected host metabolism is needed to enable full methylotrophic growth using this pathway.

2.2.5. Reductive Glycine pathway

Unlike Serine cycle and its derivatives, Reductive Glycine pathway is a linear pathway that produces C3 product (pyruvate) directly from three C1 moieties (two methylene-THF and a CO_2) with minimal overlap with the central metabolism[66] (Fig. 1-7). It cannot be completely decoupled from the host metabolism as it involves essential amino acids, glycine and serine, as intermediates but carbon fluxes at the “branching nodes” are presumably less affected compared to other pathways that directly overlap with central metabolism. rGlyP has high ATP efficiency and consists of oxygen-tolerant enzymes, which makes it a great platform for synthetic C1-trophy[15]. Notably, substantially fewer rounds of ALE were required for *E. coli* harboring rGlyP to grow on C1 source, formate and CO_2 or methanol[15], than did for RuMP cycle[13, 14] or Serine-threonine cycle[86] presumably due to the linearity and independency of the pathway, as well as not involving the toxic intermediate, formaldehyde. Interestingly, significantly higher

specific growth rate at doubling time (DT) of 7.7 hour was observed from formate and 10% CO_2 than from energy-rich methanol (DT = 54h) due to slow methanol oxidation rate, again from NAD^+-dependent MDH[15]. One disadvantage of rGlyP is that it produces pyruvate as precursor to enter central metabolism, which inevitably undergoes loss of one carbon during cleavage to acetyl-CoA. However, formate or CO_2 produced from either pyruvate formate lyase (PFL) or pyruvate dehydrogenase (PDH) can be fixed back using the rGlyP as it utilizes both as substrates. An oxygen-sensitive route that directly produces acetyl-CoA from glycine, bypassing C3 node was also proposed[85]. Recently, there was a preprint demonstrating production of lactate using engineered *E. coli* harboring rGlyP utilizing formate and CO_2 as only carbon sources[131]. Although modest titer (1.2 mM) and yield (10% of the theoretical maximum) were achieved, it is the first report on engineered C1-troph showing product synthesis only using C1 feedstock. rGlyP was successfully demonstrated in other organisms including chemolithoautotrophic *Cupriavidus necator*[132] and *Desulfovibrio desulfuricans*[133] showing its cross-platform capabilities.

2.3. Pathways independent from the central metabolism

C1 utilization pathways described in the previous two sections are based on already existing pathways and enzymes in nature which provide a great platform for synthetic C1-trophy as many pathway enzymes are available endogenously and have decent kinetic parameters. However, as seen from many examples, integrating such pathways for synthetic C1-trophy requires significant engineering not only in the C1 assimilation pathway itself but also many regulatory machineries and interconnected endogenous pathways (Fig. 1-7) to build the desired phenotype. This section focuses on native and synthetic pathways that are independent from the central metabolism as in the carbon flux from C1 substrate to product has minimal overlap with the host metabolism and could be completely decoupled from the central metabolism.

2.3.1. Wood-Ljungdahl pathway

Wood-Ljungdahl pathway (WLP), also known as the reductive acetyl-CoA pathway, is a highly energy- and carbon-efficient C1 utilization pathway that is known to operate at the "thermodynamic limit of life"[46]. Sophisticated electron bifurcation mechanisms are involved to efficiently interconvert reducing cofactors such as ferredoxin, NADH and NADPH with minimal energy loss, which ultimately enable production of C2 molecule, acetate from two C1 molecules (CO_2, CO or formate) with no expense of ATP (or acetyl-CoA with one ATP equivalent). Although it is not completely independent from the host metabolism as C1-THFs are common metabolites, the pathway from C1 to acetyl-CoA operates independently from the central metabolism (Fig. 1-7). Key enzymes in the pathway including CODH-ACS complex (Section 1.3.1) as well as various electron bifurcating enzymes are oxygen-sensitive; and the pathway itself is only found in obligate anaerobes like acetogens and methanogens but the high efficiency of the pathway as well as product acetyl-CoA as universal precursor make WLP an attractive platform for C1 bioconversion. Moreover, virtually all C1 compounds could be harnessed through WLP (although only a few reports[54] are available on methane utilization via reverse methanogenesis/WLP) and acetogens growing on C1 compounds via WLP have competitive growth rate and biomass yield[119, 134]. Although engineering obligate anaerobes are slower and more challenging than engineering model organisms, recent advance in synthetic biology tools and automation system greatly expanded the opportunities in utilizing acetogens[9, 135]. Recent success from Lanzatech in scaling up and commercializing acetogen-based bioproduction paves the way for utilizing more non-model C1-trophs for biomanufacturing[9].

Although it would be a highly attractive and impactful goal to implement WLP in model organisms like *E. coli*, requirement of diverse non-native (mostly oxygen-sensitive) enzymes and

cofactors, such as cobalamin-binding corrinoid protein (CoP) required for CODH-ACS, electron bifurcating enzymes and other oxygen-sensitive ferredoxins, makes heterologous implementation of WLP challenging. Nevertheless, the possibility of finding or engineering enzymes and cofactors that can replace WLP cannot be neglected. Recent success with harnessing oxygen-sensitive enzymes in aerobic fermentation through compartmentalization[103] could open new possibilities for utilizing WLP in heterologous host.

2.3.2. Formolase-based pathways

Enzymes catalyzing C1-C1 coupling are particularly interesting as linear pathways can be built upon the reaction as shown from rGlyP and WLP. While the two are the only found examples from nature, protein engineering has made synthetic pathways possible, building upon engineered enzymes with new-to-nature biochemistries including C1-C1 coupling reaction. Formolase (FLS), a synthetic formaldehyde aldolase[93] is the pioneering work in such approach to build a simple, linear pathway from C1 to multi-carbon products with minimal overlap with the host metabolism. As described in Section 1.3.4, FLS catalyzes C-C coupling of three molecules of formaldehyde to form dihydroxyacetone (DHA) (Fig. 1-6). While a carbon loss occurs when C3 DHA enters glycolysis and further catabolized to acetyl-CoA, starting from C3 ketose could be advantageous in reverse direction following gluconeogenesis to form sugars and carbohydrates. Artificial Starch Anabolic Pathway (ASAP)[33] precisely follows this direction, aiming to produce starch directly from CO_2 bypassing plant growth and processing. ASAP involves electrochemical reduction of CO_2 to methanol, oxidation to formaldehyde, and subsequent production of DHA catalyzed by FLS. DHA is further anabolized to sugars via gluconeogenesis upon phosphorylation, then ultimately to amylose catalyzed by starch synthase[33]. Because ASAP is demonstrated only in the cell-free system, there could be questions regarding the cost of enzymes and cofactors. High

concentration of FLS is required due to poor catalytic efficiency of this enzyme as discussed in Section 1.3.4, which would exacerbate the cost analysis considering relatively low price of the product. Moreover, toxicity of formaldehyde hinders continuous operation for an extended period even in the cell-free system, mandating physical separation of methanol to DHA module from other enzyme cascades[33]. Nevertheless, this research serves as a good example of chemoenzymatic approach for C1 utilization, where electrochemical reduction powered by renewable electricity generates reduced C1, such as methanol and formate from CO_2, combined with enzymatic reaction to generate multi-carbon products from C1.

As discussed, C3 molecules like DHA is not the best starting point for carbon-efficient synthesis of various bioproducts including fatty acids, alcohols, isoprenoids, polyketides, polyhydroxyalkanoates (PHAs), all of which are derived from acetyl-CoA as precursor. Synthetic Acetyl-CoA (SACA) pathway[97] and glycolaldehyde-allose 6-phosphate assimilation (GAPA)[136] are different variants of glycolaldehyde synthase (GALS)-based pathway using FLS-type enzyme engineered to specifically produce C2 product, glycolaldehyde. SACA consists of only two enzymes, GALS and acetyl-phosphate synthase (ACPS), to acetyl-phosphate (AcP), which can be converted to acetyl-CoA via phosphoacetyltransferase (PTA)[97] (Fig. 1-7). Although enzyme efficiency and orthogonality are the key features of this pathway, poor kinetics of the two non-native enzymes make the pathway incapable of producing high titer, rate, and yield[97]. Assuming significant improvement of both enzymes can be achieved via protein engineering, this pathway has huge potential not only in synthetic C1-trophy but also in production of diverse acetyl-CoA derived bioproducts in an orthogonal manner to the host metabolism. GAPA pathway is an alternative route to generate AcP from formaldehyde-derived glycolaldehyde, which was selected after investigating 28 new-to-nature aldolase-based pathways[136]. It is a cyclic

pathway that involves C4 and C6 phosphate intermediates with overlapping intermediates with the pentose phosphate pathway. GAPA can achieve substantially higher yield of AcP from glycolaldehyde (94%), than ACPS (50%[97]), which is the key advantage of this pathway despite higher number of enzymes and pathway complexities.

Glycolaldehyde, instead of being converted to acetyl-CoA, could be directly reduced to produce ethylene glycol, which comprises a large market in chemical industry. *E. coli* glyoxylate carboligase (EcGCL) was successfully engineered to have GALS activity which was coupled to aldehyde reductase (EcFucO) to produce ethylene glycol from formaldehyde at titer of 6.6 mM via whole-cell bioconversion[98]. Alternatively, glycolaldehyde can undergo another aldose reaction to form C4 sugar erythrulose. In a separate study, formolase was engineered for enhanced activity for C2-C2 coupling to produce erythrulose at 98% yield from glycolaldehyde[137]. The formolase-based pathways have the advantage of synthesizing multi-carbon products from C1 with minimal enzymes and interaction with host metabolism. This allows facile implementation in different platforms (cell-free, whole-cell and growth-coupled) and potentially different hosts as well, to produce desired products at high yield directly from C1 substrates.

2.3.3. Formyl-CoA elongation (FORCE) pathways

Formyl-CoA Elongation (FORCE) pathways[25] build upon promiscuous 2-hydroxyacyl-CoA synthase (HACS) activities[68] as described in Section 1.3.3. Because HACS has wide range of substrate carbon chain lengths, it can be used to construct an iterative pathway which elongates C1 unit, formyl-CoA every turn of the pathway (Fig. 1-7). Leveraging this feature, product profile of FORCE pathways can be expanded to include 2-hydroxy acids, 1,2-diols, n-alcohols at varying chain lengths and functional groups. Although several intermediates in the pathway, such as glycolate and acetaldehyde, can be recognized by organisms like *E. coli* as carbon source, genes

responsible for entry to central metabolism (i.e. connecting metabolic nodes) can simply be deleted to make the pathway independent from the host metabolism. The orthogonality of the pathway was demonstrated via C^{13} labeling experiment under resting cell and growing cell cultures showing complete decoupling of the carbon flux between central metabolism and FORCE pathway by deleting connecting metabolic nodes (*E. coli glcD* for glycolate utilization)[25]. Importantly, native[68] and engineered[73] HACS variants exhibit orders of magnitude better catalytic efficiency than FLS-type enzymes, which make them better candidates for utilizing highly toxic formaldehyde as substrate. By design, FORCE pathways operate in multiple iterations and can generate multi-carbon products of C3 and higher by tailoring pathway enzymes without interacting with the host metabolism, which is a key distinguishing feature to the Wood-Ljungdahl pathway and formolase-based pathways.

3. Synthetic, orthogonal C1 utilization pathways

Cellular metabolism has canonical "bow-tie" structure where substrates are initially converted to universal metabolic precursors through the central metabolism, which are then converted to diverse products via anabolic pathways[138]. Native C1 utilization pathways are no exception as seen from heavily interconnected pathways such as RuMP and Serine cycles. As explored in the previous sections, there are considerable challenges in engineering non-methylotrophs to utilize C1 substrates. Even though only a few foreign enzymes are needed to introduce C1 utilization pathways to non-C1-trophs like *E. coli*, numerous endogenous mutations had to be introduced via adaptive laboratory evolution (ALE) to reconstruct the host metabolism and associated regulatory mechanisms before the cells are "evolved" to build the phenotype under selective pressure[13-15, 86, 139]. Furthermore, *E. coli* as a synthetic C1-troph is not an attractive chassis anymore as it does not grow as fast as native methylotrophs, nor it has been demonstrated

to produce any value-added products from C1 sources at industrially relevant titer, yield and rate[119].

On the other hand, utilizing native C1-trophic microorganisms including cyanobacteria, microalgae, acetogens, methanotrophs and methylotrophs have shown remarkable progress with success stories in scale-up and commercialization. However, there are limitations on each of these chasses: photoautotrophs like algae and cyanobacteria suffer from low efficiency of photosynthesis[11] and CBB cycle[12]; obligate anaerobes like methanogens and acetogens require strictly anoxic conditions which bring in operational complexity and prevent use of oxygen-dependent genetic tools and enzymes (e.g., GFP) [140]; methanotrophs and methylotrophs suffer from inefficient methane and methanol oxidation leading to loss of reducing power, high oxygen demand for oxygen-dependent oxidation and heat generation requiring extra cooling[20]. While advance in synthetic biology and genetic tools made such non-model organisms more tractable than ever before, it is still extremely challenging to makes these organisms as efficient as model organisms in terms of growth rate, cell density and diverse product profiles.

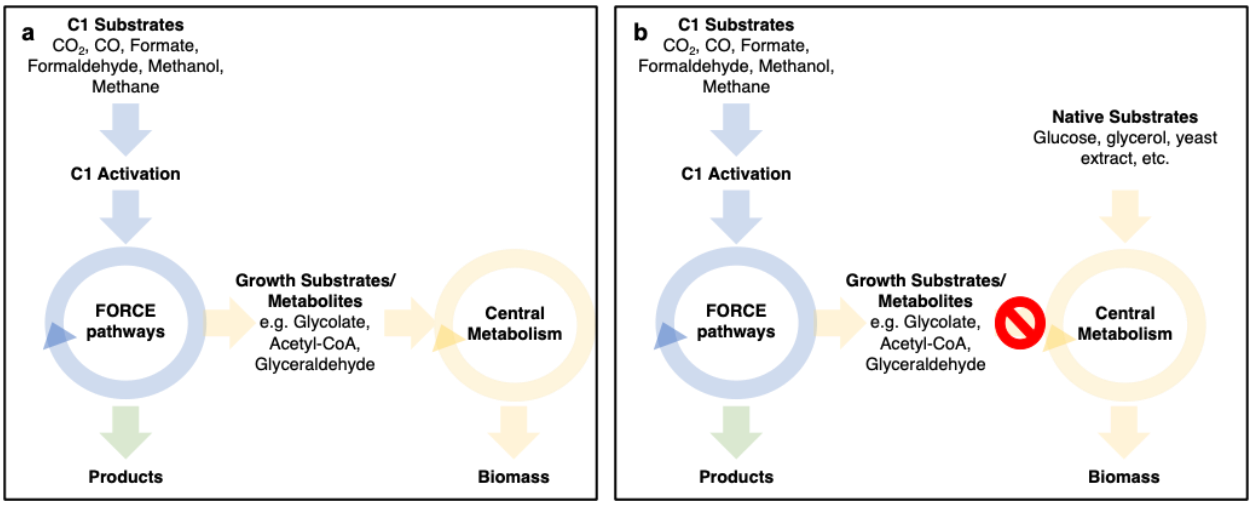


Figure 1-8. FORCE pathways as an example for synthetic, orthogonal C1 utilization pathway. (a) FORCE pathways can be implemented for synthetic C1-trophy, whereas orthogonally produced growth

substrates can be natively utilized by host for growth. (b) FORCE pathways can operate solely as product synthesis pathway with 100% theoretical carbon yield by deleting metabolic nodes connecting to central metabolism. The pathways can operate in cell-free or in vivo growth-decoupled system where cells grow on native substrates like sugars.

Synthetic C1 utilization pathways that builds upon efficient, new-to-nature C1 utilization enzymes could be a solution for addressing issues observed from synthetic C1-trophs while leveraging beneficial traits of model chassis such as high growth rate, abundant engineering tools and diverse product profiles. Incorporation of synthetic, orthogonal C1 utilization pathways, such as FORCE pathways[25], can lead to synthetic C1-trophy without extensive engineering or evolution because C1 assimilation and substrate utilization can be decoupled by nature (Figure 1-8a). Alternatively, the metabolic node that connects the C1 utilization pathway and the host metabolism can be deleted to make the carbon flux for cell growth (from native C sources) and product synthesis (from orthogonal C1 sources) independent (Figure 1-8b). Although it would require secondary carbon source(s) to maintain biocatalyst (cells) to be “active”, key advantages of model organisms, such as efficient cell growth and high cell density can still be leveraged while utilizing C1 compounds as main feedstock for product synthesis. There could be issues with incorporation of these pathways in the desired host because they require either oxygen-sensitive enzymes with uncommon cofactors (Wood-Ljungdahl pathway), or synthetic or engineered enzymes with new-to-nature chemistries and poor kinetics (FLS and FORCE). Nonetheless, these direct C1 to product pathways have high potential in realizing efficient utilization of C1 feedstock, which will be expedited with rapid advance in synthetic biology and protein engineering.

While this chapter discusses the orthogonality of metabolic pathways with focus on carbon flux, other factors ranging from energy (ATP) and redox cofactors to other cellular activities for homeostasis can affect the insulation of the pathways from cellular activities. For example, net

ATP and redox neutral orthogonal pathways can still suffer from loss of flux due to cofactors being drained by cellular activities. Enzymes engineered to utilize orthogonal cofactors, such as nicotinamide mononucleotide (NMN^+) orthogonal to NAD^+[141], could be the solution for accomplishing cofactor orthogonality, which could have synergistic effects when combined with orthogonal pathways. Alternatively, approaches to physically divide product synthesis from cellular activities were proposed, which include cell-free systems[142], compartmentalization[143], synthetic cells[144] among others. Physical segregation allows implementation of pathways that are not necessarily orthogonal to the central metabolism while achieving the same objective as orthogonal pathways. Nonetheless, orthogonal pathways still have advantages under these approaches because enzymes, cofactors, and regulations from substrate to product is clearly defined unlike pathways relying on native pathways where multiple different routes to the precursor metabolites and downstream products exist. Ultimately, we can envision an "orthogonal central dogma[145]" where engineered system operates under complete insulation from intrinsic regulations.

Chapter 2. An orthogonal metabolic framework for one-carbon utilization

Introduction

Canonical architecture of cellular metabolism has a “bow-tie” structure where universal metabolic precursors in the central metabolism serve as the link between substrate degradation (or assimilation for C2 and smaller substrates) and product synthesis[138]. The complex metabolic network is also tightly controlled by numerous local and global regulators which are triggered by different stresses. Moreover, intracellular molecular-level reactions including gene expressions are stochastic, rather than deterministic, leading to cell-to-cell phenotypic and metabolic variabilities known as “metabolic noises”[146]. Native assimilation pathways of one-carbon (C1) compounds, which has gained much attention recently as next-generation feedstock for sustainable biomanufacturing, are no exceptions from this canonical architecture as seen from Figure 2-1a. Challenges with engineering organisms that harbor canonical metabolic pathways or introducing them into foreign host is tied with managing not only the product synthesis or substrate utilization pathways but also other interconnected pathways and regulations in the native or non-native hosts. For example, engineering *E. coli* to grow on methanol by introducing heterologous C1 assimilation pathways (Ribulose Monophosphate cycle) that largely overlap with the central metabolism (Pentose Phosphate pathway) requires rounds of adaptive laboratory evolution (ALE) and numerous mutations in the genome to develop the phenotype even though only three foreign enzymes are needed for the full pathway construction in *E. coli*[13, 14]. Moreover, the synthetic methylotrophs from these studies display suboptimal growth rate, biomass yield and product synthesis which questions the approach of building synthetic C1-trophs on model organisms[119].

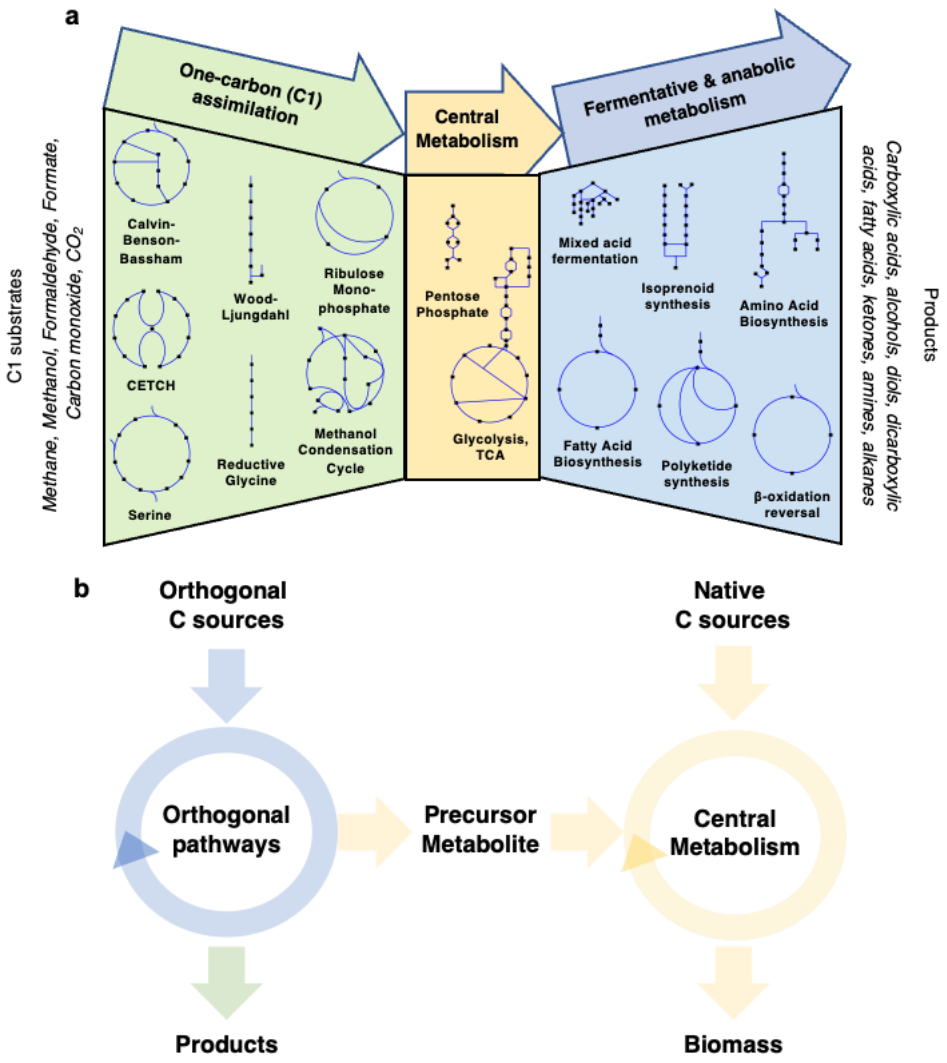


Figure 2-1. (a) Canonical architectures for biological C1 utilization. (b) A consolidated illustration of the orthogonal pathway concept. (a) 'Bowtie' architecture of metabolism in which carbon substrates are consolidated into central metabolites from which a host of products can be produced through fermentative and biosynthetic pathways. Metabolic engineering typically operates within this framework by manipulating either one or all the three components of the bowtie. (b) The orthogonal pathways serve as a platform for both product synthesis and for providing substrates/metabolites for growth. This is an alternative framework to the traditional approach, which feeds all carbon through central metabolism, and from which both products and biomass are derived.

As an alternative solution, orthogonal metabolic pathways that channel substrates directly to products bypassing the central metabolism is interesting as they can not only avoid engineering the complex cellular networks for implementation but also allow decoupling of carbon fluxes

between cell growth and product synthesis (Figure 2-1b). As a result, key advantages of industrial chasses, such as high growth rate and cell density under native substrates like glucose can still be leveraged while non-native (orthogonal) carbon sources can be utilized solely for product synthesis. In other words, model organisms like *E. coli* harboring an orthogonal product synthesis pathway from non-native substrates can simply be the biocatalysts that need to be regenerated in a regular basis for maintenance of catalytic activities, where the additional carbon source for regeneration (at substantially smaller amount than the orthogonal feedstock) can be sugars or glycerol. Under the condition that the carbon flux from orthogonal feedstock to product is completely decoupled from the host metabolism, this system can be operated not only under the two-phase fermentation where the growth phase and production phase are divided[147], but also in an actively growing culture where native (minor) carbon substrate for cell maintenance and orthogonal (major) carbon substrate for product synthesis are co-fed in the culture. Alternatively, if the product of the orthogonal pathway is the native substrate for the host metabolism, cell growth on non-native (orthogonal) carbon substrate as sole carbon source can become possible if sufficient flux can be achieved to support cell growth. Unlike examples of other engineered growth on non-native substrates like C1 compounds, minimal intervention and engineering on the host metabolism would be required in this case because there is no overlap or interconnection between the non-native substrate to the metabolic precursor and from the precursor to cell growth (Fig. 2-1b).

This chapter introduces the design and implementation of the synthetic, orthogonal C1 utilization pathway based on Formyl-CoA Elongation (FORCE) reactions (Fig. 2-2a). This approach relies on acyloin condensations between formyl-CoA and carbonyl-containing molecules, which was recently reported to be catalyzed by 2-hydroxyacyl-CoA synthase (HACS)[68]. HACS (and the related enzyme oxalyl-CoA decarboxylase, OXC[73]) can utilize carbonyl-containing

acceptors of broad chain length and functionalization, including the C1 compound formaldehyde, to generate acyl-CoAs amenable to a wide-range of biochemical conversion. Upon validation of the pathways in the cell-free format[25], the orthogonality of the pathways was demonstrated in vivo under resting cell (two-phase fermentation) and growing cell (co-feeding C1 substrate and native substrate) format. Moreover, possibility utilizing FORCE pathways for synthetic C1-trophy was examined by two-strain co-culture system where producer strain produces multi-carbon precursor metabolite via FORCE pathways and the sensor strain utilizes the precursor for growth.

Results

Design of an orthogonal C1 metabolic architecture

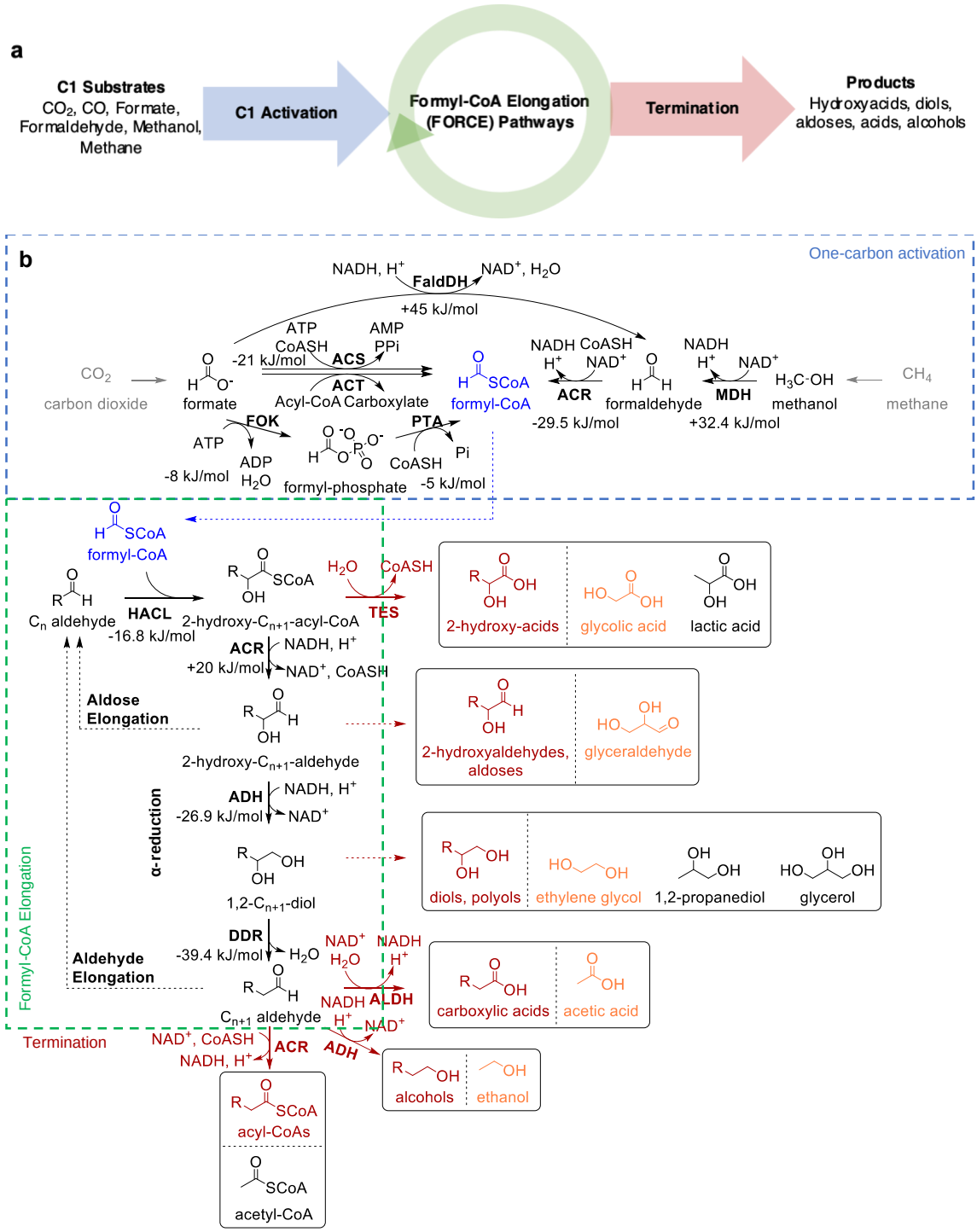


Figure 2-2. FORCE pathways for product synthesis from C1 substrates. (a) A synthetic, orthogonal architecture for C1 utilization based on formyl-CoA elongation (FORCE) pathways. Carbon skeletons are directly built from activated C1 units in the form of formyl-CoA, thus bypassing the "bowtie" architecture of metabolism for product synthesis. (b) One-carbon substrates are activated to the C1 elongation unit formyl-CoA through various redox reactions (blue box). Formyl-CoA serves to elongate an aldehyde in a reaction catalyzed by HACL/HACS, resulting in the production of 2-hydroxyacyl-CoA. 2-Hydroxyacyl-CoA can be further reduced to a 2-hydroxyaldehyde. The 2-hydroxyaldehyde can be further elongated by formyl-CoA, which we refer to as aldose elongation. Alternatively, α-reduction can take place via reduction to a 1,2-diol and dehydration to a nonfunctionalized aldehyde. The resulting aldehyde can then be further elongated. These collective routes for elongation, referred to as formyl-CoA elongation (FORCE), are boxed in green. The various intermediates of these elongation pathways can be converted to desirable chemical products (red) including 2-hydroxy-acids, aldoses, diols, polyols, carboxylic acids, and alcohols. A number of these products and intermediates can also serve as substrates for growth (highlighted in orange), such as glycolic acid, glyceraldehyde, and acetyl-CoA. Abbreviations: MDH: methanol dehydrogenase; ACR: acyl-CoA reductase; FaldDH: formaldehyde dehydrogenase; ACS: acyl-CoA synthetase; ACT: acyl-CoA transferase; FOK: formate kinase; PTA: phosphotransacylase; HACL: 2-hydroxyacyl-CoA lyase/synthase; ADH: alcohol dehydrogenase; DDR: diol dehydratase; TES: thioesterase; ALDH: aldehyde dehydrogenase. Standard Gibbs free energies of reactions are given for each pathway reaction in the direction indicated by the arrow.

The orthogonal metabolic architecture developed here has three primary features (Fig. 2-2a): 1) activation of C1 substrates into a suitable building block for carbon chain elongation; 2) iterative elongation of a carbon chain by one carbon per cycle; and 3) termination of the pathway resulting in product accumulation. Based on our previous findings[68], we investigated whether such a design could be developed using formyl-CoA as the activated C1 unit.

The role of formyl-CoA in metabolism is most well-established in the degradation of multi-carbon compounds and reports of the generation of formyl-CoA from C1 molecules are sparse. Acyl-CoAs, though, are a convenient intermediate between the carboxylate and aldehyde forms enabling formyl-CoA generation from both oxidized and reduced C1 substrates (Fig. 2-2b: one-

carbon activation panel). From formaldehyde, formyl-CoA can be produced via acyl-CoA reductase (ACR)[68] activity, and methanol oxidation to formaldehyde by methanol dehydrogenase (MDH) has been the subject of numerous studies[27, 148, 149]. Formyl-CoA may be produced from formate by CoA transferases[150] or CoA ligases, such as the promiscuous activity *Escherichia coli* acetyl-CoA synthetase (EcACS)[93]. While the latter is AMP forming (consuming 2 ATP equivalents), evidence of an ADP forming route exists via the intermediate formyl-phosphate through formate kinase (FOK) and phosphotransacylase (PTA)[74]. ATP-independent conversion of formate to formyl-CoA via reduction of formate to formaldehyde by formaldehyde dehydrogenase (FaldDH) is also possible[151], albeit thermodynamically challenging (Fig. 1b). Furthermore, CO_2 can be converted to formate by the reverse activity of formate dehydrogenase (or carbon dioxide reductase)[152, 153] and methane to methanol by methane monooxygenase[16], which when coupled to the reactions described above can lead to formyl-CoA formation.

The orthogonal, de novo construction of diverse carbon skeletons by C1 elongation necessitates an iterative pathway similar to those found in nature that construct carbon skeletons from C2-C5 metabolites[154], yet existing outside of central metabolism. Because 2-hydroxyacyl-CoA synthase (HACS) has broad carbon chain length specificity[68], it is a good candidate for establishing an iterative pathway. We evaluated reaction pathways potentially enabling iteration by converting the product of the HACS-catalyzed reaction, 2-hydroxyacyl-CoA, to an aldehyde that can be further extended by formyl-CoA. At the α-carbon, dehydration is possible, transforming the 2-hydroxyacyl-CoA to a 2-enoyl-CoA[155] similar to the well-established acrylate pathway[156]. 2-enoyl-CoA generation is also convenient as these intermediates are involved in β-oxidation, potentially allowing the use of the enzymatic toolkit and knowledge established for

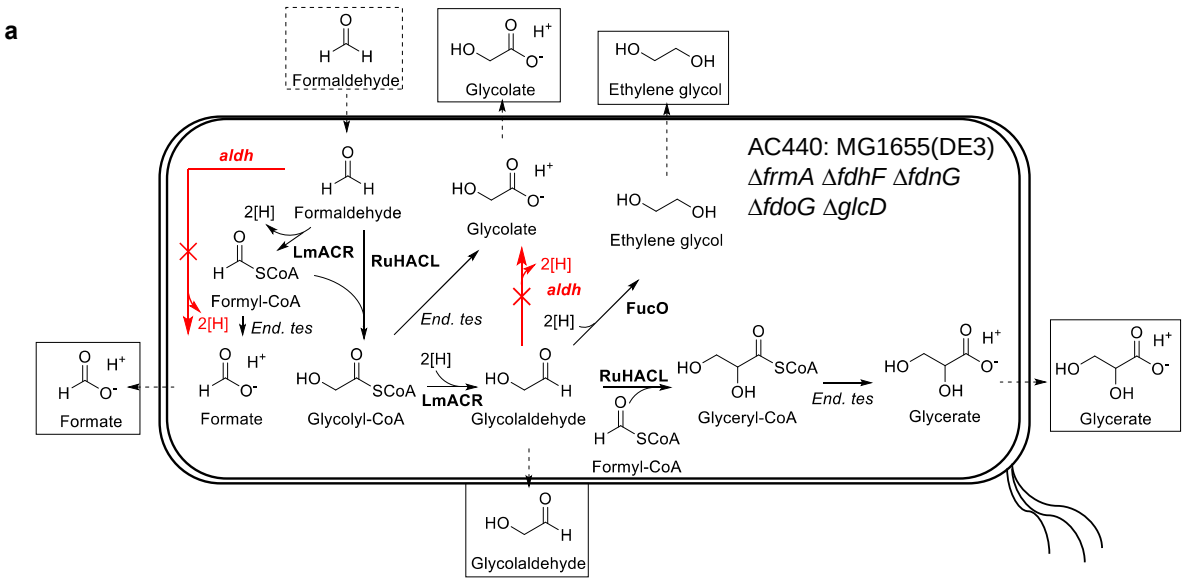
the β-oxidation reversal platform[157-159]. Dehydration of 2-hydroxyacyl-CoA, however, is much more challenging than dehydration of 3-hydroxyacyl-CoA, thus requiring an oxygen-sensitive radical mechanism[160]. It also requires the existence of a β-carbon thus limiting pathway implementation to intermediates 3 carbons or larger.

Due to these issues, we investigated transformations of the thioester. Reduction of the CoA-thioester gives a 2-hydroxyaldehyde (Fig. 2-2b: formyl-CoA elongation panel), which is possible due to the non-specific activity of certain acyl-CoA reductases (ACRs)[68]. Ligation of 2-hydroxyaldehydes with formyl-CoA by HACS gives polyhydroxyacyl-CoAs and further polyhydroxyaldehydes, commonly known as aldoses. Polyhydroxyaldehydes can in principle serve as substrates of the HACS-catalyzed reaction, which we refer to as aldose elongation (Fig. 2-2b: formyl-CoA elongation panel).

Reduction of the 2-hydroxyaldehyde via diol oxidoreductase (DOR) activity to give a 1,2-diol is also possible. For example, *E. coli* FucO catalyzes the interconversion of 1,2-diols with 2-hydroxyaldehydes[161]. 1,2-diol dehydration to an aldehyde can be catalyzed by the activity of diol dehydratase (DDR), effectively accomplishing α-reduction. While diol dehydration also requires a radical mechanism, the B12-dependent DDR is oxygen tolerant and has been the subject of numerous protein and metabolic engineering studies[162-164]. Elongation of this aldehyde by formyl-CoA, which we refer to as aldehyde elongation, enables extension of an alkyl chain, analogous to the two-carbon elongation in fatty acid biosynthesis[165] or reverse β-oxidation[166] pathways. We collectively refer to these pathways (aldose elongation, α-reduction, and aldehyde elongation) as formyl-CoA elongation (FORCE) pathways, as they facilitate the use of formyl-CoA as a carbon chain elongation unit (Fig. 2-2b: formyl-CoA elongation panel).

Various product classes can be produced as intermediates or from derivatives of intermediates of FORCE pathways (Fig. 2-2b), some of which also support microbial growth. Aldose sugars, for example, are a direct result of the 2-hydroxyaldehyde node. Diols, including major industrial chemicals such as ethylene glycol, are a result of the 1,2-diol node. Derivatives of the 2-hydroxyacyl-CoA node include 2-hydroxyacids, such as industrial products glycolic and lactic acids, produced by a thioesterase catalyzed reaction. Numerous chemical classes can be derived from the aldehyde node[167], including carboxylic acids, alcohols, and acyl-CoAs that can serve as precursors of other products.

In vivo implementation of FORCE pathways



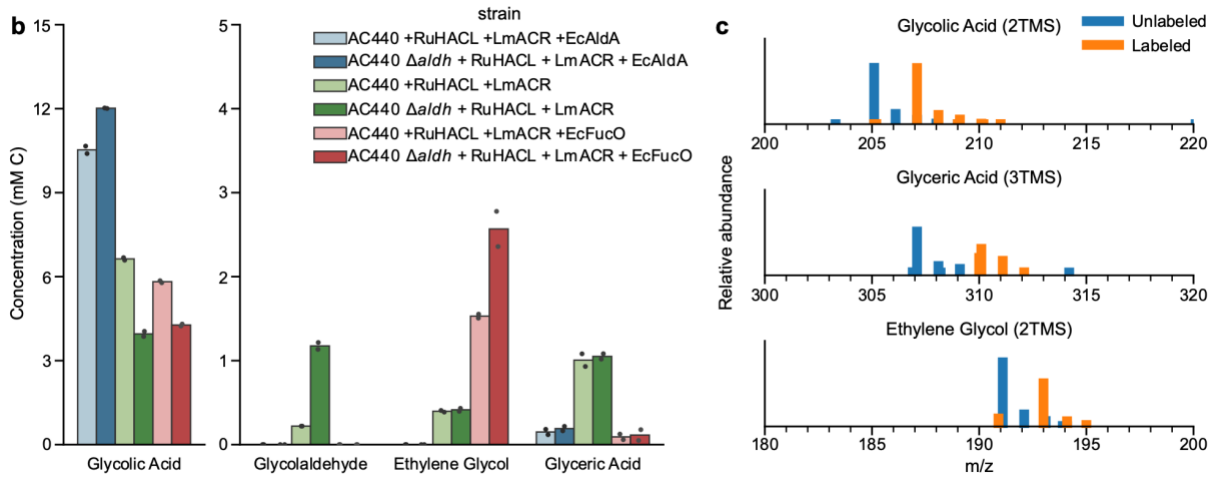


Figure 2-3. Resting cell bioconversions of C1 substrate formaldehyde using the aldose elongation and α-reduction variants of the FORCE pathways. (a) Strategies used in this work to demonstrate diverse product synthesis using FORCE pathways from formaldehyde. Detected products and byproducts are boxed with a solid outline. Knockout strategies to reduce byproduct synthesis are indicated in red. b) Metabolite profiles for strains engineered for product synthesis from formaldehyde using FORCE pathways after 24 hour resting cell bioconversions with OD_{600} = 10 ($5*10^9$ CFU/mL) and two additions of 10 mM formaldehyde at 0 and 1.5 hours. In the legend, + refers to (over)expression of the indicated enzyme. Δ*aldh* refers to knockouts of aldehyde dehydrogenases: Δ*aldA* Δ*aldB* Δ*patD* Δ*puuC*. *End. tes* refers to endogenous thioesterases and spontaneous thioester hydrolysis. No multi-carbon products were observed in a strain that was expressing LmACR and EcAldA only without RuHACL (data not shown). Concentrations are given on a carbon basis and were determined by HPLC under conditions in which carboxylates are detected in their acid form. All data points are shown for duplicate technical replicates. Bars are drawn to the mean values. c) Spectra of multi-carbon products generated from experiments using 13C-labeled formaldehyde in comparison to products from unlabeled formaldehyde. The $[M-15]^+$ ion is shown. A +2 shift in m/z is observed for glycolic acid and ethylene glycol, and a +3 shift in m/z is observed for glyceric acid.

We sought to demonstrate key features of the designed platforms, as well as the synthesis of additional products and utilization of various C1 substrates using both resting and growing *E. coli* cultures (Fig. 2-3 and 2-4). A key feature of the FORCE pathway design is iteration, which can be achieved through aldose or aldehyde elongation (Fig. 2-2b: formyl-CoA elongation panel). To demonstrate iterative aldose elongation *in vivo* we targeted the synthesis of three carbon product

glycerate from formaldehyde (Fig. 2-3a). We started with a previously developed strain having C1 dissimilation and glycolate consumption knockouts (AC440: MG1655(DE3) Δ*frmA* Δ*fdhF* Δ*fdnG* Δ*fdoG* Δ*glcD*) and overexpressing RuHACL^{G390N}, LmACR, and EcAldA[68]. To promote glycolaldehyde accumulation and condensation with formyl-CoA, we removed EcAldA from the expression vector. While formaldehyde consumption was significantly reduced, accumulation of glycolaldehyde and glycerate was observed (Fig. 2-3b), demonstrating the iterative aldose elongation pathway. To increase the production of these compounds, we deleted genes encoding aldehyde dehydrogenases (Δ*aldA* Δ*aldB* Δ*patD* Δ*puuC*, collectively referred to as Δ*aldh*), resulting in lower glycolate and higher glycolaldehyde when EcAldA was not overexpressed. However, these knockouts did not impact the accumulation of glycerate, perhaps indicating a limitation on the condensation reaction between glycolaldehyde and formyl-CoA catalyzed by RuHACL. We also extended the pathway to the next reduction product, ethylene glycol, by overexpressing *E. coli fucO*[168], which led to increased accumulation of ethylene glycol in the extracellular medium, with the Δ*aldh* background further improving production (Fig. 2-3b). To verify that the observed products were derived from formaldehyde and not from residual multi-carbon substrates or biomass components, 13C-labeled formaldehyde was used as the substrate. Glycolic acid, ethylene glycol, and glyceric acid were found to be fully 13C labeled based on the characteristic $[M-15]^+$ ions of the TMS derivatives of the products (Fig. 2-3c).

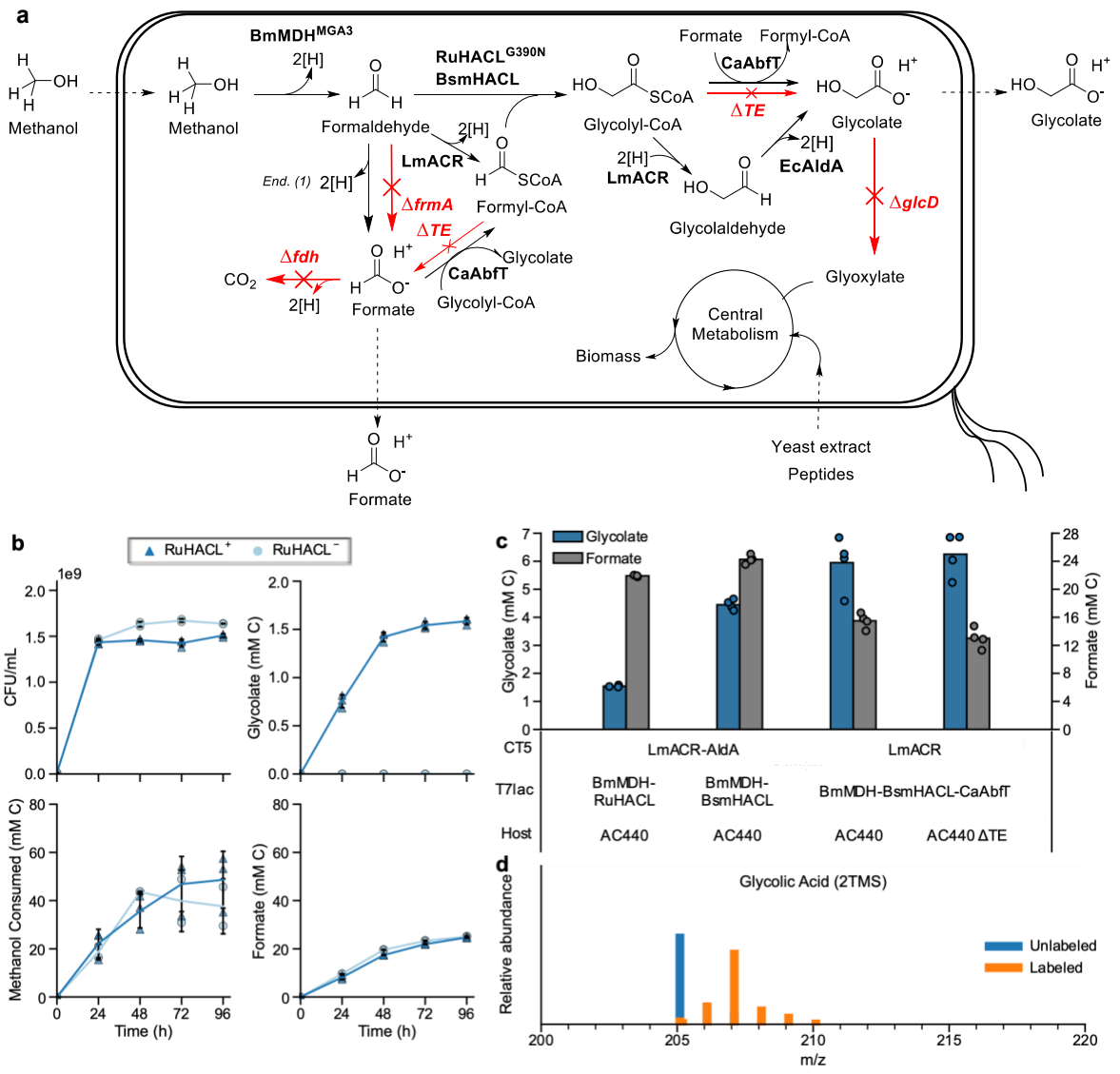


Figure 2-4. FORCE pathway implementation in growing cell cultures using methanol as the C1 substrate. (a) Host and pathway designs for the production of glycolate from methanol in actively growing *E. coli* cultures. Knockout strategies to reduce byproduct synthesis and prevent glycolate utilization are indicated in red and correspond to host strain AC440. *ΔTE* refers to knockouts of endogenous thioesterases (Δ*yciA* Δ*tesA* Δ*tesB* Δ*ybgC* Δ*ydiI* Δ*fadM*). *End. (1)* refers to endogenous aldehyde oxidation activity. Enzyme abbreviations: BmMDH2^{MGA3}: *Bacillus methanolicus* MGA3 NAD$^+$-dependent methanol dehydrogenase; LmACR: *Listeria monocytogenes* acyl-CoA reductase; RuHACL^{G390N}: *Rhodospirillales bacterium URHD0017* HACL (G390N); BsmHACL: Beach sand metagenome HACL; EcAldA: *E. coli* aldehyde dehydrogenase A; CbAbfT: *Clostridium aminobutyricum* CoA transferase. (b) Time course of production of glycolate and formate from methanol. FORCE pathway designs were implemented by overexpressing LmACR, EcAldA, and BmMdh^{MGA3} with or

without RuHACL^{G390N}. All data is shown for biological replicates (n=3 for samples with RuHACL^{G390N}; n=2 for samples without RuHACL^{G390N}). Lines are drawn to the mean values with error bars indicating the standard deviation. Concentrations are given on a carbon basis. (c) Improvement of glycolate production from methanol in growing *E. coli* cultures via rational engineering. Glycolate and formate concentrations are given on a carbon basis for the 72-hour time point. All data is shown for biological replicates (n=3 for samples with RuHACL^{G390N}; n=4 for others). Bars are drawn to the mean values. d) Spectra of the $[M-15]^+$ ion of the 2TMS derivative of glycolic acid produced by *E. coli* incubated with either 12C (unlabeled) or 13C (labeled) methanol.

To extend the above established formaldehyde utilization pathway to methanol, we expressed a well-studied MDH variant from *Bacillus methanolicus* MGA3 (BmMDH2^{MGA3})[148] for conversion of methanol to formaldehyde, in combination with RuHACL^{G390N}, LmACR, and EcAldA. Unlike formaldehyde, where toxicity necessitates the use of resting cells, methanol can also be directly added to growing *E. coli* cultures. When the engineered methanol utilizing strain was grown in the presence of complex nutrients and 500 mM methanol, glycolate formation was observed only in the strain expressing RuHACL (Fig. 2-4b). The conversion of methanol to glycolate by this strain was inefficient, however, with substantial accumulation of formate.

Seeking to improve performance, we replaced RuHACL^{G390N} with a newly identified HACS sourced from beach sand metagenome referred to here as BsmHACL (UniProt-: A0A3C0TX30). BsmHACL increased glycolate accumulation about 3-fold (Fig. 2-4c). Despite improved glycolate production, formate accumulation remained high. In an effort to address this issue, the termination enzyme EcAldA was replaced with a CoA-transferase from *Clostridium aminobutyricum* (CaAbfT) previously found to have better properties than OfFrc[73]. CaAbfT serves to both release glycolate from glycolyl-CoA and reactivate formate to formyl-CoA for further condensation. When CaAbfT was expressed, glycolate accumulation increased around 33%, while formate accumulation was reduced by approximately 36%. Finally, with CaAbfT serving to terminate the pathway via the release of glycolate, endogenous thioesterases were not expected to

be needed and were presumed to be in part responsible for the observed formate. Using a host strain deficient in thioesterases (Δ*yciA* Δ*tesA* Δ*tesB* Δ*ybgC* Δ*ydiI* Δ*fadM*), formate accumulation was further reduced. To verify that glycolate was derived from methanol, we used 13C-labeled methanol and observed that the $[M-15]^+$ ion of the TMS derivative of glycolic acid was fully derived from 13C-methanol (Fig. 2-4d).

Having established CaAbfT as a promising route for formate activation, we evaluated whether CaAbfT could be used to incorporate exogenously supplied formate. Here, CaAbfT was expressed to activate formate without LmACR overexpression as no interconversion of formaldehyde and formyl-CoA is needed upon addition of formaldehyde. In the engineered strain expressing BsmHACL, a 12-fold increase in glycolate was observed when formate was included in the media compared to when formaldehyde was supplied alone with the total carbon accumulated as glycolate greater than the amount originally added as formaldehyde.

Two-strain co-culture to evaluate synthetic methylotrophy

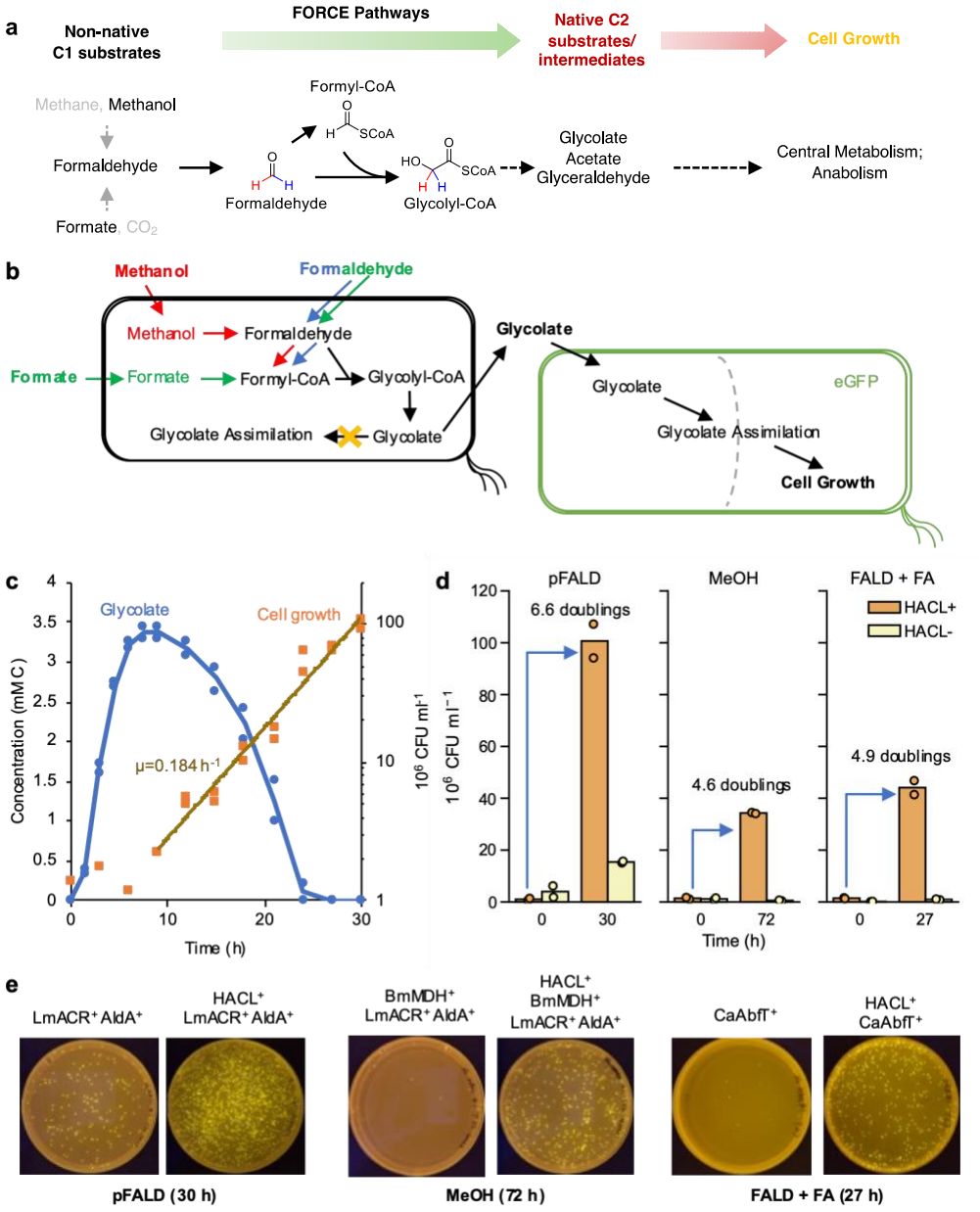


Figure 2-5. Two-strain system for evaluating the ability of FORCE pathways to enable growth on C1 substrates. (a) FORCE pathways can enable synthetic methylotrophy by converting non-native C1 substrates into native multi-carbon substrates that serve as carbon and energy sources. (b) Conceptual scheme of the two-strain system. Producer strains (yellow outline) that are unable to consume glycolate were engineered to produce glycolate from one of three C1 substrates: methanol (red), (para)formaldehyde (blue), or formate and formaldehyde (green). A second consumer strain capable of consuming glycolate was added to the culture, acting as a detectable signal to evaluate growth. (c) Time

course of glycolate concentration (blue) and cell-growth (orange) in the two-strain system with (para)formaldehyde as the sole source of carbon. 5 mM (mass equivalent) paraformaldehyde added to AC440 ($3*10^9$ CFU/mL) expressing LmACR, AldA, and BsmHACL. All data points are shown for duplicate replicates. The line for glycolate concentration is drawn to the mean values. The line for cell growth is the fit of the data to exponential growth by least squares regression, which was used to calculate the specific growth rate (μ). (d) Growth of the consumer strain when incubated for the indicated time with the relevant producer strain with (+) or without (-) HACL and the indicated C1 substrate (pFALD: paraformaldehyde; MeOH: methanol; FALD: formaldehyde; FA: sodium formate). All data is shown for duplicate technical replicates with bars drawn to the mean values. e) Plate images demonstrating growth of the consumer strain corresponding to the conditions in panel d.

The orthogonality of FORCE pathways to metabolism also allows full decoupling of the C1 conversion pathway from growth. This enables unique designs to evaluate the methylotrophic potential of the pathway (Fig. 2-5a; Extended Data Figure 1b). One potentially advantageous implementation might employ division of labor by separating multi-carbon compound generation and cell growth into two hosts, which would not be possible if the pathway directly interfaced with central metabolism, for example via aldose phosphates or acetyl-CoA, two common products of C1 assimilation pathways. Using this concept, we evaluated the ability for FORCE pathways to support *E. coli* growth on C1 substrates formaldehyde, formate, and methanol.

A two-strain *E. coli* system was designed and constructed to work in co-culture (Fig. 2-5b). The first strain, referred to as the producer strain, contained constructs to express the FORCE pathway for conversion of C1 substrates to the native C2 growth substrate glycolate but was deficient in the ability to consume glycolate. The second strain, referred to as the sensor strain, retained the ability to grow on glycolate and additionally constitutively expressed eGFP as a signal but did not express the FORCE pathway for glycolate production. These strains could thus be differentiated by both selection on glycolate minimal media plates and by detection of fluorescent colonies. To assess the feasibility of different substrates, three different producer strains were

devised: for formaldehyde utilization the producer strain expressed LmACR, BsmHACL, and EcAldA; for evaluating formate utilization with formaldehyde, BsmHACL was expressed with CaAbfT; and for methanol utilization a thioesterase deficient background expressing $BmMdh^{MGA3}$, LmACR, BsmHACL, and CaAbfT (Fig. 2-5b) was utilized.

To enable growth conditions with formaldehyde, paraformaldehyde was used. Paraformaldehyde gradually depolymerizes to give formaldehyde in aqueous media, with the ability to control the solubilization rate through the selection of particle size and concentration. This enabled a system in which formaldehyde could be kept at sub-millimolar concentrations, avoiding accumulation to toxic levels, with significant glycolate production still observed. In minimal media with (para)formaldehyde (the equivalent of 5 mM) as the sole carbon substrate, growth of the sensor strain was observed as indicated by the increase in colony-forming units (CFUs) relative to a control system in which the producer strain did not express BsmHACL (Fig. 2-5c). Glycolate accumulated rapidly in the first 8 hours with sustained exponential growth of the sensor strain occurring after an initial lag phase. The sensor strain was found to have undergone around 6.6 doublings in 30 hours.

With methanol, growth of the sensor strain was observed only when the producer strain expressed BsmHACL (Fig. 2-5d), however compared to the case for paraformaldehyde utilization the growth kinetics of the sensor strain differed, reflecting an approximately linear increase in CFUs over time. The difference in observed dynamics might reflect the limitation imposed by the rate of glycolate production from methanol by the producer strain, analogous to the phenomenon observed in constant feed-rate fed-batch culture[169]. The utilization of methanol was substantially slower than the utilization of (para)formaldehyde, resulting in approximately 4.6 doublings in 72 hours.

A similar experiment was performed using 1 mM formaldehyde and 10 mM formate co-substrate system. Here, more carbon was observed in glycolate than was added as formaldehyde, indicating the incorporation of formate. Growth of the sensor strain was faster than growth on methanol but did not result in as many doublings as on 5 mM (para)formaldehyde. In 27 hours, around 4.9 doublings were observed (Fig. 2-5d).

Discussion

In the canonical architecture of metabolism, substrates are funneled into central metabolism with biosynthetic building blocks and products of interest derived from the resulting central metabolites. To date, attempts to engineer C1 bioconversion, even those exploiting synthetic pathways[15, 93, 97, 126, 170] or novel enzyme designs[93, 97], have relied on central carbon metabolism. These designs, which exhibit minimal orthogonality, require optimizing a host's metabolic network to accommodate C1 bioconversion, which has proven challenging.

In this work, we present the design, analysis, and implementation of formyl-CoA elongation (FORCE) pathways, enabling C1 utilization and bioconversion in a manner orthogonal to the host metabolism. FORCE pathways are based on using formyl-CoA as an anabolic metabolite, which is enabled by 2-hydroxyacyl-CoA synthase (HACS) catalyzed acyloin condensation between formyl-CoA and carbonyl-containing substrates. Product synthesis is achieved with relatively high orthogonality to central metabolism compared to other approaches. Our thermodynamic analysis suggested favorable driving forces for FORCE pathway conversions of formate, formaldehyde, and methanol to glycolate or acetate as exemplary products. We demonstrate the potential of the self-contained, orthogonal pathway in both *in vitro* (purified enzymes and cell extracts) and *in vivo* (resting and growing cells) implementations, in which products of diverse functionality (e.g. glycolate, glycolaldehyde, ethylene glycol, ethanol,

glycerate) could be produced in a growth and host metabolism independent manner using formaldehyde, formate or methanol as the sole C1 substrates. One can envision potential bioprocesses in which growth and maintenance is performed with a multi-carbon substrate, while the biocatalyst is used for C1 bioconversions. Bioprocesses of this nature, based on multi-enzyme cascades and two phase fermentations, have been the subject of recent reviews[171, 172].

While product synthesis from C1 substrates is a defining feature of FORCE pathways, they also have the potential to enable growth on non-native C1 substrates (e.g. synthetic methylotrophy) via the production of multi-carbon compounds naturally consumed by heterotrophs, such as glycolate, acetate, or glyceraldehyde. Genome scale modeling and flux balance analysis revealed that FORCE pathways are comparable or better than alternative approaches and guided the design. While the current pathway performance could not support the growth of a single strain of *E. coli* on C1 substrates, the orthogonal nature of the pathway allowed us to separate and evaluate the pathway limitations to growth on formate, formaldehyde, and methanol in separate strains of *E. coli*. The potential for FORCE pathways to enable methylotrophy allows for possible bioprocess implementations more similar to traditional fermentations based on C1 as a sole carbon source for both growth and product synthesis. Because the FORCE pathway is the branch point for fluxes toward product synthesis and growth, there is significant potential for facile control over flux partitioning (Fig. 2-1b), especially with recent developments in the area of dynamic metabolic control[173, 174].

Further FORCE pathways development should enable more efficient designs for synthetic methylotrophy and diverse product synthesis, especially via pathway iteration. We assess the primary bottleneck to be the HACS catalyzed acyloin condensation reaction of formyl-CoA with aldehydes. The observation of formate as a byproduct throughout various implementations using

formaldehyde or methanol is likely due to an imbalance between the rate of production of formyl-CoA and the rate of its utilization by HACS. We have also observed formyl-CoA hydrolysis[68], which is probably exacerbated *in vivo* by endogenous thioesterases. Strategies to address this limitation include re-activating formate to formyl-CoA using a CoA-transferase, as done here using the CoA-transferase CaAbfT, and identification or engineering of an HACS enzyme with better characteristics, shown here via the identification of BsmHACL. Finally, host-strain modifications such as the deletion of endogenous aldehyde dehydrogenases and thioesterases were also explored for this purpose.

As HACS-catalyzed condensation and enzyme activity was only recently described, we expect that further genome mining, bioprospecting, enzyme engineering, and biochemical characterization will result in better performing variants, ultimately overcoming pathway bottlenecks. HACS variants with well-defined chain length and functional group specificities, in combination with compatible, specific termination enzymes, will allow for the production of specific products, analogous to what has been demonstrated with other platform pathways[175-177]. These studies will also shed additional light on the role of formyl-CoA in metabolism, which is likely greater than the synthetic pathway described here. Recent reports have already contributed to the advancement of knowledge in this area[73, 178-180], and further studies are likely to follow.

Methods:

Reagents

All chemicals were obtained from Fisher Scientific Co. and Sigma-Aldrich Co. unless otherwise specified. Primers were synthesized by Integrated DNA Technologies or by Eurofins Genomics. Restriction enzymes were obtained from New England Biolabs unless otherwise specified.

Genetic methods

Genes non-native to *E. coli* were codon-optimized and synthesized by GeneArt (Thermo Fisher). *E. coli* genes were amplified from the chromosomal DNA. Plasmid-based gene expression was achieved by cloning the desired gene(s) into pCDFDuet-1 or pETDuet-1 (Novagen) digested with appropriate restriction enzymes and by using In-Fusion cloning technology (Clontech Laboratories, Inc.). Gene knockouts and genomic modifications were created using a CRISPR-Cas9-based system developed for *E. coli*. pCas and pTargetF were gifts from S. Yang (Addgene plasmids nos. 62225 and 62226, respectively)[181]. Plasmids and strains used in this study are listed in Supplementary Table 1.

Resting cell bioconversions

Bioconversions using resting cells were performed as described previously[16] with slight modification. The basal salts media used was M9 (6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl, 2 mM $MgSO_4$, 100 µM $CaCl_2$, and 15 µM thiamine-HCl) additionally supplemented with the micronutrient solution of Neidhardt[68]. An overnight LB culture of each strain was used to inoculate (1%) a 250 mL flask containing 50 mL of the above media further supplemented with 20 g/L glycerol, 10 g/L tryptone, 5 g/L yeast extract, and appropriate antibiotics (50 µg/mL

carbenicillin, 50 µg/mL spectinomycin). The flask cultures were incubated at 30°C and 250 rpm in an NBS I24 Benchtop Incubator Shaker (New Brunswick Scientific Co.). After 2.5 hours, gene expression was induced by addition of 0.1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) and 0.04 mM cumate (0.2 mM IPTG and 0.1 mM cumulate was used for the experiment with formaldehyde and formate).

The cells from the above cultures were harvested by centrifugation (5000×g, 22°C, 5 min), and washed twice with the above M9 media without any carbon source. The final cell pellet was resuspended in M9 with the appropriate carbon source (~10 OD_{600} with 10 mM formaldehyde or ~5 OD_{600} with 1 mM formaldehyde and 10 mM formate). 5 mL of the cell suspension was added to a 25 mL Erlenmeyer flask (Corning Inc.) and topped with a foam plug. Flasks were incubated at 30°C and 200 rpm in an NBS I24 Benchtop Incubator Shaker (New Brunswick Scientific Co.). An additional 10 mM formaldehyde was added after 1.5 hours when formaldehyde was the sole carbon source. Samples were taken after 24 hours for HPLC analysis as described previously[68]. When 13C-labeled formaldehyde was used as the substrate, the samples were analyzed by GC-MS after extraction and derivatization as described previously[68].

Fermentation experiments

The growth media used was M9 (6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl, 2 mM $MgSO_4$, 100 µM $CaCl_2$, and 15 µM thiamine-HCl) additionally supplemented with 500 mM methanol, 10 g/L tryptone, 5 g/L yeast extract and micronutrient solution of Neidhardt[68]. An overnight LB culture of each strain was used to inoculate (1%) a 50 mL closed-cap conical tube (Genesee Scientific Co.) containing 5 mL of the above media further supplemented with appropriate antibiotics (50 µg/mL carbenicillin, 50 µg/mL spectinomycin). After approximately 3 hours, gene expression was induced by addition of 0.04 mM isopropyl β-d-1-

thiogalactopyranoside (IPTG) and 0.04 mM cumate. Tubes were incubated at 30°C and 200 rpm in an NBS I24 Benchtop Incubator Shaker (New Brunswick Scientific Co.). Samples (100 µL) were taken every 24, 48, 72 and 96 hours after inoculation for OD_{600} measurement and HPLC analysis as described previously[16]. When 13C-methanol was used as the substrate, the samples were analyzed by GC-MS after extraction and derivatization as described previously[68].

Two-strain *E. coli* system for growth on C1 substrates

Two-strain experiments were conducted using strains cultured and induced as described previously using M9 medium[68]. The induced cells were resuspended to an initial concentration of $3*10^9$ CFU (colony forming unit)/mL (equivalent to OD_{600} of ~5) in M9 medium. 20 mL of the suspension was added into 25 mL flask containing 3 mg paraformaldehyde (equivalent to 5 mM), or 10 mL of the suspension was added into 25 mL flask with the addition of 500 mM methanol, or 1 mM formaldehyde and 10 mM sodium formate. A second *E. coli* strain, AC763, capable of consuming glycolate, was added to an initial concentration of $5*10^6$ CFU/mL (equivalent to OD_{600} of ~0.005). AC763 additionally harbored a chromosomal copy of constitutively expressed eGFP to assist in distinguishing the two strains. Prior to its addition to the culture, AC763 was pre-grown in 25 mL Erlenmeyer flasks (from a single colony inoculation) at 200 rpm and 30°C for 24 hours in 5 mL of the above M9 minimal media supplemented with 5 g/L glycolate and 2 g/L tryptone. Cells were then centrifuged (5000×g, 22°C, 5 min), washed twice with the media supplemented with 5 g/L glycolate, and resuspended to an optical density of ~0.05. Following 24 hours of incubation at 200 rpm and 30°C (5 mL in 25 mL Erlenmeyer flasks), cells were centrifuged (5000×g, 22°C), washed twice with media without any carbon source and an appropriate volume added to the two-strain system. The flasks containing both strains were further incubated at 200 rpm and 30°C. Samples were taken at various times for HPLC and cell growth analysis.

Colony forming units per mL of culture was utilized as a measurement of cell growth. Appropriate volumes of culture were diluted in the above described minimal media without any carbon source and 50 μL of various dilutions plated on minimal media plates containing 2.5 g/L glycolate. Following plate incubation at 37 °C, colonies were counted manually, aided by visualization using a blue-light transilluminator (Vernier, Beaverton, OR) to illuminate the eGFP expressing strain AC763.

Supplementary Materials

Supplementary Table 2-1. Host strains and plasmids used in this study. Uniprot accession numbers for heterologous enzymes used in this work are given in parenthesis.

Host Strains/ Plasmids	Description/Genotype/Usage	Source
BL21(DE3)	*E. coli* B F$^-$ *ompT gal dcm lon hsdS$_B$*($r_B^- m_B^-$) [*malB*$^+$]$_{K\text{-}12}$(λ^S) λ(DE3) - Host for protein expression for *in vitro* studies	Studier et al. [182]
MG1655	*E. coli* K-12 F- l- *ilvG- rfb*-50 *rph*-1	Blattner et al.[183]
AC440	MG1655 λ(DE3) Δ*frmA* Δ*fdhF* Δ*fdnG* Δ*fdoG* Δ*glcD::FRT* - Engineered host for resting cell studies	Chou et al.[68]
AC877	AC440 Δ*aldA* Δ*aldB* Δ*patD* Δ*puuC* - Engineered host for resting cell studies	This study
AC878	AC440 Δ*yciA* Δ*tesA* Δ*tesB* Δ*ybgC* Δ*ydiI* Δ*fadM* - Engineered host for methanol utilization	This study
AC763	MG1655 λ(DE3) Δ*frmA* Δ*fdhF* Δ*fdnG* Δ*fdoG* Δ*tesB*::P^{M193}-*eGFP* - C2-utilizing (sensor) strain for two-strain pathway evaluation. The *tesB* open reading frame was replaced with *eGFP* controlled by constitutive promoter M193[71]	This study
pCDFDuet-1	CloDF13, *lacI*, Sm^R	Novagen (Darmstadt, Germany)
pCDFDuet-1-P1-ntH6-RuHACL^{G390N}	pCDFDuet-1 with codon optimized 6xHis-tagged *Rhodospirillales bacterium URHD0017 HACL* (Uniprot: A0A1H8YFL8) with a G390N mutation under control of the *T7lac* promoter and *lacI*	Chou et al. [68]
pCDFDuet-1-P1-ntH6-RuHACL^{G390N}-P2-BmMDH2^{MGA3}	pCDFDuet-1 with codon optimized 6xHis-tagged *Rhodospirillales bacterium URHD0017 HACL* with a G390N mutation in the P1 cloning site and codon optimized *Bacillus methanolicus MGA3 MDH2* (Uniprot: I3E2P9) in the P2 cloning site both under control of the *T7lac* promoter and *lacI*	This study
pCDFDuet-1-P1-P2-BmMDH2^{MGA3}	pCDFDuet-1 with codon optimized *Bacillus methanolicus MGA3 MDH2* in the P2 cloning site both under control of the *T7lac* promoter and *lacI*	This study
pCDFDuet-1-P1-ntH6-BsmHACL	pCDFDuet-1 with codon optimized 6xHis-tagged *HACL* isolated from beach sand metagenome (UniProt: A0A3C0TX30) under control of the *T7lac* promoter and *lacI*	This study
pCDFDuet-1-P1-ntH6-BsmHACL-P2-BmMDH2^{MGA3}	pCDFDuet-1 with codon optimized 6xHis-tagged *HACL* isolated from beach sand metagenome in the P1 cloning site and codon optimized *Bacillus methanolicus MGA3 MDH2* in the P2 cloning site both under control of the *T7lac* promoter and *lacI*	This study
pCDFDuet-1-P1-ntH6-BsmHACL-P2-BmMDH2^{MGA3}-CaAbfT	pCDFDuet-1 with codon optimized 6xHis-tagged *HACL* isolated from beach sand metagenome in the P1 cloning site and a synthetic operon of codon optimized *Bacillus methanolicus MGA3 MDH2* and codon optimized *Clostridium aminobutyricum abfT* (UniProt: Q9RM86) in the P2 cloning site under control of the *T7lac* promoter and *lacI*	This study

pCDFDuet-1-P1-ntH6-LmACR	pCDFDuet-1 with codon optimized 6xHis-tagged *Lysteria monocytogenes acr* (Uniprot: Q8Y7U1) under control of the *T7lac* promoter and *lacI*	Chou et al. [68]
pCDFDuet-1-P1-ntH6-OfFrc	pCDFDuet-1 with codon optimized 6xHis-tagged *Oxalobacter formigenes frc* (Uniprot: O06644) under control of the *T7lac* promoter and *lacI*	This study
pETDuet-1	pBR322-derived ColE1 origin, *lacI*, Amp^R	Novagen (Darmstadt, Germany)
pETDuet-1-P1-EcFucO	pETDuet-1 with *Escherichia coli fucO* in the P1 cloning site under control of the *T7lac* promoter and *lacI*	This study
pETDuet-1-P^{CT5}-LmACR	pETDuet-1 with codon optimized *Lysteria monocytogenes acr* expressed under control of the cumate inducible CT5 promoter and *cymR*	Chou et al. [68]
pETDuet-1-P^{CT5}-LmACR-EcAldA	pETDuet-1 with codon optimized *Lysteria monocytogenes acr* and *Escherichia coli aldA* in a synthetic operon under control of the cumate inducible CT5 promoter and *cymR*	Chou et al. [68]
pETDuet-1-P^{CT5}-LmACR-EcFucO	pETDuet-1 with codon optimized *Lysteria monocytogenes acr* and *Escherichia coli fucO* in a synthetic operon under control of the cumate inducible CT5 promoter and *cymR*	This study
pETDuet-1-P^{CT5}-CaAbfT	pETDuet-1 with codon optimized *Clostridium aminobutyricum abfT* under control of the cumate inducible CT5 promoter and *cymR*	This study
pRSFDuet-1	pRSF1030-derived RSF origin, lacI, KanR	Novagen (Darmstadt, Germany)
pRSFDuet-1-P1-ntH6-KoPddABC-P2-KoDdrAB-EcYciK-EcBtuR	pRSFDuet-1 with codon optimized 6xHis-tagged *Klebsiella oxytoca pddA* (Uniprot: Q59470) in a synthetic operon with codon optimized *Klebsiella oxytoca pddB* (Uniprot: Q59471) and *pddC* (Uniprot: Q59472) in the P1 cloning site and codon optimized *Klebsiella oxytoca ddrA* (Uniprot: O68195) and *ddrB* (Uniprot: O68196) in a synthetic operon with *Escherichia coli yciK* and *btuR* in the P2 cloning site both operons under control of the T7lac promoter and lacI	This study

Chapter 3. Combinatorial sequence and structure analysis to identify and engineer efficient one-carbon utilization enzymes

Introduction

One-carbon (C1) compounds have become attractive feedstock for sustainable biomanufacturing with advances in electrochemical reduction of CO_2 powered by rapid deployment of carbon-free electricity. While electrochemistry offers an efficient conversion from CO_2 to other C1 compounds including carbon monoxide[184], formic acid (formate)[185], formaldehyde[186], methanol[187] and methane[188], biochemistry has a great potential in creating carbon-carbon (C-C) bonds between the C1 compounds to produce value-added multi-carbon compounds. Unfortunately, industrial chasses such as *Escherichia coli* or *Saccharomyces cerevisiae* cannot harness C1 compounds natively which brings up a question whether engineering model organisms to become synthetic C1-trophs or engineering native C1-trophs to produce desired products would be the a better strategy[119]. Although the latter has had some great success thanks to advances in synthetic biology[9], there is still huge potential in the former approach considering the high growth rate, cell density, engineering throughput, as well as the diversity of products that have been demonstrated at industrial-scale using these chasses. Several native and synthetic C1 utilization pathways, such as Calvin-Benson-Bassham (CBB) cycle[189], Ribulose Monophosphate (RuMP) cycle[13, 14], Serine cycle[67] and Reductive Glycine pathway (rGlyP)[15] have successfully been demonstrated in *E. coli* by introducing only a few heterologous enzymes as these pathways share most enzymes and metabolites in common with the host central metabolism. However, it is important to note that in all cases: (1) moderate to extensive adaptive laboratory evolution was required for cells to develop the C1-trophy; (2) even the evolved strain cannot compete with native C1-trophs in growth rate and biomass yield; and (3) no demonstration of value-added product synthesis has been reported using these strains[119].

As an alternative, we proposed a synthetic, orthogonal C1 metabolism[25] which operates independently from the host metabolism. While introducing non-native metabolic pathways normally requires further engineering of the host metabolism in parallel to rewire flux toward product synthesis, orthogonal pathways only require engineering of the metabolic nodes (valves) to control flux toward cell growth or product synthesis. However, because orthogonal pathways build upon novel chemistries that rely on enzyme promiscuity and/or protein engineering, suboptimal kinetics of these enzymes cause bottleneck in the pathway flux[190]. Recently reported C1-C1 coupling enzymes, such as formolase (FLS)[93], glycolaldehyde synthase (GALS)[97], glyoxylate carboligase (GCL)[98] and 2-hydroxyacyl-CoA synthase/glycolyl-CoA synthase (HACS/GCS)[68, 73] are examples of such enzymes that can work as core enzymes for orthogonal C1 utilization pathways. FLS and GALS/GCL are engineered aldolases that catalyze C-C coupling of three or two formaldehyde molecules into dihydroxyacetone or glycolaldehyde, respectively. Although there have been efforts to engineer these enzymes via directed evolution, all variants still suffer from poor turnover and affinity with formaldehyde[96]. Considering the toxicity of formaldehyde causing DNA-protein cross-linking[13], these enzymes are not suitable for in vivo implementation.

HACS/GCS, on the other hand, catalyzes the C-C coupling between formaldehyde and a C1 moiety, formyl-CoA to form glycolyl-CoA[68]. 2-hydroxyacyl-CoA lyase/synthase (HACL/HACS) is originally identified from mammalian α-oxidation and has better catalytic efficiency toward longer chain substrates. We identified a HACS variant of prokaryotic origin, *Rhodospirillales bacterium URHD0017* HACL (RuHACL), through bioprospecting which showed improved affinity toward formaldehyde (K_m = 29 mM)[68]. Although RuHACL still has better catalytic efficiency with longer chain aldehydes, its promiscuous activity with formaldehyde

exhibits substantially higher turnover number (k_{cat} = 3.3 s^{-1}) than FLS or GALS, which leads to up to two orders of magnitude higher catalytic efficiency with formladehyde. Superior kinetics allowed the demonstration of the HACS-based pathways, formyl-CoA elongation (FORCE) pathways in vivo including growing cell cultures using methanol as C1 substrate[25]. However, the level of flux in the FORCE pathways is still insufficient to fully support cell growth or show industrially relevant titer and rate from C1 substrates.

In this study, we used gene homolog bioprospecting and clustering based on sequence similarity to identify variants with improved activity for C1-C1 condensation. Structures of high performing variants were modeled using AlphaFold[191], followed by combinatorial sequence and structure analysis to understand the key catalytic residues in the active site pocket. Furthermore, we showed the improvement of substrate affinity by rationally engineering the active site residues of a variant with high turnover (k_{cat}) using another variant with high affinity (low K_m) as a template, which resulted in improved glycolate productivity. In parallel, exploring more in depth in the phylogenetic tree of the active variants identified from the first-round variants revealed variants with improved catalytic efficiencies and better understanding of the active clusters for C1-C1 condensation activities. Using the newly identified variants with better performances, we show significantly improved titer and yield of FORCE pathway products. Overall, we believe this approach could be readily applied in other synthetic reactions and pathways to identify and engineer new-to-nature chemistries with limited access to the crystal structure of the related enzymes.

Results

Establishing HACS screening platform

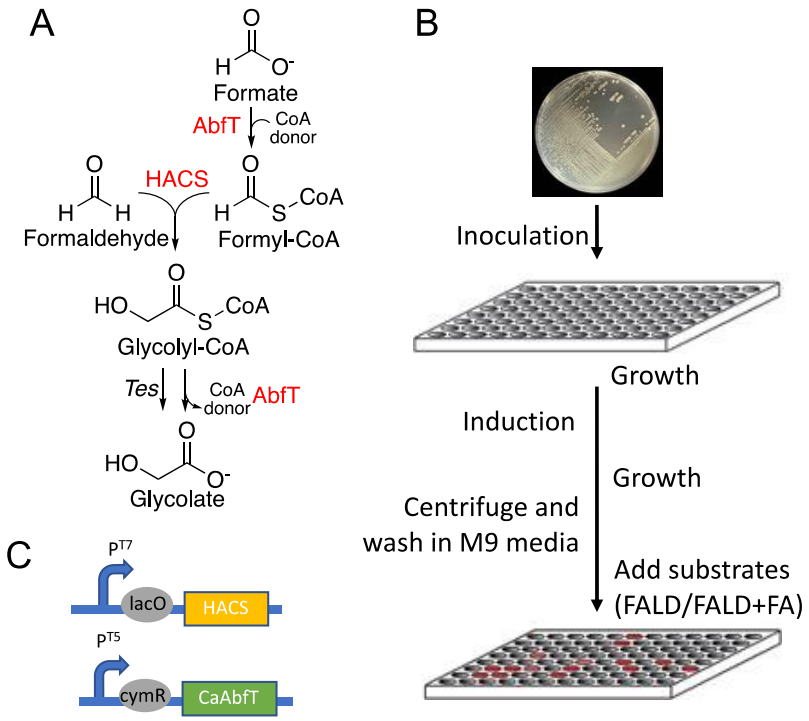


Figure 3-1. Establishing 96 well-based resting-cell bioconversion platform. A) Glycolate production from formaldehyde and formate via formate activation enzyme (AbfT) and 2-hydroxyacyl-CoA synthase (HACS). B) 96 well-based resting cell bioconversion platform for screening HACS. C) HACS and AbfT are expressed under independently inducible promoters: HACS expressed under IPTG-inducible T7 promoter and AbfT under cumate-inducible T5 promoter.

We have previously established a resting-cell bioconversion platform[25] to convert formaldehyde to glycolic acid (glycolate) by overexpressing HACS, acyl-CoA reductase (ACR) that can oxidize formaldehyde to formyl-CoA and reduce glycolyl-CoA to glycolaldehyde and aldehyde dehydrogenase (ALDH) that oxidizes glycolaldehyde to glycolate. Alternatively, formyl-CoA can be generated by activation of formate to formyl-CoA via ATP-consuming synthetases, kinase-phosphoacyltransferase pairs or acyl-CoA transferases requiring specific CoA donor molecules[25]. Among them, acyl-CoA transferase from *Clostridium aminobutyricum* (AbfT) was

chosen for rapid screening of HACS activities as it can recycle CoA by utilizing glycolyl-CoA as CoA donor[73] (Figure 3-1A). Decoupling source for formaldehyde and formyl-CoA with formate and AbfT provides three advantages: (1) formyl-CoA flux is not affected by formaldehyde concentrations; (2) only two enzymes need to be expressed for screening reducing complexity in multiple enzyme expressions in vivo; (3) moles of formaldehyde consumption per mole glycolate production is halved to prevent rapid depletion of formaldehyde causing underestimation of enzyme activities. As a host for screening, we used an engineered *E. coli* strain MG1655(DE3) with knockouts for formaldehyde (Δ*frmA*) and formate (Δ*fdhF* Δ*fdnG* Δ*fdoG*) oxidation as well as for glycolate utilization (Δ*glcD*), which we expected could compete or interfere with the analysis of our pathway. The resting-cell bioconversion setting scaled down from shake-flask (5-10 mL reaction)[25] to 96 deep-well plates (1 mL reaction) to improve the throughput for testing multiple variants, pathways, and relative expression levels of the pathway enzymes (Figure 3-1B).

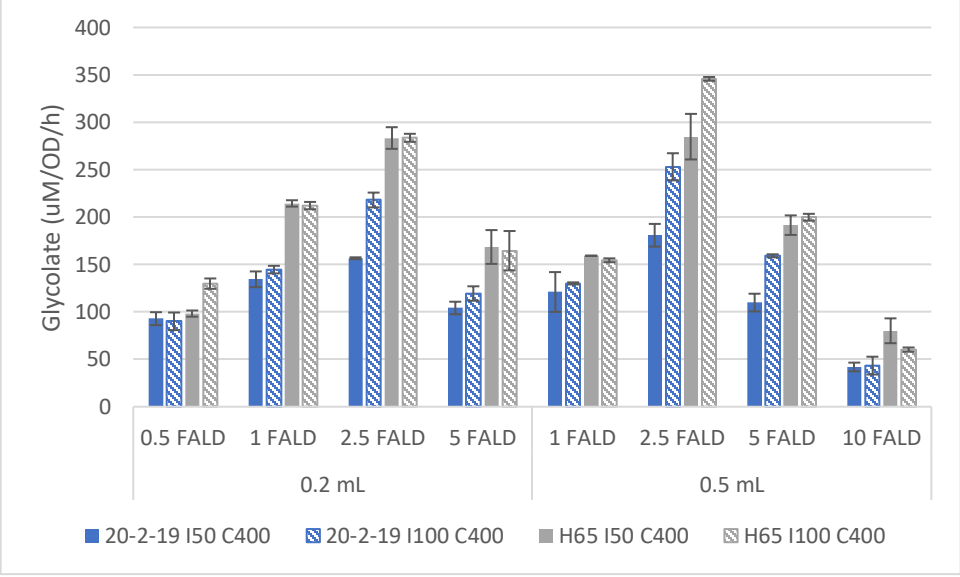


Figure 3-2. Impact of varying formaldehyde concentrations and inducer concentrations on glycolate productivity. Four different formaldehyde (FALD) concentrations (0.5, 1, 2.5, 5 mM), two preculture volumes (0.2 and 0.5 mL) and two different inducer concentrations (50 and 100 µM for IPTG and cumate fixed at 400 µM) are tested for two different HACS variants: JGI20 R480ins L549H T550G

R551del (20-2-19) and JGIH65 (H65). Bars drawn at the mean of duplicate biological replicates and error bars drawn to the standard deviation.

Expression of HACS and AbfT are controlled independently using orthogonal inducible promoters, HACS under IPTG-inducible T7 promoter and AbfT under cumate-inducible T5 promoter system (Figure 3-1C). We first tested varying levels of IPTG and cumate to find the best inducer matrix that yields optimal relative expression levels between HACS and AbfT (Figure 3-2). Level of formaldehyde concentration is another important factor and the relationship between varying cell density and formaldehyde concentrations are investigated to conclude that 2.5 mM formaldehyde under 8-10 OD600 cell density showed the best representation of the glycolate productivity. Formaldehyde concentrations higher than 2.5 mM showed substantial drop in glycolate productivity possibly due to the toxicity of formaldehyde (Figure 3-2).

Identification and screening of the first-round HACS homologs

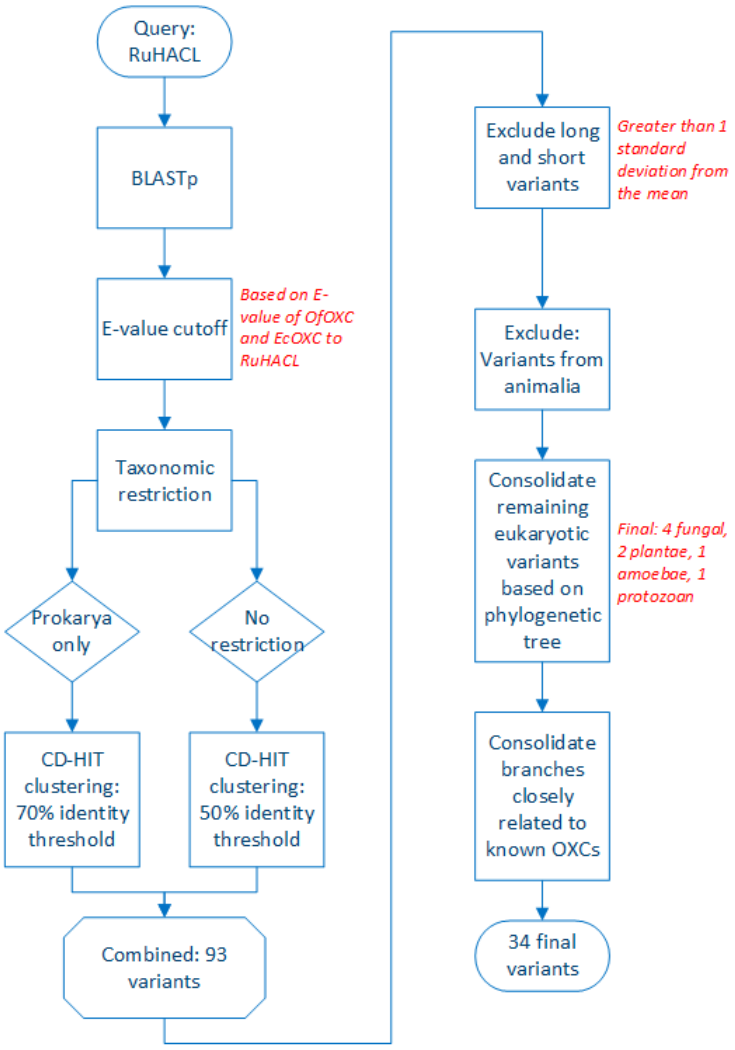


Figure 3-3. Gene homolog bioprospecting strategy used for identification of first-round HACS variants.

2-hydroxyacyl-CoA lyase (HACL) from *Rhodospirillales bacterium URHD0017* (RuHACL) was previously identified to demonstrate substantially improved affinity toward formaldehyde and expression level in *E. coli* than the human enzyme (HsHACL)[68]. Using RuHACL as a reference, unsorted gene homolog database was collected using Basic Local Alignment Search Tool (BLAST)[192] with E-value cutoff set based on the E-value between RuHACL and oxalyl-CoA decarboxylase (OXC) from *Escherichia coli* (EcOXC) and *Oxalobacter*

formigenes (OfOXC) (Figure 3-3). CD-HIT web server[193] was used to sort out and down select the representative variants by clustering query results into groups with specific identity threshold. Restriction of 70% identity threshold was imposed for genes from prokaryotic origin whereas 50% was used for no taxonomic restriction (Figure 3-3). Clustering and selecting representative genes using CD-HIT gave 93 HACS variants similar to RuHACL. The list was further curated by removing too long or too short sequences and variants from animalia, which could result in poor expression levels in *E. coli*. As a result, 34 remaining variants after curation were chosen to be the initial round of HACS variants for synthesis and testing in *E. coli* as a host (Figure 3-3). The selected genes were then codon-optimized for expression in *E. coli* and synthesized in collaboration with Joint Genome Institute (JGI). 5 variants failed during the synthesis, which left 29 first-round JGI HACS variants (JGI1 to JGI29) (Supplementary Figure 3-17).

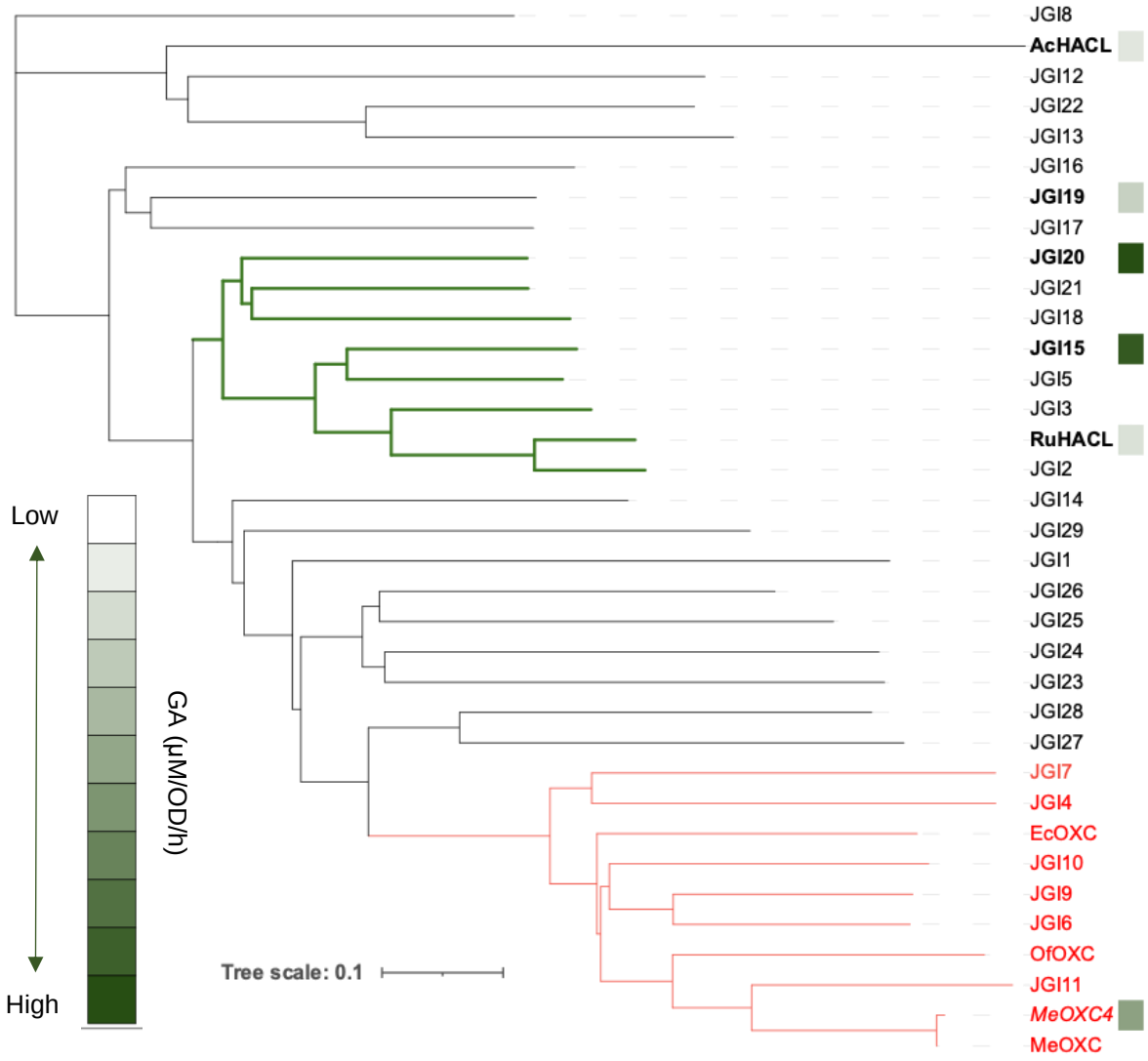


Figure 3-4. Phylogenetic tree diagram for the first-round HACS variants with heat map diagram showing glycolate (GA) productivity. Nodes and subclades that branch out RuHACL, JGI15 and JGI20 are highlighted in dark green. Variants highlighted in bold represent the selected variants as reference for the second-round gene homolog bioprospecting. Branches and leaves that belong to the OXC cluster are highlighted in red. MeOXC4 is an engineered version of MeOXC for HACS activities via directed evolution[73]. Phylogenetic tree with heatmap was constructed using Interactive Tree Of Life (iTOL)[194] .

29 first-round JGI variants, RuHACL, recently reported 2-hydroxyacyl-CoA lyase from *Actinomycetospora chiangmaiensis* (AcHACL)[179] as well as known oxalyl-CoA

decarboxylases (OXC) including EcOXC, OfOXC, *Methylobacterium extorquens* OXC (MeOXC) and its mutant MeOXC4[73] are screened in the 96 well plate-based resting-cell bioconversion described in Figure 1. Phylogenetic tree for the first-round variants show that the distances of the variants are well spread out from very close ones like JGI2 to distant ones like JGI8 and JGI11 on either end of the tree (Figure 3-4). Only three (JGI15, 19 and 20) out of 29 first round variants showed glycolate production with JGI15 and JGI20 showing substantially higher glycolate productivity than RuHACL and MeOXC4 (Figure 3-4). It is worth noting that both JGI15 and JGI20 belong to the same node as RuHACL with phylogenetic proximity (highlighted in dark green in Figure 3-4). Interestingly, AcHACL, which is phylogenetically distant from all first round variants and was reported to catalyze reversible cleavage of 2-hydroxyisobutyryl-CoA to acetone and formyl-CoA[179], also showed activity for C1-C1 condensation at similar level as RuHACL (Figure 3-4, 3-5).

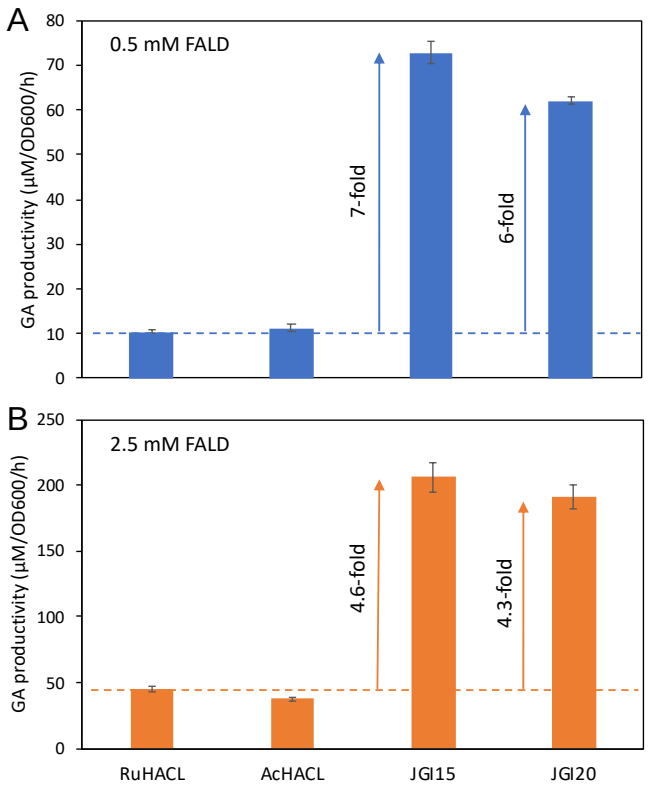


Figure 3-5. Glycolate (GA) productivity comparison between RuHACL, AcHACL, JGI15 and JGI20 at two different formaldehyde concentrations (0.5 and 2.5 mL). JGI15 and JGI20 have 7-fold and 6-fold higher glycolate productivities under 0.5 mM formaldehyde and 4.6-fold and 4.3-fold higher activities under 2.5 mM formaldehyde than RuHACL, respectively. Bars drawn to the mean of triplicate biological replicates with error bars representing standard deviation.

We further investigated the active variants at low formaldehyde concentration (0.5 mM) to see the performance of these enzymes below the toxicity limit. The fold increase in glycolate productivity of JGI15 and 20 further increased to 7-fold and 6-fold respectively, compared to RuHACL (Figure 3-5A), indicating better affinity of these variants with formaldehyde. Kinetic characterization of the purified enzymes further corroborates this result, where JGI15 (K_m = 8.8 mM) has better affinity toward formaldehyde than JGI20 (K_m = 12.1 mM), while JGI20 (k_{cat} = 7.8 s^{-1}) has better turnover number than JGI15 (k_{cat} = 4.7 s^{-1}) (Table 3-1). Both variants exhibit substantially improved affinity and turnover number compared to RuHACL with more than 5-fold

and 6-fold improvement in catalytic efficiency (k_{cat}/K_m) for JGI15 and JGI20, respectively (Table 3-1).

Table 3-1. Kinetic parameters of newly identified HACS variants and mutants.

Substrate	$k_{cat,app}$ (s^{-1})	$K_{m,app}$ (mM)	$k_{cat,app}/K_{m,app}$ (M^{-1} s^{-1})	Reference
RuHACL				
Formaldehyde	3.3 ± 0.3	29 ± 8	110	[68]
Formyl-CoA	1.5 ± 0.1	0.20 ± 0.05	7200	
MeOXC4				
Formaldehyde	2.0 ± 0.2	5 ± 1	400	[73]
Formyl-CoA	2.4 ± 0.2	0.22 ± 0.03	11000	
JGI15				
Formaldehyde	4.7 ± 0.1	8.8 ± 0.6	534	This study
Formyl-CoA	3.9 ± 0.2	0.17 ± 0.02	23000	
JGI20				
Formaldehyde	7.8 ± 0.1	12.1 ± 0.7	645	This study
Formyl-CoA	7.6 ± 0.3	0.14 ± 0.02	54000	

Understanding key catalytic residues via sequence and structure analysis

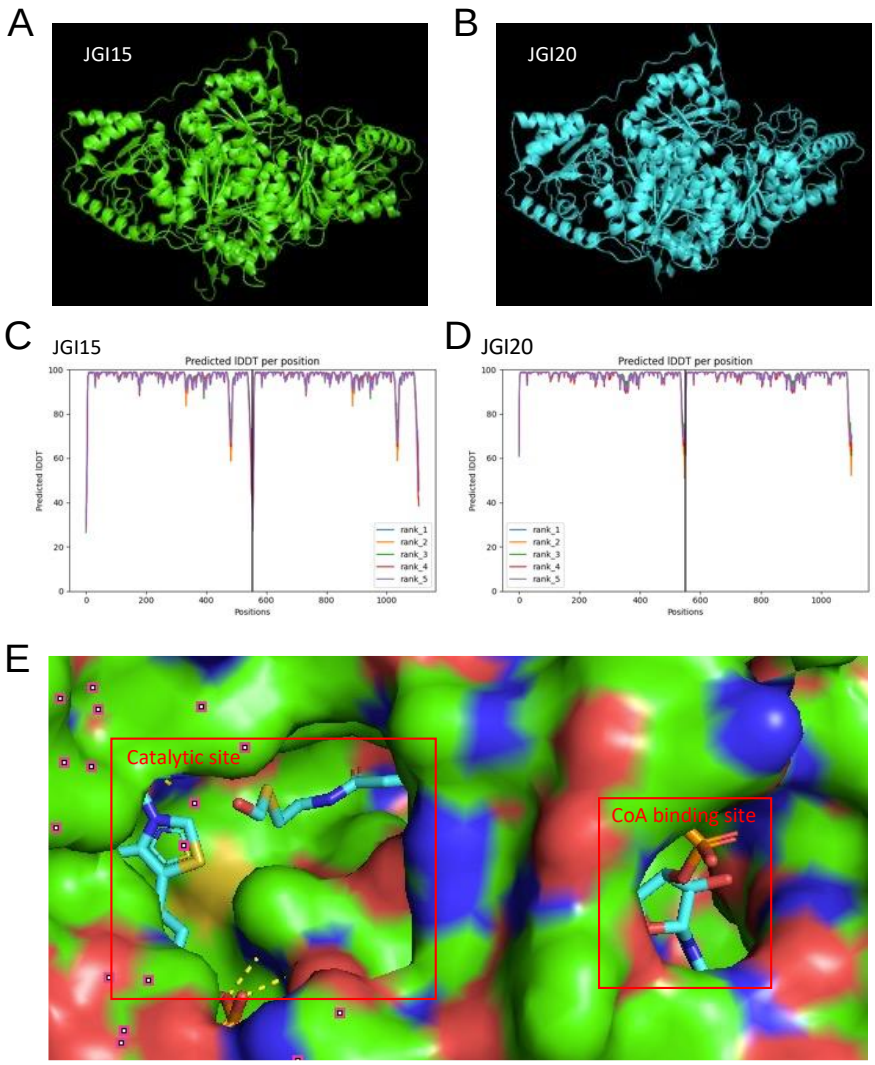


Figure 3-6. Protein structure modeling of JGI15 and JGI20 using AlphaFold. (A) JGI15 and (B) JGI20 dimeric structure predicted by AlphaFold[191] using ColabFold[195] platform. Prediction pLDDT scores of (A) JGI15 and (B) JGI20. (E) Catalytic site and CoA binding site exposed to the solvent space from surface plot of ligands (formyl-CoA and TPP) aligned to JGI15 and JGI20

To better understand the relationship between sequence and activity of HACS homologs, we investigated multiple sequence alignment (MSA) of all first-round variants. To narrow down the sequence space for active site residues, AlphaFold[191] predicted protein structures of high performing variants were used. The full homodimeric structure of JGI15 (Figure 3-6A) and JGI20

(Figure 3-6B) are constructed using the ColabFold platform[195], which showed high pLDDT score (> 90) except for the n-terminal and c-terminal regions (Figure 3-6C and D). The structures are aligned with the crystal structure of oxalyl-CoA decarboxylase from *Oxalobacter formigenes* (OfOXC) in complex with formyl-CoA (PDB code: 2JI8)[196] to understand the position and orientation of the two key ligands, thiamine diphosphate (TPP) and formyl-CoA, in the active site. The AlphaFold-generated structures of JGI15 and JGI20 exhibit high structural similarity with the OfOXC crystal structure with root-mean-square distance (RMSD) values of 1.185 Å and 0.981 Å for JGI15 and JGI20, respectively. The catalytic site (where TPP and the formyl residue of formyl-CoA interface for catalysis with carbonyl) and the CoA binding site are both exposed to the solvent space indicating correct orientation of ligands docked on to the JGI15 and JGI20 structures (Figure 3-6E).

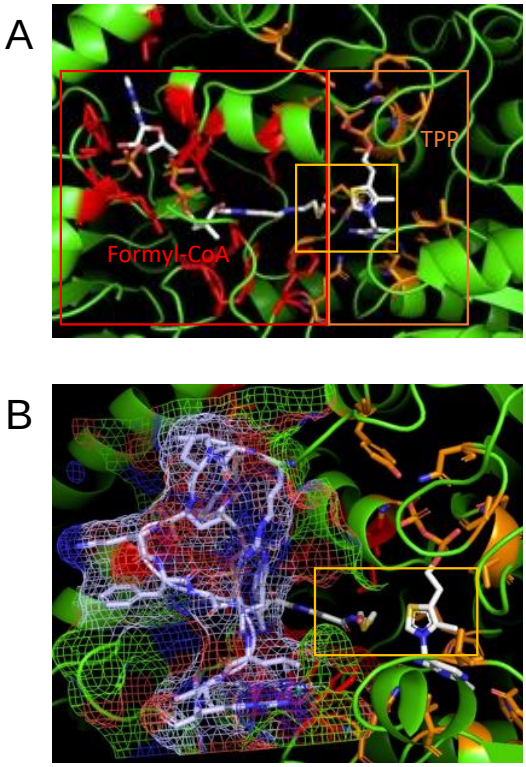


Figure 3-7. Active site of JGI20 showing (A) formyl-CoA and TPP binding region and (B) c-terminal covering loop on the catalytic site. Residues in the formyl-CoA binding region are highlighted

in red, residues in the TPP binding region are highlighted in orange in (A). Catalytic site is highlighted in yellow boxes in (A) and (B)

From the two best variants from the first round, we selected JGI20 for further understanding and engineering as it exhibits higher k_{cat} (7.8 s^{-1}) than RuHACL and JGI15 (Table 3-1). To identify JGI20 residues responsible for catalytic function and substrate binding, we decided to focus on three specific regions in the active site: TPP binding, formyl-CoA binding, and c-terminal tail regions (Figure 3-7). Literature suggests that TPP binding region is highly important for the catalytic activity of HACS and OXC family enzymes[73, 180]. Formyl-CoA binding would be important in determining not only the affinity toward formyl-CoA but also the catalytic site pocket size that affects chain-length specificity of the carbonyl substrate. Lastly, c-terminal tail was found to be folding over the active site forming a "covering loop[196]" having an important role in determining the substrate size[180] (Figure 3-7B). Subsequently, residues within 3.5 Å from TPP and formyl-CoA based on the ligands-bound AlphaFold structure of JGI20 (Figure 3-7A), as well as the c-terminal residues are selected for analysis (Supplementary Figure 3-19).

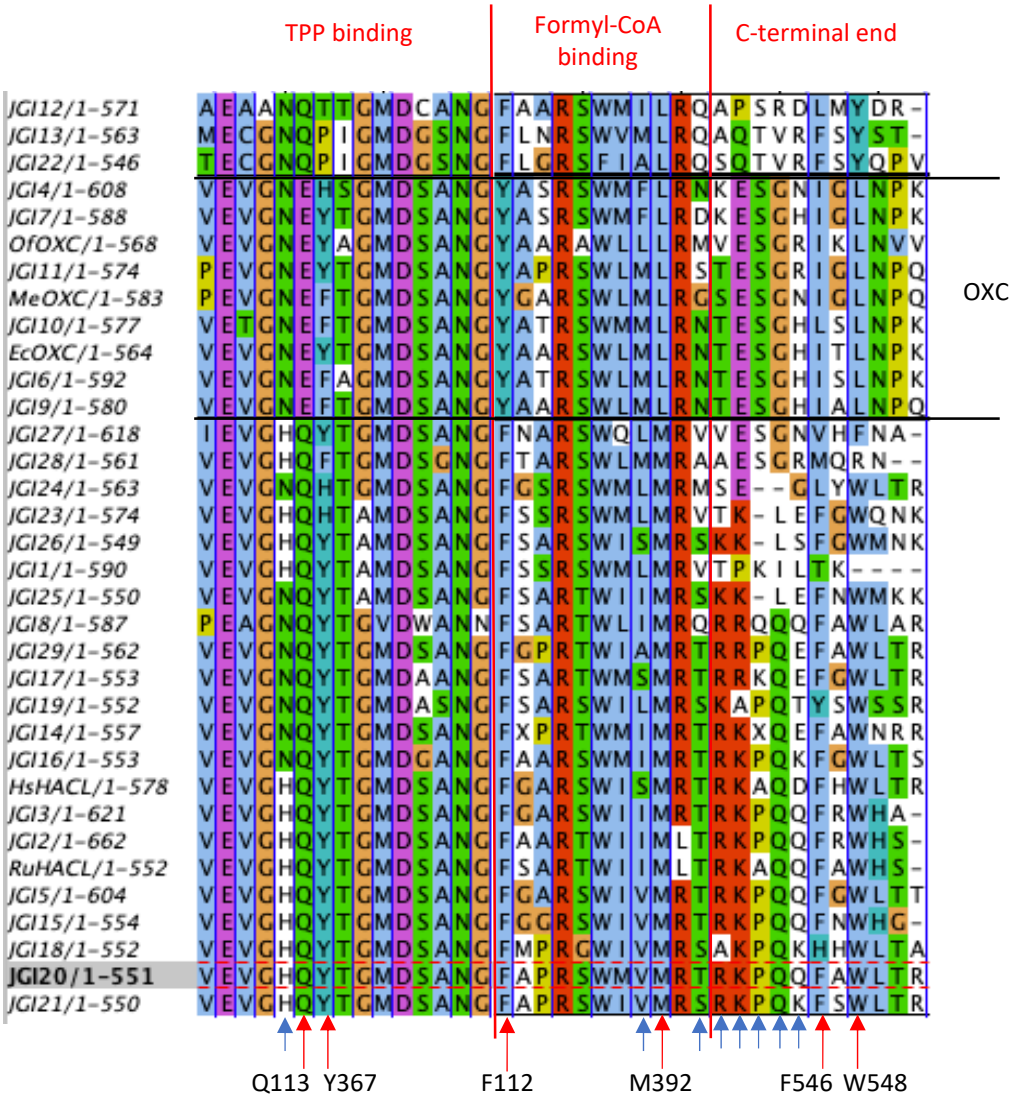


Figure 3-8. Multiple sequence alignment (MSA) of first-round variants showing TPP binding, formyl-CoA binding and c-terminal residues. Arrows at the bottom represent residues chosen for alanine scanning and residues hypothesized to have catalytic functions are highlighted in red. Rows with OXC variants with signature "YE" residues are divided for reference.

In parallel, the sequence of all 29 first round HACS variants (JGI1 to 29) and known enzymes from OXC/HACL family (i.e. HsHACL, RuHACL, EcOXC, OfOXC and MeOXC) are aligned to explore conservation of residues from the three regions (Figure 3-8). Majority of the residues in the TPP binding regions are fully conserved among HACS and OXC variants as expected but much less conserved residues are observed in the formyl-CoA binding and c-terminal

end regions. From the unconserved residues, ones that are uniquely conserved among active variants (RuHACL, JGI15 and JGI20) were chosen for further analysis to understand their role in the catalysis: 3 residues from TPP binding region (H80, Q113 and Y367), 6 from formyl-CoA binding region (F112, V354, M392, T397) and the 7 residues from the conserved motif at the c-terminal end of JGI20 (R541, K542, P543, Q544, Q545, F546 and W548) (Figure 3-8).

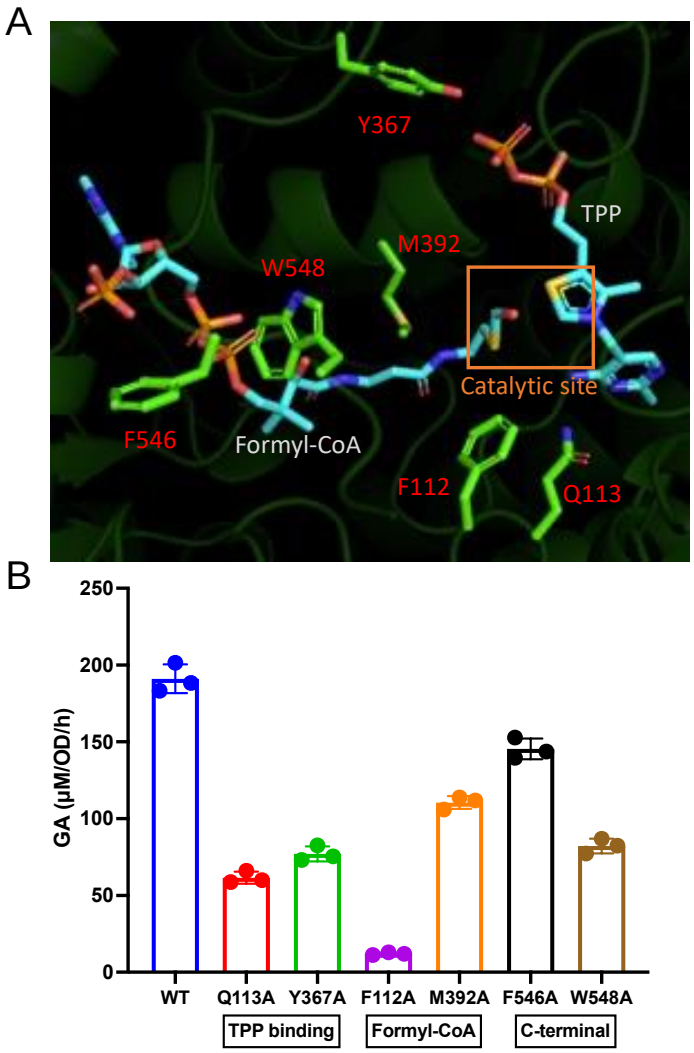


Figure 3-9. Identification of key active site residues of JGI20 via alanine scanning. A) Active site of JGI20 with residues that are hypothesized to have catalytic functions. B) Alanine scanning result of six identified residues. Activities represented by glycolate (GA) productivity under in vivo bioconversion. Bars drawn at the mean of triplicate biological replicates with error bars representing standard deviation.

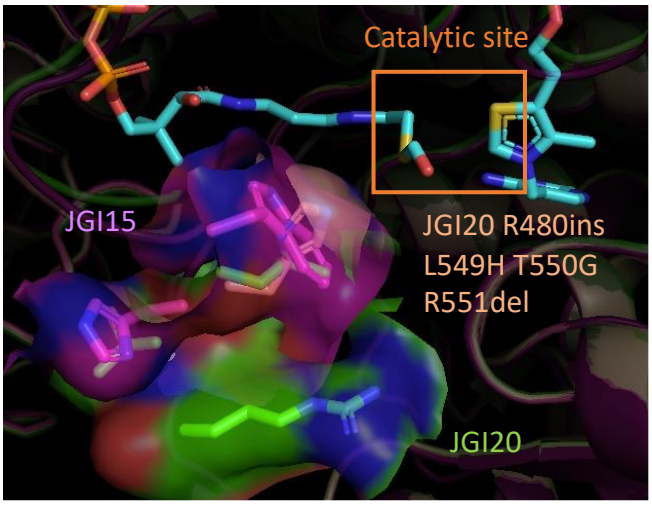
All mutants are lower than wildtype (WT) with statistical significance based on Student's t-test ($p < 0.0001$).

We used Alanine scanning[197] to rapidly identify the importance of these residues in catalytic function and stability of these enzymes. The result shows that Glutamine113 (Q113) and Tyrosine367 (Y367) from TPP binding, Phenylalanine112 (F112) and Methionine392 (M392) from formyl-CoA binding, and Phenylalanine546 (F546) and Tryptophan548 (W548) from c-terminal regions are important for the HACS activity on formaldehyde-formyl-CoA condensation (Figure 3-9, Supplementary Figure 3-16). From the analysis of AcHACL crystal structure, the Q residue in the position of JGI20 Q113 was reported to have a key catalytic function[180]. Also in RuHACL, mutagenesis of F and Q in the position of F112 Q113 in JGI20 abolished C1-C1 condensation activity[68]. The alanine scanning result (Figure 3-9B) further corroborates the previous findings in catalytic importance of FQ residues. In case of JGI15 and JGI20, F to A mutation had greater impact in activity than Q to A (Figure 3-9B and Supplementary Figure 3-16A). Based on the sequence alignment, oxalyl-CoA decarboxylase (OXC) variants including EcOXC, OfOXC and MeOXC have conserved "Tyrosine-glutamic acid (YE)" residues in the place of FQ (Figure 3-8). Notably, variants from the first round HACS variants that has YE instead of FQ residues (JGI4, 6, 7, 9, 10, 11) showed no glycolate productivity and are clustered together with OXCs in the phylogenetic tree (Figure 3-4). Taken together, FQ and YE could be the key catalytic residues that distinguish HACS and OXC type of enzymes. Interestingly, the two additional residues (Y367 and M392) found to be important for activity (Figure 3-9B) are also only conserved among HACS but not OXC. Residues in the position corresponding to Y367 are not conserved among OXC and OXCs have conserved leucine (L) residue in the place of M392 (Figure 3-8). Therefore, the two residues might also be important in distinguishing catalytic mechanisms of HACS and OXC family of enzymes.

From the conserved c-terminal motif (RKPQQF-W) among active variants, none of the c-terminal residues completely abolished the activity from point mutation to alanine (Supplementary Figure 3-16A). Based on literature, c-terminal tail of HACS serves as a covering loop stabilizing substrate (formyl-CoA) binding (Figure 3-7B) having an important role in limiting the substrate size in the binding pocket[180]. Alanine substitution of the last two residues, F546A and W548A showed considerable impact in the activity (Figure 3-9B). This could be because the last few residues of the c-terminus (F546, A547, W548, L549, R550 of JGI20) are located closer to the catalytic site, whereas RKPQQ residues that come prior to them are covering the CoA moiety, according to the AlphaFold-predicted structure of the c-terminal tail (Figure 3-7B, Figure 3-10A).

Protein engineering of JGI20 using JGI15 as a template

A



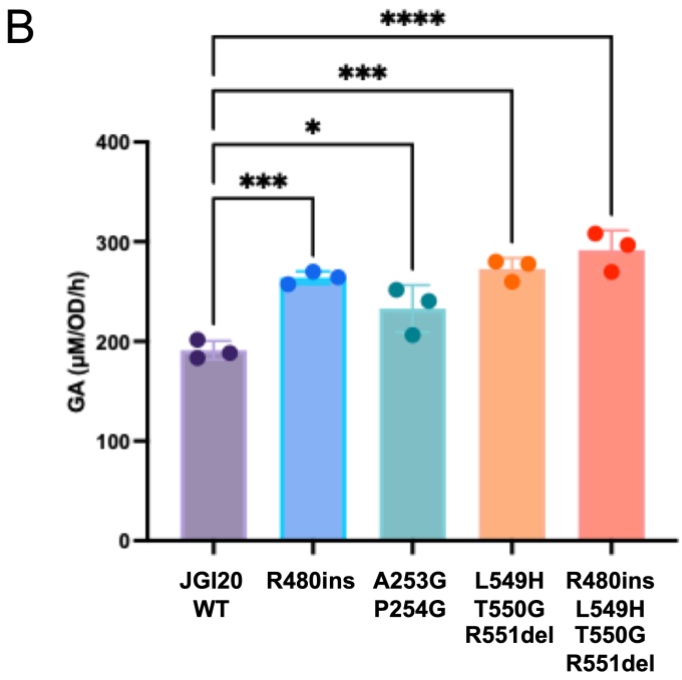


Figure 3-10. Rational engineering of JGI20 using JGI15 as a template. (A) Orientation of the C-terminal covering loop of JGI15 (violet pink), JGI20 (green) and JGI20 R480ins L549H T550G R551del (peach). (B) Activities of "JGI15-like" JGI20 mutants represented by glycolate (GA) productivity under in vivo bioconversion. Bars drawn at the mean of triplicate biological replicates with error bars representing standard deviation. All mutants are higher than JGI20 wildtype (WT) with statistical significance based on Student's t-test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

JGI20 has higher k_{cat} but also higher K_m with formaldehyde than JGI15. We hypothesized that we could improve either the affinity of JGI20 or the turnover of JGI15 by creating a hybrid protein between the two. To identify the structural difference between the two proteins we used "Pairwise Structure Alignment" function in the Protein Data Bank (PDB) website (www.rcsb.org). JGI15 and JGI20 structures modeled by AlphaFold were used for structure comparison and the result shows that there are two residues that are not aligned between the two protein structures (Supplementary Figure 3-17A). The JGI15-20 hybrid protein was constructed by inserting or deleting a residue to align the two structures without gaps. As a result, JGI15 N465ins, R493del and N465 R493del are constructed to make JGI15 "JGI20-like" whereas JGI20 N461del, R480ins

and N461del R480ins are constructed to make JGI20 “JGI15-like”. While all three JGI20-like JGI15 mutants showed decreased activities, JGI15-like JGI20 showed notable improvement in activities, especially R480ins showing up to 38% increase in activity (Supplementary Figure 3-17B, Figure 3-10B).

An alternative approach intended to improve the formaldehyde binding affinity of JGI20 by engineering the active site of JGI20 to mimic JGI15. Comparing the active site residues of the two enzymes, residues in the TPP binding region are fully conserved and only unmatching residues in the formyl-CoA binding region are A253, P254 and M276 of JGI20 (Supplementary Table 3-5). Consequently, JGI20 A253G P254G and JGI20 M276I mimicking the active site of JGI15 were constructed. Another target region was the c-terminal end, where JGI15 and 20 have unmatching residues at the end, which were shown to be important for enzyme activity from alanine scanning (Figure 3-9, 3-10A). Based on the protein structure of JGI15 and 20, the difference in c-termini shows slightly different orientation of the closing loop (Figure 3-10A). We hypothesized that this difference could contribute to the difference in formaldehyde binding affinity between JGI15 and JGI20 and constructed JGI20 mutant using JGI15 as template: JGI20 L549H T550G R551del.

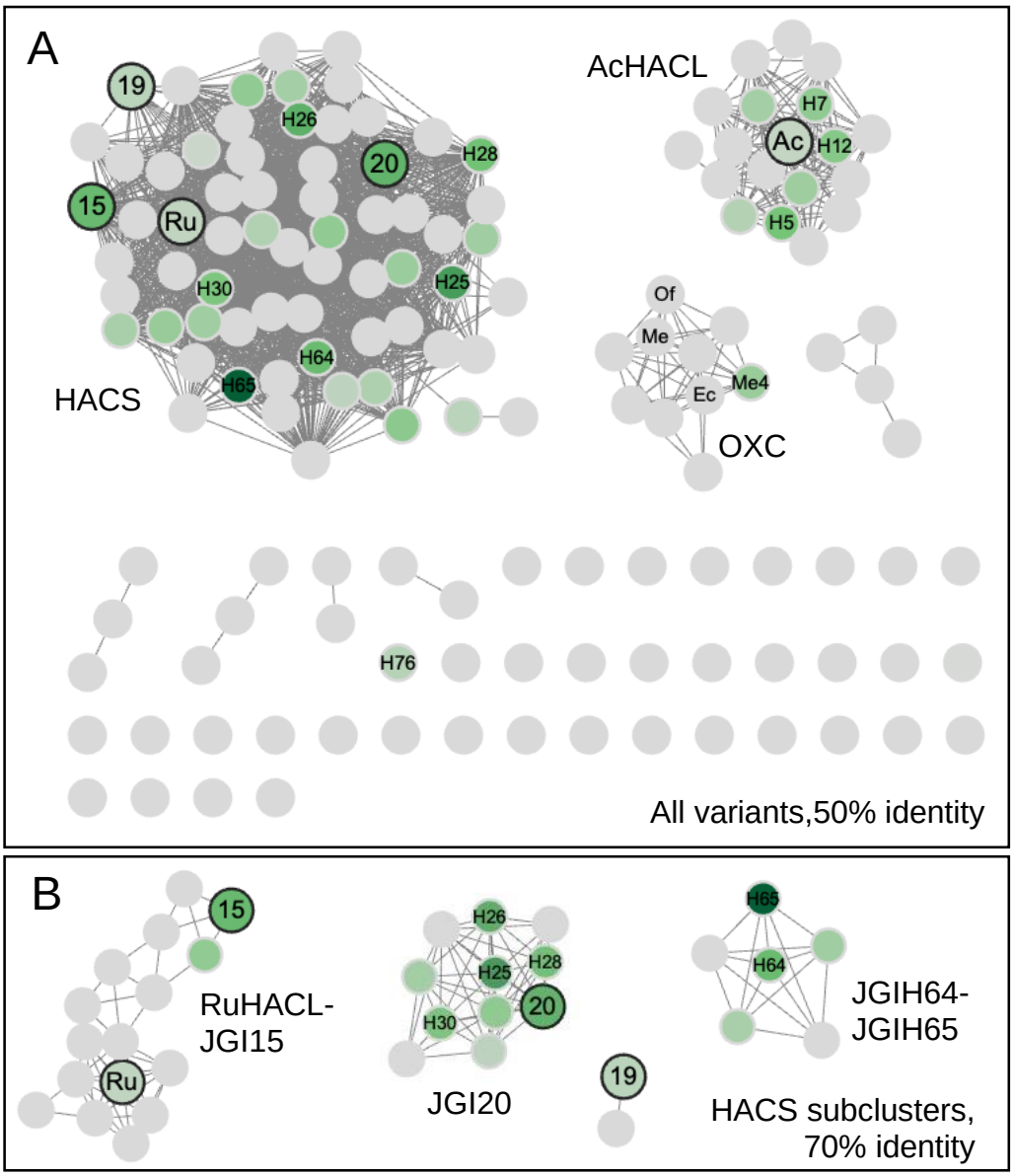
Mutations on JGI20 formyl-CoA binding region (A253G P254G) and c-terminus (L549H T550G R551del) showed significant improvement in glycolate productivity (Figure 3-10B). We also tried combining beneficial mutations and found JGI20 R480del L549H T550G R551del showed the best activity of up to 50% improvement compared to the wildtype JGI20 (Figure 3-10B). To understand the impact of mutagenesis on the structure of c-terminal covering loop, we generated AlphaFold structures of JGI20 L549H T550G R551del and R480del L549H T550G R551del and aligned the structures with wildtype JGI20 and JGI15 (Figure 3-10A, Supplementary Figure 3-18). Based on the AlphaFold models, Tryptophan (W) residue of JGI15 is facing toward

the catalytic site, whereas the corresponding W of JGI20 is facing away from it, making the closing loop farther away from the catalytic site (Supplementary Figure 3-18A). We hypothesize that such orientation causes change in the pocket size of the catalytic site, leading to the difference in affinity with small substrates like formaldehyde between JGI15 and JGI20. Interestingly, L549H T550G R551del mutation in JGI20 moved the covering loop closer to the catalytic site than the wildtype because bulky histidine (H) residue of the mutant is positioned in the same orientation as W of JGI15 (Supplementary Figure 3-18B). R480del L549H T550G R551del moves the covering loop further toward the catalytic site making the pocket size even smaller than the wildtype JGI15, presumably attributing to the proximity between formyl-CoA, TPP and a short-chain carbonyl, formaldehyde (Figure 3-10A).

Superior variants identified from the second-round HACS variants

From the first-round variants, we found JGI15, JGI19 and JGI20 to be active for glycolyl-CoA synthase activity exceeding the activity of RuHACL (Figure 3-4, 3-5). In addition, AcHACL, a phylogenetically distant variant from RuHACL and other first round JGI variants, showed notable glycolate productivity (Figure 3-4). Therefore, the four variants are used as references to identify 99 second-round HACS variants (labeled as JGIH1-99 in Supplementary Table 3-4) following similar approach as the first round (Figure 3-3). Learnings from the sequence and structure analysis of the first-round variants are applied to narrow down active variants and eliminate OXC variants that showed no HACS activity. Moreover, we explored enzymes that have structural similarity irrespective of the sequence similarity using I-TASSER[198] that gave us 9 additional variants (JGIH100-108 in Supplementary Table 3-4). Total 108 genes are codon-optimized and synthesized in collaboration with Joint Genome Institute and 99 variants are

successfully constructed for testing following the same procedure as 1st round variants screening (Figure 3-1).



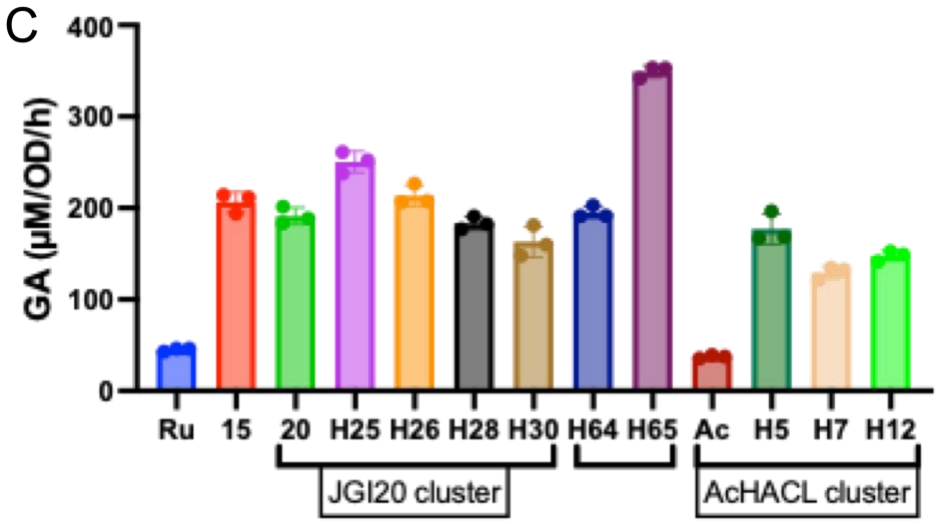


Figure 3-11. Sequence Similarity Network (SSN) for all HACS and OXC variants tested for glycolate (GA) productivity. (A) SSN generated with edges connecting nodes (variants) with 50% and higher identity (sequence similarity). (B) Subclusters within the HACS cluster in panel (A) only showing clusters formed with two or more variants. SSN generated with edges connecting nodes (variants) with 70% and higher identity (sequence similarity). Variants used as starting reference for the first and second-round homologs are highlighted with bigger circles with black border. Darker green represents higher glycolate productivity and gray circles represent no glycolate productivity. Mean glycolate productivity (μM/OD/h) from triplicate biological replicate was used for heat map. No sample has coefficient of variance exceeding 15%. (C) Glycolate productivity of select variants from the panel (A). Bars drawn at the mean of triplicate biological replicates with error bars representing standard deviation. Ru: RuHACL; Ac: AcHACL; Of: OfOXC; Ec: EcOXC; Me: MeOXC; Me4: MeOXC4.

All variants including known enzymes from HACL/OXC family (RuHACL, AcHACL, EcOXC, OfOXC, MeOXC and MeOXC4), 1st round (JGI1-29), and 2nd round (JGIH1-108) HACS variants are organized in the sequence similarity network with the heatmap showing relative glycolate productivities (Figure 3-11A and B). Clustering variants by connecting nodes with 50% and higher sequence identity resulted in three large clusters (Figure 3-11A): the biggest cluster representing HACS cluster that incorporates all HACS variants related to RuHACL, JGI15, JGI19 and JGI20; a medium cluster representing AcHACL cluster identified from the second round; and

a small cluster for OXC variants which include EcOXC, OfOXC, MeOXC and OXCs identified from the first-round variants (Figure 3-4, 3-8). Remarkably, variants showing activities are only found from the two big clusters, except for MeOXC4, an engineered variant of MeOXC for HACS activity[73], and JGIH76, which does not belong to any clusters (Figure 3-11A). When more stringent identity threshold of 70% is imposed, HACL cluster is divided into subclusters including RuHACL-JGI15 cluster, JGI20 cluster and JGIH64-65 cluster (Figure 3-11B, Supplementary Figure 3-19). JGI20 cluster shows more than half of the variants in the cluster having decent activities including JGIH25, which shows 30% increase in glycolate productivity than JGI20 (Figure 3-11B and C). Interestingly, JGIH65, which stands out among all variants with close to two-fold higher glycolate productivity than JGI20, belongs to a new cluster which does not include any of the starting references (RuHACL, AcHACL, JGI15, JGI19 and JGI20) (Figure 3-11B and C). Several variants identified from AcHACL cluster also showed significantly higher glycolate productivity than AcHACL, with JGIH5 showing up to 5-fold difference (Figure 3-11C).

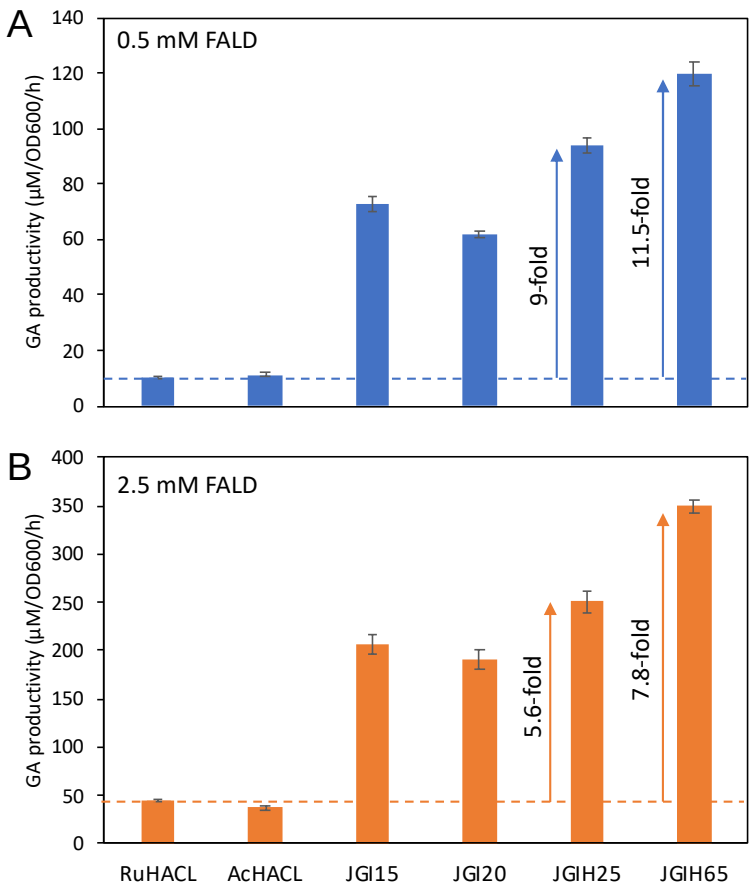


Figure 3-12. Glycolate (GA) productivity comparison between RuHACL, AcHACL, JGI15, JGI20, JGIH25 and JGIH65 at two different formaldehyde concentrations (0.5 and 2.5 mL). JGIH25 and JGIH65 from the second-round variants have 9-fold and 11.5-fold higher glycolate productivities under 0.5 mM formaldehyde and 5.6-fold and 7.8-fold higher activities under 2.5 mM formaldehyde than RuHACL, respectively. Bars drawn to the mean of triplicate biological replicates with error bars representing standard deviation.

JGIH25 and JGIH65 showed significantly better glycolate productivity than JGI15 and JGI20 under both low (0.5 mM) and high (2.5 mM) formaldehyde concentrations (Figure 3-12). Subsequently, we performed rational mutagenesis on these enzymes using active first-round variants as templates. In case of JGIH25, residues in the active site are almost identical to JGI20, except for the unconserved A253 (JGI20) in the formyl-CoA binding site (Supplementary Table 3-5). Hence, mutagenesis using JGI15 as a template on the same residues as JGI20 on formyl-CoA

binding site (S253G P254G) and c-terminal end (L549H T550G R551del) were performed. Mutations in c-terminal residues using JGIH65 as a template (T552ins N553ins E554ins) was also tested (Supplementary Figure 3-20). Improvement in glycolate productivity was observed only from L549H T550G R551del mutant under 2.5 mM formaldehyde but the activity dropped slightly below wildtype under 0.5 mM formaldehyde (Supplementary Figure 3-21). JGIH65, on the other hand, shares less sequence similarity with JGI15 or JGI20 as expected from the different clustering under 70% identity (Figure 5A). The differences were also observed from the active site residues with multiple unmatching residues in the TPP binding, formyl-CoA binding, and c-terminal residues (Supplementary Table 3-5). Consequently, JGIH65 mutants using other active variants as templates were constructed and tested (Supplementary Figure 3-20). Mutagenesis caused substantial drop in glycolate productivity of this variant, showing the importance of these residues and possibility in engineering of other variants using JGIH65 as a template on these residues. An exception was G447S mutant, which showed a significant improvement in activity under low formaldehyde concentration (0.5 mM) (Figure 3-13B). According to the AlphaFold structure of JGIH65 aligned with JGI15 and 20, replacing glycine with serine having polar side chain may contribute to improved stability in TPP binding (Figure 3-13A).

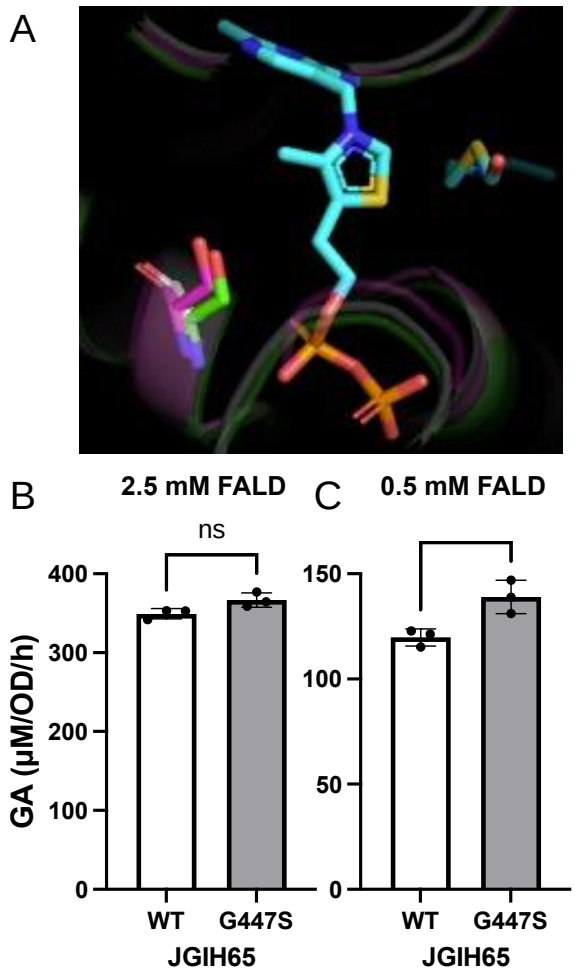


Figure 3-13. JGIH65 protein engineering. (A) Position of JGIH65 G447 (white) aligned with corresponding S residues of JGI15 (violet pink) and JGI20 (green) next to TPP. (B)Glycolate productivity of JGIH65 wildtype and G447S mutant. Two different formaldehyde concentrations (0.5 and 2.5 mM) are tested for glycolate productivity. Mutant shows significantly higher glycolate at 0.5 mM formaldehyde (FALD). Bars drawn to the mean of triplicate biological replicates with error bars representing standard deviation. (*$p < 0.05$)

Product titer, rate, and yield improvement using the new HACS variants

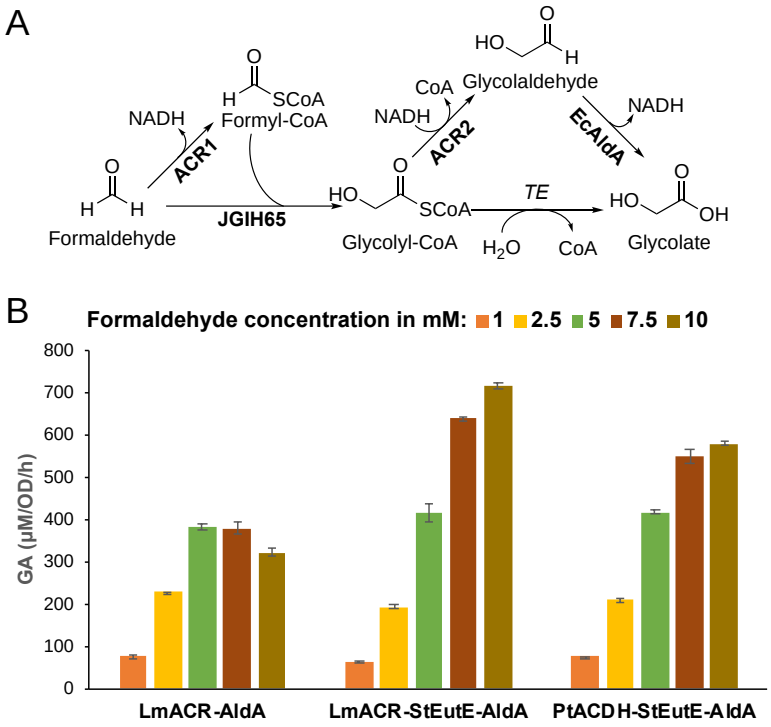


Figure 3-14. Optimizing FORCE pathways to improve glycolate productivity. (A) FORCE pathways for glycolate production from formaldehyde. (B) Glycolate productivity from varying formaldehyde concentrations using the strain expressing JGIH65 and various auxiliary enzyme combinations. From left: LmACR as both ACR1 and ACR2; LmACR as ACR1 and StEutE as ACR2; and PtACDH as ACR1 and StEutE as ACR2 are tested. Bars are drawn at the mean of duplicate biological replicates with error bars representing standard deviation.

We identified highly efficient HACS variants, one of which showed up to 12-fold improvement in glycolate productivity from the starting reference, RuHACL (Figure 3-12A). Rapidly converting formaldehyde and formyl-CoA to glycolyl-CoA also contributes to detoxification of formaldehyde improving cell viability and activity in vivo, enabling bioconversion under higher formaldehyde concentrations to achieve higher titer and productivity. Under low formaldehyde concentrations at 2.5 mM or lower, formyl-CoA flux generated from overexpression of AbfT and 20 mM formate was shown sufficient for investigating HACS

activities. However, with 5 mM or higher formaldehyde concentrations, we saw the activity drops substantially possibly due to the combined effect of saturating formyl-CoA flux and formaldehyde toxicity (Figure 3-2). Saturating formyl-CoA flux could be the result of AbfT having low turnover number (0.4 s^{-1})[73] for formate activation. As such, we hypothesized formyl-CoA generation from formaldehyde by acylating aldehyde dehydrogenase/acyl-CoA reductase (ACDH/ACR) could be a better option at high formaldehyde concentrations because of higher turnover of the enzyme, such as acyl-CoA reductase from *Listeria monocytogenes* (LmACR) (k_{cat} = 1.0 s^{-1})[68] that also mitigates toxicity effect by rapid consumption of formaldehyde. We previously showed that overexpression of LmACR and *E. coli* AldA enhances flux toward glycolate production as LmACR can serve as both ACR1 (formaldehyde to formyl-CoA) and ACR2 (glycolyl-CoA to glycolaldehyde)[25] (Figure 3-14A). Using RuHACL as HACS, we achieved glycolate production of up to 6 mM in 24 hours in the resting-cell bioconversion platform (~25 μM glycolate/OD/h)[25]. In the same system, replacing RuHACL with JGIH65 showed more than 15-fold increase in glycolate productivity reaching 400 μM glycolate/OD/h upon addition of 5 mM formaldehyde (Figure 3-14B). However, glycolate productivity did not increase further by increasing formaldehyde concentrations. We hypothesized that ACR1 and ACR2 are limiting the flux and explored different ACR1 and ACR2 candidates. To address this, we found that *Salmonella typhimurium* EutE (StEutE) has better activity for glycolaldehyde oxidation to glycolyl-CoA than LmACR[68]. Introducing StEutE as ACR2 indeed improved glycolate productivity at high formaldehyde concentrations suggesting that LmACR as ACR2 becomes limiting under high glycolyl-CoA flux. We also tested acylating aldehyde dehydrogenase from *Parageobacillus thermoglucosidasius* (PtACDH) reported to catalyze ACR1 reaction[69]. While it slightly improved glycolate productivity at low formaldehyde concentrations, LmACR was still better at

high formaldehyde concentrations for ACR1 activity (Figure 3-14B). After optimizing relative enzyme expression levels by exploring inducer matrix (Supplementary Figure 3-22), we obtained glycolate productivity of up to 700 μM glycolate/OD/h (2.5 mmol/gDCW/h) from formaldehyde as the C1 carbon source (Figure 3-14B).

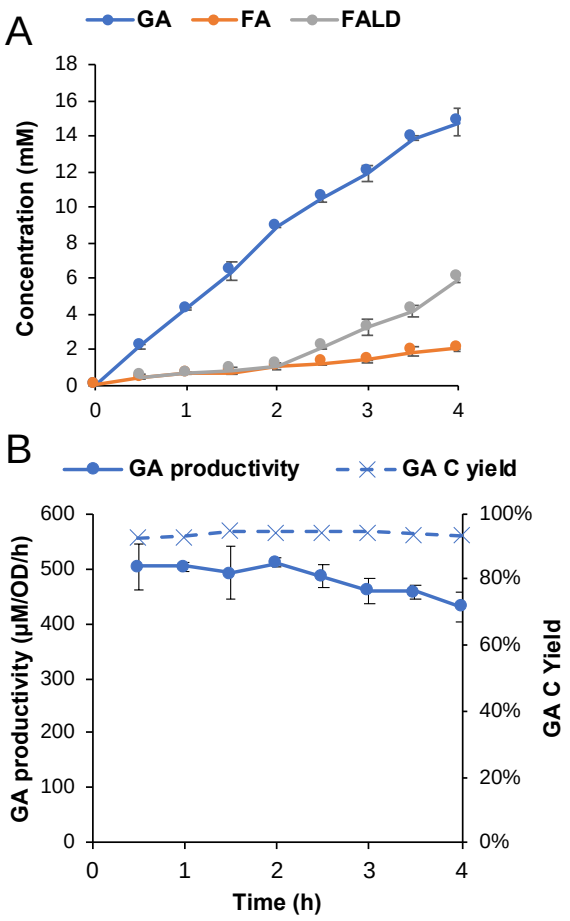


Figure 3-15. High titer glycolate production via extended resting-cell bioconversion. (A) Time course profile of formaldehyde to glycolate bioconversion for 4 hours under 10 mM formaldehyde per hour feed rate. (B) Change in glycolate productivity and carbon yield over time. Dots are drawn at the mean of duplicate biological replicates with error bars representing standard deviation.

Next, we extended the duration of bioconversion to demonstrate high titer product synthesis from formaldehyde. 5 / 7.5 mM formaldehyde was added every 30 minutes (10 / 15

mM/h feed rate) to ensure low formaldehyde concentration throughout the bioconversion and avoid toxicity effect. Remarkably, we saw only minimal decrease in glycolate productivity of the biocatalyst for up to four hours under 10 mM/h formaldehyde feed rate (Figure 3-15A and B). However, as activity slowed down due to cells under starvation, formaldehyde accumulation was observed which further exacerbated the glycolate productivity (Figure 3-15A). This phenomenon was more evident under formaldehyde feed rate of 15 mM/h (Supplementary Figure 3-23). Under 10 mM/h formaldehyde feed rate (resting cells at OD600 of ~9) for four hours, 15 mM glycolate (1.1 g/L) was accumulated at the rate of 0.28 g/L/h with consistent carbon yield at 93-95% (formate as only byproduct) (Figure 3-15A and C).

Discussion

2-hydroxyacyl-CoA synthase (HACS) catalyzes acyloin condensation between various carbonyl compounds including aldehydes and ketones and C1 moiety, formyl-CoA. Formyl-CoA Elongation (FORCE) pathways built upon this core enzyme has huge potential in biological utilization of C1 compounds as diverse products can be directly synthesized from C1 molecules. We demonstrate an in-depth analysis of HACS enzyme family by synthesizing and testing more than 130 enzyme homologs and analyzing sequence and structural properties of active and inactive variants for the condensation reaction between formaldehyde and formyl-CoA. Establishing 96-well plate-based in vivo prototyping platform greatly improved the throughput of the system allowing rapid screening of variants and mutants. Multiple rounds of enzyme homolog bioprospecting were used to eliminate clusters with inactive variants from large sequence space and explore more depth of the phylogenic branches with active variants. As a result, we identified more than 25 HACS variants with improved kinetics toward formaldehyde than the initial reference, RuHACL, including JGIH65 which showed up to 12-fold increase in glycolate

productivity. Using JGIH65 as HACS in FORCE pathways, glycolate production from formaldehyde with 1.1 g/L titer, 0.28 g/L/h rate and 95% carbon yield was achieved in vivo using *E. coli* as host.

Structure analysis in parallel with the multiple sequence alignment (MSA) analysis was shown crucial for understanding characteristic residues of active variants and choosing residues for protein engineering. Machine learning-guided protein structure prediction using AlphaFold enables fast and easy generation of protein structure saving significant time and cost required for X-ray crystallography. Hence, it was shown particularly effective in generating large quantity of protein structures[199]. Engineering active site residues of a variant with high turnover number (k_{cat}) using another variant with low Michaelis-Menten constant (K_m) as a template showed significant improvement in activity with only a handful of mutants constructed and tested. Moreover, AlphaFold could also be used to analyze the impact of mutagenesis. For example, in case of JGI20 R480ins L549H T550G R551del, which showed up to 50% higher glycolate productivity than the wildtype, AlphaFold-generated structures show apparent change in the orientation of c-terminal covering loop, causing the catalytic site pocket to shrink (Figure 3-10A, Supplementary Figure 3-18). It is important to note that AlphaFold is not validated to predict the impact of point mutations[191, 200]. Moreover, the c-terminal residues of HACS/OXC family enzymes are known to be disordered unless saturated with substrates (formyl-CoA)[196] leading to incomplete crystal structures missing this region in the database (e.g., PDB code: 6XN8). This lack of structural information on c-terminal covering loop leads to low confidence of AlphaFold prediction as seen from low pLDDT score in the c-termini for both JGI15 and 20 (Figure 3-6D and E). However, our success in engineering this region with reasonable explanations based on AlphaFold-generated structures might indicate that there is an opportunity for using modeling even

for understanding mutagenesis on regions with low confidence. Solving the crystal structures of JGI15, JGI20, and its mutants capturing the conformational changes of the c-terminal covering loop will validate the hypothesis.

Detailed analysis on the sequence, structure and activity of the first-round variants was instrumental not only for protein engineering approach but also in identifying second-round variants which showed substantially higher ratio of active variants (26/99) than the first-round (3/29). The high success rate attributes to the learnings from the first-round variants narrowing down phylogenetic branches with active variants and eliminating variants with no activities (e.g., OXCs). For example, more than half of the variants having high sequence similarity with JGI20 (JGI20 cluster in Figure 3-11B) showed decent glycolate productivities. However, the best variant was identified from a different subcluster, which still belongs to the big HACS cluster under 50% similarity nodes but has less than 70% similarity with any of the active first-round variants (Figure 3-11), indicating the significance of balance between width and depth of the phylogenetic space. Bioinformatic tools, such as CD-HIT[201] and EFI-Web[202], allow rapid selection of representative enzyme candidates from large protein, genome and metagenome databases via clustering based on user-directed parameters. Just like the case of AlphaFold, these tools are accessible to researchers with minimal computational background. As cost for DNA synthesis drops, even experimentalists in academic lab will gain access to testing hundreds, if not thousands, of enzymes for specific reaction and the approach discussed in this study could be readily applied in any other objective to facilitate such process.

The best HACS identified from this study reaches 10^4 M^{-1} s^{-1} catalytic efficiency toward formaldehyde as substrate. To the best of our knowledge, this is the best kinetics reported for enzymes that catalyze C1-C1 condensation utilizing formaldehyde as substrate and has orders of

magnitude higher values compared to engineered formolases[33, 93] and glycolaldehyde synthases[97] and with fold improvements compared to other native or engineered HACS/OXCs[73] . Formaldehyde is an attractive carbon source due to its low cost, water solubility and decent stability as a solution[203], and it can be sustainably generated with high efficiency from electrochemical reduction of CO_2[186]. In biological C1 utilization, direct utilization of formaldehyde circumvents problematic methanol oxidation step, which either relies on NAD^+-dependent methanol dehydrogenase with poor kinetics or requires specific cofactors not present in model organisms like *E. coli*[24]. However, its extreme toxicity causing DNA-protein crosslinking [13] prevents it from being considered as carbon source for living organisms. This study shows that formaldehyde can be used as substrate in vivo to demonstrate g/L scale production of multi-carbon products if the feeding strategy is properly optimized to ensure low formaldehyde concentrations throughout the duration of bioconversion. However, slowdown of activities was observed likely due to cells under starvation unable to replenish enzymes for biocatalysis. To further extend the duration of bioconversion, constant decrease in formaldehyde feeding rate or co-feeding nutrient for cell maintenance could be considered. Advances in dynamic regulation of metabolism[204] and growth and production decoupled two-phase fermentation[147] will allow more autonomous and effective adaptation of formaldehyde utilization strategies in vivo. Superior kinetics of the core enzymes and its ability to operate independently from the host metabolism make FORCE pathways a promising platform for utilizing formaldehyde as feedstock for C1 bioeconomy.

Methods:

Reagents

All chemicals were obtained from Fisher Scientific Co. and Sigma-Aldrich Co. unless otherwise specified. Primers were synthesized by Integrated DNA Technologies. Restriction enzymes were obtained from New England Biolabs unless otherwise specified.

Genetic methods

Plasmid-based gene expression was achieved by cloning the desired gene(s) into pCDFDuet-1 or pETDuet-1 (Novagen) digested with appropriate restriction enzymes and by using In-Fusion cloning technology (Clontech Laboratories, Inc.). HACS variants are codon-optimized, synthesized and integrated into pCDFDuet-1 vector by Joint Genome Institute.

HACS mutants were prepared by cloning wild type JGI15 and JGI20 into the vector pUC19 (Clontech Laboratories, Inc., Mountain View, CA). Primers containing the desired mutation were designed following the 'In Vivo assembly' (IVA) protocol for mutagenesis[205]. PCR products containing the mutations were generated following the IVA protocol and used to transform *E. coli* Stellar cells (Clontech Laboratories). The desired mutant sequence was confirmed by DNA sequencing. The mutant genes were then cloned into final expression vector (pCDFDuet-1) using restriction enzyme digestion and ligation.

Resting cell bioconversions for HACS screening

In vivo product synthesis was conducted using M9 minimal media (6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl, 2 mM $MgSO_4$, 100 µM $CaCl_2$, and 15 µM thiamine-HCl) unless otherwise stated. Cells were initially grown in 96-deep well plates (USA Scientific, Ocala, FL) containing 0.5 mL of the above media further supplemented with 20 g/L glycerol, 10 g/L tryptone, and 5 g/L yeast extract. A single colony of the desired strain was cultivated overnight

(14-16 hrs) in LB medium with appropriate antibiotics and used as the inoculum (1%). Antibiotics (100 μg/mL carbenicillin, 100 μg/mL spectinomycin) were included when appropriate. Cultures were then incubated at 30°C and 1000 rpm in a Digital Microplate Shaker (Fisher Scientific) until an OD600 of ~0.4 was reached, at which point appropriate amounts of inducer(s) (isopropyl β-D-1-thiogalactopyranoside (IPTG) and cumate) were added. Plates were incubated for a total of 24 hrs post-inoculation.

Cells from the above pre-cultures were then centrifuged (4000 rpm, 22°C), washed with the above minimal media without any carbon source, and resuspended with 1 mL of above minimal media containing indicated amounts of carbon source. The above minimal media containing 2.5 mM formaldehyde and 20 mM formate was added and incubated at 30°C and 1000 rpm in Digital Microplate Shaker (Fisher Scientific). After incubation at 30°C for 30 minutes, the cells were pelleted by centrifugation and the supernatant was analyzed by HPLC. Quantification of product and substrate concentrations (formic acid, formaldehyde and glycolic acid) were determined via HPLC using a Shimadzu Prominence SIL 20 system (Shimadzu Scientific Instruments, Inc., Columbia, MD) equipped with a refractive index detector and an Shim-pack Fast-OA column (Shimadzu) with operating conditions to optimize peak separation (0.4 ml/min flowrate, 5 mM p-toluenesulfonic acid mobile phase, column temperature 45°C).

Supplementary Materials

Supplementary Table 3-2. Host strains and plasmids used in this study. Uniprot accession numbers for heterologous enzymes used in this work are given in parenthesis.

Host Strains/ Plasmids	Description/Genotype/Usage	Source
BL21(DE3)	*E. coli* B F^- *ompT gal dcm lon* $hsdS_B(r_B^-m_B^-)$ $[malB^+]_{K\text{-}12}(\lambda^S)$ λ(DE3) - Host for protein expression for *in vitro* studies	Studier et al.[182]
MG1655	*E. coli* K-12 F- l- *ilvG- rfb*-50 *rph*-1	Blattner et al.[183]
AC440	MG1655 λ(DE3) Δ*frmA* Δ*fdhF* Δ*fdnG* Δ*fdoG* Δ*glcD::FRT* - Engineered host for resting cell studies	Chou et al.[68]
pCDFDuet-1	CloDF13, *lacI*, Sm^R	Novagen (Darmstadt, Germany)
pETDuet-1	pBR322-derived ColE1 origin, *lacI*, Amp^R	Novagen (Darmstadt, Germany)
pETDuet-1-P^{CT5}-CaAbfT	pETDuet-1 with codon optimized *Clostridium aminobutyricum AbfT* in a synthetic operon under control of the cumate inducible CT5 promoter and *cymR*	This study
pETDuet-1-P^{CT5}-LmACR-EcAldA	pETDuet-1 with codon optimized *Lysteria monocytogenes acr* and *Escherichia coli aldA* in a synthetic operon under control of the cumate inducible CT5 promoter and *cymR*	Chou et al.[68]
pETDuet-1-P^{CT5}-LmACR-StEutE-EcAldA	pETDuet-1 with codon optimized *Lysteria monocytogenes acr*, *Salmonella typhimurium eutE* and *Escherichia coli aldA* in a synthetic operon under control of the cumate inducible CT5 promoter and *cymR*	This study
pETDuet-1-P^{CT5}-PtACDH-StEutE-EcAldA	pETDuet-1 with codon optimized *Parageobacillus thermoglucosidasius acdh Salmonella typhimurium eutE* and *Escherichia coli aldA* in a synthetic operon under control of the cumate inducible CT5 promoter and *cymR*	This study

Supplementary Table 3-3. First-round HACS variants with corresponding GenBank Accession Number

JGI#	GenBank Accession Number
1	XP_012756082.1
2	TMK01573.1
3	PYM26381.1
4	EEG70177.1
5	MBH80817.1
6	WP_030891887.1
7	AGK93615.1
8	MAX57815.1
9	WP_068916287.1
10	WP_062165271.1
11	MBB43458.1
12	PCJ72347.1
13	TMQ19149.1
14	MAX11513.1
15	HAK63664.1
16	MBG92919.1
17	PZC46201.1
18	MBB84818.1
19	OGA51379.1
20	PWB41796.1
21	MAE93843.1
22	OGP60024.1
23	OWB57166.1
24	KXN72624.1
25	PVU86112.1
26	ORZ16580.1
27	XP_005644825.1
28	KZV27770.1
29	EJY87672.1

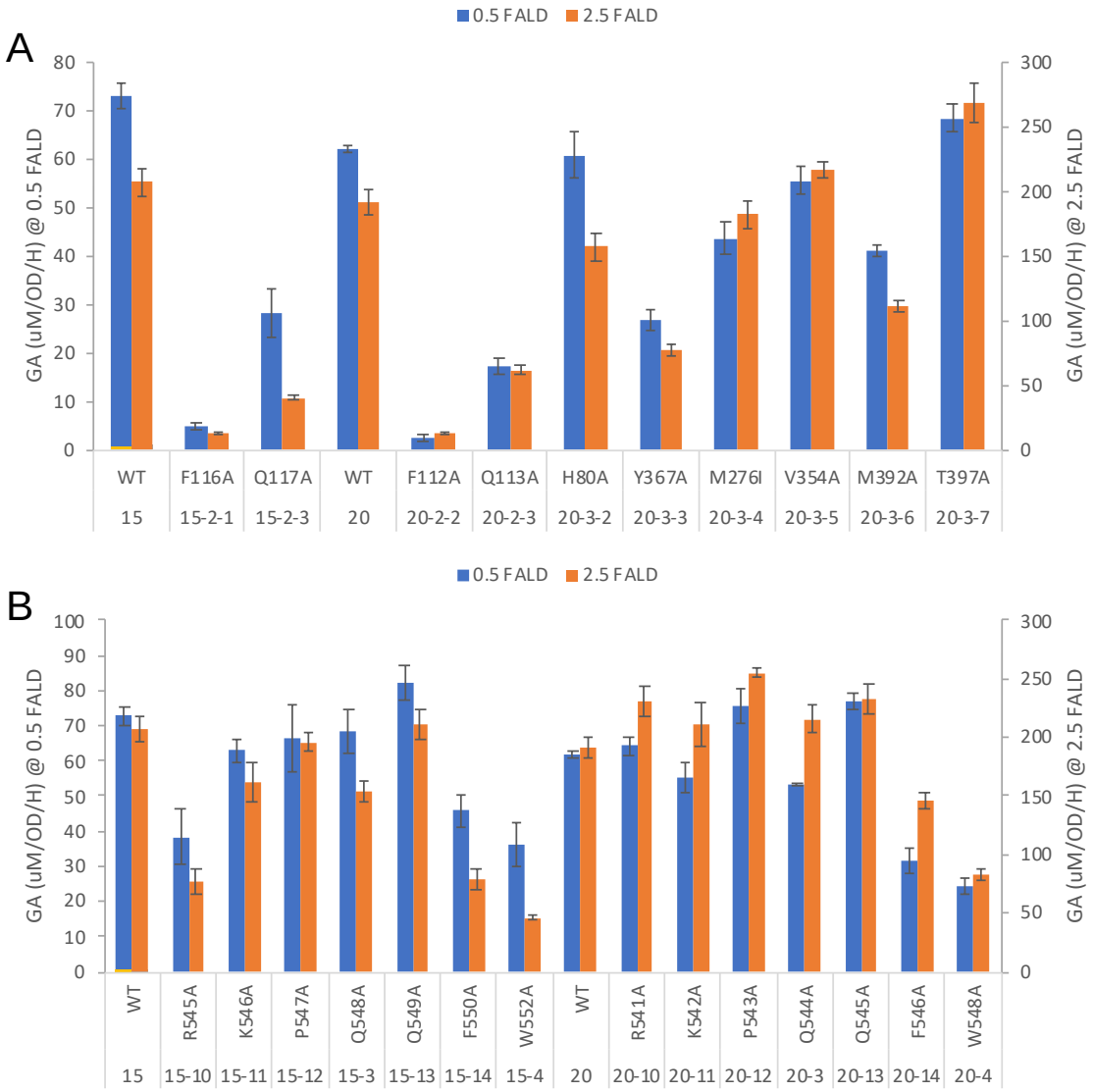
Supplementary Table 3-4. Second-round HACS variants with corresponding GenBank Accession Number

JGIH#	GenBank Accession Number
1	HIG47824.1
2	TMD03111.1
3	MBJ56818.1
4	WP_095860310.1
5	MBL8483477.1
6	WP_058697592.1
7	WP_130292058.1
8	WP_207956071.1
9	WP_132429652.1
10	WP_060575023.1
11	WP_068796145.1
12	OJY48151.1
13	WP_062397209.1
14	WP_169186431.1
15	WP_133828190.1
16	MBS0560157.1
17	PCJ59575.1
18	MXY78649.1
19	MBA01399.1
20	MXX31676.1
21	MXV80929.1
22	MBI4083577.1
23	MBK6319978.1
24	MBI5948182.1
25	PFG74273.1
26	WP_158065972.1
27	MBN9492325.1
28	MBK6663287.1
29	MBI2766664.1
30	HEM18354.1
31	GBD22648.1
32	MBF6599205.1
33	MXW00101.1
34	MYA07641.1
35	REJ76484.1
36	HDY15625.1
37	MBW2231087.1
38	NRA08835.1
39	NQZ98823.1
40	MBI3918747.1
41	MBI2761137.1
42	MBE0608783.1
43	MYA54281.1
44	NRA01576.1
45	MBW2623123.1
46	MBI5615765.1
47	MSR14309.1
48	XP_004342722.2
49	MSP42197.1
50	TDI61101.1
51	MBO0741576.1
52	MBO0736096.1
53	MBV9828771.1
54	MAW55136.1

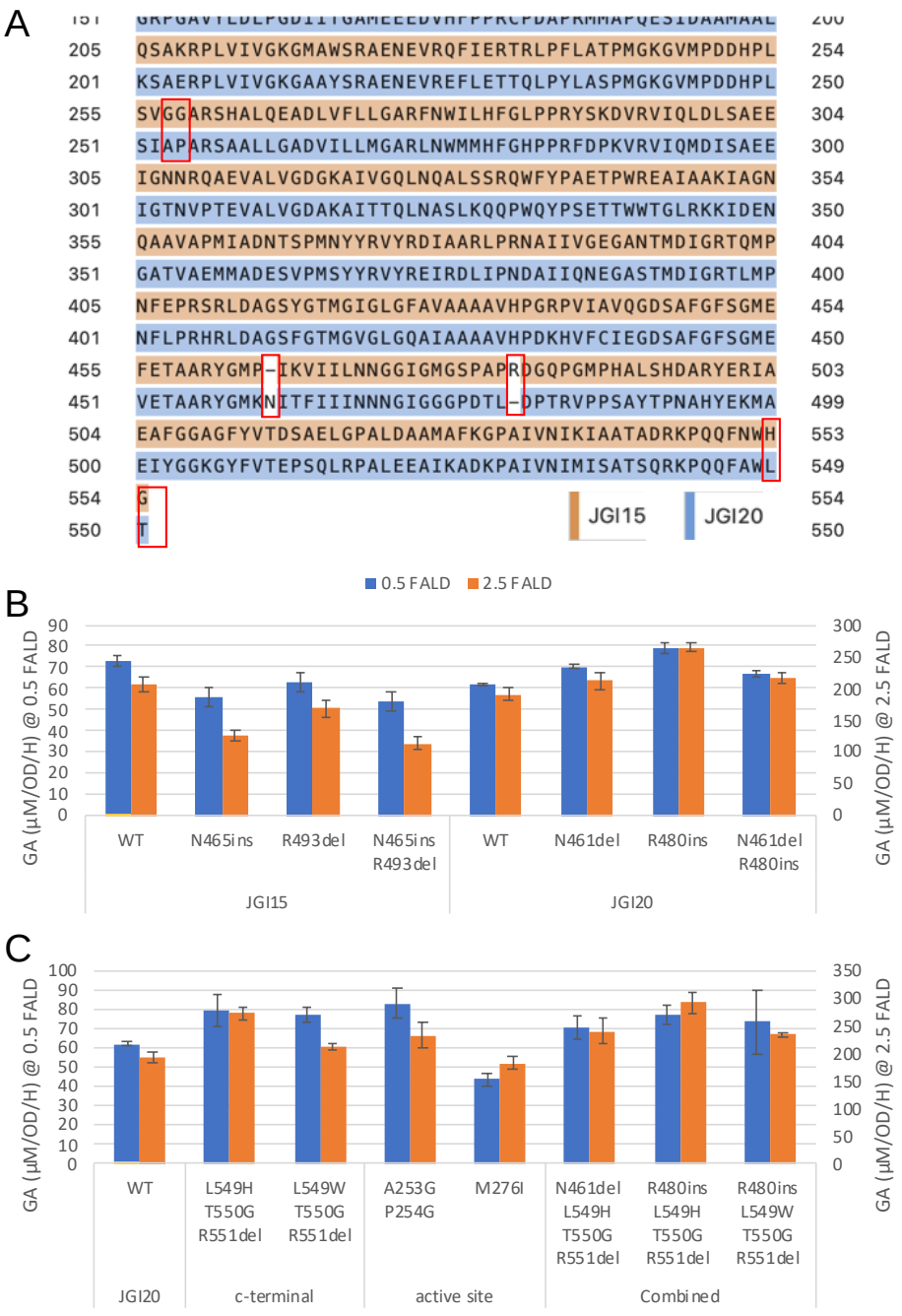
JGIH#	GenBank Accession Number
55	MBV38827.1
56	TMJ68231.1
57	TMJ64557.1
58	MBV9815528.1
59	MYH41266.1
60	MPZ97997.1
61	MBT5774752.1
62	XP_014714961.1
63	TAK78428.1
64	TAJ19927.1
65	PKN81274.1
66	RLT34960.1
67	MBT5775398.1
68	TMD99851.1
69	MSQ12864.1
70	MBL0714078.1
71	WP_114297888.1
72	MAK25262.1
73	WP_068138361.1
74	RMG94145.1
75	MBA4180234.1
76	MBM3723043.1
77	ABF11225.1
78	TAL98798.1
79	NNN20496.1
80	MBP1761901.1
81	PPQ43247.1
82	MSQ25793.1
83	TMK28344.1
84	HIB12002.1
85	WP_179589464.1
86	MXY42918.1
87	WP_184156128.1
88	HET53513.1
89	TMK22624.1
90	MXX66290.1
91	GIS94895.1
92	MBN1557905.1
93	MSV30368.1
94	MBN2179295.1
95	TDI90456.1
96	OGN76415.1
97	WP_102074055.1
98	PZC47999.1
99	HHH88785.1
100	OLB93949.1
101	PKB76696.1
102	HED24197.1
103	WP_066960443.1
104	WP_169259343.1
105	WP_201494572.1
106	MBN9621549.1
107	OZG26106.1
108	WP_016501746.1

Supplementary Table 3-5. Active site residues of JGI15, JGI20, JGIH25 and JGIH65 categorized by TPP binding, Formyl-CoA binding and c-terminal residues. Active site residues (3.5 Å within TPP and formyl-CoA based on the AlphaFold structure of JGI20) and corresponding residues of high performing variants on C1-C1 condensation. Highlighted in yellow indicates unconserved residues among active variants. Residues with asterisk(*) indicate key catalytic residues that are hypothesized to distinguish between HACS and OXC.

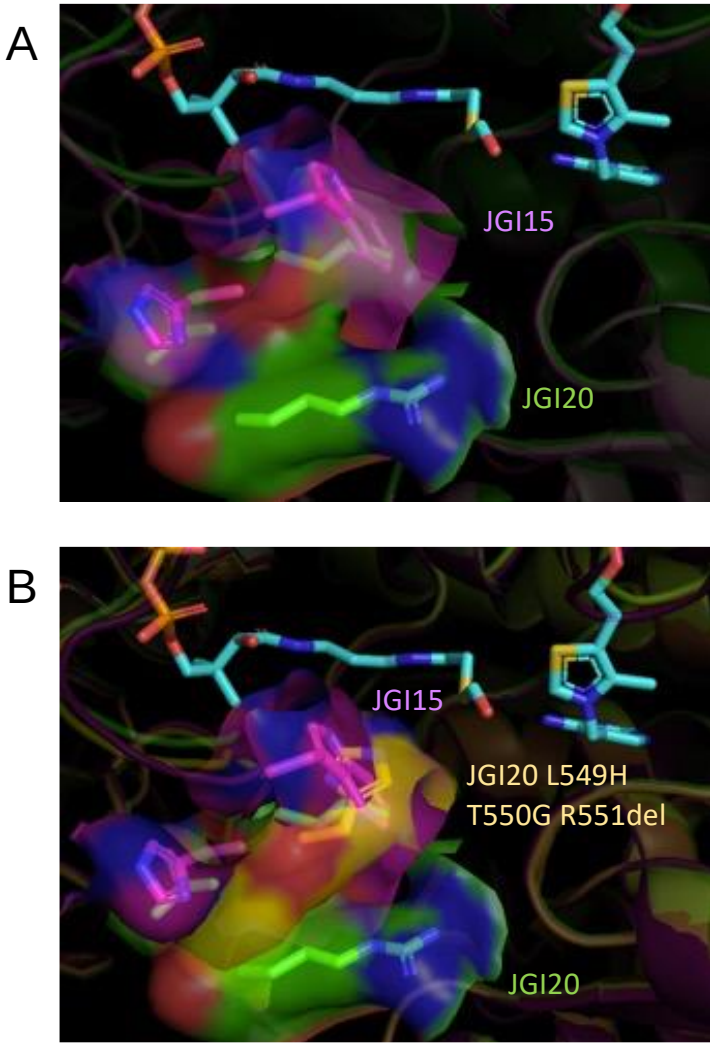
	JGI20	JGI15	JGIH25	JGIH65
TPP binding	V26	V	V	V
	E50	E	E	E
	V73	V	V	V
	G77	G	G	G
	H80	H	H	N
	Q113*	Q	Q	Q
	Y367	Y	Y	Y
	T391	T	T	T
	G414	G	G	G
	M416	M	M	M
	D441	D	D	D
	S442	S	S	G
	A443	A	A	A
	N469	N	N	N
	G471	G	G	G
Formyl-CoA binding	F112*	F	F	F
	A253	G	S	A
	P254	G	P	A
	R256	R	R	R
	S257	S	S	S
	W275	W	W	W
	M276	I	I	M
	V354	V	V	V
	M392*	M	M	M
	R396	R	R	R
	T397	T	T	T
	Q544	Q	Q	Q
	W548	W	W	W
c-terminal end	L549	H	L	L
	T550	G	T	T
	R551	-	R	R
	-	-	-	TNE



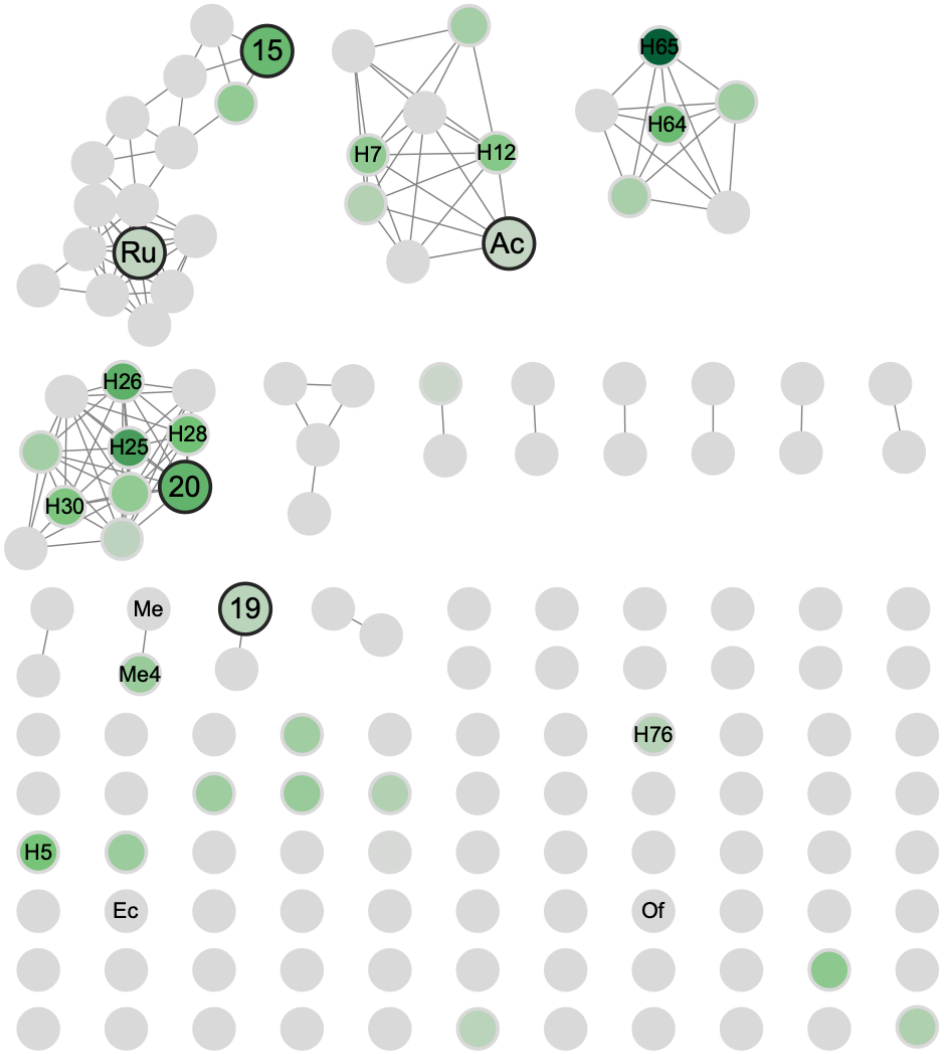
Supplementary Figure 3-16. Alanine scanning result for potential active site residues of JGI15 and 20 on (A) formyl-CoA and TPP binding region and (B) c-terminal residues. Two different formaldehyde concentrations (0.5 and 2.5 mM) are tested for glycolate productivity. Bars drawn to the mean of triplicate biological replicates with error bars representing standard deviation.



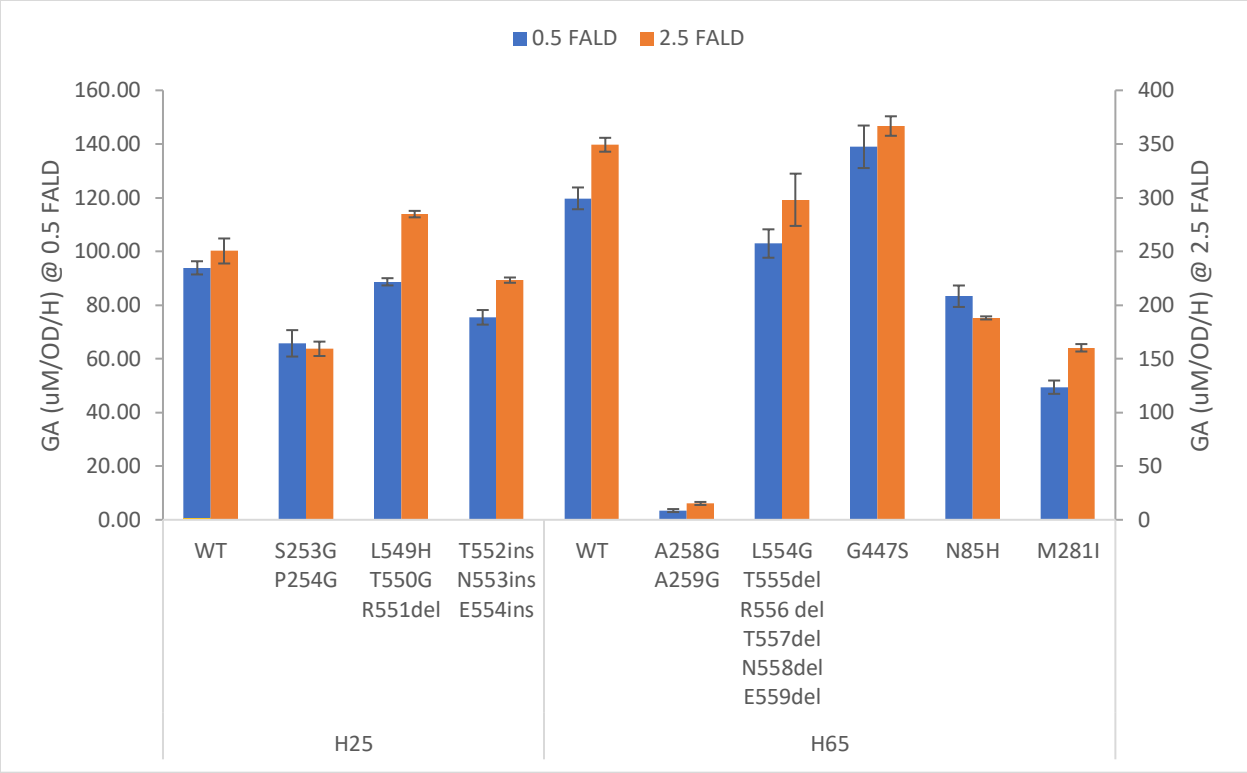
Supplementary Figure 3-17. JGI15-JGI20 hybrid protein engineering. (A) Pairwise sequence alignment of AlphaFold-generated structures of JGI15 and 20. Highlighted in red boxes are residues targeted for mutagenesis (B) JGI15-JGI20 hybrid based on structure alignment. (C) JGI20 active site engineering using JGI15 as a template and combined beneficial mutants. Two different formaldehyde concentrations (0.5 and 2.5 mM) are tested for glycolate productivity. Bars drawn to the mean of triplicate biological replicates with error bars representing standard deviation.



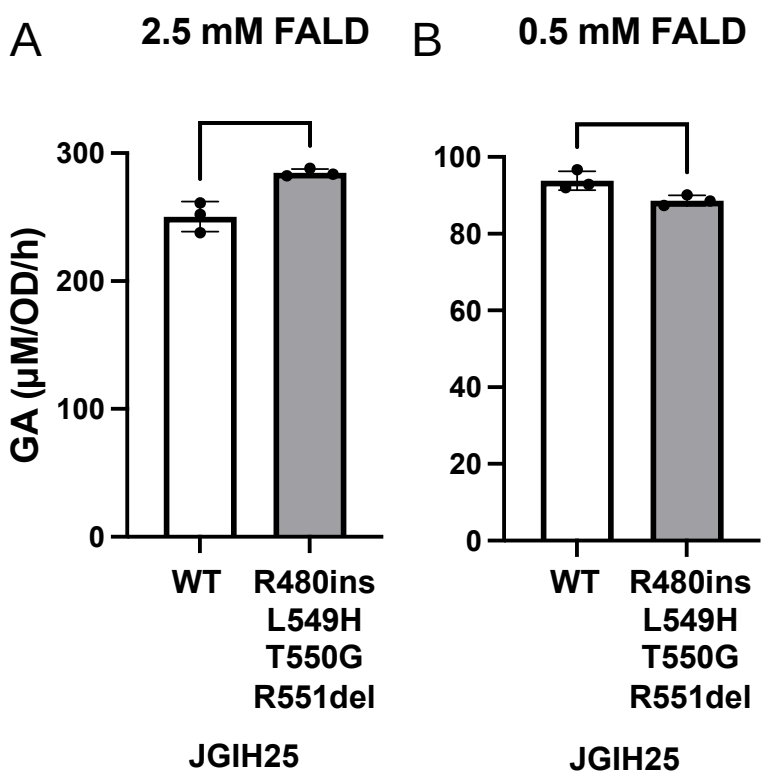
Supplementary Figure 3-18. Orientation of the C-terminal covering loop of (A) JGI15 (violet pink) and JGI20 (green) and (B) JGI20 L549H T550G R551del (yellow) predicted by AlphaFold.



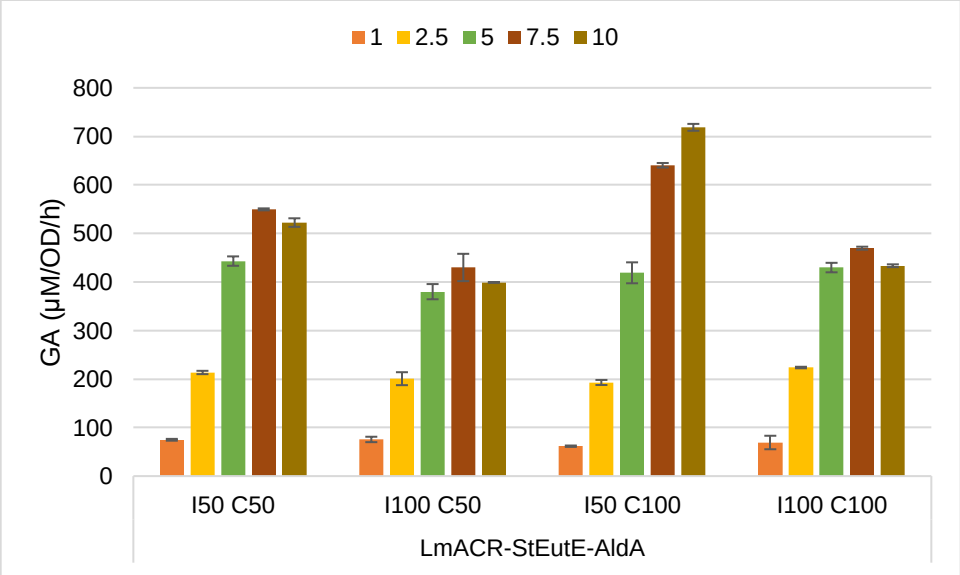
Supplementary Figure 3-19. Sequence similarity network (SSN) for all HACS and OXC variants tested for glycolate (GA) productivity under 70% identity threshold. Variants used as starting reference for the first and second-round homologs are highlighted with bigger circles with black border. Darker green represents higher glycolate productivity and gray circles represent no glycolate productivity. Mean glycolate productivity (μM/OD/h) from triplicate biological replicate was used for heat map. No sample has coefficient of variance exceeding 15%.



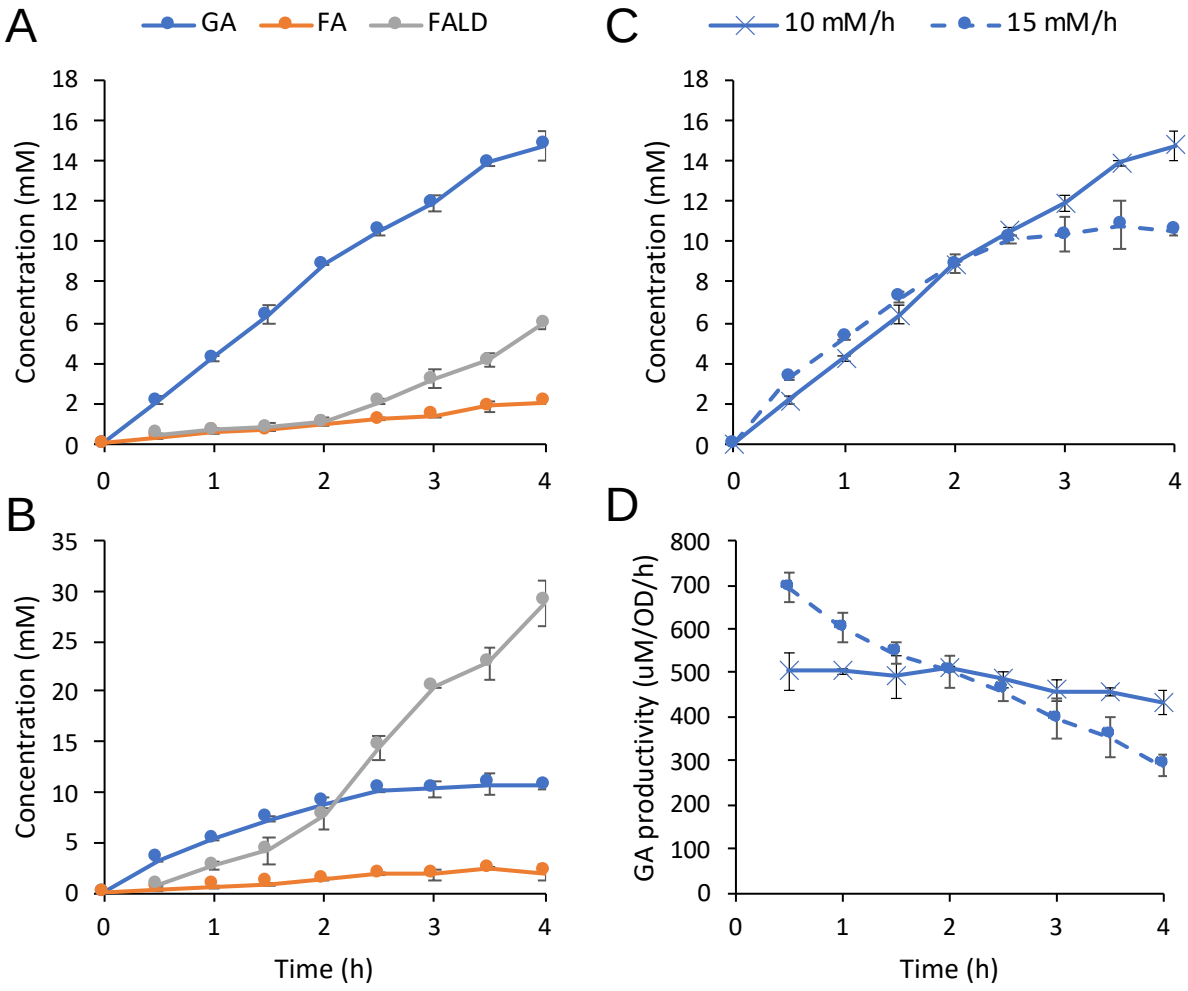
Supplementary Figure 3-20. JGIH25 and JGIH65 protein engineering. JGIH25 (H25) and JGIH65 (H65) mutants using other active variants as templates. Two different formaldehyde concentrations (0.5 and 2.5 mM) are tested for glycolate productivity. Bars drawn to the mean of triplicate biological replicates with error bars representing standard deviation.



Supplementary Figure 3-21. JGIH25 protein engineering. Glycolate productivity of JGIH25 wildtype and R480ins L549H T550G R551del mutant. Two different formaldehyde concentrations (0.5 and 2.5 mM) are tested for glycolate productivity. Mutant shows significantly higher glycolate at 2.5 mM formaldehyde (FALD) but lower at 0.5 mM FALD. Bars drawn to the mean of triplicate biological replicates with error bars representing standard deviation. (*$p < 0.05$, **$p < 0.01$)



Supplementary Figure 3-22. Inducer matrix of IPTG-inducible JGIH65 and cumate-inducible LmACR-StEutE-AldA at varying IPTG (I) and cumate (C) concentrations. Five different formaldehyde concentrations (1, 2.5, 5, 7.5 and 10 mM) are tested and four different inducer concentrations are tested. Bars drawn to the mean of duplicate biological replicates with error bars representing standard deviation.



Supplementary Figure 3-23. High titer glycolate production via extended resting-cell bioconversion. Time course profile of formaldehyde to glycolate bioconversion for 4 hours under (A) 10 mM (B) 15 mM formaldehyde per hour feed rate. (C) Comparison between 10 mM/h and 15 mM/h formaldehyde feed rate for glycolate production. (D) Glycolate productivity comparison between 10 mM/h and 15 mM/h formaldehyde feed rate over time. Dots are drawn at the mean of duplicate biological replicates with error bars representing standard deviation.

Chapter 4. Engineering *Escherichia coli* for utilization of methylsuccinic acid, a key metabolic precursor from the oxygen-independent activation of methane

Introduction

Branched five-carbon (C5) dicarboxylic acids and corresponding CoA thioesters (acyl-CoAs) are the key metabolites in various native and synthetic C1 and C2 assimilation pathways, including Ethylmalonyl-CoA pathway[206], 2-Hydroxypropionate bicycle[207], Methylasparate cycle[208] and Crotonyl-CoA/ethylmalonyl-CoA/hydroxybutyryl-CoA (CETCH) cycle[170]. Among them, methylsuccinate is particularly interesting as its possibility of being generated from anaerobic activation of methane has been demonstrated[20]. Oxygen-independent activation of methane is highly sought-after especially from an energy efficiency perspective, as the electrons possessed by methane can retain in the product which can subsequently be utilized as reducing power to generate energy-containing (reduced) products or transfer electrons to terminal electron acceptor to generate energy and driving force for the downstream reactions.

Methane activation is extremely challenging due to its high dissociation energy of 440 kJ/mol[209]. Methane monooxygenases, most widely studied methane activation enzymes, overcome the activation energy by coupling it with highly exergonic water forming reaction using molecular oxygen as an electron acceptor[16]. As a result, the reaction suffers from substantial energy loss as heat during the activation process, which not only loses two pairs of electrons in the form of (NAD(P)H) for irreversible water forming reaction but also generates significant amount of heat mandating extra cooling demand for large scale biocatalysis or fermentation[20]. The only oxygen-independent enzymatic activation of methane demonstrated to date is via reverse methanogenesis in methanogenic archaea[53], which suffers from extremely slow kinetics likely due to the unfavorable thermodynamics of the methane activation (ΔG = + 30 kJ/mol)[20]. Moreover, various cofactors involved in the reverse methanogenesis only existing in the anaerobic archaea make it hard to transfer this pathway to more tractable organisms[210].

Methane activation via fumarate addition is postulated as an alternative oxygen-independent activation method. Based on the potential energy surface (PES) analysis, this activation method is overall thermodynamically downhill (ΔG = - 15 kJ/mol) as carbon-carbon bond formation between activated methyl radical and the secondary carbon of fumarate is highly exergonic compensating high activation energy required for methane activation via methyl thiyl radical[105]. Numerous studies are available for alkylsuccinate synthases (ASS) and (1-methylalkyl)succinate synthase (MAS), which catalyze the alkane activation via fumarate addition, with carbon chain lengths ranging from C3 to C16[211]. Although most ASS and MAS enzymes are known to activate the secondary (subterminal) carbon of the C3 and longer-chain alkanes, a study reports terminal carbon activation of propane from the anaerobic culture of sulfate-reducing bacteria[104], which opens possibility for ethane and methane activations. Recently, we showed the functional expression of (1-methylalkyl)succinate synthase (Mas) from *Azoarcus sp.* Strain HxN1 in *Escherichia coli,* activating hexane, butane and propane via fumarate addition[108]. This research shows the feasibility of transferring the short-chain alkane metabolism in a model organism if an appropriate downstream metabolism of alkylsuccinate is in place.

Here, we show the construction and implementation of a synthetic metabolic pathway in *E. coli* for utilization of methylsuccinate, a key metabolic precursor from the oxygen-independent activation of methane via fumarate addition. We constructed the methylsuccinate metabolism pathway bottom up by identifying and screening putative enzymes with activities with methylsuccinate and downstream metabolites. After the full pathway demonstration in vitro, we introduced the pathway in *E. coli,* showing consumption of methylsuccinate in vivo. Through combinatorial approach of rational pathway design and adaptive laboratory evolution, we obtained a strain that can grow efficiently on methylsuccinate as sole carbon source. This strain can serve

as a selection platform for screening potential methylsuccinate synthase, generated via genome mining and/or protein engineering. Moreover, upon identification of efficient methylsuccinate synthase, it opens the possibility for a synthetic methanotrophs which has not been successful up to date.

Results

Metabolic pathways for methylsuccinyl-CoA and related compounds

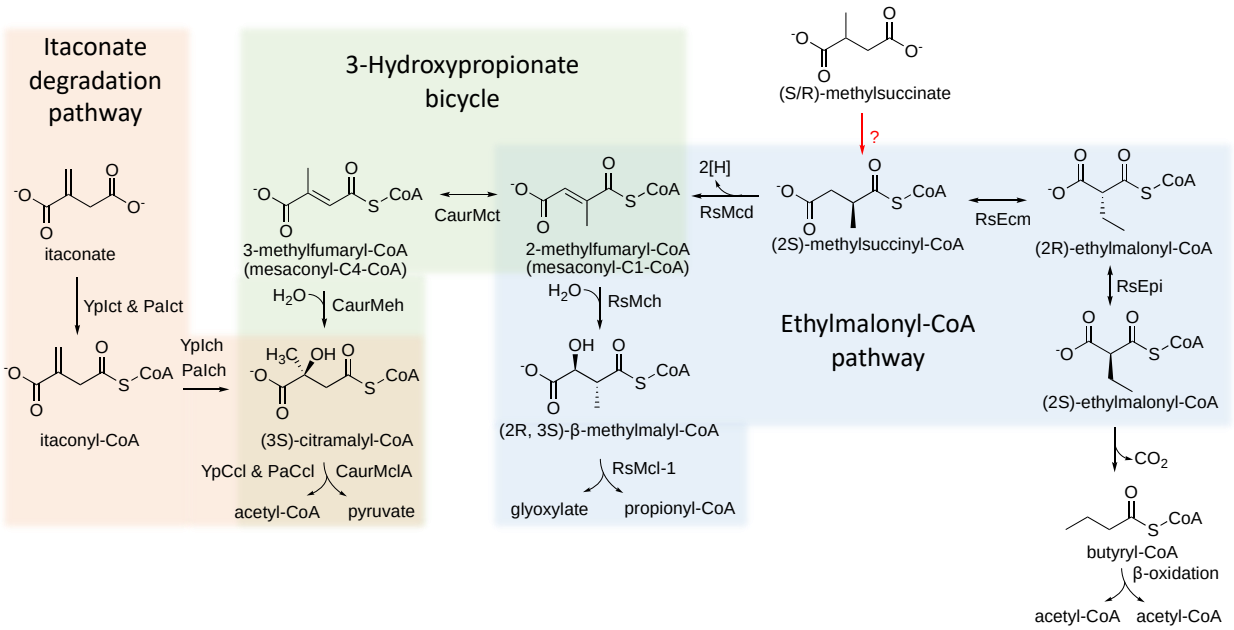


Figure 4-1. Overview of metabolic pathways that involve methylsuccinyl-CoA as an intermediate. Orange: Itaconate degradation pathway from pathogens, *Yersinia pestis* and *Pseudomonas aeruginosa[212]*; Green: part of CO_2-fixing 3-Hydroxypropionate bicycle from *Chloroflexus aurantiacus[207]*; Blue: part of C2 unit utilizing Ethylmalonyl-CoA pathway from *Rhodobacter sphaeroides[206]*.

There is limited information about the metabolism of methylsuccinate. However, its activated form, methylsuccinyl-CoA appears in the C2 (acetyl-CoA) assimilation pathway of purple non-sulfur bacteria such as *Rhodobacter sphaeroides*, known as Ethylmalonyl-CoA pathway[206] (Figure 1). (2R)-Ethylmalonyl-CoA mutase (Ecm)[213] and (2S)-methylsuccinyl-CoA dehydrogenase (Mcd)[214] are the pathway enzymes that catalyze conversion of (2S)-

methylsuccinyl-CoA to (2R)-ethylmalonyl-CoA and mesaconyl-(C1)-CoA, respectively. It was found that both enzymes are highly specific toward (2S) isomer of the methylsuccinyl-CoA[213, 214]. Ethylmalonyl-CoA epimerase converts (2R)-ethylmalonyl-CoA to (2S)-ethylmalonyl-CoA which could be decarboxylated to form butyryl-CoA, catalyzed by (2S)-ethylmalonyl-CoA decarboxylase[215]. Butyryl-CoA can be metabolized by *E. coli* via β-oxidation pathway to generate energy and carbon building block for biomass and bioproducts. On the other direction from the methylsuccinyl-CoA node of the ethylmalonyl-CoA pathway, mesaconyl-(C1)-CoA generated from (2S)-methylsuccinyl-CoA dehydrogenation can further be hydrated to form (2R, 3S)-β-methylmalyl-CoA, which can be cleaved into glyoxylate and propionyl-CoA, both of which are native metabolites of *E. coli*[214]. Alternatively, an intramolecular CoA transferase interconverts between mesaconyl-(C1)-CoA and mesaconyl-(C4)-CoA, where mesaconyl-(C4)-CoA can be hydrated to form (3S)-citramalyl-CoA, which subsequently divides into pyruvate and acetyl-CoA as demonstrated in 3-hydroxypropionate bicycle[207] (Figure 4-1). Ethylmalonyl-CoA pathway enzymes from *R. sphaeroides* and 3-hydroxypropionate bicycle from *Chloroflexus aurantiacus* are functionally expressed and characterized using *E. coli* as a host[213, 214, 216].

Methylsuccinate activation to methylsuccinyl-CoA

Table 4-1. Potential methylsuccinate CoA-transferases and ligases with their primary substrates, catalytic efficiencies, and CoA donors.

Name	Acronym	Organism	Primary substrate	k_{cat}/K_m with methylsuccinate ($mM^{-1} s^{-1}$)	CoA donor	Ref.
Itaconate CoA-transferase	YpIct	*Y. pestis*	Itaconate	6.1	Various acyl-CoAs including acetyl-CoA	[212]
Itaconate CoA-transferase	PaIct	*P. aeruginosa*	Itaconate	60.4	Succinyl-CoA	[212]
Mesaconate CoA-transferase	HhMct	*H. hispanica*	Mesaconate	12.4	Succinyl-CoA	[217]
Benzylsuccinate CoA-transferase	TaBbsEF	*T. aromatica*	Benzylsuccinate	N/A (58% of activity with benzylsuccinate)	Succinyl-CoA	[218]

Succinate-CoA ligase	EcSucCD	*E. coli*	Succinate	N/A (3-15 with itaconate)	CoA and ATP (ADP forming)	[219]
Adipyl-CoA synthetase	Tfu2576-7	*T. fusca*	Adipate	N/A	CoA and ATP (ADP forming)	[220]

While methylsuccinate activation to methylsuccinyl-CoA has not been the subject of study, promiscuous activities of several CoA transferases with methylsuccinate as substrate have been demonstrated. Notably, itaconate CoA-transferase (Ict), which catalyzes CoA activation of itaconate is shown to have activity with methylsuccinate[212]. This is not surprising due to the structural similarity of the two compounds, itaconate having methylene instead of methyl branch at the secondary carbon. Two different Icts from pathogens, one from *Yersinia pestis* (YpIct) and the other from *Pseudomonas aeruginosa* (PaIct), are characterized to have activities not only with itaconate but also with methylsuccinate[212]. While PaIct has an order of magnitude better catalytic efficiency (k_{cat}/K_m = 60.4 mM^{-1} s^{-1}) than YpIct (k_{cat}/K_m = 6.1 mM^{-1} s^{-1}) with methylsuccinate, YpIct can accept wide range of CoA donors including acetyl-CoA while PaIct can only accept succinyl-CoA as CoA donor[212] (Table 4-1). Interestingly, itaconate degradation metabolism follows hydration of itaconyl-CoA to (3S)-citramalyl-CoA and cleavage to pyruvate and acetyl-CoA, which is identical to the 3-Hydroxypropionate bicycle downstream of mesaconyl-C4-CoA (Figure 4-1).

Apart from the itaconate CoA transferases, there are a couple more CoA transferases that are identified to activate methylsuccinate using succinyl-CoA as donor. One is mesaconate CoA-transferase identified from haloarchaea *Haloarcula hispanica*[217] (Table 4-1). This enzyme is part of the Methylaspartate cycle found in haloarchaea which involves hydration of mesaconyl-C1-CoA to β-methylmalyl-CoA and cleavage into propionyl-CoA and glyoxylate identical to the Ethylmalonyl-CoA pathway (Figure 4-1). Problem with this enzyme is that haloarchaea grow in high salt concentrations and their proteins are often misfolded and aggregated when heterologously

expressed under low salt concentration[221]. Indeed, we were not able to obtain a soluble fraction of this enzyme (Supplementary Figure 4-1). Another enzyme, benzylsuccinate CoA-transferase (Table 4-1) is part of the anaerobic toluene degradation pathway, which involves the activation of toluene via fumarate addition, catalyzed by glycyl radical enzyme called benzylsuccinate synthase[218]. Although anaerobic toluene degradation pathway has high similarity with the activation and metabolism of methane via fumarate addition, there is no report of *T. aromatica* growing on methane or any short hydrocarbon via anaerobic oxidation. Also, this operon is induced strictly by toluene, which makes the benzylsuccinate synthase unlikely for utilizing methane as co-substrate.

Although CoA transferases allow activation of acids without the direct cost of ATP, one key disadvantage is the requirement of an appropriate CoA donor, which necessitates the rewiring of downstream pathway to regenerate the donor. While acetyl-CoA is an abundant metabolite in the cells, succinyl-CoA is relatively less available which could potentially decrease the activity of CoA transferases in vivo[222]. Hence, we explored other putative CoA synthetases and ligases, which require ATP and free CoA instead of specific acyl-CoA as CoA donor. Although there is no report on CoA ligases having activities with methylsuccinate, succinate-CoA ligase such as SucCD from *E. coli* is a common enzyme part of the TCA cycle and its promiscuity on various dicarboxylic acids such as malate and itaconate have been discussed[219]. Moreover, adipyl-CoA synthetase (Tfu2576, 2577) from adipate degradation pathway of *Thermobifida fusca* is shown to have activities with C5 dicarboxylate, glutarate[223], and C6 dicarboxylate, adipate[220], and is successfully expressed in *E. coli*, which makes it a potential candidate for methylsuccinate activation as well.

Identification of novel methylsuccinyl-CoA dehydrogenase

As shown in Figure 4-1, methylsuccinyl-CoA can be isomerized to ethylmalonyl-CoA or dehydrogenated to mesaconyl-CoA in either direction of the Ethylmalonyl-CoA pathway. A caveat is that the activated methylsuccinyl-CoA should specifically be the (2S)-methylsuccinyl-CoA out of four different possible stereo and structural isomers of methylsuccinyl-CoA ((2S), (2R), (3S) and (3R)). After the activation to (2S)-methylsuccinyl-CoA, the pathway toward isomerization followed by decarboxylation to form butyryl-CoA does not require specific cofactor other than vitamin B12 and the functional expression of all enzymes are demonstrated in *E. coli*[213, 215]. The dehydrogenation step in the opposite direction, however, could be troublesome as it involves heterologous electron transfer flavoprotein (ETF) as an electron shuttle[214]. There is limited information on this type of electron transfer chain and no report of transferring the entire ETF-mediated electron transfer chain into the heterologous host. Due to these challenges, the authors who built a synthetic CO_2 fixation pathway[170] involving this pathway engineered the (2S)-methylsuccinyl-CoA dehydrogenase (RsMcd) into an oxidase (RsMco) that directly utilizes oxygen as electron acceptor[224]. However, they later found that the engineered oxidase is not as efficient as the wildtype utilizing ETF, causing bottleneck in the pathway and addressed the issue by introducing purified ETF in the *in vitro* system[225].

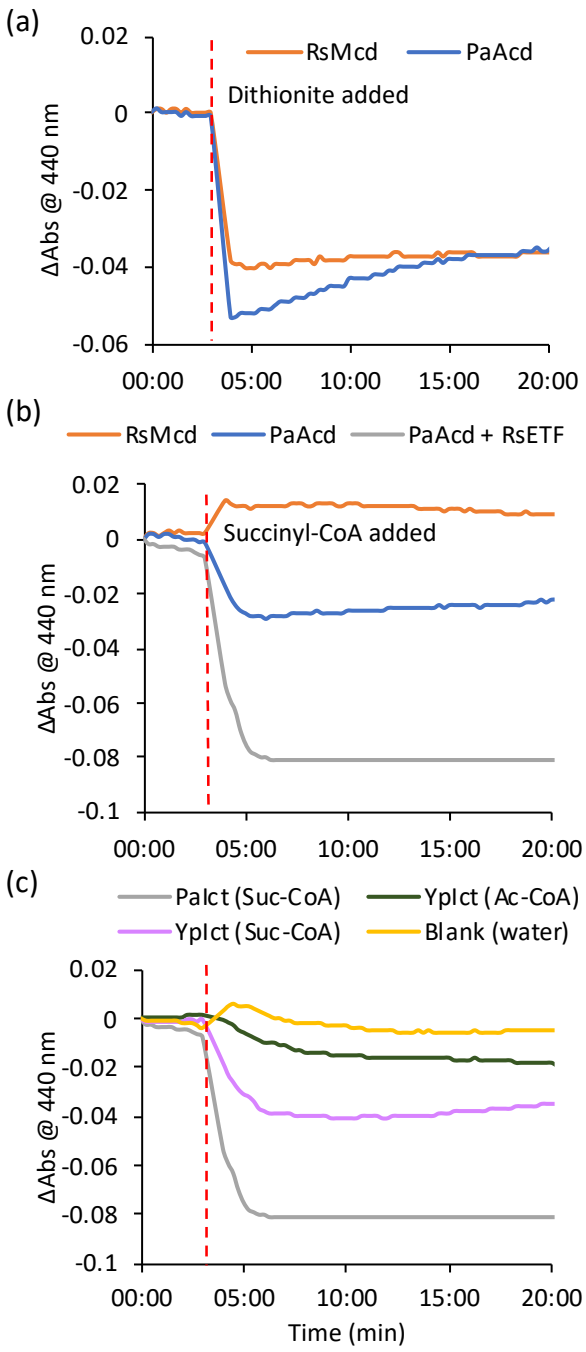


Figure 4-2. Colorimetric assays with FAD-containing dehydrogenases and electron transfer flavoprotein (ETF) by measuring change in absorbance at 440 nm. (a) Reduction of RsMcd and PaAcd when unspecific reducing agent, dithionite is added at 3 min. (b) Testing different methylsuccinyl-CoA dehydrogenases and compatibility with RsETF via coupled enzyme assay with PaIct. Reaction initiated upon addition of succinyl-CoA at 3 min. (c) Testing different methylsuccinate CoA transferases and corresponding CoA donors. Reaction initiated upon addition of succinyl-CoA (Suc-CoA) or acetyl-CoA (Ac-CoA) at 3 min.

As a method to screen methylsuccinate CoA transferases and ligases identified from the previous section for the formation of (2S)-methylsuccinyl-CoA, we decided to use the coupled enzyme assay with RsMcd which was demonstrated to utilize only the (2S) isomer. It was demonstrated that RsMcd activity can be monitored by colorimetric assay as there is change in absorbance at 440 nm from oxidized state to reduced state of FAD-containing dehydrogenases and electron transfer flavoproteins such as RsMcd[224] (Figure 4-2a). Using this characteristics, various methylsuccinate CoA-transferases and ligases can be tested in a coupled enzymatic assay with RsMcd to see if (2S)-methylsuccinyl-CoA is generated. However, we were not able to see the activity of RsMcd in combination with any of the putative methylsuccinate CoA-transferases and ligases in Table 1 (Figure 4-2b). This means that we cannot construct the methylsuccinate metabolism pathway on either direction of the Ethylmalonyl-CoA pathway because (2S)-methylsuccinyl-CoA cannot be generated from methylsuccinate.

Therefore, we had to find an alternative methylsuccinyl-CoA dehydrogenase that shows activity in a coupled assay with the potential CoA transferases and ligases we identified. While exploring, we found a putative acyl-CoA dehydrogenase (PaAcd, Uniprot accession: Q9I566) in the same operon as the itaconate degradation pathway genes of *P. aeruginosa*. The authors who identified this operon comment that this enzyme could be a methylsuccinyl-CoA dehydrogenase based on the substrate specificities of other enzymes in the operon[212]. Purified PaAcd showed the change in absorbance at 440 nm, when unspecific reducing agent dithionite is added (Figure 4-2a), indicating the presence of FAD. Remarkably, PaAcd showed activity with the product of PaIct and methylsuccinate using succinyl-CoA as donor, while RsMcd did not (Figure 4-2b). We also tested the compatibility between PaAcd and the purified electron transfer flavoprotein (ETF) from *R. sphaeroides*. Although the two enzymes are from different hosts, we saw the reduction of

not only PaAcd but also RsETF upon addition of succinyl-CoA, suggesting the electron transfer activity from PaAcd to RsETF (Figure 4-2b).

We further tested different CoA transferases and ligases in a coupled assay with PaAcd and RsETF. The only other enzyme from Table 1 that showed activity was YpIct, which generated methylsuccinyl-CoA from both acetyl-CoA and succinyl-CoA as donors, as reported from literature[212] (Figure 4-2c). However, the change in absorbance was notably smaller and slower than PaIct with succinyl-CoA (Figure 4-2c), which is also consistent with the reported catalytic efficiency of the two enzymes with methylsuccinate as substrate (Table 4-1). Based on the molecular structure of itaconyl-CoA from the itaconate degradation pathway (Figure 4-1), it is hypothesized that PaIct and YpIct generate 3-methylsuccinyl-CoA, which explains why RsMcd could not recognize it as substrate. Taken together, we have identified enzymes that can convert methylsuccinate to mesaconyl-C4-CoA, which can further be converted to pyruvate and acetyl-CoA following the downstream route of 3-hydroxypropionate bicycle (Figure 4-1).

Demonstration of the full pathway in vitro

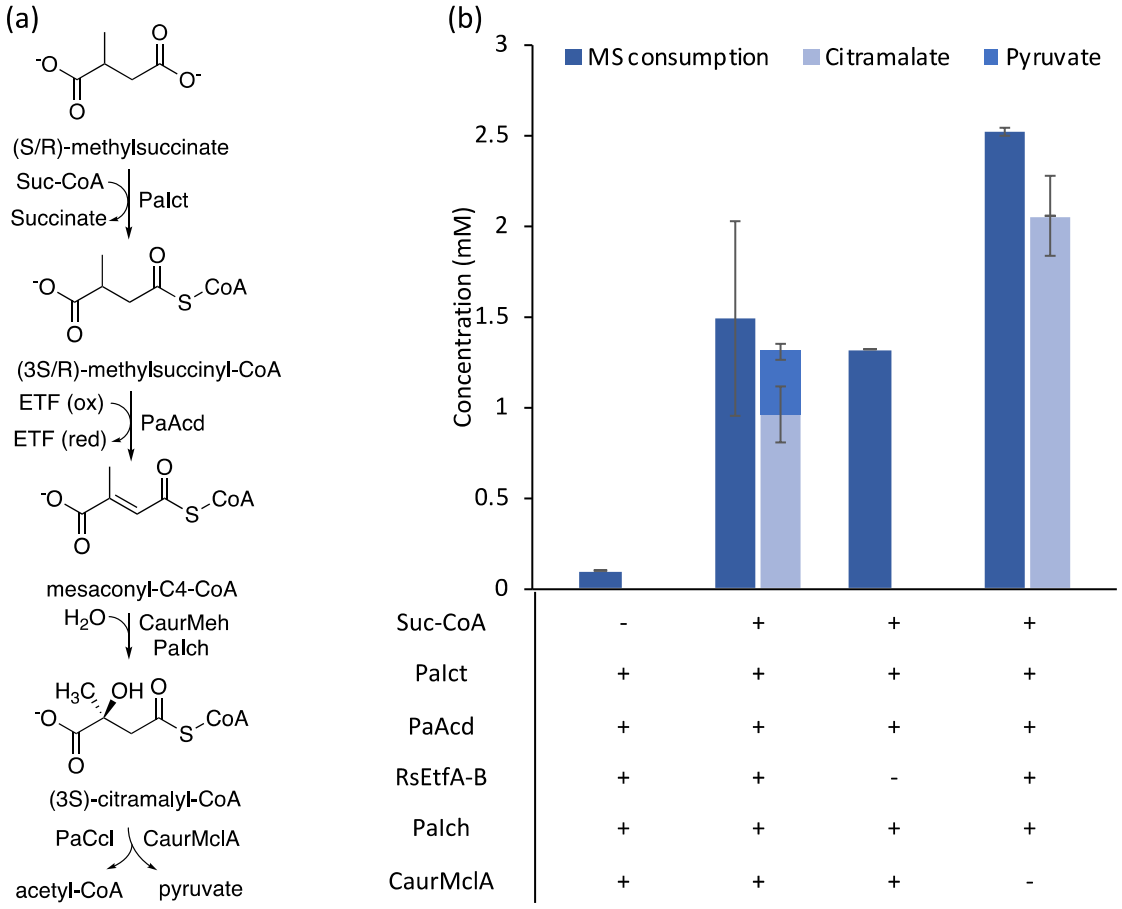


Figure 4-3. In vitro prototyping of methylsuccinate degradation pathway. (a) Proposed methylsuccinate metabolism pathway consisting of itaconate degradation pathway genes from *P. aeruginosa* (PaIct, PaIch and PaCcl), 3-hydroxypropionate bicycle genes from *C. aurantiacus* (CaurMeh and CaurMclA) and newly identified methylsuccinyl-CoA dehydrogenase from *P. aeruginosa* (PaAcd) using electron transfer flavoprotein (ETF) from *R. sphaeroides* as electron acceptor. (b) Full pathway prototyping with purified enzymes. Methylsuccinate consumption and two major products, citramalate and pyruvate production are measured. Bars drawn to the mean of triplicate technical replicates and error bars represent standard deviation.

Heterologous expression of 3-Hydroxypropionate bicycle pathway genes, mesaconyl-CoA hydratase (CaurMeh) and (3S)-citramalyl-CoA lyase (CaurMclA) from *C. aurantiacus* was demonstrated in *E. coli*[216]. Moreover, itaconyl-CoA hydratase (PaIch) of itaconate degradation

pathway genes from P. aeruginosa has very high activity with mesaconyl-C4-CoA (k_{cat}/K_m = 4650.8 mM^{-1} s^{-1}) exceeding its activity with itaconyl-CoA by more than an order of magnitude[212]. Therefore, the full pathway can be constructed with two enzyme options each for mesaconyl-C4-CoA hydratase and (3S)-citramalyl-CoA lyase (Figure 4-3a). Full methylsuccinate metabolism pathway was tested in vitro with purified enzymes. It was found that ETF is essential to produce products downstream of methylsuccinyl-CoA because it serves as the electron acceptor for the PaAcd driving reaction (Figure 4-3b). We also saw significant accumulation of citramalate both with and without CaurMclA addition, likely due to (3S)-citramalyl-CoA serving as CoA donor for methylsuccinate accumulation (Figure 4-3b). Indeed, PaIct is shown to activate (3S)-citramalate to (3S)-citramalyl-CoA indicating its possibility to be used as CoA donor in the reverse direction[212]. This also explains citramalate titer exceeding the stoichiometric amount of succinyl-CoA added (1 mM). We saw pyruvate titer of 0.34 mM in the presence of all pathway enzymes, 10 mM methylsuccinate and 1 mM succinyl-CoA, indicating the full pathway is functional in vitro.

In vivo implementation of the pathway using itaconate as proxy

The full pathway enzymes are constructed in expression vectors and transformed into *E. coli* strain MG1655(DE3) but no consumption of methylsuccinate was observed in the MOPS minimal media supplemented with 10 g/L tryptone, 5 g/L yeast extract and 5 g/L sodium methylsuccinate. We hypothesized the non-native electron transfer pathway mediated by heterologous ETF to be the reason. Hence, we decided to use itaconate as a proxy to demonstrate the entire pathway except for the potentially problematic dehydrogenation step. Although itaconate has methylene branch while methylsuccinate metabolism involves mesaconyl-CoA having double bond between secondary and tertiary carbons (Figure 4-1), the same enzymes identified for

methylsuccinate metabolism (PaIct, YpIct and PaIch) have activities with both substrates (Figure 4-1). Therefore, the strategy was to build an "itaconate degradation module (IDM)" first to engineer strain that grows on itaconate and introduce the "methylsuccinyl-CoA dehydrogenase module (MDM)" that contains PaAcd, ETF and appropriate electron transfer chain later to complete the full methylsuccinate metabolism in vivo.

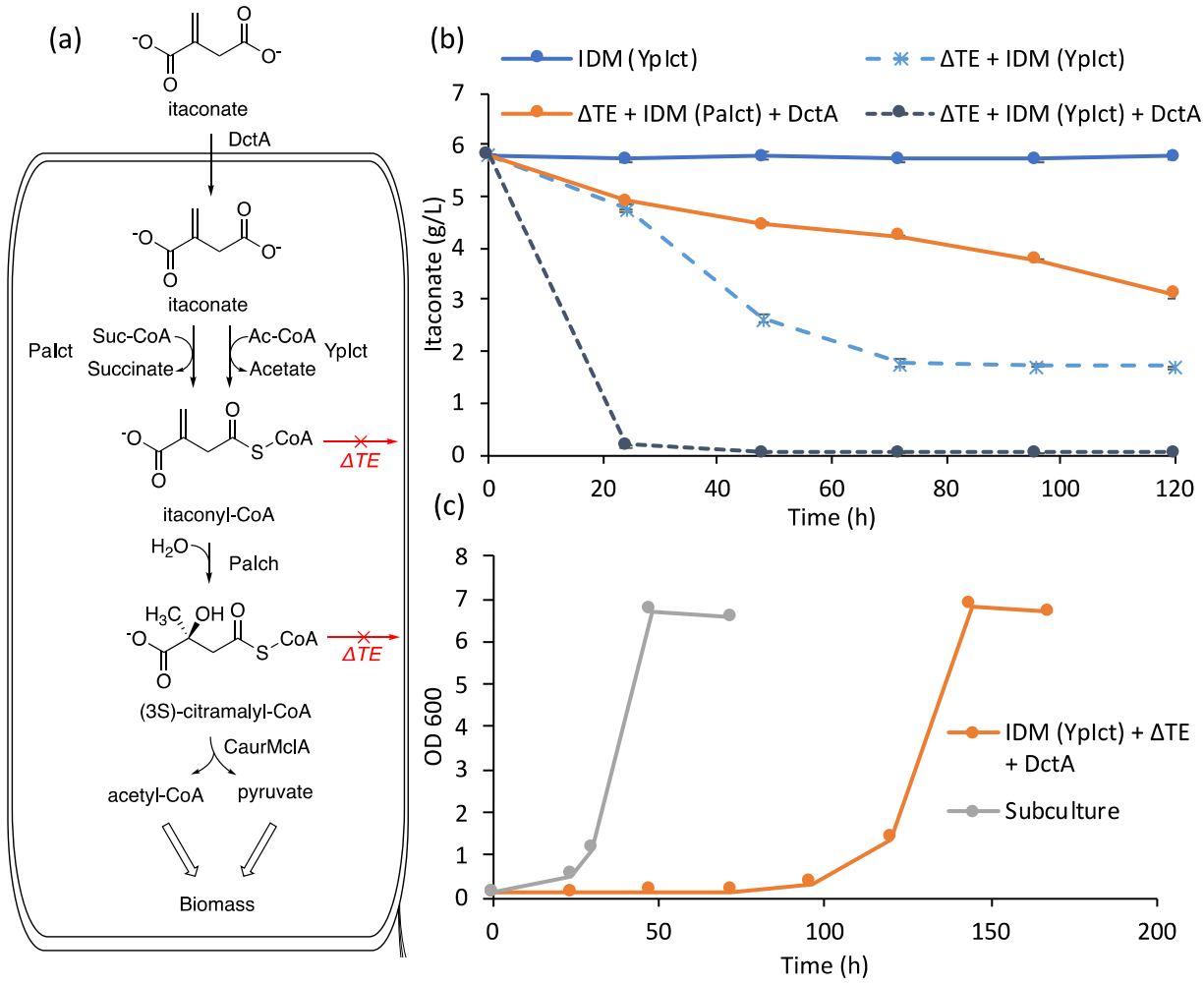


Figure 4-4. In vivo prototyping of the itaconate degradation module (IDM) in E. coli. (a) Itaconate degradation pathway incorporated into *E. col*i Strain MG1655 (DE3) with endogenous thioesterases (*yciA, tesA, tesB, ybgC, ydiI, fadM*) deleted (*ΔTE*). Two different itaconate CoA transferases (PaIct and YpIct) are tested utilizing different CoA donors. Final products, acetyl-CoA and pyruvate can be metabolized by *E. coli* for cell growth and maintenance. (b) Demonstration of itaconate consumption of *E. coli* in growing cell cultures. Cells are grown in the itaconate containing MOPS minimal media supplemented with 10 g/L tryptone and 5 g/L yeast extract. Dots and lines are drawn at the mean of

triplicate biological replicates and error bars represent standard deviation. (c) IDM (YpIct) + ΔTE + DctA from (b) is sub-cultured after washing in the MOPS minimal media with 6 g/L itaconate as sole carbon source. Exponentially growing IDM (YpIct) + ΔTE + DctA strain is sub-cultured to the fresh media with initial OD600 of 0.05.

For the CoA activation enzyme, we tested PaIct and YpIct both of which showed methylsuccinate activation activity in vitro (Figure 4-2c). Although PaIct has better kinetic parameters with methylsuccinate, it can only utilize succinyl-CoA as donor and causes CoA recycling from citramalyl-CoA leading to citramalate accumulation in vitro (Figure 4-3b). YpIct, on the other hand, can accept acetyl-CoA as donor (Figure 4-2c) and has no activity with (S)-citramalate[212]. Two IDMs were then constructed in expression vectors and transformed into *E. coli* strain MG1655(DE3) to test itaconate consumption in rich media but no consumption of itaconate was observed for 120 hours (Figure 4-4b). We hypothesized two possibilities: (1) endogenous thioesterases hydrolyzing the intermediate itaconyl-CoA and/or citramalyl-CoA and (2) absent or uninduced transporter gene responsible for itaconate import. To address them, we used a strain with six native thioesterases deleted (*ΔTE: ΔyciA ΔtesA ΔtesB ΔybgC ΔydiI ΔfadM*) and co-expressed *E. coli* dicarboxylate transporter (DctA) along with the IDMs (Figure 4-4a). It was found that the deletion of native thioesterases is critical for the itaconate degradation pathway to be functional in *E. coli* and overexpression of DctA greatly enhanced the consumption of itaconate showing full consumption of 5 g/L within 24 hours (Figure 4-4b). Native thioesterase deletion was shown crucial in other metabolic pathways that involve acyl-CoAs as intermediates, such as reverse β-oxidation cycle[135] and formyl-CoA elongation pathways[25]. YpIct showed better itaconate consumption than PaIct, which is consistent with the kinetic parameters available, where YpIct (k_{cat}/K_m = 170.9 mM^{-1} s^{-1}) has up to two-fold higher catalytic efficiency than PaIct (k_{cat}/K_m = 82.2 mM^{-1} s^{-1})[212] (Figure 4-4b). Subsequent inoculation of itaconate consuming

strains in minimal media with itaconate showed growth indicating the strain can grow on the pyruvate and acetyl-CoA, the two products from itaconate degradation pathway (Figure 4-4c). Long lag phase of 100 hours observed from the initial strain decreased substantially to 24 hours after subculturing showing the requirement for adaptation period from rich media to minimal media (Figure 4-4c). However, introduction of MDM containing PaAcd and RsEtfA-B on either of the strain containing IDM (YpIct) or IDM (PaIct) did not show methylsuccinate consumption, nor growth on methylsuccinate indicating further optimization of MDM is needed.

Engineering methylsuccinate consuming strain via adaptive laboratory evolution

Full pathway from methylsuccinate to pyruvate was demonstrated in vitro (Figure 4-3) and the same enzymes (except for the enzymes in MDM) are used to enable consumption and growth on itaconate in *E. coli* in vivo (Figure 4-4). However, the fact that the strain cannot consume methylsuccinate means there could be issues with the MDM, possibly with the heterologous electron transfer chain. Electron transfer mediated by ETFs are commonly found in acyl-CoA dehydrogenases (Acd) like PaAcd. ETFs are soluble proteins that transfer electron from Acd to membrane-bound electron transfer flavoprotein:ubiquinone oxidoreductases (ETF-QO)[226]. While PaAcd activity mediated by RsEtfA-B was observed during in vitro prototyping because ETF served as the terminal electron acceptor (close to stoichiometric amount of ETF (654 μM) was added), it could not be functional in vivo due to the absence of ETF-QO that channels electrons to the ubiquinone pool and ultimately to the terminal electron acceptor. However, introducing a new vector harboring MDM, which contains PaAcd, RsEtfA-B and RsEtfQO did not support methylsuccinate consumption.

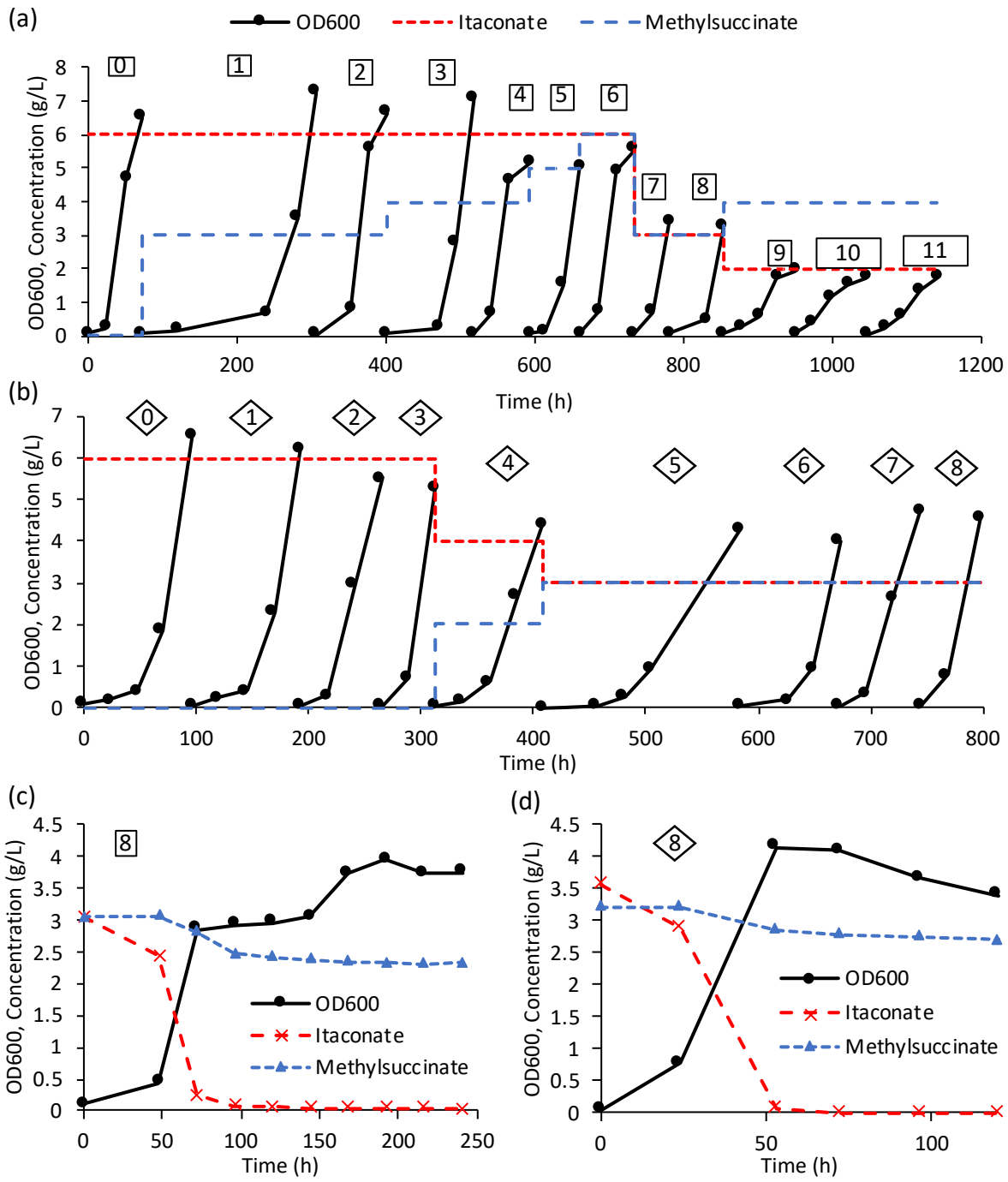


Figure 4-5. Adaptive laboratory evolution (ALE) feeding itaconate and methylsuccinate as mixed carbon substrates. (a) ALE of strain overexpressing IDM under expression vector harboring T7lac promoter. Numbers in rectangle represents generation number. (b) ALE of strain expressing IDM under constitutive promoter integrated into the genome. Numbers in rhombus represents generation number. (c) Growth and substrate consumption profile of the 8th generation from (a). (d) Growth and substrate consumption profile of the 8th generation from (b).

We hypothesized that either heterologous RsEtfA-B and RsEtfQO are not functional or "compatible" with *E. coli* possibly because *R. sphaeroides* have Ubiquinone-10 (CoQ10) while *E. coli* has Ubiquinone-8 (CoQ8)[226]. Nonetheless, we hypothesized that the phenotype could be obtained via adaptive laboratory evolution (ALE) since *E. coli* has native acyl-CoA dehydrogenases and ETFs which could compensate methylsuccinyl-CoA dehydrogenase and/or ETF and ETF-QO activity. Leveraging the overlapping pathways of itaconate and methylsuccinate metabolism, we designed an ALE strategy that uses itaconate and methylsuccinate as mixed carbon substrates with increasing methylsuccinate concentration and decreasing itaconate concentrations simultaneously (Figure 4-5a). Through the course of evolution, however, we saw that the cell growth aligns with the available itaconate concentration in the media with minimal methylsuccinate consumption of up to 0.7 g/L observed from analysis of the 8th generation (Figure 4-5c).

With unsuccessful initial attempt for ALE, we slightly modified the strategy and made two changes to the host strain: (1) brought back one of the six thioesterases, *ydiI* (*menI*) gene, back in the genome as it was found to be essential for menaquinone synthesis[227]. Although ETF-QO is known to interact with ubiquinone pool and not menaquinone, *ydiI* knockout could still have deleterious effect for the pathway; (2) Replaced IDM under the control of constitutive promoter instead of IPTG-inducible T7 promoter and integrated into the genome to mitigate burdens from strong T7 promoter and plasmid maintenance. We found that *ydiI* deletion is not crucial for the pathway and overexpression of DctA is not necessary as the strain was able to grow on itaconate as sole carbon source after 50 hours of lag phase. Similar ALE strategy of mixed carbon sources with increasing methylsuccinate and decreasing itaconate concentrations was used (Figure 4-5b). While the evolved strain at the 8th generation (Figure 4-5d) was able to grow on the mixed carbon

media with reduced lag phase and to slightly higher OD600 than the strain with plasmids (Figure 4-5c), it still failed to consume more than 0.5 g/L methylsuccinate (Figure 4-5d).

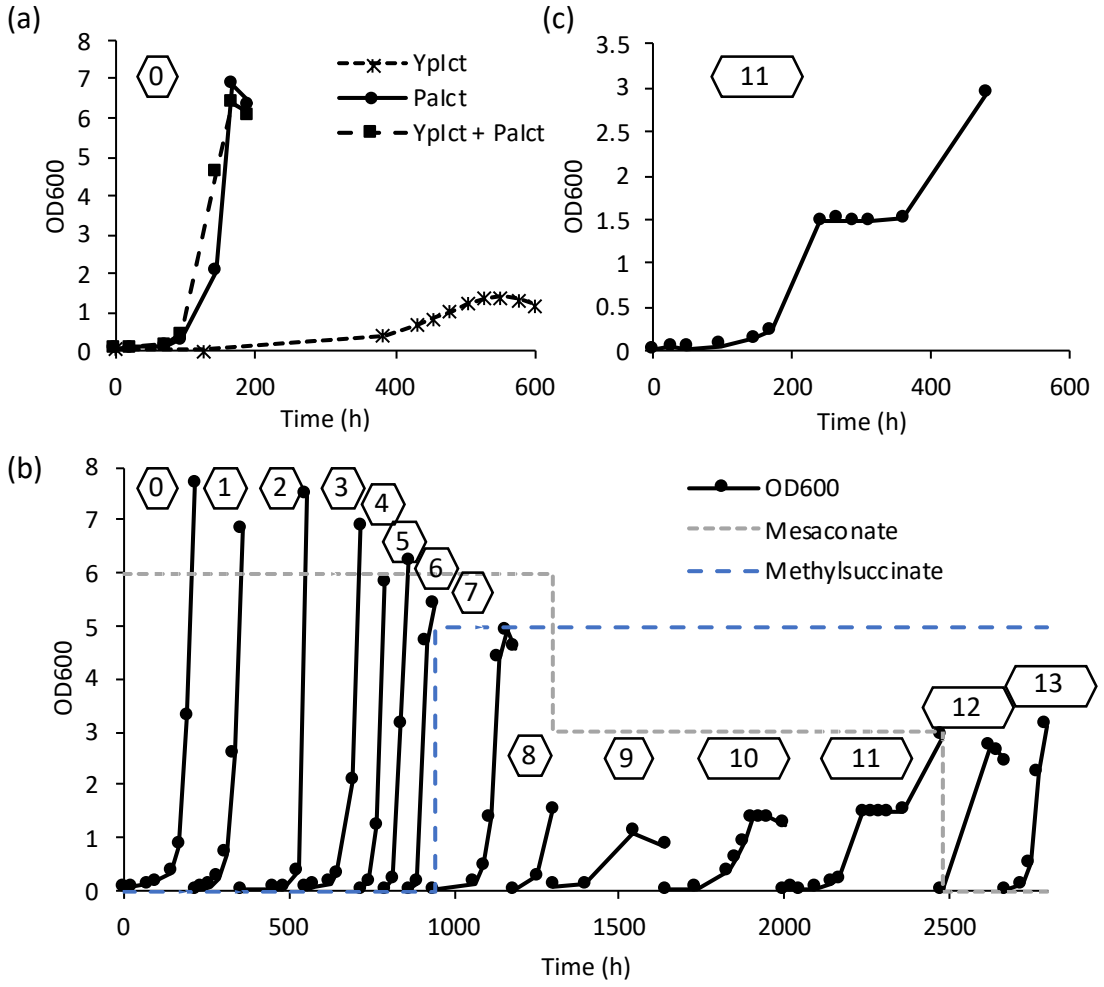


Figure 4-6. Adaptive laboratory evolution (ALE) feeding mesaconate and methylsuccinate as mixed carbon substrates. (a) Growth on mesaconate minimal media of strains harboring YpIct, PaIct or YpIct + PaIct in the genome-integrated IDM. (b) ALE of strain expressing IDM (PaIct) under constitutive promoter integrated into the genome. Numbers in hexagon represents generation number. (c) Growth profile of 11th generation under 3 g/L mesaconate and 6 g/L methylsuccinate showing diauxic growth behavior.

Since the mixed carbon substrate-based strategy with itaconate was not successful, we decided to try mesaconate as proxy since its CoA thioester, mesaconyl-CoA is the direct intermediate of the methylsuccinate metabolism pathway (Figure 4-3a). We also brought back

PaIct as it showed substantially better activity with methylsuccinate from in vitro assay (Figure 4-2c). Two additional strains were constructed: one harboring IDM with PaIct and the other harboring two copies of IDM operon one with YpIct and the other with PaIct integrated in the genome. Surprisingly, only the strains harboring PaIct fully consumed mesaconate and grew up to OD600 of 7 while the strain harboring YpIct grew only up to OD600 of 1.5 after 600 hours (Figure 4-6a). This is contrary to what is described in literature, where YpIct is reported to have activity with mesaconate (k_{cat}/K_m = 10.5 mM^{-1} s^{-1}) while PaIct is not[212]. The strain harboring genome integrated IDM with PaIct and with PaIct and YpIct continued to evolve under mesaconate as sole carbon source. More generations were required to reduce the lag phase possibly due to unfavorable kinetics of PaIct with mesaconate than YpIct with itaconate. However, when we started co-feeding methylsuccinate at the 8^{th} generation, we already saw methylsuccinate consumption of 1 g/L within 192 hours which was higher than the previous two strains evolved under itaconate as co-carbon substrate. A strain harboring both PaIct and YpIct failed to grow when the carbon source mixture switched from 6 g/L mesaconate and 6 g/L methylsuccinate to 3 g/L mesaconate and 6 g/L methylsuccinate at the 11^{th} generation possibly due to substrate competition between the two CoA transferases. The strain harboring IDM with only PaIct continued to grow (Figure 4-6b) and after a couple more passages under 3 g/L mesaconate and 6 g/L methylsuccinate, we incubated the strain for a prolonged time after it reached the stationary phase at around OD600 of 1.5. Interestingly, we saw the cells start to grow again between 360 hours and 480 hours, exhibiting a diauxic growth behavior (Figure 4-6c). We hypothesized that there was a critical genotypic change in the process that made the cells to start consuming the remaining methylsuccinate. To test the hypothesis, we subsequently passaged the strain to a fresh media containing only methylsuccinate after washing. To our excitement, we saw the cell growing after about 120-hour lag phase. After several more

passages in methylsuccinate minimal media, the final evolved strain had significantly reduced lag phase and was able to grow up to OD600 of 2.7 with specific growth rate of 0.05 h^{-1} (Figure 4-7b).

Engineering methylsuccinyl-CoA dehydrogenation module

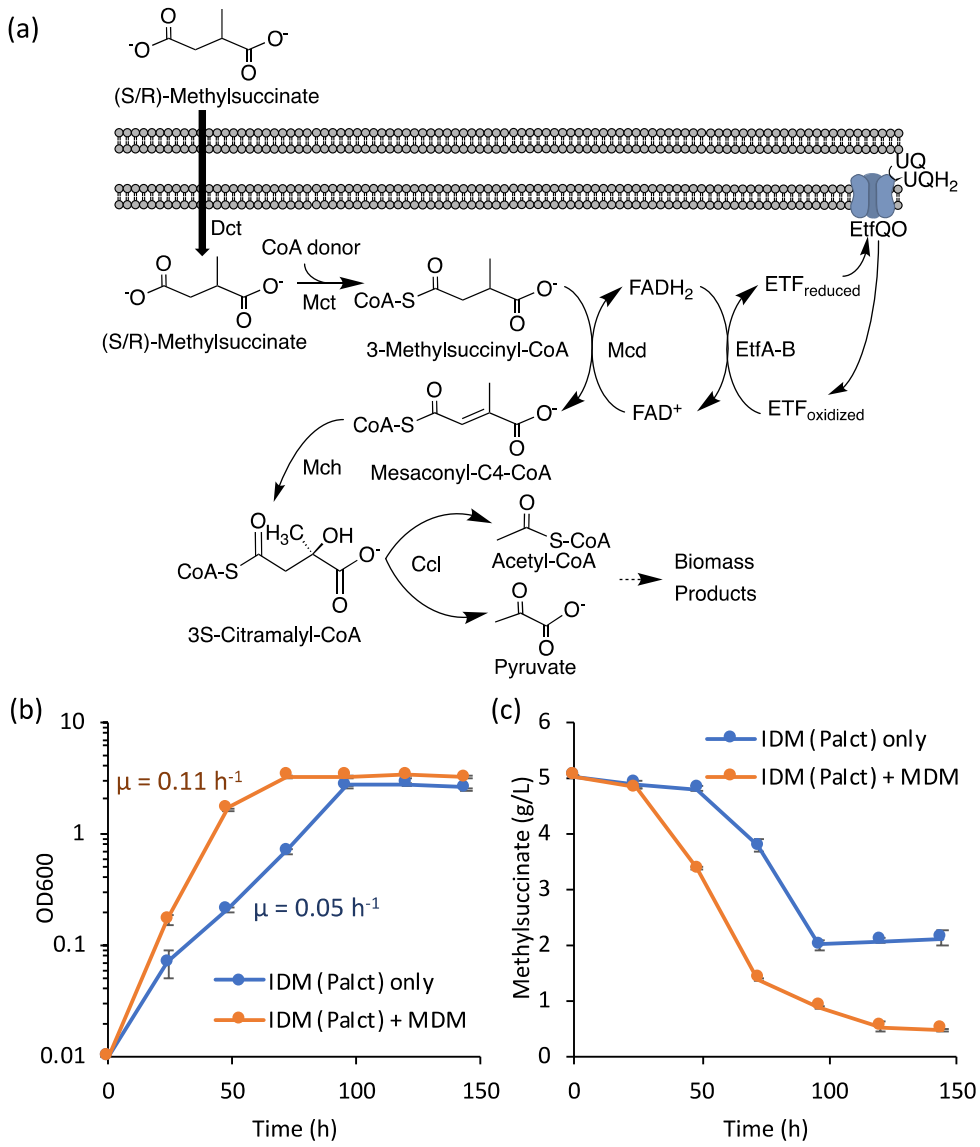


Figure 4-7. Synthetic methylsuccinate metabolism in *E. coli*. (a) Methylsuccinate metabolism demonstrated in *E. coli*. Dct: dicarboxylate transporter; Mct: methylsuccinyl-CoA transferase; Mcd: methylsuccinyl-CoA dehydrogenase; Mch: methylsuccinyl-CoA hydratase; EtfA-B: electron transfer flavoprotein; EtfQO: electron transfer flavoprotein-ubiquinone oxidoreductase; Ccl: (S)-citramalyl-CoA lyase. (b) Growth profile of strains growing on methylsuccinate as sole carbon source. Strain harboring IDM (PaIct) only shows specific growth rate of 0.05 h^{-1} while strain harboring both IDM (PaIct) and

MDM shows specific growth rate of 0.11 h^{-1}. (c) Methylsuccinate consumption profile of strains growing on methylsuccinate as sole carbon source. Dots and lines are drawn at the mean of triplicate biological replicates. Error bars indicate standard deviation.

Although selection pressure forced to unveil *E. coli*'s dormant methylsuccinyl-CoA dehydrogenation capability, it might not be as efficient as heterologous enzymes with innate activities and specificity toward desired substrate. We have already identified PaAcd to have activities with methylsuccinyl-CoA generated by PaIct using RsETF as an electron acceptor via in vitro assay (Figure 4-2b). We hypothesized ETF-QO is the critical component that connects electron transfer from RsETF to *E. coli* ubiquinone pool (Figure 4-7a) and tried to find enzymes via genome mining. As RsETF-QO was not compatible possibly due to incompatibility between *E. coli* native system and the membrane-bound protein, we focused on finding ETF-QO candidates from organisms closely related to *E. coli*. Interestingly, we found an enzyme annotated as an electron transfer flavoprotein-ubiquinone oxidoreductase (ETF-QO) from *E. coli* strain O157:H7 (GenBank accession: MHO03096.1, hereby named EcEtfQO). Although there is no literature information about this gene, it has high similarity with ETF-QO from *Aeromonas salmoncida*, which belongs to the *Aeromonas* genus employing ubiquinone-8 (CoQ8) like *E. coli*[228]. Subsequently, we constructed a new MDM operon that contains PaAcd, RsEtfA-B and EcEtfQO under the constitutive promoter and integrated into the genome of the strain harboring IDM (PaIct). Unlike previous strains that required secondary carbon sources, the strain harboring both IDM (PaIct) and MDM started to grow after 11 days lag phase (264 hours) with methylsuccinate as sole carbon source. After 5 subsequent passages for adaptation in methylsuccinate minimal media, the strain was able to grow without the lag phase and final OD600 of 3.4 with specific growth rate of 0.11 h^{-1} (Figure 4-7b).

Discussion

We have successfully engineered *E. coli* strain to grow on methylsuccinate as sole carbon source by integrating two operons (modules) encompassing 7 heterologous genes into the genome, followed by adaptive laboratory evolution (ALE). We identified a novel enzyme (3-methylsuccinyl-CoA dehydrogenase from *Pseudomonas aeruginosa*) that can catalyze the dehydrogenation reaction from 3-methylsuccinyl-CoA to mesaconyl-C4-CoA (Figure 4-7a). This enzyme can utilize electron transfer flavoprotein (ETF) from another host (RsETF from *Rhodobacter sphaeroides*) as an electron acceptor both in vitro and in vivo. We also identified a membrane-bound ETF-ubiquinone oxidoreductase (ETF-QO) that channels electrons from RsETF to *E. coli* ubiquinone pool, demonstrating the full heterologous ETF-mediated electron transfer pathway for the first time. *E. coli* also has its own acyl-CoA dehydrogenases and ETFs[229] and under rounds of ALE, we were able to see growth on methylsuccinate without the methylsuccinyl-CoA dehydrogenase module (MDM) showing its native capability for catalyzing the reaction. Full genome sequencing of the evolved strain supplemented with transcriptomic analysis will provide more information about which native genes are responsible for the methylsuccinyl-CoA dehydrogenase, ETF and ETF-QO activities. In addition, it may provide some insights to the relationship between native *E. coli* native acyl-CoA dehydrogenases (FadE and YdiO) and the three putative ETF and ETF-QO complexes (FixABCX, YdiQRST and YgcRQNO)[158, 230, 231].

Multiple ALE strategies were used to engineer strains to grow on methylsuccinate. Initial approach implemented plasmid-based expression of heterologous genes under the inducible promoter which mandated addition of antibiotics and inducers every passage to the fresh media. Main advantage with this approach is high expression levels of key pathway enzymes leveraging

high-copy plasmids and strong promoters like T7 promoter. Therefore, it would be useful for a metabolic pathway that requires high expression of certain gene as demonstrated in the recent study[14]. All enzymes in the methylsuccinate metabolism pathway expressed very well in *E. coli* (Supplementary Figure 4-1) and the enzymes in the itaconate degradation module (IDM) have decent kinetic parameters, and hence, genome integration under constitutive promoter was sufficient to engineer growth phenotype. In fact, genome integration under constitutive promoter reduced the lag phase and improved biomass yield of the evolved strain presumably by mitigating metabolic burden from maintaining plasmids and overexpressing genes under the strong T7 promoter (Figure 4-5). The mixed carbon substrate-based ALE was not successful with itaconate but successful with mesaconate as the co-carbon substrate. Based on the result, the key change was replacing YpIct with PaIct, which was shown to have higher activity not only with methylsuccinate but also with mesaconate (but not with itaconate). Even though YpIct has the advantage in accepting various CoA donors including acetyl-CoA[212], kinetic advantage of PaIct was apparently more crucial as the required succinyl-CoA flux for PaIct activity could possibly be enhanced via metabolic rewiring during the evolutionary process.

The engineered strain growing on methylsuccinate as carbon substrate can serve as a growth selection platform to screen potential fumarate-addition enzymes with methane (methylsuccinate synthase, MSS). A key advantage of the growth selection platform is its outstanding throughput[232], which could be particularly useful for multi-subunit enzymes like MSS that possess many potential sites for protein engineering increasing the library size. Recent advance in machine learning-powered protein structure modeling tools, such as AlphaFold[191], renders ability to construct the protein model and analyze the key residues with no crystal structure available. Library of mutants generated from alkylsuccinate synthases can then be tested using our

engineered strain as the growth selection platform. Upon successful construction of the full pathway starting from the MSS, our pathway holds a huge promise for biological utilization of methane as the two products from the pathway, pyruvate and acetyl-CoA are the key metabolic precursors for a myriad of bioproducts[233].

Methods:

Reagents

All chemicals were obtained from Fisher Scientific Co. and Sigma-Aldrich Co. unless otherwise specified. Primers were synthesized by Integrated DNA Technologies. Restriction enzymes were obtained from New England Biolabs unless otherwise specified.

Plasmids, strains and genetic methods

Plasmid-based gene expression was achieved by cloning the desired gene(s) into pETDuet-1, pCDFDuet-1, and pRSFDuet-1 (Novagen) digested with appropriate restriction enzymes and by using In-Fusion cloning technology (Clontech Laboratories, Inc.). Linear DNA fragments for insertion were created via PCR of the open reading frame of interest (for genes native to *E. coli*) or codon-optimized and synthesized by GeneArt (Thermo Fisher) (for genes non-native to *E. coli*). Resulting In-Fusion reaction products were used to transform *E. coli* Stellar cells (Clontech Laboratories, Inc.), and clones identified by PCR screening were further confirmed by DNA sequencing. Plasmids used in this study are listed in Supplementary Table 1.

Itaconate degradation module (IDM) was constructed in a low-copy plasmid (pZS backbone[234]) under constitutive M1-93 promoter[235] and B1002 terminator with RBS upstream of each gene designed to the maximum strength using RBS calculator[236]. Methylsuccinyl-CoA dehydrogenation module (MDM) was constructed in the pSL1521[237] under constitutive D/E20

promoter[238] and B1006 terminator with RBS upstream of each gene designed to the maximum strength using RBS calculator[236].

Genome integration of IDM was done using a CRISPR-Cas9-based system developed for *E. coli[181]* in the previously identified "safe site" (SS3) with high integration efficiency[239]. Genome integration of MDM was done using the INTEGRATE system in the lacZ locus described in the study[237]. pCas and pTargetF were gifts from Sheng Yang (Addgene plasmid # 62225 and 62226). pSL1521 (pSPIN, pSC101* backbone) was a gift from Samuel H. Sternberg (Addgene plasmid # 160729). Strains used in this study are listed in Supplementary Table 4-1.

Enzyme expression and purification

Enzymes with 6X n-terminal His-tag were cloned into vectors as described above, which were then transformed into *E. coli* BL21(DE3) for expression. Overnight cultures of the expression strains were grown in LB, which was used to inoculate 50 ml Terrific Broth (TB) medium at 1% and sealed with foam plugs filling the necks. After inoculation, flasks were incubated at 30°C and 250 rpm in an NBS C24 Benchtop Incubator Shaker (New Brunswick Scientific Co., Inc., Edison, NJ) until an optical density of ~0.3–0.5 was reached, at which point varying concentrations of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added. Flasks were incubated for a total of 24 hrs post-inoculation. Cells with overexpressed proteins were then pelleted and stored in -80 °C.

The frozen cell pellets were resuspended in cold lysis buffer (50 mM NaPi pH 7.4, 300 mM NaCl, 10 mM imidazole, 0.1% Triton-X 100) to an approximate OD600 of 40, to which 1 mg/mL of lysozyme and 250 U of Benzonase nuclease (Sigma) was added. The mixture was further treated by sonication on ice using a Branson Sonifier 250 (5 minutes with a 25% duty cycle and output control set at 3), and centrifuged at 7500×g for 15 minutes at 4°C. The supernatant was applied to a chromatography column containing 1 mL TALON metal affinity resin (Clontech Laboratories,

Inc., Mountain View, CA), which had been pre-equilibrated with the lysis buffer. The column was then washed first with 10 mL of the lysis buffer and then twice with 20 mL of wash buffer (50 mM NaPi pH 7.4, 300 mM NaCl, 20 mM imidazole). The his-tagged protein of interest was eluted with 1-2 applications of 4 mL elution buffer (50 mM NaPi pH 7.4, 300 mM NaCl, 250 mM imidazole). The eluate was collected and applied to a 10,000 MWCO Amicon ultrafiltration centrifugal device (Millipore, Billerica, MA), and the concentrate (~100 µL) was washed twice with 4 mL of 50 mM KPi pH 7.4 for desalting. Protein concentrations were measured by the Bradford assay. Purified protein was saved in 20 µL aliquots at -80°C until needed.

In vitro pathway testing using purified enzymes

Reactions were run in 1.5 mL Eppendorf tubes for HPLC assay or in 96-well plates for colorimetric assay. Reaction mixture contains 50 mM Tris-HCl (pH 7.4), 5 mM $MgCl_2$, 10 mM methylsuccinate and 1 mM CoA donor (succinyl-CoA, acetyl-CoA or free CoASH with ATP) and variable concentration of enzymes, unless otherwise indicated.

For FAD colorimetric assay with dehydrogenases (RsMcd/PaAcd) and ETF (RsETF), the same reaction mixture was used with addition of 1 µM CoA transferases and ligases, 5 µM RsMcd/PaAcd, and 5 µM RsETF. The absorption at 440 nm was measured at 30 °C using BioTek plate reader. Initial readings were measured for first 3 minutes, CoA donor or blank (water) was added and resumed reading until 20 minutes.

For the full methylsuccinate metabolism pathway prototyping, 1 µM CoA transferase (PaIct), 5 µM acyl-CoA dehydrogenase (PaAcd), 654 µM electron transfer flavoprotein (RsEtfAB), 200 µg/ml hydratase (PaIch) and 1000 µg/ml lyase (CaurMclA) were added in addition to the reaction mixture with methylsuccinate and succinyl-CoA. The reaction was started by the addition of succinyl-CoA. The reaction was stopped by addition of NaOH to hydrolyze CoA as needed, and a

saturated ammonium sulfate solution acidified with 1% sulfuric acid to precipitate proteins. The proteins were removed by centrifugation at 20000×g for 15 minutes. The supernatant was analyzed by HPLC.

In vivo pathway characterization in growing cell culture

Itaconate, mesaconate and/or methylsuccinate consumption in rich media was conducted using the minimal medium designed by Neidhardt et al.[240], modified to contain 125 mM MOPS and supplemented with indicated amounts of carbon source(s), 10 g/L tryptone, and 5 g/L yeast extract, 1.48 mM K_2HPO_4, 5 mM $(NH_4)_2SO_4$, 30 mM NH_4Cl, 2 mM $MgSO_4$ unless otherwise stated. A single colony of the desired strain was cultivated overnight (14-16 hrs) in LB medium with appropriate antibiotics and used as the inoculum (1%). Antibiotics were included when appropriate. Cultures were then incubated at 30°C and 200 rpm in an NBS I24 Benchtop Incubator Shaker (New Brunswick Scientific Co., Inc., Edison, NJ) until an OD600 of ~0.4 was reached, at which point appropriate amounts of IPTG was added. Samples were collected every 24 hours for HPLC analysis.

In vivo growth was conducted using the above MOPS media (containing IPTG if needed) without 10 g/L tryptone, and 5 g/L yeast extract supplementation. Pre-induced cells grown in the rich media were centrifuged (5000×g, 22°C), washed twice with the above minimal media without any carbon source, and used as inoculum to the fresh media with appropriate amounts of carbon source (e.g. itaconate, mesaconate and/or methylsuccinate) at initial OD600 of 0.05 in 5 mL culture in the 50 mL closed-cap conical tube (Genesee Scientific Co.) to prevent evaporation of the media. Flasks were incubated at 37°C and 200 rpm in an NBS I24 Benchtop Incubator Shaker (New Brunswick Scientific Co., Inc., Edison, NJ). Samples were collected for OD600 measurement and HPLC analysis as needed.

Adaptive laboratory evolution (ALE) of the strain growing on itaconate, mesaconate and/or methylsuccinate

Cells actively growing at exponential phase grown in the MOPS minimal media supplemented with varying concentrations of itaconate, mesaconate and/or methylsuccinate are sub-cultured to the fresh media after washing (if different carbon source or mixture) or without washing at the initial OD600 of 0.05. Every generation during the mesaconate and methylsuccinate mixed media ALE was streaked out on the MOPS with mesaconate or methylsuccinate agar plate for single colony, which was then grown in the liquid MOPS with mesaconate or methylsuccinate media for glycerol stock in -80°C.

Time profile assessment of strains growing on methylsuccinate

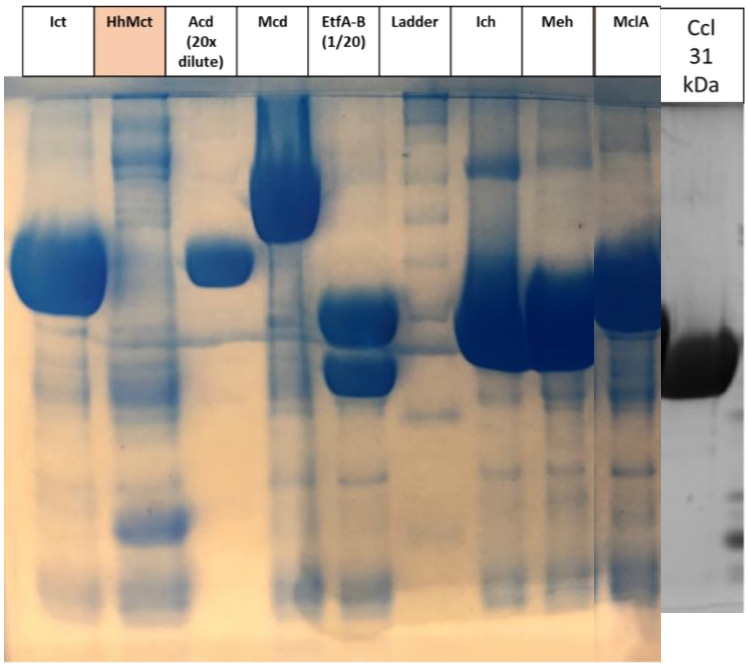
Single colonies from MOPS with 5 g/L methylsuccinate agar plate were inoculated to MOPS with 5 g/L methylsuccinate minimal media. Cells actively growing at exponential phase are sub-cultured to the fresh media at the initial OD600 of 0.01. Samples were taken every 24h after inoculation for OD600 measurement and HPLC analysis as described previously[25].

Supplementary Materials

Supplementary Table 4-2. Host strains and plasmids used in this study. Uniprot accession numbers for heterologous enzymes used in this work are given in parenthesis.

Host Strains/ Plasmids	Description/Genotype/Usage	Source
BL21(DE3)	*E. coli* B F⁻ *ompT gal dcm lon hsdS*$_B$($r_B^-m_B^-$) [*malB*$^+$]$_{K\text{-}12}$(λ^S) λ(DE3) - Host for protein expression for *in vitro* studies	Studier et al.[182]
MG1655 (DE3)	*E. coli* K-12 F- l- *ilvG- rfb*-50 *rph*-1 λ(DE3)	Blattner et al.[183]
MG1655 ΔTE	MG1655 λ(DE3) *ΔyciA*::FRT *ΔybgC*::FRT *ΔydiI*::FRT *ΔtesA*::FRT *ΔfadM*::FRT *ΔtesB*::FRT (*ΔfrmA ΔfdhF ΔfdnG ΔfdoG*)	Chou et al.[25]
MG1655 ΔTE2	MG1655 λ(DE3) *ΔyciA*::FRT *ΔybgC*::FRT *ΔtesA*::FRT *ΔfadM*::FRT *ΔtesB*::FRT	This study
MG1655 ΔTE2 + IDM (YpIct)	MG1655 ΔTE2 SS3::M193-YpIct-PaIch-CaurMclA	This study
MG1655 ΔTE2 + IDM (PaIct)	MG1655 ΔTE2 SS3::M193-PaIct-PaIch-CaurMclA - Evolved to grow on methysuccinate after ALE in mesaconate and methylsuccinate mixed carbon substrates	This study
MG1655 ΔTE2 + IDM (YpIct) + IDM (PaIct)	MG1655 ΔTE2 *ΔfadM*::M193-YpIct-PaIch-CaurMclA SS3::M193-PaIct-PaIch-CaurMclA	This study
MG1655 ΔTE2 + IDM (PaIct) + MDM	MG1655 ΔTE2 SS3::M193-PaIct-PaIch-CaurMclA *ΔlacZ*::D/E20-PaAcd-RsEtfA-B-EcEtfQO - Able to grow on methylsuccinate	This study
pCDFDuet-1	CloDF13, *lacI*, Sm^R	Novagen (Darmstadt, Germany)
pETDuet-1	pBR322-derived ColE1 origin, *lacI*, Amp^R	Novagen (Darmstadt, Germany)
pRSFDuet-1	pBR322-derived ColE1 origin, *lacI*, Kan^R	Novagen (Darmstadt, Germany)
pCDFDuet-p1-H6-PaAcd	pCDF carrying His-tagged putative acyl-CoA dehydrogenase (*P. aeruginosa*) (Uniprot: Q9I566)	This study
pCDFDuet-p1-H6-PaIct	pCDF carrying His-tagged Ict (*P. aeruginosa*) (Uniprot: Q9I563)	This study
pCDFDuet-p1-H6-YpIct	pCDF carrying His-tagged Ict (*Y. pestis*) (Uniprot: Q9ZC36)	This study
pCDFDuet-p1-H6-RsEtfA-B	pCDF carrying His-tagged EtfA-B (*R. sphaeroides*) (Uniprot: Q9AQC3, Q9AQC4)	This study
pCDFDuet-p1-H6-RsMcd	pCDF carrying His-tagged Mcd (*R. sphaeroides*) (Uniprot: D3JV03)	This study

pCDFDuet-p1-H6-HhMct	pCDF carrying His-tagged Mct (*H. hispanica*) (Uniprot: G0HQ31)	This study
pCDFDuet-p1-H6-EcSucCD	pCDF carrying His-tagged SucCD (*E. coli*)	This study
pCDFDuet-p1-H6-TaBbsEF	pCDF carrying His-tagged BbsEF (*T. aromatica*) (Uniprot: Q9KJF0, Q9KJE9)	This study
pCDFDuet-p1-H6-Tfu2526-2527	pCDF carrying His-tagged Tfu2576-2577 (*T. fusca*) (Uniprot: Q47LR3, Q47LR2)	This study
pCDFDuet-p1-H6-PaIch	pCDF carrying His-tagged Ich (*P. aeruginosa*) (Uniprot: Q9I567)	This study
pCDFDuet-p1-H6-PaCcl	pCDF carrying His-tagged Ccl (*P. aeruginosa*) (Uniprot: Q9I562)	This study
pCDFDuet-p1-H6-CaurMeh	pCDF carrying His-tagged Meh (*C. aurantiacus*) (Uniprot: A9WC41)	This study
pCDFDuet-p1-H6-CaurMclA	pCDF carrying His-tagged MclA (*C. aurantiacus*) (Uniprot: A9WC35)	This study
pCDFDuet-p1-PaIct-p2-CaurMclA-PaIch	Expression vector harboring itaconate degradation module (PaIct) for in vivo testing	This study
pCDFDuet-p1-YpIct-p2-CaurMclA-PaIch	Expression vector harboring itaconate degradation module (PaIct) for in vivo testing	This study
pRSFDuet-P1-EcDctA	Expression vector harboring *E. coli* DctA for in vivo testing	This study
pETDuet-p1-PaAcd-p2-RsEtfA-B	Expression vector harboring methylsuccinyl-CoA dehydrogenation module (without ETF-QO) for in vivo testing	This study
pETDuet-p1-PaAcd-p2-RsEtfA-B-RsEtfQO	Expression vector harboring methylsuccinyl-CoA dehydrogenation module (with RsETF-QO) for in vivo testing	This study



Supplementary Figure 4-8. Heterologous expression of synthetic methylsuccinate metabolism pathway enzymes. Concentrations of purified enyzme and corresponding molecular weights are summarizied in the table on the left and SDS-PAGE for purified enzymes are shown on the right. HhMct is not purified possibly due to misfolding.

General discussion and future work

In this thesis, two synthetic metabolic pathways are discussed: FORCE pathways utilizing activated C1, formyl-CoA as a C1 elongation unit, and methylsuccinate metabolism centered around new-to-nature methane activation via fumarate addition. Although the two pathways involve completely different enzymes and intermediates, common strategies were employed to address the problems and the approaches used from either project can be applied to the other project as a future work. Both pathways started from enzymes, 2-hydroxyacyl-CoA synthase (HACS) and methylsuccinate synthase (MSS), that were proposed to catalyze C-C coupling reactions involving C1 compound or moiety, based on new-to-nature biochemistries inspired by known enzyme activities with longer chain substrates. In case of HACS, not only the reversibility of 2-hydroxyacyl-CoA lyase (HACL) but also its utilization of C1 aldehyde, formaldehyde, as substrate were verified from initial testing of a variant from human HACL[68]. On the other hand, functional expression of a (1-methylalkyl)succinate synthase (MAS) variant in *E. coli* was demonstrated with activities with alkanes down to C3, but no activity on methane or ethane was observed[108]. Therefore, the pathway downstream of C1 activation product, methylsuccinate became the initial focus for this project because no reported pathways exist for methylsuccinate metabolism.

In both studies, we initially validated individual enzyme activities in vitro using purified enzymes followed by cell-free prototyping of the full pathways for product synthesis. As FORCE pathways are designed to generate diverse products independently from the host metabolism, we focused on showing synthesis of different products from various C1 compounds. Leveraging the orthogonality of the pathways, we were able to show product synthesis in diverse platforms including resting cells and growing cells in vivo. FORCE pathway products, such as glycolate, can

be utilized as native carbon source for growth in *E. coil* and the possibility of synthetic C1-trophy was demonstrated via two-strain co-culture system. Methylsuccinate metabolism directly feeds its product into the host central metabolism allowing it as a platform for growth as well. Using proxy metabolite (itaconate) and pathway intermediate (mesaconate) as sole carbon source in minimal media, thereby imposing selective pressure on cells to utilize the non-native substrates, enabled identification of bottlenecks in the pathway and addressing the issues via ALE. Although the same strategy cannot be used for FORCE pathway using formaldehyde as substrate due to its toxicity, methanol or formate could be used as C1 substrates to further improve FORCE pathways via ALE, ultimately leading to synthetic C1-trophy as demonstrated in different C1 utilization pathways[13-15].

Finding HACS variants with improved kinetics for formaldehyde-formyl-CoA condensation led to significant improvement in the FORCE pathway flux. Combinatorial approach of HACS homolog library screening and rational protein engineering guided by sequence analysis of active variants led to identification of variants with more than an order of magnitude better activities. Similar approach can be taken for identification of MSS. The key difference is that HACS bioprospecting started with initial reference that already exhibited activities with C1 substrate, whereas in case of MSS, there is no reported enzyme with methane activation activity. This will likely require more protein engineering in the process, both rational and random, targeting active site residues of a MAS variant that shows activities with short-chain (C3-C4) alkanes. MAS from *Azoarcus sp.* Strain HxN1, reported to be expressed in *E. coli* and have activities with short-chain alkanes[108] could be a good starting point. MSS/MAS is a multi-subunit enzyme with little known functions of each subunit, which will expand the sequence space to be targeted for mutagenesis. This will lead to increase in the size of mutant library to be screened

requiring high throughput screening method, where the engineered strain growing efficiently on methylsuccinate can be used as the selection platform. Approaches and tools discussed in HACS bioprospecting and engineering, such as AlphaFold, can be useful in engineering MSS variants with no crystal structure available.

Both FORCE pathways and methylsuccinate utilization pathway are great platforms for diverse product synthesis. Utilization of C1 molecules for synthesis of diverse C2 and C3 products including glycolate, ethylene glycol, ethanol and glycerate is already demonstrated using FORCE pathways, and it can be further extended to longer chain molecules either by operating multiple iterations of the pathway or by utilizing longer chain carbonyl as co-substrates for condensation with formyl-CoA. RuHACL, for example, showed improved kinetic properties toward C2 and C3 aldehydes[68] and AcHACL activity for condensation of acetone and formyl-CoA to generate 2-hydroxyisobutyryl-CoA was also reported[179]. Utilization of various carbonyl compounds for HACS condensation followed by α-reduction pathways can greatly expand the product profile of the FORCE pathways. In case of methylsuccinate metabolism, the two products are pyruvate and acetyl-CoA, key metabolic precursors for numerous bioproduct synthesis demonstrated in *E. coli*[233]. Moreover, methane activation via fumarate addition retains the reducing power possessed by methane unlike oxygen-dependent oxidation catalyzed by methane monooxygenases (MMO). This means the additional reducing power can be supplemented in the downstream pathways for generating reduced, energy-rich products at high yields. These products with high degree of reduction, such as alcohols and hydrocarbons, can be readily used as fuels in heavy-duty transportation and aviation, which are hard to be replaced by sustainable electricity.

References

1. Nolan, C., et al., *Past and future global transformation of terrestrial ecosystems under climate change.* Science, 2018. **361**(6405): p. 920-923.
2. van den Bergh, J.C.J.M. and W.J.W. Botzen, *A lower bound to the social cost of CO2 emissions.* Nature Climate Change, 2014. **4**(4): p. 253-258.
3. Adger, W.N., et al., *Cultural dimensions of climate change impacts and adaptation.* Nature Climate Change, 2012. **3**(2): p. 112-117.
4. Diaz, D. and F. Moore, *Quantifying the economic risks of climate change.* Nature Climate Change, 2017. **7**(11): p. 774-782.
5. Global, E., *Outlook (2021),"Accelerating ambitions despite the pandemic".* International Energy Agency (IEA), April, 2021.
6. Jullesson, D., et al., *Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals.* Biotechnology Advances, 2015. **33**(7): p. 1395-1402.
7. Nielsen, J., C.B. Tillegreen, and D. Petranovic, *Innovation trends in industrial biotechnology.* Trends Biotechnol, 2022. **40**(10): p. 1160-1172.
8. Kemf, E., *GCO--Global Chemicals Outlook: Towards Sound Management of Chemicals*. 2013: United Nations Environment Programme.
9. Liew, F.E., et al., *Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale.* Nature Biotechnology, 2022. **40**(3): p. 335-344.
10. Sullivan, K.P., et al., *Mixed plastics waste valorization through tandem chemical oxidation and biological funneling.* Science, 2022. **378**(6616): p. 207-211.
11. Hann, E.C., et al., *A hybrid inorganic–biological artificial photosynthesis system for energy-efficient food production.* Nature Food, 2022. **3**(6): p. 461-471.

12. Claassens, N.J., et al., *Making quantitative sense of electromicrobial production.* Nature Catalysis, 2019. **2**(5): p. 437-447.

13. Chen, F.Y., et al., *Converting Escherichia coli to a Synthetic Methylotroph Growing Solely on Methanol.* Cell, 2020. **182**(4): p. 933-946 e14.

14. Keller, P., et al., *Generation of an Escherichia coli strain growing on methanol via the ribulose monophosphate cycle.* Nat Commun, 2022. **13**(1): p. 5243.

15. Kim, S., et al., *Growth of E. coli on formate and methanol via the reductive glycine pathway.* Nat Chem Biol, 2020. **16**(5): p. 538-545.

16. Sirajuddin, S. and A.C. Rosenzweig, *Enzymatic oxidation of methane.* Biochemistry, 2015. **54**(14): p. 2283-94.

17. Kwon, M., A. Ho, and S. Yoon, *Novel approaches and reasons to isolate methanotrophic bacteria with biotechnological potentials: recent achievements and perspectives.* Appl Microbiol Biotechnol, 2019. **103**(1): p. 1-8.

18. Ge, X., et al., *Biological conversion of methane to liquid fuels: status and opportunities.* Biotechnol Adv, 2014. **32**(8): p. 1460-75.

19. Atkinson, J.T., et al., *Cellular Assays for Ferredoxins: A Strategy for Understanding Electron Flow through Protein Carriers That Link Metabolic Pathways.* Biochemistry, 2016. **55**(51): p. 7047-7064.

20. Haynes, C.A. and R. Gonzalez, *Rethinking biological activation of methane and conversion to liquid fuels.* Nat Chem Biol, 2014. **10**(5): p. 331-9.

21. Kim, H.J., et al., *Biological conversion of methane to methanol through genetic reassembly of native catalytic domains.* Nature Catalysis, 2019. **2**(4): p. 342-353.

22. Bennett, R.K., et al., *Expression of soluble methane monooxygenase in Escherichia coli enables methane conversion.* bioRxiv, 2021: p. 2021.08.05.455234.

23. Le, T.K., et al., *Methanol Dehydrogenases as a Key Biocatalysts for Synthetic Methylotrophy.* Front Bioeng Biotechnol, 2021. **9**: p. 787791.

24. Whitaker, W.B., et al., *Synthetic methylotrophy: engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization.* Curr Opin Biotechnol, 2015. **33**: p. 165-75.

25. Chou, A., et al., *An orthogonal metabolic framework for one-carbon utilization.* Nat Metab, 2021. **3**(10): p. 1385-1399.

26. Bozdag, A., C. Komives, and M.C. Flickinger, *Growth of Bacillus methanolicus in 2 M methanol at 50 °C: the effect of high methanol concentration on gene regulation of enzymes involved in formaldehyde detoxification by the ribulose monophosphate pathway.* J Ind Microbiol Biotechnol, 2015. **42**(7): p. 1027-38.

27. Wu, T.Y., et al., *Characterization and evolution of an activator-independent methanol dehydrogenase from Cupriavidus necator N-1.* Appl Microbiol Biotechnol, 2016. **100**(11): p. 4969-83.

28. Lee, J.Y., et al., *Discovery and Biochemical Characterization of a Methanol Dehydrogenase From Lysinibacillus xylanilyticus.* Front Bioeng Biotechnol, 2020. **8**: p. 67.

29. Good, N.M., et al., *Contrasting in vitro and in vivo methanol oxidation activities of lanthanide-dependent alcohol dehydrogenases XoxF1 and ExaF from Methylobacterium extorquens AM1.* Sci Rep, 2019. **9**(1): p. 4248.

30. Koch, C., et al., *Crystal Structure of Alcohol Oxidase from Pichia pastoris.* PLoS One, 2016. **11**(2): p. e0149846.

31. Dai, Z., et al., *Metabolic construction strategies for direct methanol utilization in Saccharomyces cerevisiae.* Bioresour Technol, 2017. **245**(Pt B): p. 1407-1412.

32. Wang, G., et al., *Engineering Yeast Yarrowia lipolytica for Methanol Assimilation.* ACS Synthetic Biology, 2021.

33. Cai, T., et al., *Cell-free chemoenzymatic starch synthesis from carbon dioxide.* Science, 2021. **373**(6562): p. 1523-1527.

34. Gonzalez, C.F., et al., *Molecular basis of formaldehyde detoxification. Characterization of two S-formylglutathione hydrolases from Escherichia coli, FrmB and YeiG.* J Biol Chem, 2006. **281**(20): p. 14514-22.

35. Fernández, M.R., et al., *Formaldehyde dehydrogenase from yeast and plant. Implications for the general functional and structural significance of class III alcohol dehydrogenase.* Adv Exp Med Biol, 1997. **414**: p. 373-81.

36. Newton, G.L. and R.C. Fahey, *Mycothiol biochemistry.* Arch Microbiol, 2002. **178**(6): p. 388-94.

37. Newton, G.L., et al., *Bacillithiol is an antioxidant thiol produced in Bacilli.* Nat Chem Biol, 2009. **5**(9): p. 625-7.

38. Liao, Y., et al., *Structure of formaldehyde dehydrogenase from Pseudomonas aeruginosa: the binary complex with the cofactor NAD+.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2013. **69**(Pt 9): p. 967-72.

39. Karzanov, V.V., et al., *Evidence for the presence of a new NAD+-dependent formate dehydrogenase in Pseudomonas sp. 101 cells grown on a molybdenum-containing medium.* FEMS Microbiol Lett, 1989. **51**(1): p. 197-200.

40. Andreesen, J.R. and L.G. Ljungdahl, *Nicotinamide adenine dinucleotide phosphate-dependent formate dehydrogenase from Clostridium thermoaceticum: purification and properties.* J Bacteriol, 1974. **120**(1): p. 6-14.

41. Wang, S., et al., *Clostridium acidurici electron-bifurcating formate dehydrogenase.* Appl Environ Microbiol, 2013. **79**(19): p. 6176-9.

42. Jormakka, M., et al., *Molecular Basis of Proton Motive Force Generation: Structure of Formate Dehydrogenase-N.* Science, 2002. **295**(5561): p. 1863-1868.

43. da Silva, S.M., I. Pacheco, and I.A. Pereira, *Electron transfer between periplasmic formate dehydrogenase and cytochromes c in Desulfovibrio desulfuricans ATCC 27774.* J Biol Inorg Chem, 2012. **17**(5): p. 831-8.

44. Schauer, N.L., et al., *Mechanistic studies of the coenzyme F420 reducing formate dehydrogenase from Methanobacterium formicicum.* Biochemistry, 1986. **25**(22): p. 7163-8.

45. Axley, M.J., D.A. Grahame, and T.C. Stadtman, *Escherichia coli formate-hydrogen lyase. Purification and properties of the selenium-dependent formate dehydrogenase component.* J Biol Chem, 1990. **265**(30): p. 18213-8.

46. Schuchmann, K. and V. Muller, *Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria.* Nat Rev Microbiol, 2014. **12**(12): p. 809-21.

47. Deppenmeier, U., *Redox-driven proton translocation in methanogenic Archaea.* Cell Mol Life Sci, 2002. **59**(9): p. 1513-33.

48. Timmers, P.H., et al., *Reverse Methanogenesis and Respiration in Methanotrophic Archaea.* Archaea, 2017. **2017**: p. 1654237.

49. Chistoserdova, L.V. and M.E. Lidstrom, *Genetics of the serine cycle in Methylobacterium extorquens AM1: identification of sgaA and mtdA and sequences of sgaA, hprA, and mtdA.* Journal of Bacteriology, 1994. **176**(7): p. 1957-1968.

50. Claassens, N.J., *Reductive Glycine Pathway: A Versatile Route for One-Carbon Biotech.* Trends Biotechnol, 2021. **39**(4): p. 327-329.

51. Yan, Z. and J.G. Ferry, *Electron Bifurcation and Confurcation in Methanogenesis and Reverse Methanogenesis.* Front Microbiol, 2018. **9**: p. 1322.

52. Boetius, A., et al., *A marine microbial consortium apparently mediating anaerobic oxidation of methane.* Nature, 2000. **407**(6804): p. 623-626.

53. Scheller, S., et al., *The key nickel enzyme of methanogenesis catalyses the anaerobic oxidation of methane.* Nature, 2010. **465**(7298): p. 606-8.

54. Soo, V.W., et al., *Reversing methanogenesis to capture methane for liquid biofuel precursors.* Microb Cell Fact, 2016. **15**: p. 11.

55. Moore, S.J., et al., *Elucidation of the biosynthesis of the methane catalyst coenzyme F430.* Nature, 2017. **543**(7643): p. 78-82.

56. Balch, W.E. and R.S. Wolfe, *Specificity and biological distribution of coenzyme M (2-mercaptoethanesulfonic acid).* J Bacteriol, 1979. **137**(1): p. 256-63.

57. Krishnakumar, A.M., et al., *Getting a handle on the role of coenzyme M in alkene metabolism.* Microbiol Mol Biol Rev, 2008. **72**(3): p. 445-56.

58. Partovi, S.E., et al., *Coenzyme M biosynthesis in bacteria involves phosphate elimination by a functionally distinct member of the aspartase/fumarase superfamily.* J Biol Chem, 2018. **293**(14): p. 5236-5246.

59. Kremp, F., et al., *Methanol metabolism in the acetogenic bacterium Acetobacterium woodii.* Environ Microbiol, 2018. **20**(12): p. 4369-4384.

60. Kremp, F. and V. Muller, *Methanol and methyl group conversion in acetogenic bacteria: biochemistry, physiology and application.* FEMS Microbiol Rev, 2021. **45**(2).

61. Flaiz, M., et al., *Production of the biocommodities butanol and acetone from methanol with fluorescent FAST-tagged proteins using metabolically engineered strains of Eubacterium limosum.* Biotechnol Biofuels, 2021. **14**(1): p. 117.

62. Kallen, R.G. and W.P. Jencks, *The Mechanism of the Condensation of Formaldehyde with Tetrahydrofolic Acid.* Journal of Biological Chemistry, 1966. **241**(24): p. 5851-5863.

63. Escalante-Semerena, J.C., K.L. Rinehart, Jr., and R.S. Wolfe, *Tetrahydromethanopterin, a carbon carrier in methanogenesis.* J Biol Chem, 1984. **259**(15): p. 9447-55.

64. Vorholt, J.A., et al., *Novel Formaldehyde-Activating Enzyme in Methylobacterium extorquens AM1 Required for Growth on Methanol.* Journal of Bacteriology, 2000. **182**(23): p. 6645-6650.

65. Chistoserdova, L., M.G. Kalyuzhnaya, and M.E. Lidstrom, *The expanding world of methylotrophic metabolism.* Annu Rev Microbiol, 2009. **63**: p. 477-99.

66. Yishai, O., et al., *In Vivo Assimilation of One-Carbon via a Synthetic Reductive Glycine Pathway in Escherichia coli.* ACS Synth Biol, 2018. **7**(9): p. 2023-2028.

67. Yu, H. and J.C. Liao, *A modified serine cycle in Escherichia coli coverts methanol and CO2 to two-carbon compounds.* Nat Commun, 2018. **9**(1): p. 3992.

68. Chou, A., et al., *2-Hydroxyacyl-CoA lyase catalyzes acyloin condensation for one-carbon bioconversion.* Nat Chem Biol, 2019. **15**(9): p. 900-906.

69. Wang, J., et al., *Enzyme engineering and in vivo testing of a formate reduction pathway.* Synth Biol (Oxf), 2021. **6**(1): p. ysab020.

70. Yishai, O., et al., *Engineered Assimilation of Exogenous and Endogenous Formate in Escherichia coli.* ACS Synth Biol, 2017. **6**(9): p. 1722-1731.

71. Marx, C.J., et al., *Purification of the formate-tetrahydrofolate ligase from Methylobacterium extorquens AM1 and demonstration of its requirement for methylotrophic growth.* J Bacteriol, 2003. **185**(24): p. 7169-75.

72. Berthold, C.L., et al., *Reinvestigation of the catalytic mechanism of formyl-CoA transferase, a class III CoA-transferase.* J Biol Chem, 2008. **283**(10): p. 6519-29.

73. Nattermann, M., et al., *Engineering a Highly Efficient Carboligase for Synthetic One-Carbon Metabolism.* ACS Catal, 2021. **11**(9): p. 5396-5404.

74. Sly, W.S. and E.R. Stadtman, *FORMATE METABOLISM. II. ENZYMATIC SYNTHESIS OF FORMYL PHOSPHATE AND FORMYL COENZYME A IN CLOSTRIDIUM CYLINDROSPORUM.* J Biol Chem, 1963. **238**: p. 2639-47.

75. Starai, V.J., J.G. Gardner, and J.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, 2005. **280**(28): p. 26200-5.

76. Watanabe, T., et al., *Three-megadalton complex of methanogenic electron-bifurcating and CO₂-fixing enzymes.* Science, 2021. **373**(6559): p. 1151-1156.

77. de Poorter, L.M., et al., *Bioenergetics of the formyl-methanofuran dehydrogenase and heterodisulfide reductase reactions in Methanothermobacter thermautotrophicus.* Eur J Biochem, 2003. **270**(1): p. 66-75.

78. Bernhardsgrutter, I., et al., *CO2-converting enzymes for sustainable biotechnology: from mechanisms to application.* Curr Opin Biotechnol, 2021. **67**: p. 80-87.

79. Erb, T.J., *Carboxylases in natural and synthetic microbial pathways.* Appl Environ Microbiol, 2011. **77**(24): p. 8466-77.

80. Adam, P.S., G. Borrel, and S. Gribaldo, *Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes.* Proc Natl Acad Sci U S A, 2018. **115**(6): p. E1166-E1173.

81. Riordan, C.G., *Synthetic chemistry and chemical precedents for understanding the structure and function of acetyl coenzyme A synthase.* J Biol Inorg Chem, 2004. **9**(5): p. 542-9.

82. Gencic, S., E.C. Duin, and D.A. Grahame, *Tight coupling of partial reactions in the acetyl-CoA decarbonylase/synthase (ACDS) multienzyme complex from Methanosarcina thermophila: acetyl C-C bond fragmentation at the a cluster promoted by protein conformational changes.* J Biol Chem, 2010. **285**(20): p. 15450-15463.

83. Grahame, D.A., S. Gencic, and E. DeMoll, *A single operon-encoded form of the acetyl-CoA decarbonylase/synthase multienzyme complex responsible for synthesis and cleavage of acetyl-CoA in Methanosarcina thermophila.* Arch Microbiol, 2005. **184**(1): p. 32-40.

84. Ragsdale, S.W., *Enzymology of the wood-Ljungdahl pathway of acetogenesis.* Ann N Y Acad Sci, 2008. **1125**: p. 129-36.

85. Bar-Even, A., *Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways.* Biochemistry, 2016. **55**(28): p. 3851-63.

86. Wenk, S., et al., *Synthetic carbon fixation via the autocatalytic serine threonine cycle.* bioRxiv, 2022: p. 2022.09.28.509898.

87. Rozova, O.N., et al., *Characterization of Two Recombinant 3-Hexulose-6-Phosphate Synthases from the Halotolerant Obligate Methanotroph Methylomicrobium alcaliphilum 20Z.* Biochemistry (Mosc), 2017. **82**(2): p. 176-185.

88. Nobuo, K., et al., *Purification and properties of a transketolase responsible for formaldehyde fixation in a methanol-utilizing yeast, Candida boidinii (Kloeckera sp.) No. 2201.* Biochimica et Biophysica Acta (BBA) - General Subjects, 1982. **715**(2): p. 143-150.

89. Contestabile, R., et al., *<scp>l</scp>-Threonine aldolase, serine hydroxymethyltransferase and fungal alanine racemase.* European Journal of Biochemistry, 2001. **268**(24): p. 6508-6525.

90. Hernandez, K., et al., *Combining Aldolases and Transaminases for the Synthesis of 2-Amino-4-hydroxybutanoic Acid.* ACS Catalysis, 2017. **7**(3): p. 1707-1711.

91. Wang, C., et al., *An Aldolase-Catalyzed New Metabolic Pathway for the Assimilation of Formaldehyde and Methanol To Synthesize 2-Keto-4-hydroxybutyrate and 1,3-Propanediol in Escherichia coli.* ACS Synth Biol, 2019. **8**(11): p. 2483-2493.

92. Meng, H., et al., *An Aldolase-Based New Pathway for Bioconversion of Formaldehyde and Ethanol into 1,3-Propanediol in Escherichia coli.* ACS Synth Biol, 2021. **10**(4): p. 799-809.

93. Siegel, J.B., et al., *Computational protein design enables a novel one-carbon assimilation pathway.* Proc Natl Acad Sci U S A, 2015. **112**(12): p. 3704-9.

94. Li, T., et al., *Totally atom-economical synthesis of lactic acid from formaldehyde: combined bio-carboligation and chemo-rearrangement without the isolation of intermediates.* Green Chemistry, 2020. **22**(20): p. 6809-6814.

95. Hu, G., et al., *Light-driven CO2 sequestration in Escherichia coli to achieve theoretical yield of chemicals.* Nature Catalysis, 2021. **4**(5): p. 395-406.

96. Massad, N. and S. Banta, *Development of a kinetic model and figures of merit for formaldehyde carboligations catalyzed by formolase enzymes.* Biotechnol Bioeng, 2022.

97. Lu, X., et al., *Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design.* Nat Commun, 2019. **10**(1): p. 1378.

98. Jo, H.-J., et al., *Glyoxylate Carboligase-based Whole-Cell Biotransformation of Formaldehyde into Ethylene Glycol via Glycolaldehyde.* Green Chemistry, 2021.

99. Shisler, K.A. and J.B. Broderick, *Glycyl radical activating enzymes: structure, mechanism, and substrate interactions.* Arch Biochem Biophys, 2014. **546**: p. 64-71.

100. Backman, L.R.F., et al., *New tricks for the glycyl radical enzyme family.* Crit Rev Biochem Mol Biol, 2017. **52**(6): p. 674-695.

101. Hesslinger, C., S.A. Fairhurst, and G. Sawers, *Novel keto acid formate-lyase and propionate kinase enzymes are components of an anaerobic pathway in Escherichia coli that degrades L-threonine to propionate.* Mol Microbiol, 1998. **27**(2): p. 477-92.

102. Zelcbuch, L., et al., *Pyruvate Formate-Lyase Enables Efficient Growth of Escherichia coli on Acetate and Formate.* Biochemistry, 2016. **55**(17): p. 2423-6.

103. Kirst, H., et al., *Toward a glycyl radical enzyme containing synthetic bacterial microcompartment to produce pyruvate from formate and acetate.* Proceedings of the National Academy of Sciences, 2022. **119**(8).

104. Kniemeyer, O., et al., *Anaerobic oxidation of short-chain hydrocarbons by marine sulphate-reducing bacteria.* Nature, 2007. **449**(7164): p. 898-901.

105. Beasley, K.K. and M.A. Nanny, *Potential energy surface for anaerobic oxidation of methane via fumarate addition.* Environ Sci Technol, 2012. **46**(15): p. 8244-52.

106. Callaghan, A.V., *Enzymes involved in the anaerobic oxidation of n-alkanes: from methane to long-chain paraffins.* Front Microbiol, 2013. **4**: p. 89.

107. Thauer, R.K. and S. Shima, *Methane as fuel for anaerobic microorganisms.* Ann N Y Acad Sci, 2008. **1125**: p. 158-70.

108. Wang, Y., et al., *Engineering Escherichia coli for anaerobic alkane activation: Biosynthesis of (1-methylalkyl)succinates.* Biotechnol Bioeng, 2021.

109. Qiao, W., et al., *Challenges and opportunities in C1-based biomanufacturing.* Bioresource Technology, 2022. **364**.

110. Tuyishime, P. and J.P. Sinumvayo, *Novel outlook in engineering synthetic methylotrophs and formatotrophs: a course for advancing C1-based chemicals production.* World J Microbiol Biotechnol, 2020. **36**(8): p. 118.

111. Barenholz, U., et al., *Design principles of autocatalytic cycles constrain enzyme kinetics and force low substrate saturation at flux branch points.* Elife, 2017. **6**.

112. Alfarouk, K.O., et al., *The Pentose Phosphate Pathway Dynamics in Cancer and Its Dependency on Intracellular pH.* Metabolites, 2020. **10**(7): p. 285.

113. Garg, S., J.M. Clomburg, and R. Gonzalez, *A modular approach for high-flux lactic acid production from methane in an industrial medium using engineered Methylomicrobium buryatense 5GB1.* J Ind Microbiol Biotechnol, 2018. **45**(6): p. 379-391.

114. Garg, S., et al., *Bioconversion of methane to C-4 carboxylic acids using carbon flux through acetyl-CoA in engineered Methylomicrobium buryatense 5GB1C.* Metab Eng, 2018. **48**: p. 175-183.

115. Muller, J.E., et al., *Methylotrophy in the thermophilic Bacillus methanolicus, basic insights and application for commodity production from methanol.* Appl Microbiol Biotechnol, 2015. **99**(2): p. 535-51.

116. Chen, C.T., et al., *Synthetic methanol auxotrophy of Escherichia coli for methanol-dependent growth and production.* Metab Eng, 2018. **49**: p. 257-266.

117. Keller, P., et al., *Methanol-dependent Escherichia coli strains with a complete ribulose monophosphate cycle.* Nat Commun, 2020. **11**(1): p. 5403.

118. He, H., et al., *Ribulose Monophosphate Shunt Provides Nearly All Biomass and Energy Required for Growth of E. coli.* ACS Synth Biol, 2018. **7**(6): p. 1601-1611.

119. Sanford, P.A. and B.M. Woolston, *Synthetic or natural? Metabolic engineering for assimilation and valorization of methanol.* Current Opinion in Biotechnology, 2022. **74**: p. 171-179.

120. Yurimoto, H., M. Oku, and Y. Sakai, *Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis.* Int J Microbiol, 2011. **2011**: p. 101298.

121. Krainer, F.W., et al., *Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway.* Microbial Cell Factories, 2012. **11**(1): p. 22.

122. Gao, J., et al., *Rescuing yeast from cell death enables overproduction of fatty acids from sole methanol.* Nature Metabolism, 2022.

123. Cai, P., et al., *Methanol biotransformation toward high-level production of fatty acid derivatives by engineering the industrial yeast Pichia pastoris.* Proceedings of the National Academy of Sciences, 2022. **119**(29): p. e2201711119.

124. Cai, P., et al., *Microbial synthesis of long-chain α-alkenes from methanol by engineering Pichia pastoris.* Bioresources and Bioprocessing, 2022. **9**(1).

125. Espinosa, M.I., et al., *Adaptive laboratory evolution of native methanol assimilation in Saccharomyces cerevisiae.* Nat Commun, 2020. **11**(1): p. 5564.

126. Bogorad, I.W., et al., *Building carbon-carbon bonds using a biocatalytic methanol condensation cycle.* Proc Natl Acad Sci U S A, 2014. **111**(45): p. 15928-33.

127. Crowther, G.J., G. Kosaly, and M.E. Lidstrom, *Formate as the main branch point for methylotrophic metabolism in Methylobacterium extorquens AM1*. J Bacteriol, 2008. **190**(14): p. 5057-62.

128. Šmejkalová, H., T.J. Erb, and G. Fuchs, *Methanol Assimilation in Methylobacterium extorquens AM1: Demonstration of All Enzymes and Their Regulation.* PLoS ONE, 2010. **5**(10): p. e13001.

129. Liang, W.F., et al., *Biosensor-assisted transcriptional regulator engineering for Methylobacterium extorquens AM1 to improve mevalonate synthesis by increasing the acetyl-CoA supply.* Metab Eng, 2017. **39**: p. 159-168.

130. He, H., et al., *An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in E. coli.* Metab Eng, 2020. **60**: p. 1-13.

131. Kim, S., et al., *Optimizing E. coli as a formatotrophic platform for bioproduction via the reductive glycine pathway.* bioRxiv, 2022: p. 2022.08.23.504942.

132. Claassens, N.J., et al., *Replacing the Calvin cycle with the reductive glycine pathway in Cupriavidus necator.* Metab Eng, 2020. **62**: p. 30-41.

133. Sanchez-Andrea, I., et al., *The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans.* Nat Commun, 2020. **11**(1): p. 5090.

134. Cotton, C.A., et al., *Renewable methanol and formate as microbial feedstocks.* Curr Opin Biotechnol, 2020. **62**: p. 168-180.

135. Vogeli, B., et al., *Cell-free prototyping enables implementation of optimized reverse beta-oxidation pathways in heterotrophic and autotrophic bacteria.* Nat Commun, 2022. **13**(1): p. 3058.

136. Mao, Y., et al., *Non-natural Aldol Reactions Enable the Design and Construction of Novel One-Carbon Assimilation Pathways in vitro.* Front Microbiol, 2021. **12**: p. 677596.

137. Güner, S., et al., *Design of a synthetic enzyme cascade for the in vitro fixation of a C1 carbon source to a functional C4 sugar.* Green Chemistry, 2021. **23**(17): p. 6583-6590.

138. Nielsen, J. and J.D. Keasling, *Engineering Cellular Metabolism.* Cell, 2016. **164**(6): p. 1185-1197.

139. Bang, J., et al., *Escherichia coli is engineered to grow on CO2 and formic acid.* Nat Microbiol, 2020. **5**(12): p. 1459-1463.

140. Liew, F., et al., *Gas Fermentation-A Flexible Platform for Commercial Scale Production of Low-Carbon-Fuels and Chemicals from Waste and Renewable Feedstocks.* Front Microbiol, 2016. **7**: p. 694.

141. Black, W.B., et al., *Engineering a nicotinamide mononucleotide redox cofactor system for biocatalysis.* Nat Chem Biol, 2020. **16**(1): p. 87-94.

142. Swartz, J.R., *Expanding biological applications using cell-free metabolic engineering: An overview.* Metab Eng, 2018. **50**: p. 156-172.

143. Huttanus, H.M. and X. Feng, *Compartmentalized metabolic engineering for biochemical and biofuel production.* Biotechnol J, 2017. **12**(6).

144. Elani, Y., *Interfacing Living and Synthetic Cells as an Emerging Frontier in Synthetic Biology.* Angew Chem Int Ed Engl, 2021. **60**(11): p. 5602-5611.

145. Liu, C.C., et al., *Toward an orthogonal central dogma.* Nat Chem Biol, 2018. **14**(2): p. 103-106.

146. Vasdekis, A.E. and A. Singh, *Microbial metabolic noise.* WIREs Mech Dis, 2021. **13**(3): p. e1512.

147. Ye, Z., et al., *Two-stage dynamic deregulation of metabolism improves process robustness & scalability in engineered E. coli.* Metab Eng, 2021. **68**: p. 106-118.

148. Muller, J.E.N., et al., *Engineering Escherichia coli for methanol conversion.* Metab Eng, 2015. **28**: p. 190-201.

149. Whitaker, W.B., et al., *Engineering the biological conversion of methanol to specialty chemicals in Escherichia coli.* Metab Eng, 2017. **39**: p. 49-59.

150. Jonsson, S., et al., *Kinetic and mechanistic characterization of the formyl-CoA transferase from Oxalobacter formigenes.* J Biol Chem, 2004. **279**(34): p. 36003-12.

151. Singh, R.K., et al., *Insights into Cell-Free Conversion of CO2 to Chemicals by a Multienzyme Cascade Reaction.* ACS Catalysis, 2018. **8**(12): p. 11085-11093.

152. Roger, M., et al., *Efficient Hydrogen-Dependent Carbon Dioxide Reduction by Escherichia coli.* Curr Biol, 2018. **28**(1): p. 140-145 e2.

153. Schuchmann, K. and V. Müller, *Direct and Reversible Hydrogenation of CO₂ to Formate by a Bacterial Carbon Dioxide Reductase.* Science, 2013. **342**(6164): p. 1382-1385.

154. Felnagle, E.A., et al., *Engineering synthetic recursive pathways to generate non-natural small molecules.* Nat Chem Biol, 2012. **8**(6): p. 518-26.

155. Kim, J., et al., *Dehydration of (R)-2-hydroxyacyl-CoA to enoyl-CoA in the fermentation of α-amino acids by anaerobic bacteria.* FEMS Microbiology Reviews, 2004. **28**(4): p. 455-468.

156. Kandasamy, V., et al., *Engineering Escherichia coli with acrylate pathway genes for propionic acid synthesis and its impact on mixed-acid fermentation.* Appl Microbiol Biotechnol, 2013. **97**(3): p. 1191-200.

157. Cheong, S., J.M. Clomburg, and R. Gonzalez, *Energy- and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions.* Nat Biotechnol, 2016. **34**(5): p. 556-61.

158. Clomburg, J.M., et al., *A synthetic biology approach to engineer a functional reversal of the beta-oxidation cycle.* ACS Synth Biol, 2012. **1**(11): p. 541-54.

159. Tarasava, K., et al., *Reverse beta-oxidation pathways for efficient chemical production.* J Ind Microbiol Biotechnol, 2022.

160. Buckel, W. and B.T. Golding, *Radical enzymes in anaerobes.* Annu Rev Microbiol, 2006. **60**: p. 27-49.

161. Obradors, N., et al., *Site-directed mutagenesis studies of the metal-binding center of the iron-dependent propanediol oxidoreductase from Escherichia coli.* Eur J Biochem, 1998. **258**(1): p. 207-13.

162. Tobimatsu, T., et al., *Heterologous expression, purification, and properties of diol dehydratase, an adenosylcobalamin-dependent enzyme of Klebsiella oxytoca.* Arch Biochem Biophys, 1997. **347**(1): p. 132-40.

163. Jain, R., et al., *Systematically engineering Escherichia coli for enhanced production of 1,2-propanediol and 1-propanol.* ACS Synth Biol, 2015. **4**(6): p. 746-56.

164. Wang, J., et al., *Rational engineering of diol dehydratase enables 1,4-butanediol biosynthesis from xylose.* Metab Eng, 2017. **40**: p. 148-156.

165. Lennen, R.M. and B.F. Pfleger, *Engineering Escherichia coli to synthesize free fatty acids.* Trends Biotechnol, 2012. **30**(12): p. 659-67.

166. Dellomonaco, C., et al., *Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals.* Nature, 2011. **476**(7360): p. 355-9.

167. Kunjapur, A.M. and K.L. Prather, *Microbial engineering for aldehyde synthesis.* Appl Environ Microbiol, 2015. **81**(6): p. 1892-901.

168. Caballero, E., et al., *Identification of lactaldehyde dehydrogenase and glycolaldehyde dehydrogenase as functions of the same protein in Escherichia coli.* J Biol Chem, 1983. **258**(12): p. 7788-92.

169. YAMANE, T. and S. HIRANO, *Semi-batch Culture of Microorganisms with Constant Feed of Substrate: A Methematical Simulation: Kinetic Studies on Fed-batch Cultures (III).* Journal of fermentation technology, 1977. **55**(2): p. 156-165.

170. Schwander, T., et al., *A synthetic pathway for the fixation of carbon dioxide in vitro.* Science, 2016. **354**(6314): p. 900-904.

171. Rudroff, F., *Whole-cell based synthetic enzyme cascades-light and shadow of a promising technology.* Curr Opin Chem Biol, 2019. **49**: p. 84-90.

172. Burg, J.M., et al., *Large-scale bioprocess competitiveness: the potential of dynamic metabolic control in two-stage fermentations.* Current Opinion in Chemical Engineering, 2016. **14**: p. 121-136.

173. Dinh, C.V. and K.L. Prather, *Layered and multi-input autonomous dynamic control strategies for metabolic engineering.* Curr Opin Biotechnol, 2020. **65**: p. 156-162.

174. Shen, X., et al., *Dynamic gene expression engineering as a tool in pathway engineering.* Curr Opin Biotechnol, 2019. **59**: p. 122-129.

175. Kim, S. and R. Gonzalez, *Selective production of decanoic acid from iterative reversal of beta-oxidation pathway.* Biotechnol Bioeng, 2018. **115**(5): p. 1311-1320.

176. Hernandez Lozada, N.J., et al., *Highly Active C8-Acyl-ACP Thioesterase Variant Isolated by a Synthetic Selection Strategy.* ACS Synth Biol, 2018. **7**(9): p. 2205-2215.

177. Yan, Q. and B.F. Pfleger, *Revisiting metabolic engineering strategies for microbial synthesis of oleochemicals.* Metab Eng, 2020. **58**: p. 35-46.

178. Burgener, S., N.S. Cortina, and T.J. Erb, *Oxalyl-CoA Decarboxylase Enables Nucleophilic One-Carbon Extension of Aldehydes to Chiral alpha-Hydroxy Acids.* Angew Chem Int Ed Engl, 2020. **59**(14): p. 5526-5530.

179. Rohwerder, T., et al., *Actinobacterial Degradation of 2-Hydroxyisobutyric Acid Proceeds via Acetone and Formyl-CoA by Employing a Thiamine-Dependent Lyase Reaction.* Front Microbiol, 2020. **11**: p. 691.

180. Zahn, M., et al., *Mechanistic details of the actinobacterial lyase-catalyzed degradation reaction of 2-hydroxyisobutyryl-CoA.* J Biol Chem, 2022. **298**(1): p. 101522.

181. Jiang, Y., et al., *Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system.* Appl Environ Microbiol, 2015. **81**(7): p. 2506-14.

182. Studier, F.W. and B.A. Moffatt, *Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes.* Journal of Molecular Biology, 1986. **189**(1): p. 113-130.

183. Blattner, F.R., et al., *The Complete Genome Sequence of Escherichia coli K-12.* Science, 1997. **277**(5331): p. 1453-1462.

184. Rosen, B.A., et al., *Ionic Liquid-Mediated Selective Conversion of CO₂ to CO at Low Overpotentials.* Science, 2011. **334**(6056): p. 643-644.

185. Han, N., et al., *Promises of Main Group Metal–Based Nanostructured Materials for Electrochemical CO2 Reduction to Formate.* Advanced Energy Materials, 2019. **10**(11).

186. Bontemps, S., L. Vendier, and S. Sabo-Etienne, *Ruthenium-catalyzed reduction of carbon dioxide to formaldehyde.* J Am Chem Soc, 2014. **136**(11): p. 4419-25.

187. Din, I.U., et al., *Recent developments on heterogeneous catalytic CO2 reduction to methanol.* Journal of CO2 Utilization, 2019. **34**: p. 20-33.

188. Jayathilake, B.S., et al., *Developing reactors for electrifying bio-methanation: a perspective from bio-electrochemistry.* Sustainable Energy & Fuels, 2022. **6**(5): p. 1249-1263.

189. Gleizer, S., et al., *Conversion of Escherichia coli to Generate All Biomass Carbon from CO2.* Cell, 2019. **179**(6): p. 1255-1263 e12.

190. Erb, T.J., P.R. Jones, and A. Bar-Even, *Synthetic metabolism: metabolic engineering meets enzyme design.* Curr Opin Chem Biol, 2017. **37**: p. 56-62.

191. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.

192. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of Molecular Biology, 1990. **215**(3): p. 403-410.

193. Huang, Y., et al., *CD-HIT Suite: a web server for clustering and comparing biological sequences.* Bioinformatics, 2010. **26**(5): p. 680-2.

194. Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation.* Nucleic Acids Research, 2021. **49**(W1): p. W293-W296.

195. Mirdita, M., et al., *ColabFold: making protein folding accessible to all.* Nat Methods, 2022. **19**(6): p. 679-682.

196. Berthold, C.L., et al., *Crystallographic snapshots of oxalyl-CoA decarboxylase give insights into catalysis by nonoxidative ThDP-dependent decarboxylases.* Structure, 2007. **15**(7): p. 853-61.

197. Morrison, K.L. and G.A. Weiss, *Combinatorial alanine-scanning.* Current Opinion in Chemical Biology, 2001. **5**(3): p. 302-307.

198. Yang, J., et al., *The I-TASSER Suite: protein structure and function prediction.* Nat Methods, 2015. **12**(1): p. 7-8.

199. Tunyasuvunakool, K., et al., *Highly accurate protein structure prediction for the human proteome.* Nature, 2021. **596**(7873): p. 590-596.

200. Pak, M.A., et al., *Using AlphaFold to predict the impact of single mutations on protein stability and function.* bioRxiv, 2021: p. 2021.09.19.460937.

201. Li, W. and A. Godzik, *Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.* Bioinformatics, 2006. **22**(13): p. 1658-9.

202. Zallot, R., N. Oberg, and J.A. Gerlt, *The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways.* Biochemistry, 2019. **58**(41): p. 4169-4182.

203. Helander, K.G., *Formaldehyde prepared from paraformaldehyde is stable.* Biotech Histochem, 2000. **75**(1): p. 19-22.

204. Ni, C., C.V. Dinh, and K.L.J. Prather, *Dynamic Control of Metabolism.* Annu Rev Chem Biomol Eng, 2021. **12**: p. 519-541.

205. García-Nafría, J., J.F. Watson, and I.H. Greger, *IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly.* Sci Rep, 2016. **6**: p. 27459.

206. Erb, T.J., et al., *Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway.* Proc Natl Acad Sci U S A, 2007. **104**(25): p. 10631-6.

207. Zarzycki, J., et al., *Identifying the missing steps of the autotrophic 3-hydroxypropionate CO2 fixation cycle in Chloroflexus aurantiacus.* Proc Natl Acad Sci U S A, 2009. **106**(50): p. 21317-22.

208. Borjian, F., et al., *The methylaspartate cycle in haloarchaea and its possible role in carbon metabolism.* ISME J, 2016. **10**(3): p. 546-57.

209. Ravi, M., M. Ranocchiari, and J.A. van Bokhoven, *The Direct Catalytic Oxidation of Methane to Methanol-A Critical Assessment.* Angew Chem Int Ed Engl, 2017. **56**(52): p. 16464-16483.

210. Hallam, S.J., et al., *Reverse methanogenesis: testing the hypothesis with environmental genomics.* Science, 2004. **305**(5689): p. 1457-62.

211. Wilkes, H., et al., *Metabolism of Hydrocarbons in n-Alkane-Utilizing Anaerobic Bacteria.* J Mol Microbiol Biotechnol, 2016. **26**(1-3): p. 138-51.

212. Sasikaran, J., et al., *Bacterial itaconate degradation promotes pathogenicity.* Nat Chem Biol, 2014. **10**(5): p. 371-7.

213. Erb, T.J., et al., *Ethylmalonyl-CoA mutase from Rhodobacter sphaeroides defines a new subclade of coenzyme B12-dependent acyl-CoA mutases.* J Biol Chem, 2008. **283**(47): p. 32283-93.

214. Erb, T.J., G. Fuchs, and B.E. Alber, *(2S)-Methylsuccinyl-CoA dehydrogenase closes the ethylmalonyl-CoA pathway for acetyl-CoA assimilation.* Mol Microbiol, 2009. **73**(6): p. 992-1008.

215. Linster, C.L., et al., *Ethylmalonyl-CoA decarboxylase, a new enzyme involved in metabolite proofreading.* J Biol Chem, 2011. **286**(50): p. 42992-3003.

216. Mattozzi, M., et al., *Expression of the sub-pathways of the Chloroflexus aurantiacus 3-hydroxypropionate carbon fixation bicycle in E. coli: Toward horizontal transfer of autotrophic growth.* Metab Eng, 2013. **16**: p. 130-9.

217. Borjian, F., et al., *Succinyl-CoA:Mesaconate CoA-Transferase and Mesaconyl-CoA Hydratase, Enzymes of the Methylaspartate Cycle in Haloarcula hispanica.* Front Microbiol, 2017. **8**: p. 1683.

218. Leutwein, C. and J. Heider, *Succinyl-CoA:(R)-benzylsuccinate CoA-transferase: an enzyme of the anaerobic toluene catabolic pathway in denitrifying bacteria.* J Bacteriol, 2001. **183**(14): p. 4288-95.

219. Nolte, J.C., et al., *Novel characteristics of succinate coenzyme A (Succinate-CoA) ligases: conversion of malate to malyl-CoA and CoA-thioester formation of succinate analogues in vitro.* Appl Environ Microbiol, 2014. **80**(1): p. 166-76.

220. Zhao, M., et al., *Metabolic engineering of Escherichia coli for producing adipic acid through the reverse adipate-degradation pathway.* Metab Eng, 2018. **47**: p. 254-262.

221. Martinez-Espinosa, R.M., *Heterologous and Homologous Expression of Proteins from Haloarchaea: Denitrification as Case of Study.* Int J Mol Sci, 2019. **21**(1).

222. Bennett, B.D., et al., *Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli.* Nat Chem Biol, 2009. **5**(8): p. 593-9.

223. Zhao, M., G. Li, and Y. Deng, *Engineering Escherichia coli for Glutarate Production as the C5 Platform Backbone.* Appl Environ Microbiol, 2018. **84**(16).

224. Burgener, S., et al., *Molecular Basis for Converting (2S)-Methylsuccinyl-CoA Dehydrogenase into an Oxidase.* Molecules, 2017. **23**(1).

225. Miller, T.E., et al., *Light-powered CO2 fixation in a chloroplast mimic with natural and synthetic parts.* Science, 2020. **368**(6491): p. 649-654.

226. Watmough, N.J. and F.E. Frerman, *The electron transfer flavoprotein: ubiquinone oxidoreductases.* Biochim Biophys Acta, 2010. **1797**(12): p. 1910-6.

227. Chen, M., et al., *Identification of a hotdog fold thioesterase involved in the biosynthesis of menaquinone in Escherichia coli.* J Bacteriol, 2013. **195**(12): p. 2768-75.

228. Whistance, G.R., J.F. Dillon, and D.R. Threlfall, *The nature, intergeneric distribution and biosynthesis of isoprenoid quinones and phenols in Gram-negative bacteria.* Biochemical Journal, 1969. **111**(4): p. 461-472.

229. Garcia Costas, A.M., et al., *Defining Electron Bifurcation in the Electron-Transferring Flavoprotein Family.* J Bacteriol, 2017. **199**(21).

230. Vick, J.E., et al., *Escherichia coli Enoyl-Acyl Carrier Protein Reductase (FabI) Supports Efficient Operation of a Functional Reversal of the β-Oxidation Cycle.* Appl Environ Microbiol, 2015. **81**: p. 1406-1416.

231. Campbell, J.W. and J.E. Cronan, Jr., *The enigmatic Escherichia coli fadE gene is yafH.* J Bacteriol, 2002. **184**(13): p. 3759-64.

232. Zhang, L., et al., *Directed evolution of phosphite dehydrogenase to cycle noncanonical redox cofactors via universal growth selection platform.* Nat Commun, 2022. **13**(1): p. 5021.

233. Lee, S.Y., et al., *A comprehensive metabolic map for production of bio-based chemicals.* Nature Catalysis, 2019. **2**(1): p. 18-33.

234. Yazdani, S.S. and R. Gonzalez, *Engineering Escherichia coli for the efficient conversion of glycerol to ethanol and co-products.* Metab Eng, 2008. **10**(6): p. 340-51.

235. Lu, J., et al., *Combinatorial modulation of galP and glk gene expression for improved alternative glucose utilization.* Appl Microbiol Biotechnol, 2012. **93**(6): p. 2455-62.

236. Reis, A.C. and H.M. Salis, *An Automated Model Test System for Systematic Development and Improvement of Gene Expression Models.* ACS Synth Biol, 2020. **9**(11): p. 3145-3156.

237. Vo, P.L.H., et al., *CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering.* Nat Biotechnol, 2021. **39**(4): p. 480-489.

238. Deuschle, U., et al., *Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures.* The EMBO Journal, 1986. **5**(11): p. 2987-2994.

239. Bassalo, M.C., et al., *Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli.* ACS Synth Biol, 2016. **5**(7): p. 561-8.

240. Neidhardt, F.C., P.L. Bloch, and D.F. Smith, *Culture Medium for Enterobacteria.* Journal of Bacteriology, 1974. **119**(3): p. 736-747.