UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Expanding the synthetic capabilities of chemoautotrophic metabolism.

Permalink

https://escholarship.org/uc/item/6xh1n6j5

Author

Thevasundaram, Kersh

Publication Date

2020

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Expanding the synthetic capabilities of chemautotrophic metabolism

by

Kershanthen Thevasundaram

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy in
Molecular & Cell Biology
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Michelle C. Y. Chang, Chair
Professor David Savage
Professor Nicholas Ingolia
Professor Michi Taga

Fall 2020

Expanding the synthetic capabilities of
chemautotrophic metabolism

© 2020

by Kershanthen Thevasundaram

Abstract

Expanding the synthetic capabilities of

chemautotrophic metabolism

by

Kershanthen Thevasundaram

Doctor of Philosophy in Molecular & Cell Biology

University of California, Berkeley

Professor Michelle C. Y. Chang, Chair

The steep environmental costs of petrochemical-derived products have created a pressing need for sustainable approaches to chemical synthesis. To address this need, scientists and engineers have developed diverse methods for capturing and either storing or using greenhouse gases to make value-added chemicals. In this effort, many scientists have looked to Nature for inspiration. Nature has evolved many methods for capturing CO_2 and using it as a carbon source for autotrophic growth. Of all the diverse forms of autotrophy, chemoautotropic metabolism is particularly remarkable for its versatile use of inorganic electron donors. If this versatility could be harnessed and coupled to sustainable sources of electrons and energy, chemoautotrophs may be used as a metabolic host for sustainable chemical synthesis.

This thesis explores two approaches for the engineering of chemoautotrophic metabolism for the production of value-added chemicals. In the first approach, essential genes for the Wood-Ljungdahl carbon fixation pathway were investigated in *Moorella thermoacetica*, a model acetogenic bacteria. A panel of gene candidates were identified for the appropriate maturation of the acetyl-CoA synthase/CO dehydrogenase (ACS/CODH) complex, an important metalloenzyme complex in the Wood-Ljungdahl pathway. The most promising candidate, AcsF, was recombinantly expressed with ACS in *Escherichia coli*, showing an increase in ACS nickel content. Furthermore, methods were developed to produce knockouts in *M. thermoacetica* to assess the native role of our gene candidates and measure the *in vitro* and *in vivo* activity of a recombinant ACS/CODH complex from *E.coli*.

In our second approach, we attempted to engineer *Methanococcus maripaludis*, a model methanogenic archaea, to produce 3-hydroxybutyrate (3HB) and polyhydroxybutyrate (PHB). NAD(H) pools in *M. maripaludis* were less than 5% of the pool in anaerobically-grown *E.coli,* suggesting that low NADH availability may be a metabolic bottleneck for biosynthesis. Using transcriptomic and proteomic studies, alanine dehydrogenase was identified as a critical native enzyme for *in vivo* NADH regeneration and improvements in product titers were observed in alanine-dependent growth. Multiple engineering strategies were used to precisely target and increase the cofactor pool, including heterologous expression of NAD^+-dependent formate dehydrogenase from *Candida boidinii* and a synthetic nicotinamide salvage pathway. Engineered strains of *M. maripaludis* with improved NADH pools produced up to 150 ± 4 mg/L of PHB in McNA medium, a more than 2.5-fold improvement over the starting strain.

To further characterize heterologous gene expression, native transcript structure in *M. maripaludis* was more carefully analyzed. Methanogenic archaea display characteristics of translation that are analogous to both eukaryotic and prokaryotic processes, but some features remain to be determined. To address this, we conducted a transcriptome assembly and determined the consensus Shine-Dalgarno (SD) sequence of GGAGGA from our 5' UTR-containing sequences. Our predicted transcripts were verified using 3' UTR sequencing. In addition, we found that a large fraction of our transcripts were leaderless, motivating us to investigate translation initiation for both leaderless and leadered transcripts. To probe this question, we developed a ribosome profiling methodology that could be used to precisely identify the ribosome's position and interactions with mRNA. These developments help pave the way towards a robust heterologous gene pathway design strategy for this species. The metabolic engineering strategies presented in this study broaden the utility of *M. maripaludis* for sustainable chemical synthesis using CO_2 and may be transferable to related archaeal species.

Table of Contents

Chapter 4: Investigating translation in *Methanococcus maripaludis* for improved heterologous gene expression

Appendices

List of Figures, Schemes, and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations

3HB	3-hydroxybutyrate
ACS	acetyl-CoA synthase
ATP	adenosine triphosphate
CODH	CO dehydrogenase
CFeSP	corrinoid iron-sulfur protein
MeTr	methyltransferase
dNTP	deoxynucleotide triphosphate
Cb	carbenicillin
Cm	chloramphenicol
CoA	coenzyme A
CoM	coenzyme M
CoB	coenzyme B
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
DTT	diothiothreitol
EDTA	ethylenediaminetetraacetic acid
ESI-MS	electrospray ionization mass spectrometry
GC-MS	gass chromatography-mass spectrometry
HEPES	4-[(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HER	hydrogen evolution reaction
HPLC	high performance liquid chromatography
H_4MPT	tetrahydromethanopterin
H_4SPT	tetrahydromethanosarcinapterin
ICP-MS	inductively coupled plasma-mass spectrometry
IPTG	isopropyl β-D-1-thiogalactopyranoside
Km	kanamycin
LB	Luria-Bertani broth
LC-MS/MS	liquid chromatography-tandem mass spectrometry
McCas	undefined, rich media with hydrolyzed casein
McNA-Ala	defined mineral media with L-alanine as the N-source
McNA	defined mineral media with NH_4Cl as the N-source
MF	methanofuran
NAD^+	nicotinamide adenine dinucleotide, oxidized

NADH	nicotinamide adenine dinucleotide, reduced
$NADP^+$	nicotinamide adenine dinucleotide phosphate, oxidized
NADPH	nicotinamide adenine dinucleotide phosphate, reduced
OD_{600}	optical density at 600 nm
ORFs	open reading frames
PCR	polymerase chain reaction
PHA	polyhydroxyalkanoate
PHB	polyhydroxybutyrate
PVDF	polyvinylidene fluoride
rpm	revolutions per minute
SDS-PAGE	sodium dodecyl sulfate- polyacrylamide gel electrophoresis
TBS	tris-buffered saline
THF	tetrahydrofolate
Tris	trisaminomethane

Acknowledgments

I deeply thank the friends, family and mentors that helped me through my graduate school journey. This process would have been impossible without your support. Firstly, I would like to thank Michelle Chang for her patience, mentorship and guidance. Your support has been essential for this period of incredible personal and professional growth. I have also really appreciated the freedom you have given me to pursue projects that are very different from other efforts in the lab. I cannot imagine learning how to conduct rigorous science without this experience and your guidence. I would also like to thank my former mentor, Joe Gallagher, for teaching me foundational skills in anaerobic microbiology. Your patience and mentorship helped me conduct genetic experiments in difficult microbes with confidence.

Additionally, I would like to thank all the members of the M. Chang Lab who I have had the privilege of working with. While I am thankful for all of your technical experimental advice and insightful research discussions, the thing I appreciate most is that you all have made our lab a great place to come to work to everyday. It is wonderful to do science in a friendly, helpful and supportive environment. Outside of the M. Chang lab, I must thank many other collaborators and advisors for providing me with essential technical and strategic advice for helping me pursue my goals, including Nicholas McGlincy, Audrey Reeves, Diana Iovan, Vanha Pham and Tyler DeTomasi.

I would also like to thank the members of my thesis committee, Dave Savage, Nicholas Ingolia and Michi Taga, for helpful research discussions and creative project suggestions. During my time in graduate school, I have had the pleasure of working with Max Sosa and Aidan Cowan as rotation students. I have also worked with incredible undergraduates including Freeman Cherng and Hyery Yoo. I appreciate all of your hard work and all the contributions you have made to my projects. Finally, I have to thank all of the family and friends who have been so supportive during this time. You have been there to raise my spirits during the most challenging moments and to celebrate with me during times of success. Your support has helped me grow into what I am today. Thank you.

Chapter 1: *Introduction*

1.1. Introduction

Hydrocarbons derived from crude oil have greatly expanded the repertoire of fuels and chemicals used by humans today. However, there are growing concerns that the steep environmental costs of using these resources may threaten to radically alter global climate [*1*]. Substituting crude oil-dependent processes with more sustainable chemical synthesis methods has therefore become a pressing global challenge. To inspire new approaches, scientists have often looked to Nature.

Nature has evolved numerous methods for chemical synthesis using CO_2 as a precursor molecule. In particular, autotrophic species are capable of using CO_2 as a carbon source for the production of cellular components such as amino acids, nucleotides and lipids (i.e. biomass). Species with this form of metabolism can be further classified based on the energy source they require. For example, photoautotrophs, such as plants, microalgae and cyanobacteria, use light as a source of energy to incorporate inorganic carbon (e.g. CO_2) into biomass. This biomass is also a source of carbon and energy for heterotrophs. (*Figure 1.1)*. By breaking down the metabolic products of autotrophs, heterotrophs produce their own metabolites. For example, several heterotrophic microorganisms use fermentation to convert sugars into biomass and metabolic by-products such as ethanol.

Heterotrophs are often considered suitable hosts for metabolic engineering, particularly due to our extensive technical and genetic understanding of how to utilize these species. Heterotrophic microbes may be genetically engineered to expand their product scope, improve their productivity and maximize product titers. Plant feedstocks can be derived from a variety of crops, including corn, sugarcane, sorghum, cassava and sugar beets [*2*], which fix CO_2 to produce sugars that are extracted by processing. These sugars can be used as growth substrates by heterotrophs, including *Escherichia coli* and *Saccharomyces cerevisiae*, which have long been used to convert hexose sugars such as glucose into fuels, polymers and other petrochemical alternatives [*3*, *4*] Thus, this approach to chemical manufacturing may have lower net CO_2 emissions compared to processes that use crude oil-derived hydrocarbon feedstocks. Understanding and harnessing these processes at industrial scale may help address the challenge of sustainable chemical synthesis.

On the other hand, issues have been raised about the feasibility of this approach due to the massive volumes of plant biomass that would be required at industrial scale. The large demand for energy crops may lead to competition with the food industry for a finite amount of arable land to grow these plants [*2*]. Solutions have been proposed to circumvent these issues, including the use of waste agricultural residues that are not typically used in the food industry [*2*, *5*]. For example, plant polysaccharides such as cellulose and hemicellulose typically compose a large fraction of total plant biomass and often remain underutilized. These complex sugars can be broken down into simple sugars for fermentation [*6*]. Alternatively, plants and other photosynthetic organisms can be directly engineered to produce value-added chemicals from CO_2 using heterologous pathways [*7*]. However, even this promising approach faces its own challenges. For example, light penetration into microbial cultures can be low and theoretical solar-to-chemical energy conversion efficiency of C3 photosynthesis is only 4-6% [*8*], which might raise the possibility of prohibitively large energy and space requirements for a photoautotroph-based chemical synthesis approach.

A third approach is the use of chemoautotrophic species as engineering hosts. Chemoautotrophic metabolism is a form of metabolism that requires chemical sources of energy for CO_2, including energy-dense molecules such as hydrogen and formate. Using these compounds

for fermentation may reduce the energy and space requirements for sustainable chemical synthesis compared to a photoautotrophic approach. Furthermore, these chemical energy sources can be electrochemically derived using inorganic heterocatalysts, which allows for the coupling of any renewable source of electricity to carbon fixation [*9*]. A hybrid process that incorporates both chemoautotrophic microorganisms and inorganic heterocatalysts may have the potential to overcome the limitations of other approaches to sustainable chemical synthesis. However, metabolic engineering of chemoautotrophs is a continuing field of research. A better understanding of gene expression, metabolism and important regulatory mechanisms in this group of species would help expand the versatility and robustness of these species as engineering hosts.

Here we will review progress that has been made using each of the approaches described above, which will help frame the work that follows on the investigation and use of chemoautotrophic metabolism for sustainable chemical synthesis.

1.2 Using model heterotrophs for metabolic engineering

Since the inception of synthetic biology, one of its aims has been the engineering of microorganisms for sustainable chemical synthesis [*10*, *11*]. Initial efforts focused on using model heterotrophs such as *Saccharomyces cerevisiae* and *Escherichia coli* because of the available genetic tools and their well characterized metabolism [*12*, *13*]. Gene expression levels were tuned in these species using non-native promoters, modified ribosome binding sites (RBS), metabolite-responsive riboswitches and RNA interference for gene downregulation [*3*]. Aside from tuning gene expression, several biochemical strategies have also been implemented to improve the product yield of these model organisms. For example, engineers have improved *in vivo* enzyme kinetics using substrate channeling between pathway proteins and microcompartment-targeted heterologous gene expression [*14*, *15*].

Another strategy has been the use of computational tools and *in silico* modelling. Genome-scale metabolic models have been developed to more accurately model and predict microbial host metabolism [*1*, *16*]. The models for a particular organism are based on available knowledge about native enzymes and the corresponding reactions they catalyze. Flux balance analysis (FBA) can be performed using the metabolic model to predict the flux of intracellular reactions and the predicted growth rate. These predictions can then be compared to real growth rate measurements to test the validity of the FBA. If the FBA is robust, the predictions it makes can be used to inform gene knockouts that could improve product titers [*17*]. In addition, FBA can be used to inform adaptive evolution experiments in which microbial growth is coupled to target product biosynthesis [*18*]. Another computational strategy for finding mutant strains with improved product titers involves the use of machine learning algorithms and lab automation for strain design. By "training" a machine learning algorithm with a dataset of known mutants and their product titers, the algorithm can make predictions about optimal mutant combinations without being explicitly programmed to produce those predictions [*19*].

Using the array of tools described above, metabolic engineers have improved biosynthesis of a broad range of useful molecules such as alcohols, fatty acids, alkanes and isoprenoids in engineered heterophic hosts. These molecules have potential applications as petrochemical and fossil fuel alternatives, pharmaceuticals, flavors, chemical precursors and fragrances [*20*]. However, the metabolism of model heterotrophic species can also be a limitation; since these species require an organic carbon source for carbon and energy, plant biomass is typically grown

as a feedstock for fermentation. These plants are grown under similar conditions as other food crops and typically have similar land and nutrition requirements as well. Therefore, limitations to available arable land and competition with the food industry may undermine the economic feasibility of this approach [*4*].

On the other hand, waste plant biomass sources that are unsuitable for cultivation as food crops have also been seen as a potential feedstock for heterotrophic fermentation[*5*, *21*]. Although this avenue may hold promise, much of this waste biomass is composed of lignocellulose sugars [*22*]. The majority of lignocellulosic biomass is made of carbohydrate polymers such as cellulose, hemicellulose, and lignin. These polymers are held tightly together by forming covalent and hydrogen bonds. Lignin in particular can be difficult to solubilize, which can inhibit hydrolysis of cellulose and hemicellulose [*23*]. This can make it challenging to break down lignocellulosic biomass, although a broad range of physical, chemical and enzymatic approaches have been used for pretreatment [*2*]. Unfortunately, pretreatment can be a prohibitively expensive addition to the fermentation process [*24*]. Progress is currently being made towards integrating lignocellulose treatment and simple sugar fermentation in a one-pot process using engineered strains, which has the potential to lower or even eliminate the cost of pretreatment [*25*].

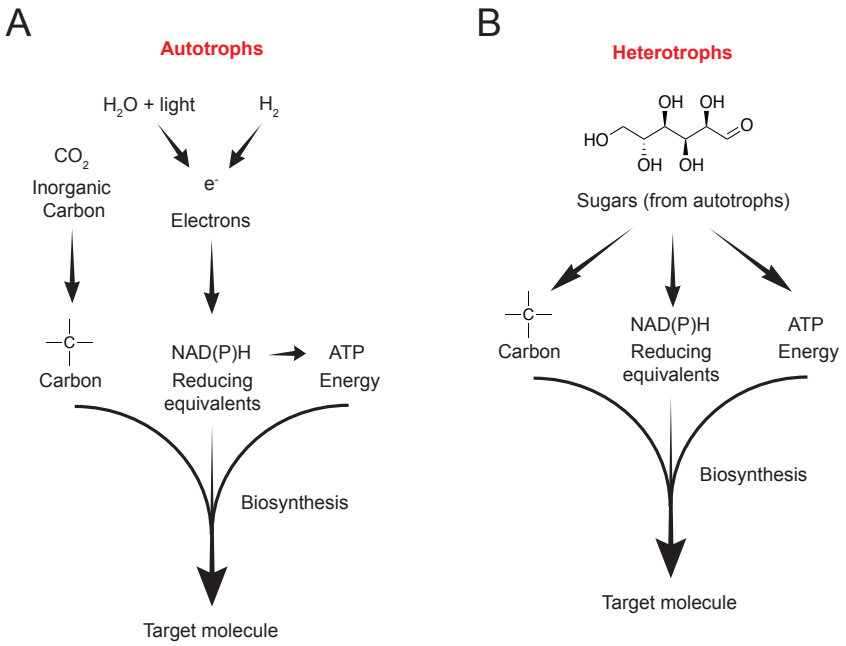


Figure 1.1 Comparing autotrophic and heterotrophic metabolism. *(A) Autotrophic metabolism uses CO_2 as a carbon source. Autotrophs can be further classified based on the energy source used to generate reducing equivalents in the form of NAD(P)H and cellular energy in the form of ATP. Sub-classifications of autotrophs include photoautotrophs (uses light energy) and chemoautotrophs (uses chemical energy). B) Heterotrophic metabolism (right) can use organic carbon sources, such as sugars derived from autotrophic biomass, as sources of carbon and energy. Sugars can be broken down via glycolysis to generate reduced NAD(P)H and ATP as sources of reducing equivalents and energy, respectively. Similar to autotrophic metabolism, these essential metabolites can be used for biosynthesis of cellular components and target molecules.*

1.3 Engineering photosynthetic organisms

To more directly couple photosynthesis to production pathways, photosynthetic organisms can be engineered as well. Instead of using expensive plant biomass pretreatment steps or competing directly with the food market for feedstocks, this method relies on species that use CO_2, water and light as carbon, electron and light sources respectively. Both multicellular and unicellular photosynthetic species have been engineered for synthetic biology applications [*26*, *27*]. When comparing these two, however, there appear to be many advantages to the unicellular approach. Namely, the product extraction process is simpler in unicellular, microbial cultures, genetic manipulation of microbes tends to be less complex and less structural biomass is required for growth, which will increase the available carbon and energy for target chemical biosynthesis [*28–30*]. Furthermore, photosynthetic microbes can have relatively simple feed requirements: sunlight and aqueous growth media [*29*].

One particularly promising branch of photoautotroph metabolic engineering has focused on the use of engineered cyanobacteria for chemical synthesis (*Figure 1.2)*. There are numerous benefits to using cyanobacteria as hosts, including a large set of known genetic tools. These have been used for genomic deletion or insertion, tuning gene expression, knocking out native enzymes, and substituting pathway enzymes to improve target pathway kinetics [*31–34*]. In addition, the large NADPH pool found in cyanobacteria can be utilized by replacing any NADH-dependent enzymes in a heterologous pathway with NADPH-dependent homologs [*35*], or by overexpressing a transhydrogenase that can transfer reducing equivalents from NAD(P)H to NAD^+ [*36*]. Furthermore, the photosynthetic efficiency of cyanobacteria can be improved by minimizing the light-harvesting antennae size of the phycobilisome [*37*, *38*]. Also, efforts are ongoing to screen mutant libraries of the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo) to identify variants with improved properties for more energy-efficient photosynthesis [*39*].

The strategies described above have played a role in expanding the biosynthetic capabilities of cyanobacteria, allowing them to produce a wide range of useful products at reasonably high titers, including carbohydrates, fatty acids, isoprenoids, vitamins, alcohols, and hydroxy acids [*40*, *41*]. However, there are still challenges that must be addressed before this approach can fully realize its commercial potential. For example, the circadian clock in cyanobacteria can regulate gene expression in these species, which may result in unexpected changes in heterologous gene expression during target chemical production [*42*]. Furthermore, open-air ponds often used to grow cyanobacteria can be easily contaminated by unwanted species [*43*]. Alternatively, photobioreactors can be used to grow these species in a controlled environment, although this approach can be excessively costly and suffer from low solar-to-chemical energy conversion efficiency [*8*]. Both of these culture methods are continuing to address challenges with low light penetration into the media and reported genetic instability in polyploid transgenic strains, which can prevent robust gene expression [*44*, *45*].

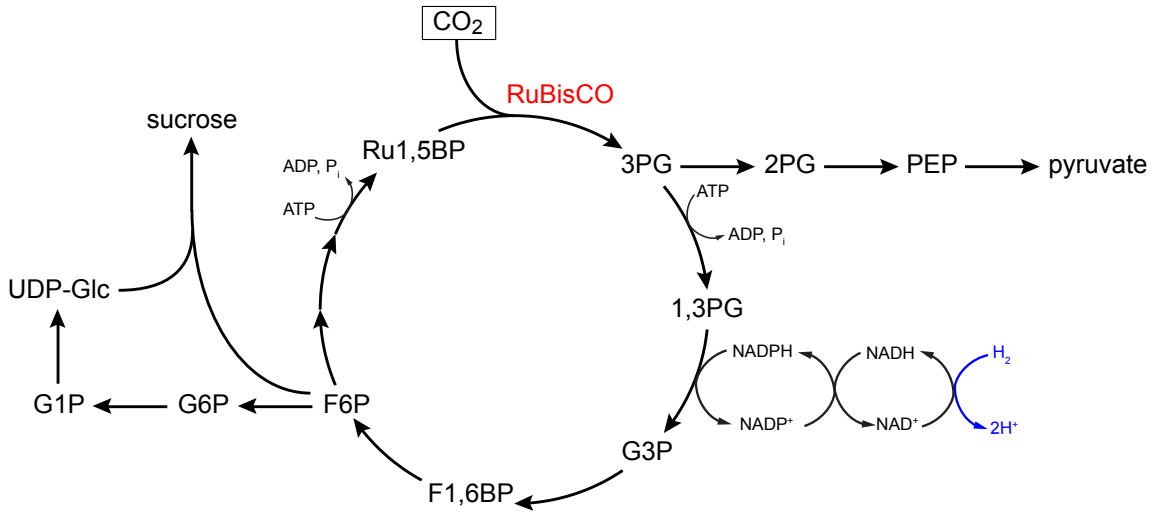


Figure 1.2 The Calvin cycle for carbon fixation. *Cyanobacteria and Knallgas bacteria such as Synechococcus elongatus PCC 7942 and Cupriavidus necator, respectively, both use the Calvin Cycle to generate biomass. Both forms of metabolism require RuBisCO (labelled red) to carboxylate ribulose 1,5-bisphosphate (Ru-1,5-BP) to make 3-phosphoglycerate (3PG). Photoautotrophs will use light-dependent reactions to produce key energy intermediates such as NADPH and ATP. Conversely, chemoautotrophs such as C. necator can use inorganic energy sources such hydrogen (H_2), which replenish energy intermediates, including NADPH, NADH and ATP, in a non-photosynthetic manner (labelled blue). Fixed carbon can be diverted to value-added chemicals via sucrose and pyruvate. UDP-Glc, UDP-glucose; G1P, glucose-1-phosphate; G6P, glucose-6-phosphate; F6P, fructose-6-phospate; Ru1,5BP, ribulose-1,5-bisphosphate; F1,6BP, fructose-1,6-bisphosphate, G3P, glyceraldehyde-3-phosphate; 1,3PG, 1,3-phosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate.*

1.4 Harnessing chemoautotrophic metabolism for chemical synthesis

Aside from photoautotrophs, there is another group of autotrophs that may have potential as microbial hosts for metabolic engineering: chemoautotrophs. These fascinating organisms use chemical sources such as energy-dense molecular hydrogen (H_2) as a source of energy and electrons. Hydrogen can be produced electrochemically using water electrolysis catalyzed with energy efficient, robust inorganic catalysts. A broad range of renewable energy sources could be used to drive the water electrolysis reaction, including but not limited to sunlight [*46*, *47*]. Electrochemically derived hydrogen can then be used by biocatalysts to make target products, leveraging the low overpotentials and product selectivity of biological enzymes. The combination of energy efficient electrochemical H_2 production with inorganic catalysts and chemoautotrophic metabolism represents a hybrid platform (*Scheme 1.1*) that leverages the strengths of inorganic and biological catalysts. This approach may also avoid the low solar-to-chemical conversion efficiency and light penetration challenges of photoautotrophic platforms [*48*].

One relatively well studied chemoautotroph is *Cupriavidus necator*, a bacterial species that uses H_2 as an energy source and the Calvin-Benson cycle for CO_2 fixation (*Figure 1.2*) [*49*]. Engineering *C. necator* has many practical advantages, including a collection of genetic tools, inexpensive media components and a high-flux carbon storage pathway used to make polyhydroxyalkanoates (PHA), a class of biodegradable polymers that may have applications as next-generation plastics [*50*]. In addition, past engineering efforts have shown that *C. necator* can be used to make other value-added chemicals using H_2 and CO_2 feedstocks, including isopropanol,

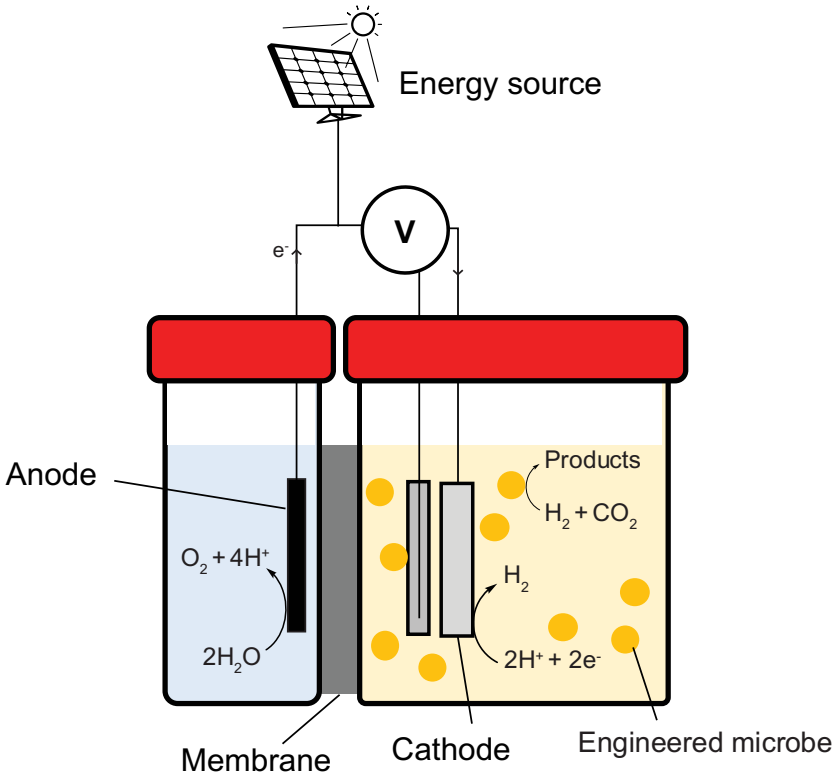
C_4/C_5 alcohols, alkanes, alkenes, and methyl ketones [*51*]. Despite this, there are still limitations to the use of *C. necator* due to their reliance on the Calvin-Benson cycle for central metabolism. Although this is the most common carbon fixation pathway observed in nature, it is also the most energetically costly biological method of fixing CO_2 (in terms of ATP required per mole of fixed CO_2) [*52*]. This may limit the theoretical yield of the target product and affect the economic feasibility of this approach. In contrast, other chemoautotrophs use a other, more energy efficient pathways for carbon fixation, including the reductive TCA cycle, dicarboxylate-4-hydroxybutyrate cycle, 3-hydroxypropionate-4-hydroxybutyrate cycle and the Wood-Ljungdahl pathway [*52*].

The Wood-Ljungdahl Pathway in acetogens. Amongst carbon fixation pathways in nature, the Wood-Ljungdahl pathway (WLP) is the most energy efficient. This pathway makes acetyl-CoA using 2 molar equivalents of CO_2 and 4 molar equivalents of H_2 (*Figure 1.3)* [*53*]. In addition to H_2, a variety of other electron sources may be used as energy sources for the pathway, including CO, formate, methanol, methylamines and methylsulfides [*54*]. This diversity allows the use of potentially low-cost energy sources, including waste industrial syngas, instead of more costly, electrochemically derived energy sources. The WLP was first characterized biochemically using the thermophilic acetogen, *Moorella thermoacetica* (formerly called *Clostridium thermoaceticum*. Acetogens are named for the acetate by-product made from acetyl-CoA, which is made using the WLP under autotrophic conditions [*55*].

Besides *M. thermoacetica,* numerous other acetogenic species from the genus *Clostridia* use the WLP for autotrophic growth and many have been used as microbial hosts for sustainable chemical synthesis from industrial syngas [*56*]. Some acetogens are natively capable of converting syngas into butanol (e.g. *C. acetobutylicum*, *C. beijerinckii*), ethanol (*C. autoethanogenum*) and 1,3-propanediol (e.g. *C. pasteurianum*). However, the product profile from these species has remained limited compared to well-studied heterotrophs and photoautotrophs. There may be numerous reasons for this, including low transformation efficiencies and a smaller set of synthetic biology tools in these organisms [*57*]. To help mitigate this problem, several genetic tools are being developed and improved, including the "ClosTron" tool for targeted gene disruption using group II mobile introns, Mariner-transposable *Himar1*-based approaches for random mutagenesis, and counter selection markers such as *pyrE* or *pyrF* [*57–59*]. Another potentially useful tool is Cas9, which can improve the rate of double strand break (DSB) formation and homology-directed repair (HDR) in these species. This can be particularly helpful in the case of acetogens since they have been reported to have low HDR rates [*60*]. A toolbox also exists for pathway engineering, containing inducible and constitutive promoters, reporter genes (e.g. the flavin mononucleotide-based fluorescent protein, iLov), conditional origins of replication and 3' terminators [*57*].

Although many of the biochemical details of the WLP have been established, an understanding of the carbon and energy metabolism in acetogens is not yet complete [*61*]. Without sufficient knowledge about the levels of relevant metabolites in chemoautotrophs and how they are coupled, it may be difficult to know whether a pathway that works well in an engineered heterotroph will also work well in an engineered chemoautotrophic species. An alternative approach for harnessing the unique central metabolism of acetogens would be to transplant the pathway into a more tractable microbial host, such as *E. coli* (*Scheme 1.2*). This approach is made more convenient by the fact that many of the enzymes in the Eastern branch of the WLP have homologs in the THF cycle used for methionine biosynthesis in *E. coli* [*54*]. Assuming this pathway could be used to produce methyl-THF, only the Western branch of the WLP would need to be heterologously

expressed in *E. coli* in order to form acetyl-CoA from CO_2. Once acetyl-CoA was formed, it could be diverted towards the production of target molecules using well-established heterologous pathways. Identifying the genes necessary for *in vivo* reconstitution of the Western branch of the WLP in *E.coli* would therefore go towards harnessing chemoautotrophic metabolism while leveraging the knowledge already available for a well-studied heterotrophic host *(Scheme 1.2)*.



Scheme 1.1 Design of a hybrid bioinorganic approach to chemical synthesis. *An electrochemical cell interfaced with microbial cultures can be used to drive target product synthesis using CO_2, water and an energy source. In the anode chamber (left), an anode coated with an oxygen evolving reaction (OER) catalyst is suspended in an electrolyte solution. The OER catalyst oxidizes water, producing molecular oxygen and dissociated protons. Protons diffuse across a cation-selective membrane to the cathode chamber, allowing for charge balancing. Electrons from water are driven from the anode to the cathode using an external applied potential. This applied potential can come from renewable sources of energy, including solar, wind and geothermal energy. The cathode is coated with a hydrogen evolution reaction (HER) catalyst that produces hydrogen from protons and electrons. Using the generated hydrogen as an energy source, CO_2 sparged into the cathode electrolyte solution can be incorporated into biomass and target products by engineered microbes.*

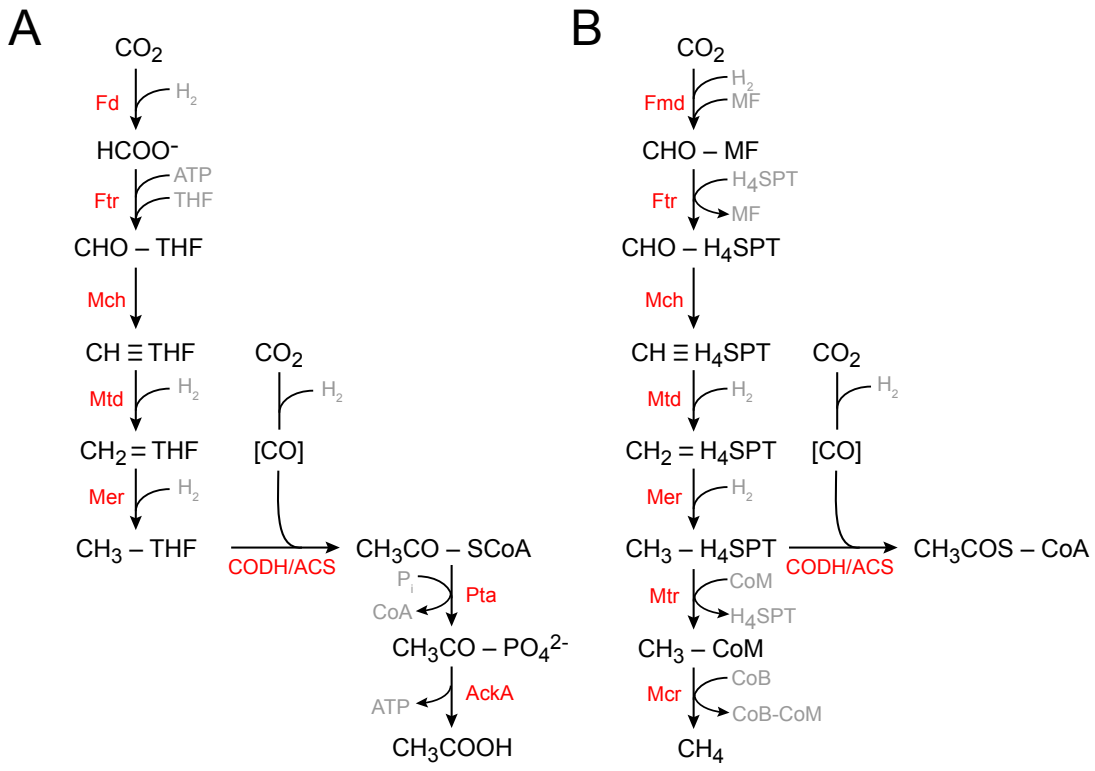
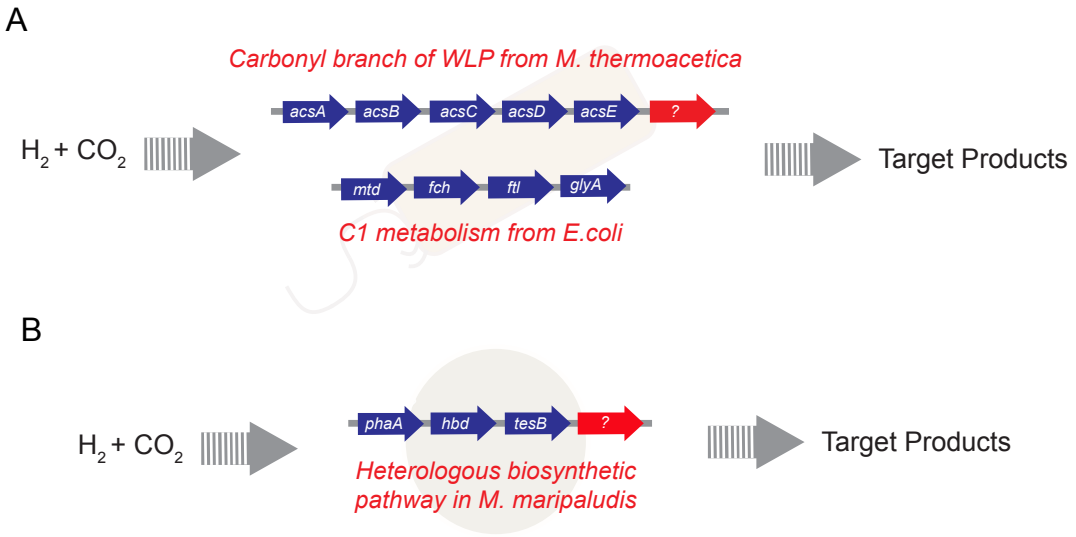


Figure 1.3 The Wood-Ljungdahl pathway in acetogens and methanogens. *(A) The canonical Wood-Ljungdahl pathway in acetogens. CO_2 is reduced via a 6-electron reduction process to methyl-THF using reducing equivalents derived from molecular hydrogen (H_2). In the other branch of the pathway, CO_2 is reduced to CO using the ACS/CODH complex and reduced ferredoxin as an electron donor. The methyl group from methyl-THF is transferred to ACS and combined with CO to form an acetyl group at the active site of ACS. Nucleophilic attack by CoA forms acetyl-CoA, which is used for cellular biomass. To generate ATP for the cell, acetyl-CoA can also be converted to acetate via acetylphosphate. (B) The modified Wood-Ljungdahl pathway in Methanococcus maripaludis. Instead of THF as a C1 unit carrier, M. maripaludis uses tetrahydromethanosarcinapterin (H_4SPT) and methanofuran (MF), while other methanogens use tetrahydromethanopterin (H_4MPT) instead of H_4SPT. Formyl-H_4SPT is made in an ATP-independent manner, in contrast to formyl-THF in acetogens. CO_2 is reduced to methyl-H_4SPT using three reducing equivalents derived from H_2. The compound is either further reduced to make methane and generate a membrane electrochemical gradient for ATP production, or it is used by the ACS/CODH complex to make acetyl-CoA. Enzyme names are shown in red and additional substrates are shown in grey. CHO-THF, 10-formyltetrahydrofolate; CH≡THF, 5,10-methenyltetrahydrofolate; CH_2=THF, 5,10-methylenetetrahydrofolate; CH_3-THF, 5-methyltetrahydrofolate; CH_3CO-SCoA, acetyl-CoA;CH_3CO-$PO4^{2-}$, acetylphosphate; CHO-MF,formyl-methanofuran; CHO-H_4MPT,5-formyl-H_4MPT; CH≡H_4MPT, 5,10-methenyl- H_4MPT; CH_2=H4MPT, 5,10 methylene- H_4MPT;CH_3-H4MPT, 5-methyl- H_4MPT; CH_3-CoM, methyl-coenzyme M; Mtr, methyl-H_4MPT: HS-CoM methyltransferase; Mcr, methyl-CoM reductase; Fd, formate dehydrogenase;Ftl,formate-THF-ligase;Mch,Metheny-THF,cylcohydrolase/methylenate-H_4MPT-cyclohydrolase;Mtd, methylene-THF dehydrogenase/ methylene-H_4MPT hydrogenase; Mer, methylene-THF reductase/ methylene-H_4MPT reductase; CODH/ACS, carbon monoxide dehydrogenase/ acetyl-CoA synthase; Pta, phosphotransacetylase; AckA, acetate kinase A; Fmd, formylmethanofuran dehydrogenase; Ftr, formyl-MF: H_4MPT formyltransferase.*

The modified Wood-Ljungdahl Pathway in methanogens. Methanogens use a modified WLP that is NADPH- and ATP-independent, unlike the canonical WLP found in acetogens. Also, the modified WLP uses unusual cofactors such as tetrahydromethanopterin (H_4MPT), tetrahydrosarcinapterin (H_4SPT) and methanofuran (MF) instead of THF as a C_1 unit carrier [*62*]. Ecologically, methanogens serve a fundamental role in the global methane cycle by producing methane as a byproduct of their metabolism [*63*]. Although methane is a greenhouse gas, it can be made in a renewable fashion via methanogenic metabolism using waste CO_2 feedstocks and renewable sources of energy. Renewable methane (also called renewable natural gas, or RNG) can be used as a heating fuel, chemical precursor and energy source [*64*]. Since it is a gas at room temperature, it can be easily separated during fermentation. Furthermore, methanogenic archaea have membranes composed of isoprenoids instead of fatty acids, unlike acetogens, photoautotrophs and heterotrophs [*65*]. Since isoprenoids can be used to make a variety of useful chemicals such as flavors, fragrances and cancer therapeutics, this metabolite pool serves as a potentially useful source of precursors for engineered pathways. This is demonstrated by the only published engineering effort in methanogens thus far, involving the conversion of endogenous isoprenoids to geraniol in the model methanogen, *Methanoccocus maripaludis* [*66*].

Another potentially interesting opportunity might be the conversion of acetyl-CoA to target products in methanogens. Acetyl-CoA is a common metabolite in WLP-containing organisms as well as heterotrophs. In addition, there are many known pathways for the conversion of acetyl-CoA to useful target compounds. Despite this opportunity, metabolic engineering efforts in methanogens have been hindered by multiple challenges. Firstly, methanogens utilize unusual cofactors like 2[4Fe-4S] ferredoxin and coenzyme F_{420} for their central metabolism instead of NAD(P)H [*67*]. As a result, the size of the NAD(P)H pool in methanogens is currently unclear. Since many heterologous pathways are NAD(P)H-dependent, it may be difficult to successfully produce target products without engineering these NAD(P)H pools. Furthermore, relatively little is known about gene regulation and relevant metabolite concentrations in these species. Gaining a broader understanding of these processes could help expand the scope of methanogen metabolic engineering. Studies that investigate the NADH pool of these methanogens and engineering efforts that establish NAD(P)H dependent pathways in a model methanogen could make progress towards making methanogens more tractable for sustainable chemical synthesis efforts.



Scheme 1.2 Engineering strategies for harnessing chemoautotrophic metabolism. *(A) Engineering E.coli to make target products using CO_2 and H_2. Natively, an industrial tractable heterotrophic microbe, such as E.coli, uses organic carbon sources during aerobic fermentation to produce biomass. It has also been routinely engineered to divert its acetyl-CoA pool towards target product synthesis. The methyl (Eastern) branch of the WLP is homologous to key genes involved in C1 metabolism from E.coli [54]. To form a hybrid Wood-Ljungdahl pathway in E.coli, essential genes for carbonyl (Western) branch of the Wood-Ljungdahl pathway (WLP) can be identified from a model acetogen such as M. thermoacetica and transplanted into E.coli to allow this model host to use CO_2 and H_2 as precursors to make target products. Additional genes may be necessary for the reconstitution of the carbonyl branch. (B) Engineering M. maripaludis to make target products using CO_2 and H_2. Methanogens are capable of converting CO_2 and H_2 into biomass and methane via native metabolism. It may be possible to divert their endogenous acetyl-CoA pool towards target product formation using a model pathway such as phaA-hbd-tesB. Additional genes may be necessary to increase product titers.*

1.5 Specific aims and thesis organization

The following work aims to harness chemoautotrophic metabolism for the production of value-added chemicals from sustainable energy sources and carbon dioxide using two major approaches: transplanting the Wood-Ljungdahl pathway from an acetogen into a model heterotrophic host and engineering a methanogen for the production of target products. Chapter 2 details the investigation of essential genes for chemoautotrophic metabolism in *Moorella thermoacetica*, a model acetogenic species, to identify metallochaperones that may be necessary to express an active recombinant Wood-Ljungdahl pathway in *E. coli*. Chapter 3 details the second approach, in which the model methanogen, *Methanococcus maripaludis*, is engineered to expand its product scope. These efforts focus on expressing heterologous pathways and increasing the NAD(H) cofactor pool in this organism. Finally, Chapter 4 details approaches to investigating translation in *M. maripaludis*. These efforts focused on method development for ribosome profiling in this species, as well as a transcriptome assembly that produced predicted operons, 5' UTR and 3' UTR. Using this dataset, we predicted Shine-Dalgarno sequences and 3' terminators that could help tune heterologous gene expression for *M. maripaludis* and possibly other Archaea as well.

1.6 References

1. P. P. Peralta-Yahya, F. Zhang, S. B. Del Cardayre, J. D. Keasling, "Microbial engineering for the production of advanced biofuels". *Nature*. **2012**, *488*, 320–328.
2. M. E. Himmel, *et al.*, "Biomass recalcitrance: Engineering plants and enzymes for biofuels production". *Science*. **2007**, *315*, 804–807.
3. P. E. M. Purnick, R. Weiss, "The second wave of synthetic biology: From modules to systems". *Nature Reviews Molecular Cell Biology*. **2009**, *10*, 410–422.
4. S. Nonhebel, Nonhebel, Sanderine, "Renewable energy and food supply: will there be enough land?". *Renewable and Sustainable Energy Reviews*. **2005**, *9*, 191–201.
5. D. Tilman, J. Hill, C. Lehman, "Carbon-negative biofuels from low-input high-diversity grassland biomass". *Science*. **2006**, *314*, 1598–1600.
6. C. Xiros, E. Topakas, P. Christakopoulos, "Hydrolysis and fermentation for cellulosic ethanol production". *WIREs Energy and Environment*. **2013**, *2*, 633–654.
7. M. A. Gomaa, L. Al-Haj, R. M. M. Abed, "Metabolic engineering of cyanobacteria and microalgae for enhanced production of biofuels and high-value products". *Journal of Applied Microbiology*. **2016**, *121*, 919–931.
8. X. G. Zhu, S. P. Long, D. R. Ort, "What is the maximum efficiency with which photosynthesis can convert solar energy into biomass?". *Current Opinion in Biotechnology*. **2008**, *19*, 153–159.
9. S. Cestellos-Blanco, H. Zhang, J. M. Kim, Y. xiao Shen, P. Yang, "Photosynthetic semiconductor biohybrids for solar-driven biocatalysis". *Nature Catalysis*. **2020**, *3*, 245–255.
10. J. E. Bailey, "Toward a science of metabolic engineering". *Science*. **1991**, *252*, 1668–1675.
11. G. Stephanopoulos, "Synthetic biology and metabolic engineering". *ACS Synthetic Biology*. **2012**, *1*, 514–525.
12. J. Lian, S. Mishra, H. Zhao, "Recent advances in metabolic engineering of *Saccharomyces cerevisiae*: New tools and their applications". *Metabolic Engineering*. **2018**, *50*, 85–108.
13. D. Yang, S. Y. Park, Y. S. Park, H. Eun, S. Y. Lee, "Metabolic engineering of *Escherichia coli* for natural product biosynthesis". *Trends in Biotechnology*. **2020**, *38*, 745–765.
14. Y.-H. P. Zhang, "Substrate channeling and enzyme complexes for biotechnological applications". *Biotechnology Advances*. **2011**, *29*, 715–725.
15. H. Kirst, C. A. Kerfeld, "Bacterial microcompartments: Catalysis-enhancing metabolic modules for next generation metabolic and biomedical engineering". *BMC Biology*. **2019**, *17*, 1–11.
16. P. Jouhten, "Metabolic modelling in the development of cell factories by synthetic biology". *Computational and Structural Biotechnology Journal*. **2012**, *3*, e201210009.
17. E. J. O'Brien, J. M. Monk, B. O. Palsson, "Using genome-scale models to predict biological capabilities". *Cell*. **2015**, *161*, 971–987.
18. A. P. Burgard, P. Pharkya, C. D. Maranas, "OptKnock: A bilevel programming framework for identifying gene knockouts strategies for microbial strain optimization". *Biotechnology and Bioengineering*. **2003**, *84*, 647–657.
19. D. B. Kell, D. B. Kell, "Metabolomics, modelling and machine learning in systems biology-towards an understanding of the languages of cells." *The FEBS Journal*. **2006**, *273*, 873–894.
20. C. A. Rabinovitch-Deere, J. W. K. Oliver, G. M. Rodriguez, S. Atsumi, "Synthetic biology and metabolic engineering approaches to produce biofuels". *Chemical Reviews*. **2013**, *113*,

4611–4632.
21. J. G. Elkins, B. Raman, M. Keller, "Engineered microbial systems for enhanced conversion of lignocellulosic biomass". *Current Opinion in Biotechnology*. **2010**, *21*, 657–662.
22. M. C. McCann, N. C. Carpita, "Biomass recalcitrance: A multi-scale, multi-factor, and conversion-specific property". *Journal of Experimental Botany*. **2015**, *66*, 4109–4118.
23. A. Zoghlami, G. Paës, "Lignocellulosic biomass: Understanding recalcitrance and predicting hydrolysis". *Frontiers in Chemistry*. **2019**, *7*, 874.
24. G. Bokinsky, P. P. Peralta-Yahya, A. George, B. M. Holmes, E. J. Steen, J. Dietrich, T. S. Lee, D. Tullman-Ercek, C. A. Voigt, B. A. Simmons, J. D. Keasling, "Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*". *Proceedings of the National Academy of Sciences of the United States of America*. **2011**, *108*, 19949–19954.
25. R. Chen, J. Dou, "Biofuels and bio-based chemicals from lignocellulose: Metabolic engineering strategies in strain development". *Biotechnology Letters*. **2016**, *38*, 213–221.
26. J. L. Nemhauser, K. U. Torii, "Plant synthetic biology for molecular engineering of signalling and development". *Nature Plants*. **2016**, *2*, 1–7.
27. P. Xia, H. Ling, J. L. Foo, M. W. Chang, "Synthetic biology toolkits for metabolic engineering of cyanobacteria". *Biotechnology Journal*. **2019**, *14*, 1800496.
28. E. C. Tatsis, S. E. O'Connor, B. Pfeifer, Y. Tang, "New developments in engineering plant metabolic pathways". *Current Opinion in Biotechnology*. **2016**, *42*, 126–132.
29. B. Wang, J. Wang, D. R. Meldrum, "Application of synthetic biology in cyanobacteria and algae". *Frontiers in Microbiology*. **2012**, *3*.
30. C. Cook, L. Martin, R. Bastow, "Opportunities in plant synthetic biology". *Journal of Experimental Botany*. **2014**, *65*, 1921–1926.
31. S. A. Angermayr, A. Gorchs Rovira, K. J. Hellingwerf, "Metabolic engineering of cyanobacteria for the synthesis of commodity products". *Trends in Biotechnology*. **2015**, *33*, 352–361.
32. L. Al-Haj, Y. Lui, R. Abed, M. Gomaa, S. Purton, "Cyanobacteria as chassis for industrial biotechnology: Progress and prospects". *Life*. **2016**, *6*, 42.
33. B. M. Berla, R. Saha, C. M. Immedthun, C. D. Maranas, T. S. Moon, H. B. Pakrasi, "Synthetic biology of cyanobacteria: Unique challenges and opportunities". *Frontiers in Microbiology*. **2013**, *4*, 246.
34. X. Gao, T. Sun, G. Pei, L. Chen, W. Zhang, "Cyanobacterial chassis engineering for enhancing production of biofuels and chemicals". *Applied Microbiology and Biotechnology*. **2016**, *100*, 3401–3413.
35. C. Li, F. Tao, J. Ni, Y. Wang, F. Yao, P. Xu, "Enhancing the light-driven production of d-lactate by engineering cyanobacterium using a combinational strategy". *Scientific Reports*. **2015**, *5,* 1-11.
36. S. A. Angermayr, M. Paszota, K. J. Hellingwerf, "Engineering a cyanobacterial cell factory for production of lactic acid". *Applied and Environmental Microbiology*. **2012**, *78*, 7098–7106.
37. H. Kirst, C. Formighieri, A. Melis, "Maximizing photosynthetic efficiency and culture productivity in cyanobacteria upon minimizing the phycobilisome light-harvesting antenna size". *Biochimica et Biophysica Acta - Bioenergetics*. **2014**, *1837*, 1653–1664.
38. A. Melis, "Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency". *Plant Science*. **2009**, *177*, 272–280.

39. H. Aigner, *et al.*, "Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2". *Science*. **2017**, *358*, 1272–1278.
40. P. Savakis, K. J. Hellingwerf, "Engineering cyanobacteria for direct biofuel production from CO2". *Current Opinion in Biotechnology*. **2015**, *33*, 8–14.
41. A. E. Case, S. Atsumi, "Cyanobacterial chemical production". *Journal of Biotechnology*. **2016**, *231*, 106–114.
42. B. M. Berla, R. Saha, C. M. Immedthun, C. D. Maranas, T. S. Moon, H. B. Pakrasi, "Synthetic biology of cyanobacteria: Unique challenges and opportunities". *Frontiers in Microbiology*. **2013**, *4*, 246.
43. M. Hannon, J. Gimpel, M. Tran, B. Rasala, S. Mayfield, "Biofuels from algae: Challenges and potential". *Biofuels*. **2010**, *1*, 763–784.
44. S. Sevda, S. Bhattacharya, I. M. A. Reesh, S. Bhuvanesh, T. R. Sreekrishnan, "Challenges in the design and operation of an efficient photobioreactor for microalgae cultivation and hydrogen production". *Biohydrogen Production: Sustainability of Current Technology and Future Perspective*, 147–162.
45. P. R. Jones, "Genetic instability in cyanobacteria - an elephant in the room?". *Frontiers in Bioengineering and Biotechnology*. **2014**, *2*, 12.
46. A. P. Murthy, J. Madhavan, K. Murugan, "Recent advances in hydrogen evolution reaction catalysts on carbon/carbon-based supports in acid media". *Journal of Power Sources*. **2018**, *398*, 9–26.
47. J. A. Turner, "Sustainable hydrogen production". *Science*. **2004**, *305*, 972–974.
48. E. M. Nichols, J. J. Gallagher, C. Liu, Y. Su, J. Resasco, Y. Sun, P. Yang, M. C. Y. Chang, C. J. Chang, "Hybrid bioinorganic approach to solar-to-chemical conversion". *Proceedings of the National Academy of Sciences of the United States of America*. **2015**, *112*, 11461–11466.
49. M. Esparza, E. Jedlicki, M. Dopson, D. S. Holmes, "Expression and activity of the Calvin-Benson-Bassham cycle transcriptional regulator CbbR from ˆin Ralstonia eutropha". *FEMS Microbiology Letters*. **2015**, *362*, 108.
50. M. Beeby, M. Cho, J. Stubbe, G. J. Jensen, "Growth and localization of polyhydroxybutyrate granules in *Ralstonia eutropha*". *Journal of Bacteriology*. **2012**, *194*, 1092–1099.
51. S. Nangle, *et al.*, "Valorization of CO_2 through lithoautotrophic production of sustainable chemicals in *Cupriavidus necator*". *Metabolic Engineering*. **2020**, *62*, 207-220.
52. A. Bar-Even, E. Noor, R. Milo, "A survey of carbon fixation pathways through a quantitative lens". *Journal of Experimental Botany*. **2012**, *63*, 2325–2342.
53. F. E. Fontaine, W. H. Peterson, E. McCoy, M. J. Johnson, G. J. Ritter, "A new type of glucose fermentation by *Clostridium thermoaceticum*". *Journal of Bacteriology*. **1942**, *43*, 701–715.
54. S. W. Ragsdale, E. Pierce, "Acetogenesis and the Wood-Ljungdahl pathway of CO_2 fixation". *Biochimica et Biophysica Acta - Proteins and Proteomics*. **2008**, *1784*, 1873–1898.
55. G. Diekert, G. Wohlfarth, "Metabolism of homoacetogens". *Antonie van Leeuwenhoek*. **1994**, *66*, 209–221.
56. M. E. Martin, H. Richter, S. Saha, L. T. Angenent, "Traits of selected *Clostridium* strains for syngas fermentation to ethanol". *Biotechnology and Bioengineering*. **2016**, *113*, 531–539.

57. R. C. Joseph, N. M. Kim, N. R. Sandoval, "Recent developments of the synthetic biology toolkit for *Clostridium*". *Frontiers in Microbiology*. **2018**, *9*, 154.
58. S. A. Kuehne, N. P. Minton, "ClosTron-mediated engineering of *Clostridium*.". *Bioengineered*. **2012**, *3*, 247–254.
59. S. T. Cartman, N. P. Minton, "A mariner-based transposon system for *in vivo* random mutagenesis of *Clostridium difficile*". *Applied and Environmental Microbiology*. **2010**, *76*, 1103–1109.
60. H. Huang, C. Chai, N. Li, P. Rowe, N. P. Minton, S. Yang, W. Jiang, Y. Gu, "CRISPR/Cas9-based efficient genome editing in *Clostridium ljungdahlii*, an autotrophic gas-fermenting bacterium". *ACS Synthetic Biology*. **2016**, *5*, 1355–1361.
61. A. Kaspar Valgepea, R. S. P. Lemgruber, K. Meaghan, R. W. Palfreyman, T. Abdalla, B. D. Hejistra, J. B. Behrendorff, R. Tappel, M. Köpke, S. D. Simpson, L. K. Nielsen, E. Marcellin, "Maintenance of ATP homeostasis triggers metabolic shifts in gas-fermenting acetogens". **2017,** *4*, 505-515.
62. J. T. Keltjens, B. W. Brömmelstroet, S. M. Kengen, C. Drift, G. D. Vogels, "5,6,7,8-tetrahydromethanopterin-dependent enzymes involved in methanogenesis". *FEMS Microbiology Letters*. **1990**, *87*, 327–332.
63. W. S. Reeburgh, "Oceanic methane biogeochemistry". *Chemical Reviews*. **2007**, *107*, 486–513.
64. D. E. Holmes, J. A. Smith, "Biologically produced methane as a renewable energy source". *Advances in Applied Microbiology*. **2016**, *97*, 1–61.
65. S. Jain, A. Caforio, A. J. M. Driessen, "Biosynthesis of archaeal membrane ether lipids". *Frontiers in Microbiology*. **2014**, *5*, 641.
66. Z. Lyu, R. Jain, P. Smith, T. Fetchko, Y. Yan, W. B. Whitman, "Engineering the autotroph *Methanococcus maripaludis* for geraniol production". *ACS Synthetic Biology*. **2016**, *5*, 577–581.
67. U. Deppenmeier, T. Lienard, G. Gottschalk, "Novel reactions involved in energy conservation by methanogenic archaea". *FEBS Letters*. **1999**, *457*, 291–297.

Chapter 2: *Identifying candidate accessory proteins for acetyl-CoA synthase from Moorella thermoacetica*

2.1. Introduction

The Wood-Ljungdahl pathway is a common carbon fixation pathway amongst chemoautotrophic microorganisms, many of which belong to the class, *Clostridia* [*1-5*].This pathway incorporates carbon dioxide (CO_2) into biomass using molecular hydrogen (H_2), carbon monoxide (CO) and other chemical energy sources [*2*]. The pathway can be divided into two branches that are both involved in catalytic CO_2 reduction: the methyl and carbonyl branch (*Scheme 2.1*) [*3*]. In the methyl branch, CO_2 is reduced to formate, and then condensed with tetrahydrofolate (THF) to form formyl-THF. This compound undergoes further reduction and dehydration steps to produce methyl-THF. Overall, this branch uses several enzymes with homologs in the folate cycle and methionine biosynthesis pathways found in bacterial and eukaryotic species [*4*].

Conversely, the carbonyl branch of the Wood-Ljungdahl pathway appears to be particularly unique in nature. One unique enzyme complex in this pathway is the activity of the acetyl-CoA synthase/CO dehydrogenase (ACS/CODH) complex. The CODH subunit catalyzes CO_2 reduction to CO using reduced ferredoxin as an electron donor [*5*]. CO is subsequently transferred to a catalytic nickel at the active site of ACS via an internal hydrophobic tunnel [*6*]. The methyl group from methyl-THF is transferred to the catalytic nickel in the ACS active site by two consecutive methyl-transfer steps catalyzed by methyl-THF methyltransferase (MeTr) and corrinoid iron-sulfur protein (CFeSP), respectively [*7,8*]. Through migratory insertion, the coordinated CO and methyl groups are condensed to form an acetyl group. In the next step, a free Coenzyme A (CoA) thiolate performs a nucleophilic attack on the acetyl-metal intermediate to form and release acetyl-CoA [*9,10*]. Acetyl-CoA may then be used in downstream pathways to produce biomass or converted into acetate via the activity of phosphotransacetylase and acetate kinase to generate ATP [*11*].

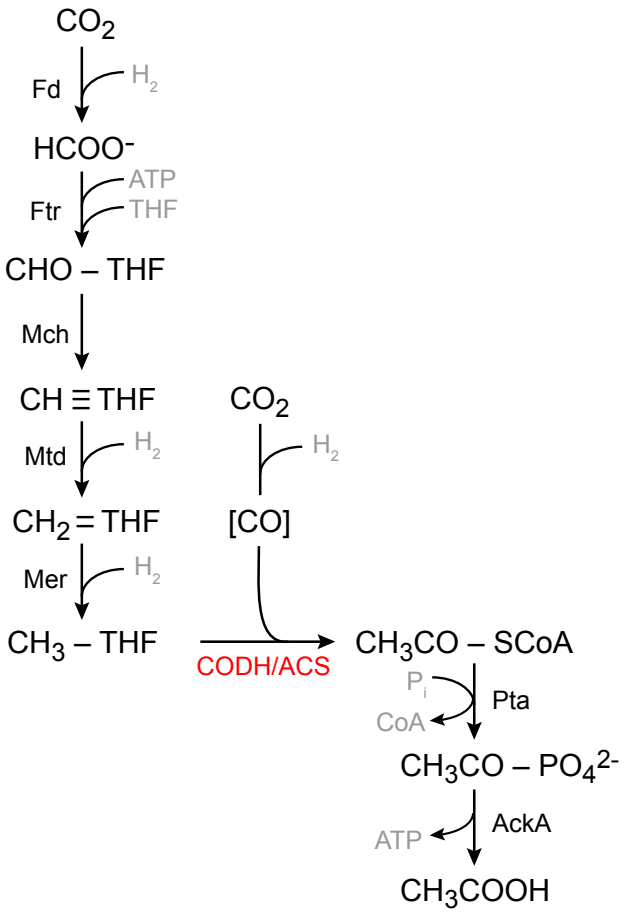
Many studies have been performed on the Wood-Ljungdahl pathway enzymes using the model acetogen, *Moorella thermoacetica.* $ACS/CODH_{Mt}$ is a heterotetrameric ($\alpha_2\beta_2$) complex in *M. thermoacetica* that contains four metal clusters, denoted clusters A-D [*12*]. The unusual A cluster, found at the active site of each ACS subunit, contains a [4Fe-4S] cluster bridged to a Ni site (the "proximal" nickel) that is thiolate-bridged by a cysteine to another nickel (the "distal" nickel) in a thiolato-and-carboxamido-type N_2S_2 coordination environment (*Scheme 2.2*) [*13*]. The B and D clusters are [4Fe-4S] clusters with putative roles in electron transfer from ferredoxin to the C cluster found at the active site of CODH [*14*]. The C cluster consists of a [3Fe-4S] cluster bridged to a binuclear NiFe involved in CO_2 reduction to CO [*15*]. The A and C metal clusters in ACS/CODH are necessary for the catalytic activity of the complex (*Scheme 2.2*) [*16,17*]. However, the maturation of these unusual metal clusters has not been fully elucidated. One strategy to probe questions about the maturation of the ACS/CODH active site *in vivo* would be to express, genetically perturb and study this complex in a recombinant host with better genetic tractability than *M. thermoacetica*. Furthermore, this engineering effort could have important industrial implications for the environmentally sustainable production of fuels and polymers from carbon dioxide. However, the complex has not yet been reported to be active *in vivo* in a recombinant host. Towards this aim, the genes corresponding to ACS/CODH complex (*acsA* and *acsB* for $CODH_{Mt}$ and ACS_{Mt}, respectively) have been expressed in *E.coli* under anaerobic conditions [*18*,*19*]. After purifying the protein in an aerobic environment, it was shown that the complex maintained the native $\alpha_2\beta_2$ quaternary structure. Also, the complex was active in a CO oxidation activity assay, suggesting that recombinant CODH is active *in vitro*.

Spectroscopic data also suggests that the iron-sulfur components found in cluster A, B and C were intact [*18*]. This suggests the possibility that *E. coli* contains accessory proteins that can complement the absence of at least some of the loading factors that may be required for *in vivo* ACS/CODH activity, if required at all. IscS and IscU are native *E.coli* enzymes that catalyze the formation of iron-sulfur clusters [*20*], and it is possible that these factors were necessary for Fe-S cluster assembly in recombinant ACS/CODH. In addition, a native metallochaperone for Ni-dependent hydrogenases in *E.coli*, HypB, is a homolog of CooC [*21*] (a known accessory protein for CODH metalation) and may serve as a surrogate of CooC for Ni insertion into the C cluster.

However, recombinant ACS activity was not observed until the complex was incubated overnight in a $NiCl_2$ solution [*22*]. Before incubation, electron paramagnetic resonance (EPR) measurements of recombinant ACS reduced with CO were indistinguishable from the NiFeC signal of Ni-replete ACS_{Mt} (*Scheme 2.2*). Also, electronic absorption spectra for thionin-oxidized and CO-reduced recombinant ACS/CODH were similar to $ACS/CODH_{Mt}$, exhibiting features typical of redox-active 4Fe-4S clusters [*18*]. After incubation with stoichiometric excess of nickel chloride, activity typical of fully active ACS_{Mt} could be observed. Since the A cluster of ACS_{Mt} requires nickel, this set of observations suggest that nickel diffused into the active site during incubation, allowing for the observed *in vitro* activity. It is also possible that nickel was loaded into the A cluster successfully *in vivo,* but may have leached out of the active site during purification, leading to a loss of activity [*23*]. However, it is unlikely that *E.coli* would be able to produce active ACS with native machinery because of the unusual metal content and geometry of the A cluster.

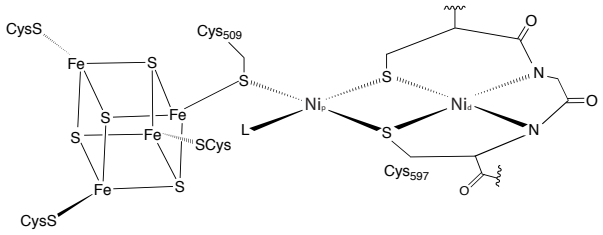
Similar to other metalloproteins, ACS and CODH have been reported to use specific accessory proteins called “metallochaperones” to traffic potentially cytotoxic metals through the cytoplasm and transfer these metals to their cognate metalloprotein. These proteins have been reported to play a role in the maturation of metalloenzymes by binding tightly to free metals intracellularly, protecting these metals from undesirable side reactions and delivering them to their appropriate target [*24*]. AcsF and CooC are metallochaperones that have been shown to load nickel into ACS_{Mt} and $CODH_{Mt}$, respectively [*25,26*]. Indeed, homologs of CooC from *Clostridium ljungdahlii* were coexpressed with $ACS/CODH_{Cj}$ in a recombinant *Clostridium acetobutylicum*, which revealed low but detectable levels of *in vivo* ACS and CODH activity [*27*]. Whether there are other unique factors from *C. acetobutylicum* that facilitate active site maturation for this complex is still unclear.

Here, we aim to elucidate the panel of necessary accessory proteins required for cofactor loading in ACS/CODH by studying the Wood-Ljungdahl pathway of the model acetogen *Moorella thermoacetica*. To elucidate the set of genes that are most likely required for ACS/CODH maturation, differential expression analysis was performed on the species after growth in either ACS/CODH-dependent or -independent conditions (*Scheme 2.3*). Over 200 genes were significantly upregulated in ACS/CODH-dependent growth conditions. Several steps of bioinformatic analysis were used to filter these genes to identify a short list of candidate accessory proteins for ACS/CODH. The top candidate, AcsF was introduced into recombinant *E.coli* with ACS/CODH and metal content analysis was performed. Although complete metallation was not observed, multiple explanations are provided for the observed results and a set of activity assays and possible follow-up experiments are presented.

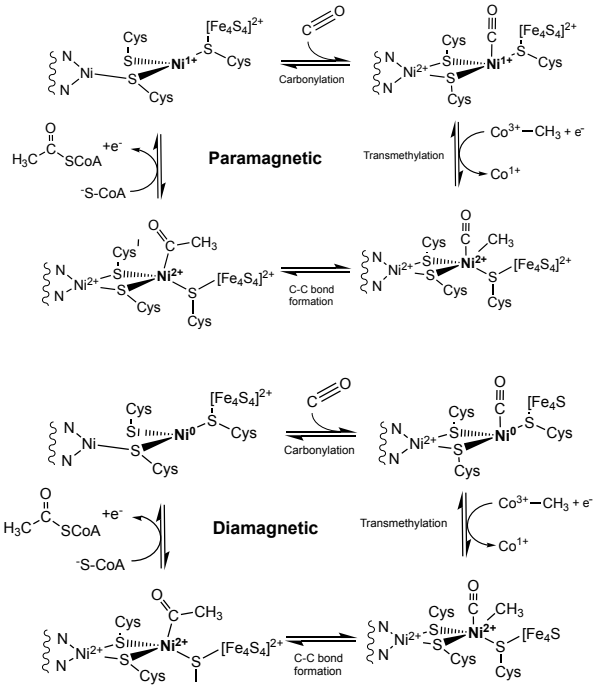


Scheme 2.1 The role of ACS/CODH in the Wood-Ljungdahl pathway of acetogens. *The canonical Wood-Ljungdahl pathway in acetogens. CO_2 is reduced via a 6-electron reduction process to methyl-THF using reducing equivalents derived from molecular hydrogen (H_2). In the other branch of the pathway, CO_2 is reduced to CO using the ACS/CODH complex (labelled red) and reduced ferredoxin as an electron donor. The methyl group from methyl-THF is transferred to ACS and combined with CO to form an acetyl group at the active site of ACS. Nucleophilic attack by CoA forms acetyl-CoA, which is used for cellular biomass. To generate ATP for the cell, acetyl-CoA can also be converted to acetate via acetylphosphate. Enzyme names are shown in black and additional substrates are shown in grey. ATP, adenosine triphosphate; CHO-THF, 10-formyltetrahydrofolate; CH≡THF, 5,10-methenyltetrahydrofolate; CH_2=THF, 5,10-methylenetetrahydrofolate; CH_3-THF, 5-methyltetrahydrofolate; CH_3CO-SCoA, acetyl-CoA; CH_3CO-$PO4^{2-}$, acetylphosphate; Fd, formate dehydrogenase; Ftl, formate-THF-ligase; Ftr, formyl-THF formyltransferase; Mch, Metheny-THF cyclohydrolase/methylenate-H_4MPT-cyclohydrolase; Mtd, methylene-THF dehydrogenase/methylene-H_4MPT hydrogenase; Mer, methylene-THF reductase/ methylene-H_4MPT reductase; CODH/ACS, carbon monoxide dehydrogenase/ acetyl-CoA synthase; Pta, phosphotransacetylase; AckA, acetate kinase A.*

A



B



Scheme 2.2. Active site structure and proposed mechanisms of ACS/CODH. *(A) The unusual A cluster, found at the active site of each ACS subunit, contains a [4Fe-4S] cluster bridged to a Ni site (the "proximal" nickel) that is thiolate-bridged by a cysteine to another nickel (the "distal" nickel) in a thiolato-and-carboxamido-type N_2S_2 coordination environment. The proximal nickel is known to be labile [8] and may be displaced during protein purification. Although the [4Fe-4S] cluster of the A site has been reconstituted in vivo in a recombinant host, the proximal nickel has not yet been shown to be intact in a recombinant host. (B) The proposed mechanisms of ACS activity are distinguished by the oxidation state of the catalytic nickel. The ACS/CODH complex plays a critical role in the Wood-Ljungdahl pathway. The paramagnetic mechanism includes a paramagnetic NiFeC intermediate, while the diamagnetic mechanism uses only diamagnetic intermediates, cycling between Ni^0 and Ni^{2+} [13]. An acetyl group is formed by migratory insertion of coordinated -CO and $-CH_3$. The sequence of substrate binding appears to be random but is shown as ordered reactions for clarity.*

2.4. Materials and Methods

Commercial materials. Ultra-high purity gases purchased from Praxair (Danbury, CT) were used for all anaerobic manipulations. Distilled water (dH_2O) was deionized to a resistivity of 18.2 MΩ·cm using a Millipore Milli-Q UF Plus system (ddH_2O). Fructose, L(+)-arabinose, glycerol, sodium bicarbonate, Tris base, kanamycin sulfate, sodium bicarbonate, sodium chloride, magnesium sulfate heptahydrate, magnesium chloride hexahydrate, ammonium chloride, potassium phosphate monobasic, potassium phosphate dibasic, sodium phosphate monobasic, sodium phosphate dibasic, potassium chloride, calcium chloride, manganese (II) chloride tetrahydrate, and pressure release aluminum crimp seals (20 mm) were purchased from Fisher Scientific (Pittsburgh, PA). Calcium chloride dihydrate, ammonium iron(II) sulfate hexahydrate, sodium selenite, nickel(II) chloride hexahydrate, boric acid, cobalt(II) chloride hexahydrate, copper(II) chloride dihydrate, sodium molybdate dihydrate, sodium sulfide, d-Desthiobiotin, L-cysteine, biotin, folic acid, pyridoxine hydrochloride, thiamine hydrochloride, riboflavin, nicotinic acid, calcium D-(+)-pantothenate, vitamin B12, p-aminobenzoic acid, PEG 3350, formic acid reagent grade, nitrilotriacetic acid, manganese(II) sulfate dihydrate, sodium tungstate dihydrate, dimethyl sulfoxide (DMSO), tris(2-carboxyethyl)phosphine (TCEP), cOmplete protease inhibitor cocktail tablets, acetyl coenzyme A sodium salt, 3'-dephosphocoenzyme A, sodium citrate dihydrate, nitrilotriacetic acid disodium salt, cyanocobalamin, MOPS and thioctic acid were purchased from Sigma Aldrich (St. Louis, MO). Zinc sulfate heptahydrate was purchased from Mallinckrodt Chemicals (St. Louis, MO). Resazurin was purchased from Eastman Kodak Company (Rochester, NY). Balch tubes (18 x 150 mm), anaerobic glass bottles (250 mL and 2L) and butyl rubber stoppers (20 mm) were purchased from Chemglass (Vineland, NJ).

Bacterial strain details and culture conditions. E. coli DH10B-T1^{R} was used for DNA construction and BL21(DE3)-T1^{R} was used for heterologous production of proteins for purification. To prepare chemically competent cells from plasmid transformations, the appropriate *E. coli* strain was cultured from a single colony in 100 mL LB broth, then harvested at an OD_{600} of 0.4-0.6 and resuspended in 10 mL of sterile transformation-storage solution (10% (*w/v*) PEG 3350, 5% (*v/v*) DMSO, 20 mM $MgCl_2$ in LB broth). 100 uL aliquots were flash frozen in liquid nitrogen and stored at -80°C for later use.

Plasmid transformations were performed by adding the plasmid(s), 20 µL 5× KCM solution (0.5 M KCl, 0.15 M $CaCl_2$, 0.25 M $MgCl_2$) and 175 µL of MilliQ-treated water to a thawed competent cell aliquot of the appropriate strain. After 10 min incubation on ice, cells were subjected to heat shock at 42°C for 90s, diluted to 1 mL with LB broth, and incubated at 37°C for 1 h before plating with sterile beads on LB agar containing appropriate antibiotics.

Acetogen strain details and culture conditions. *Moorella thermoacetica* ATCC 39073 was used for differential expression analysis and transformation experiments. The *M. thermoacetica* strains were anaerobically cultured at 55°C with shaking in 250 mL glass bottles or 18 × 150 mm Balch tubes containing 50 mL or 10 mL of acetogen growth medium, respectively, with an initial pH of 6.5. The medium was supplemented with 60 mM fructose or H_2/CO_2 as the sole carbon and energy source. To culture the strains in H_2/CO_2, the bottles were flushed with a filter-sterilized H_2/CO_2 mixture (80%, 20%) after inoculation and then adjusted to a final pressure of 200 kPa. In the fructose culture, the gas phase of the bottles was filled with N_2.

Acetogen medium preparation. Modified ATCC 1754 PETC medium [*28*] was used as the basal medium. Per L, the medium contains:

ddH_20	980 mL
NH_4Cl	1.0 g
KCl	0.1 g
$MgSO_4 \cdot 7H_2O$	0.2 g
NaCl	0.8 g
KH_2PO_4	0.1 g
$CaCl_2 \cdot 2H_2O$	0.02 g
$NaHCO_3$	2.0 g
Trace elements solution	10 mL
Wolfe's vitamin solution	10 mL
Resazurin solution (0.01% (*w/v*))	1 mL

All ingredients except $NaHCO_3$ were added to water and the pH was adjusted to 5.5. $NaHCO_3$ was added and the media was boiled, degassed and the bottles and tubes were sealed with butyl rubber stoppers and aluminum crimps. The sealed medium was autoclaved, resulting in a clear solution. After the medium cooled, 100 µL reducing agent (see preparation below) was injected per 10 mL of medium using a N_2-flushed syringe. Additionally, fructose was added anaerobically to the medium. Final pH was adjusted to 5.9-6.0.

A stock of 100 × reducing agent was prepared by adding a 0.01% (*w/v*) resazurin solution (200 µL) to MilliQ-treated water (200 mL) and sparging the solution with N_2 for 20 minutes. Under continuous N_2 flow, 6 g of L-cysteine was added, followed by 6 g of sodium sulfide nonahydrate. Sparging with N_2 was continued until the sodium sulfide had dissolved completely and the solution was completely colorless. After preparation, the reducing agent was stored under an N_2 atmosphere in a 250 mL glass bottle and sealed with butyl rubber stoppers and aluminum crimps.

Per liter, the trace elements solution contains:

ddH_2O	1000 mL
Nitrilotriacetic acid	2.0 g
$MnSO_4 \cdot 2H_2O$	1.0 g
$Fe(SO_4)_2(NH_4)_2 \cdot 6H_2O$	0.8 g
$CoCl_2 \cdot 6H_2O$	200 mg
$ZnSO_4 \cdot 7H_2O$	200 mg
$CuCl_2 \cdot 2H_2O$	20 mg
$Na_2MoO_4 \cdot 2H_2O$	200 mg
$NiCl_2 \cdot 6H_2O$	20 mg
Na_2SeO_3	20 mg
$Na_2WO_4 \cdot 2H_2O$	20 mg

Gene and plasmid construction. Standard molecular biology techniques were used for plasmid construction. *E. coli* DH10B-T1^R served as the cloning host. All PCR amplifications were carried out with Phusion polymerase (New England Biolabs, Ipswich, MA). Plasmids were assembled using the Gibson method [*29*], and sequences were verified using Sanger sequencing

(Quintara Biosciences, Berkeley, CA). Detailed descriptions of the gBLOCKs, plasmids and primers are contained in *Appendix 1*.

pETDuet-AcsB-Strep. pETDuet was digested with NcoI and BamHI. A synthetic version of the AcsB-Strep gene codon-optimized for *E.coli* was inserted into the NcoI/BamHI site of pETDuet to generate pETDuet-AcsB-Strep.

pETDuet-AcsB-Strep-CODH. pETDuet-AcsB-Strep was digested with KpnI and PacI. A synthetic version of the CODH gene codon-optimized for *E.coli* was inserted into the KpnI/PacI site of pETDuet-AcsB-Strep to generate pETDuet-AcsB-Strep-CODH.

pACYCDuet-AcsF. pACYCDuet was digested with NcoI and HindIII. A synthetic version of the AcsF gene codon-optimized for *E.coli* was inserted into the NcoI/HindIII site of pACYCDuet to generate pACYCDuet-AcsF.

pMTL83151-ΔpyrF::Km^R. The pMTL83151 plasmid was digested with FseI/PmeI to remove the *pyrF* gene. The left and right homology arms were generated by PCR using *M. thermoacetica* genomic DNA and *pyrF* LH f/r primer pair or the *pyrF* LH f/r primer pair, respectively. The P_{G3PD}-Km^R gene was synthesized as a gBLOCK and used without further clean-up. Inserts and vector were then joined using Gibson assembly.

Acetogen transformation. All transformation procedures were performed as described in published transformation protocols, with some modifications [*30*]. All procedures were performed under aerobic conditions, except for cell growth, which occurred under anaerobic conditions. pMTL83151-*ΔpyrF::Km^R* was transformed into the non-methylating *E. coli* GM272 strain containing pAN1 to obtain appropriately methylated DNA. Cells were incubated at 37°C overnight and harvested. pMTL83151-*ΔpyrF::Km^R* was gel purified for transformation.

M. thermoacetica ATCC 39073 strains were cultured in reinforced clostridial media (BD, San Jose, CA) supplemented with fructose at 55°C and harvested by centrifugation (5,800 × *g* for 10 min for 4°C) at OD_{600} = 0.1-0.2. For each transformation, 6 mL of cell suspension (OD_{600} = 0.2-0.8) was harvested stepwise in a 1.5 mL reaction tube by centrifugation (2000 × *g*, 1 min) with 4 × 1.5 mL cell suspension. Next, cell pellets were washed twice with 1.5 mL ice-cold 10% (*v/v*) glycerol (pH 6). Washed cells were then resuspended in 200 µL of ice-cold 10% (*v/v*) glycerol (pH 6) and allowed to cool for 5 min.

2 µg of methylated plasmid DNA was added and the mixture was directly transferred into a precooled electroporation cuvette with a gap-size of 0.2 cm. Electroporation was performed using a Bio-Rad Micropulser electroporator (Bio-Rad, Hercules, CA) in an anaerobic chamber. The cells were pulsed at 2.5 kV and 600 Ω. Cells were then immediately inoculated into 5 mL of pre-warmed RCM and incubated for 48 h at 55°C after pressurizing headspace with 200 kPa of 80% H_2:20% CO_2. After this outgrowth, the culture (200-1000 µL) was plated using the roll-tube method on plates containing 300 µg/mL of kanamycin. The plates were incubated at 55°C and colonies were obtained after approximately 6-8 d of incubation. Isolated colonies were inoculated into 5 mL of RCM with 300 µg/mL of kanamycin in 18 × 150 mm Balch tubes and grown anaerobically.

RNA isolation and sequencing. Cells were grown in 10 mL of modified ATCC 1754 PETC medium with either 60 mM fructose and 10 mM sodium nitrate or 200 kPa (80% H_2:20% CO_2) for heterotrophic or autotrophic growth, respectively. All measurements were collected in biological triplicates. When the OD_{600} reached mid- to late-log phase (after approximately 48 h), the cultures

were rapidly cooled in an ice-water bath. The cells were collected via centrifugation at 8,000 × *g* for 30 min at 4°C and rapidly frozen in liquid nitrogen before storage at −80°C.

RNA isolation was performed using a NucleoSpin® RNA isolation kit (Macherey-Nagel; Duren, Germany) according to the manufacturer protocol. Quality of total RNA samples was verified using an Agilent 2100 Bioanalyzer (Agilent Technologies; Palo Alto, CA). rRNA was removed using a Ribo-Zero rRNA removal kit for bacteria (Illumina, San Diego, CA) according to the manufacturer protocol and a modified RNeasy MinElute Cleanup Kit (Qiagen, Hilden, Germany) was used to recover mRNA. Sample quality was again verified using an Agilent 2100 Bioanalyzer (Agilent). Sample RNA integrity number (RIN) was typically between 8-10. Library preparation was performed by the Functional Genomics lab at the University of California, Berkeley, using the TruSeq RNA library prep kit (Illumina) according to the manufacturer's protocols. cDNA was sequenced using Illumina HiSeq 4000 PE100 at the UC Berkeley Vincent J. Coates Genomic Sequencing Laboratory. Gene-level quantification and differential expression analysis was performed using the CLC genomics workbench (Qiagen).

Expression of Strep-tagged ACS/CODH and candidate accessory proteins. LB (1 L) in a 2.5 L Ultra Yield flask was inoculated to an OD_{600} of 0.06 with an overnight culture of freshly transformed *E. coli* BL21(DE3) containing the appropriate expression plasmid. The cultures were anaerobically grown at 37°C in 250 mL serum bottles with shaking at 250 rpm to OD_{600} of 0.7 to 1.0. Cultures were cooled on ice for 20 min, after which protein expression was induced with IPTG (0.5 mM) and overnight growth at 16°C. Cell pellets were harvested by transferring saturated culture to O-ring containing centrifuge bottles in an anaerobic chamber. Centrifugation was performed at 6,000 × *g* for 7 min at 4°C and pellets were stored at -80°C in 5 mL anaerobic vials.

Anaerobic purification of Strep-tagged ACS/CODH. All columns and serum bottles that may contact ACS/CODH were washed with detergent to remove all traces of grease and surface dust left over from production. After rinsing thoroughly with ddH_2O, bottles and columns were submerged in 10% (*v/v*) nitric acid and soaked overnight at 50°C. Buffers were treated with Chelex® 100 molecular biology grade resin (Bio-Rad) for 1 h and made anaerobic by sparging the buffer with N_2 for 20 min before transferring buffers to acid-washed 250 mL serum bottles. Bugbuster® protein extraction reagent (Novagen, Gibbstown, NJ) was made anaerobic using 3 × freeze-pump-thaw cycles on a Schlenk line and transferred to an acid-washed serum bottle. Sarstedt® tubes were used to collect elution fractions for storage.

Frozen cell pellets were thawed and resuspended at 5 mL per g of cell paste with anaerobic Bugbuster supplemented with TCEP (2 mM) and one cOmplete EDTA-free protease inhibitor tablet. To reduce lysate viscosity, 25 units of Benzonase® Nuclease (EMD Millipore) were added per 1 mL Bugbuster reagent. The cell suspension was stirred at room temperature for 10-20 min. Lysate was centrifuged at 12,000 × *g* for 30 min at 4°C to separate soluble and insoluble fractions. Soluble lysate was then loaded onto a Strep-Tactin® column by gravity flow in the anaerobic chamber.

The column was washed with 10 column volumes of buffer B (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM TCEP) before eluting with Buffer C (Buffer B with 2.5 mM desthiobiotin). Fractions containing protein were pooled by A_{280nm} and concentrated using a 15 mL Amicon Ultra spin concentrator (MilliporeSigma, Burlington, MA) in an anaerobic chamber. Protein concentrations were estimated using the $\lambda_{280\ nm}$ calculated by ExPASY ProtParam 4 (97345 $M^{-1}cm^{-1}$).

ICP-MS analysis. Approximately 10 μM of purified Strep-ACS was desalted using a PD-10 desalting column (GE Healthcare, Chicago, IL) and eluted into a Starstedt® tube (Fisher Scientific). Then, samples were diluted into an equal volume of standard solution (4% HNO_3, 40 ppb Ga) to make a final volume of 2.4 mL. Samples were digested overnight at room temperature and centrifuged the next day at 12,000 × *g* to remove precipitates. 2.2 mL of supernatant was transferred into new Starstedt® tubes for ICP-MS analysis.

A standard curve of Ni, Cu, Zn and Fe was run every 20 samples with approximately 5-6 standards/curve. Concentrated standard stocks were obtained from Inorganic Ventures (Christiansburg, VA) and diluted into trace-metals grade nitric acid at a 2% final concentration. The internal standard was 20 ppb Ga, which was made using a 1 ppm Ga solution. Reagents, acids, plasticware, glassware and buffers were trace-metals grade, where possible. All dilutions were done with a microbalance for precision, so that back-calculation of the correct amount of metal was possible. All final dilutions were done on the same day as analysis in order to avoid protein precipitation. After preparation, samples were injected on to ICP-MS.

Plasmid digest analysis by HPLC-qTOF-MS. All procedures were performed as described in published protocols, with some modifications [*31*]. *pMTL83151-ΔpyrF::Km*^{R} and *pBAD33-mt1671-mt1672-mt2281* were transformed and cloned into *E.coli* DH10B or GM272 in the presence or absence of 0.05% (*w/v*) L-arabinose and grown overnight. Plasmids were extracted from *E. coli* and dried to dryness in a speed-vac without heat. 1 mL of neat formic acid was added to each tube and place in an oven at 140°C. Hydrolysis was performed for 24 h. After hydrolysis, samples were again dried to dryness in a speed-vac and samples were resuspended in water for HPLC-qTOF-MS analysis.

Samples were injected onto an Agilent 1290 HPLC equipped with an auto-sampler, a Phenomenex Rezex-ROA Organic Acid H^+ column (150 × 4.6 mm) and with a Carbo-H^+ Security Guard cartridge. 0.5% (*v/v*) formic acid was used as the mobile phase with a flow rate of 0.3 mL/min and the column temperature set to 50°C. dNTPs were quantified using mass spectrometry on an Agilent 6530 Q-TOF LC/MS with ESI source. Between 10-20 min, the following *m/z* values were monitored: *m/z* 330.06 (deoxyadenosine triphosphate), *m/z* 309.04 (deoxycytidine triphosphate), *m/z* 321.04 (deoxythymidine triphosphate), *m/z* 346.06 (deoxyguanidine triphosphate), *m/z* 344.08 (deoxymethyladenosine triphosphate), *m/z* 324.06 (deoxymethylcytidine triphosphate). Samples were quantified relative to a standard curve made using a nucleotide mix of 25, 50, 100, 250, 500 and 1000 mg/L prepared in MilliQ water.

Analytic restriction enzyme digest of plasmid DNA. Isolated plasmid DNA was mixed with restriction enzyme (DpnI or HaeIII, depending on the methyltransferase used for plasmid methylation) and the appropriate reaction buffer to a 50 μL final volume. The reaction was incubated at 37°C for one hour and purified using a PCR cleanup kit (Qiagen). 6 × gel loading dye (New England Biolabs) was added to the sample and 10 μL of each reaction mixture was loaded into the wells of a 2% (*w/v*) agarose gel, along with a DNA 1 kB plus ladder dye (New England Biolabs) in a separate well. Gels were placed in an electrophoresis apparatus and run at 120V for 25 min. Each gel was imaged using a UV light source in a Universal Hood II (Bio-Rad).

Acetyl-CoA/3′-dephospho-CoA exchange assay. The reaction was performed as described previously with minor modifications [*32*]. A reaction mixture was made using 50 mM MOPS, 100 mM KCl, 50 μM aquocobalamin and 151 μM Ti^{3+} nitrilotriacetic acid (NTA). To prepare the Ti^{3+} NTA solution, 27.5 μL of 30% (*w/w*) Ti^{3+} chloride in 2N HCl reagent (Acros Organics; Geel,

Belgium) was added to 1 mL of 0.275 M disodium NTA in 0.5 M Tris-HCl (pH 8.0) and degassed with N_2 using a Schlenk line. To dilute the stock to a working solution, 10 µL of the Ti^{3+} NTA stock was added to 310 µL of water, degassed degassed with N_2 using a Schlenk line, on the day of the assay.

The reaction mixture was incubated at 25°C for 20 minutes to allow aquocobalamin to be fully reduced. Then, 100 µM of Ti^{3+} NTA was added to the mixture, followed by 250 µM 3′-dephospho-CoA and 1 mM TCEP. The reaction was initiated with 250 µM acetyl-CoA. At time points of 0.5, 1, 2, 3, 4, 6, and 8 min, 55 µL of reaction mixture was removed and mixed with 55 µL of stop solution (2 mM Ti^{3+} NTA in 0.5 M sodium citrate pH 4.0) dispensed in anaerobic vials. Mixtures were frozen in liquid nitrogen. Samples were thawed and centrifuged at 18,000 × *g* for 10 min. The supernatant was aspirated and used for HPLC analysis.

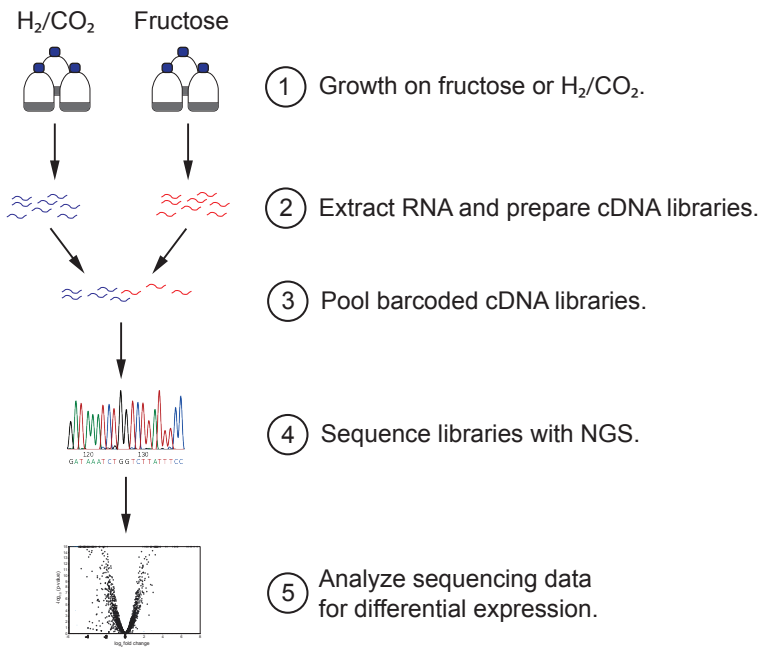
Analysis was performed using reverse phase HPLC-UV using an Agilent 1290 HPLC equipped with an autosampler, a Phenomenex Rezex-ROA Organic Acid H^+ column (150 × 4.6 mm) with a Carbo-H^+ Security Guard cartridge. After thawing, samples were immediately injected to minimize exposure to air. Aqueous component was 0.5% (*v/v*) formic acid (solvent A). Organic solvent was acetonitrile (solvent B). Linear gradient segments were initiated immediately upon injection:

Time	Solvent B (%)
0	0
5	10
20	28.7
25	100

Absorbance at 260 nm was monitored and quantified using a standard curve of each CoA using 25, 50, 100, 250, 500 and 1000 mg/L prepared in MilliQ water.

2.3. Results and discussion

Differential expression in heterotrophic or autotrophic conditions. In the model acetogen, *Moorella thermoacetica*, expression of genes from the carbonyl branch of the Wood-Ljüngdahl pathway can be regulated by the final electron acceptors available to the organism [*33*]. For instance, the presence of nitrate, an electron sink for glycolysis, has been shown to downregulate expression of the carbonyl branch in low CO_2 conditions. If the expression of carbonyl branch enzymes can be regulated in response to the presence of nitrate or CO_2, it may be possible that the expression of metallochaperones responsible for ACS/CODH maturation are also regulated by these factors. Based on this hypothesis, *M. thermoacetica* was grown either autotrophically on H_2/CO_2 or heterotrophically on fructose and sodium nitrate to compare RNA expression profiles from each condition (*Scheme 2.3*).



Scheme 2.3 RNA-seq experimental workflow for Moorella thermoacetica. *The differential expression analysis pipeline requires five key steps. The process begins with heterotrophic growth on either 60 mM fructose (with 10 mM sodium nitrate) and an inert nitrogen atmosphere or autotrophic growth with 275 kPa H_2/CO_2 (80%, 20%) without any organic carbon source. After growth to mid- to late-log phase, RNA was extracted and cDNA libraries were prepared with unique indexed adaptors for each sample. Barcoded samples were then pooled and sequenced with NGS. The resulting data was then mapped to reference genomes, transcript abundance was counted and compared across conditions to identify genes that were significantly upregulated or downregulated in either condition.*

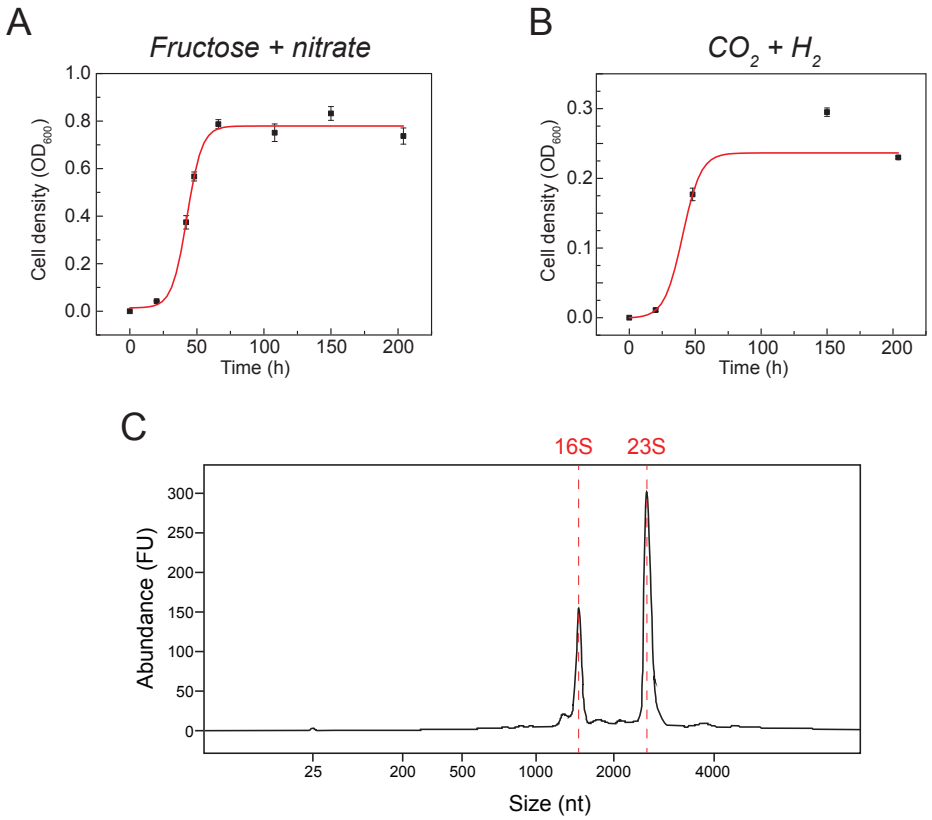


Figure 2.1 Evaluation of RNA-seq library quality. *M. thermoacetica was grown in heterotrophic or autotrophic conditions for extraction of RNA for sequencing. (A) Growth curve of M. thermoacetica cultured heterotrophically in 60 mM fructose and 10 mM sodium nitrate. (B) Growth curve of M. thermoacetica cultured autotrophically in 275 kPa 80% H_2:20% CO_2. Absorbance values were measured at 600 nm, showing that the culture reached late-log phase between 50-60 h. (C) RNA was extracted and analyzed by Agilent Bioanalyzer. The chromatogram shows sharp rRNA peaks and a 23S/16S ratio of ~2, indicating that the RNA quality is fairly intact and non-degraded. This chromatogram is representative of all total RNA extracted used for RNA sequencing.*

To prepare RNA, a gram-positive RNA extraction kit with a commercial lysozyme solution was required to effectively lyse the bacterial cell wall. After isolating RNA and confirming the quality of the RNA library (*Figure 2.1*), cDNA libraries were prepared for NGS using standard protocols at the UC Davis Functional Genomics Center. The resulting datasets were analyzed for differential expression to reveal genes that may have been downregulated during growth on an organic carbon source and nitrate. This analysis revealed 251 genes that have 2-fold or higher expression in autotrophic growth with a p-value < 0.05 (*Figure 2.2*). These genes were used as a starting panel of candidate accessory proteins.

Candidate hit refinement with bioinformatic analysis. This panel was then further refined by first using a more stringent statistical significance cutoff (Bonferroni p-value < 0.05), reducing the shortlist to 183 candidates. Next, we turned to the use of UniProt annotation scores to identify proteins with known, non-metallochaperone functions. Using the UniProt database, genes assigned to functions unrelated to metallochaperone activity with annotation confidence scores of 3/5 or more were removed from the shortlist, leaving 80 candidates (*Figure 2.2*). To further refine the candidates to a more experimentally tractable set, we attempted to identify conserved features among known accessory proteins for other metalloenzymes.

In the case of urease from *Helicobacter pylori*, appropriate metallation of the binuclear nickel active site of the enzyme requires three proteins (UreE, UreF and UreG) to form metal loading complexes [*34,35*]. GTP/ATP-hydrolysis is used to drive conformational changes that allow for the dynamic formation and disassembly of these transient complexes, enabling the protection and appropriate transfer of metal cargo [*36-38,25*]. Therefore, urease metallochaperones are ATP-dependent and contain conserved P-loop domains. A similar observation has been made for [NiFe] hydrogenase from *E.coli*, which requires the activity of HypA and HypB metallochaperones, where HypB is an ATPase that hydrolyzes ATP and induces conformational changes in HypA through complex formation, which leads to the formation of a Ni binding site [*39*]. Also, numerous metal-binding proteins for zinc, nickel and other transition metals has been shown to contain a motif with clustered cysteine residues (e.g. CxxC and CxxCxnCPxC, where x can be any amino acid) [*40,41*].

Using these known features of other metallochaperones, the NCBI conserved domain (CD) search tool was used to remove candidates that did not have a P-loop domain or common metal-binding motifs (*Figure 2.2*). This set of analyses generated thirteen final candidates (*Table 2.1*). Studies of accessory proteins for other metalloenzymes indicate that the genes corresponding to these auxiliary factors can be clustered in a genetic operon with the gene(s) of their target enzyme and expressed at similar transcript levels [*42,43*]. Using this observation, we performed a multi-geneBLAST genomic analysis that searches for syntenic genes within the ACS gene cluster amongst all known organisms containing the Wood-Ljungdahl pathway. Only two genes were consistently found: *acsF* and *cooC*. This analysis lends further support to the possibility that AcsF and CooC may indeed be the only metallochaperones required for ACS/CODH metalation. However, this would be an unusual observation, since urease, the most well-studied metalloenzyme with a binuclear Ni active site like ACS, requires at least three metallochaperones to properly load metal into the cognate enzyme. Furthermore, it is still unclear whether additional accessory proteins are necessary for trafficking nickel through the cytoplasm or facilitating the formation of a nickel loading complex for ACS/CODH.

Besides *acsF* and *cooC*, a hydrogenase maturation gene cluster was also found in our RNA seq analysis (Moth_2176-Moth_2182). The genome of *M. thermoacetica* encodes an Ech-type [NiFe]

hydrogenase [*43*], which is homologous to the hydrogenase 4 complex in *E.coli* [*44*]. This complex catalyzes the oxidation of H_2 to produce Fd_{red}^{2-}, which may play a role in energy metabolism during *M. thermoacetica* growth on H_2/CO_2. However, it has also been shown that maturation proteins for [NiFe] hydrogenase may have roles in the metallation of more than one metalloenzyme, such as the case of the HypA in *Helicobacter pylori*, which forms a ternary complex with UreE2, facilitating the trafficking of HypA-bound Ni^{2+} to UreE, which is subsequently transferred to urease [*44,45,26*]. The last two genes from the list of top candidates, Moth_1408 and Moth_1409, are putatively involved in cobalamin biosynthesis. These genes may be necessary for the biosynthesis of cobalamin for CFeSP, an essential gene in the Wood-Ljungdahl pathway. Since it does not appear to have homology to any other nickel metallochaperones, it is unlikely that this cluster of genes is involved in nickel trafficking.

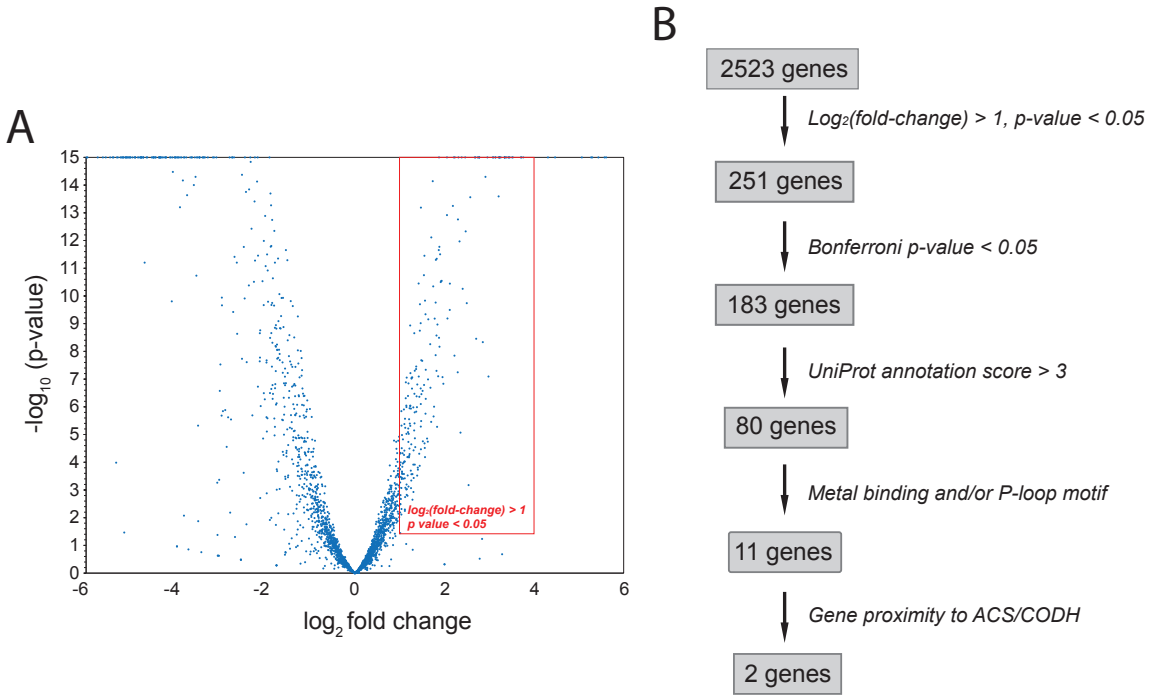


Figure 2.2 RNA-seq analysis and bioinformatic filters. *Analyzing the RNA-seq dataset through the lens of several bioinformatic filters leads to eleven metallochaperone candidates, with two candidates at the top of the list. (A) Volcano plot representation of differential expression analysis in M. maripaludis grown in H_2/CO_2 vs. fructose. Red lines indicate boundaries of significantly increased or decreased expression ($p<0.05$, log_2(fold-change) is greater than 1), respectively. (B) Initial log_2(fold-change) and significance cutoffs were used to reduce the set of candidate metallochaperones to 251 genes. From there, a series of bioinformatic filters were applied to the RNA-seq dataset using known features of other metallochaperones and online bioinformatics tools. After using a more stringent p-value cutoff (Bonferroni p-value), a high confidence UniProt annotation score to remove genes with unrelated functions, and the NCBI conserved domain (CD) search tool to keep genes with metal binding domains and/or P-loop motifs, a final list of eleven top candidates remained. From this list, two genes were identified with proximity to the ACS/CODH gene cluster: acsF and cooC.*

Gene ID	Name	Fold-change	*p*-value
Moth_1199	AcsF	9.39	1.0×10^{-16}
Moth_1204	CooC	9.23	1.0×10^{-16}
Moth_2176	Putative hydrogenase formation protein	13.28	1.0×10^{-16}
Moth_2177	Putative hydrogenase formation protein	9.32	1.0×10^{-16}
Moth_2178	Putative hydrogenase assembly chaperone	12.14	3.7×10^{-13}
Moth_2179	Putative carbamoyltransferase	11.52	1.09×10^{-13}
Moth_2180	Putative hydrogenase accessory protein	10.84	1.29×10^{-11}
Moth_2181	Putative nickel incorporation protein	7.54	1.0×10^{-13}
Moth_2182	Putative hydrogenase 3 maturation peptidase	5.65	4.67×10^{-7}
Moth_1408	Cobyrinic a,c-diamide synthase	22.14	1.0×10^{-16}
Moth_1409	Cobyrinic a,c-diamide synthase	39.32	1.0×10^{-16}

Table 2.1 Candidate metallochaperones for ACS/CODH in M. thermoacetica. *A list of the top eleven candidates identified through differential expression analysis of M. thermoacetica and a series of bioinformatic filters. Fold-change of expression shows that all of the top metallochaperone candidates were expressed at least 5-fold higher in autotrophic growth conditions compared to heterotrophic growth conditions. AcsF and CooC are known metallochaperones for the ACS/CODH complex. Multiple putative hydrogenase maturation proteins, including Moth_2176, Moth_2177, Moth_2178, Moth_2179, Moth_2180, Moth_2181 and Moth_2182 are potentially involved in maturation of [Ni,Fe] hydrogenases required for autotrophic growth, but may also have some auxiliary role in ACS maturation, similar to other examples of known multifunctional metallochaperones [44,45]. Moth_1408 and Moth_1409 are putative accessory proteins for corrinoid proteins and are unlikely to play a role in nickel metalloenzyme maturation.*

Towards making genetic knockouts in *M. thermoacetica*. There is evidence to show that knockouts of metallochaperones lead to a loss of function in the cognate metalloenzyme [*47*]. Therefore, we hypothesized that genetic knockouts of accessory protein candidates that are involved in ACS/CODH metallation would lead to a loss of ACS/CODH function, which would inhibit autotrophic growth. If a candidate accessory protein knockout eliminates autotrophic growth without eliminating the ability of *M. thermoacetica* to grow in an ACS/CODH-independent manner using fructose and nitrate, this would provide supporting evidence for the possible role of that candidate as an accessory protein for ACS/CODH.

In order to knockout candidate accessory proteins, *M. thermoacetica* must be transformed with a donor plasmid containing an antibiotic resistance marker flanked by homology arms that target the candidate gene. Although *M. thermoacetica* transformation has been successfully performed by the Nakashimada lab at Hiroshima University, a customized electroporation apparatus was used. To circumvent the need for this custom equipment, some modifications were made to the original protocol. Firstly, the pMTL83151 plasmid was chosen as a donor plasmid for transformation because of its Gram-positive origin of replication and reported use in Clostridial species that are relatively closely related to *M. thermoacetica* [*48*]. Since the restriction-modification system in *M. thermoacetica* has been reported to pose a barrier to stable plasmid transformants, donor plasmids are typically pre-methylated before transformation. According to the Nakashimada lab protocol, three methyltransferases from *M. thermoacetica* (*mt1671*, *mt1672*, and *mt2281*) were inserted into a second shuttle plasmid and place under control of an arabinose-inducible promoter. Expression of these genes was induced after transformation into *E.coli* with native methyltransferases knocked out (ie. GM272) while co-expressing the donor plasmid. This protocol is expected to produce methylated donor plasmids that can then be purified and used for transformation of *M. thermoacetica* [*48-50*].

Therefore, we constructed this second plasmid and co-expressed it with our pMTL83151 donor plasmid in *E.coli* to obtain pre-methylated donor plasmid for transformation. The methylation pattern produced by this protocol has not been documented, so we first attempted to detect methylation using a standard DpnI digest assay (*Figure 2.3*). The assay was performed on purified plasmid DNA from both GM272 and DH10B *E.coli* strains with and without the methylation plasmid (pBAD33-*mt1671-mt1672-mt2281*) or arabinose to induce methyltransferase expression. The results of this assay suggested that purified plasmid DNA was not digested by DpnI. It's possible that the plasmid methylation may not have been appropriately recognized by DpnI, which typically cleaves between $G^{m}A$ and TC.

To test the possibility that the plasmid had indeed been methylated but not recognized by DpnI, another method was used to measure methylation patterns. Specifically, plasmid DNA was purified from *E.coli* after induction of methyltransferase expression using arabinose and digested with a formic acid treatment at high temperatures. After digestion, the mixture was analyzed with HPLC-MS and deoxymethyladenine triphosphate was detected from plasmid extracted from GM272 in the presence of pBAD33-*mt1671-mt1672-mt2281* and arabinose (*Figure 2.3*). This result suggests that the methyltransferases were expressed and methylated the donor plasmid.

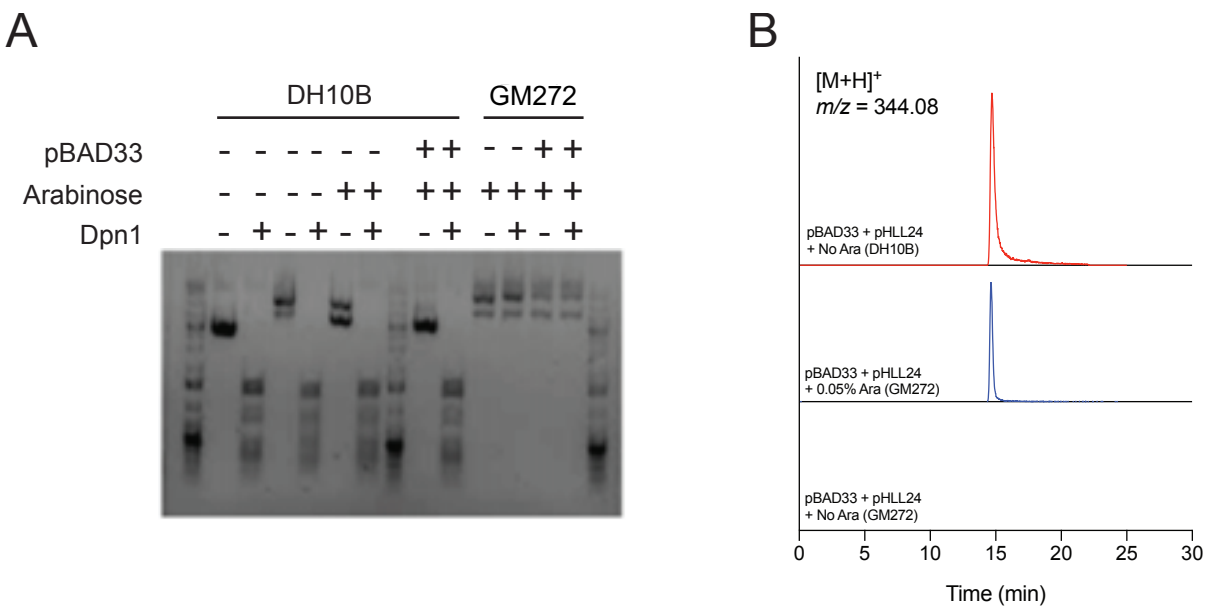


Figure 2.3 Pre-methylation of donor plasmid for M. thermoacetica transformation. *We attempted to pre-methylate the pHLL24 plasmid using a pBAD33 plasmid containing mt1671, mt1672 and mt2281, which showed methylation using HPLC-QTOF, but not by DpnI digest. (A) Plasmid digest gel assay showing DpnI digestion patterns of pHLL24 in the presence or absence of pBAD33-mt1671-mt1672-mt2281 and 0.05% arabinose using E.coli strain DH10B or GM272 (no native methyltransferases). Gel digest shows that the plasmids were digested by DpnI in DH10B independent of the presence of pBAD33 or arabinose, while DpnI digestion was not observed in GM272 with any combination of pBAD33 or arabinose. B) HPLC-QTOF analysis of nucleotides after digesting extracted plasmids with strong acid shows the presence of a deoxymethyladenine triphosphosphate (dmAMP; m/z = 344.08) in DH10B without any arabinose present, but it is seen in GM272 only when pBAD33 was present and arabinose was supplemented, suggesting that expression of methyltransferase genes from pBAD33 was necessary to see the dmAMP peak.*

However, repeated transformation attempts failed to produce the expected transformants on antibiotic selective media. Some reasons for this include an electroporator without the appropriate electroporation parameters, the use of a *M. thermoacetica* strain with low recombination activity and insufficient methylation. To address the first possibility, we considered adapting a more generic *Clostridium* protocol with standard electroporation parameters. This protocol uses the Φ3T I methyltransferase to pre-methylate donor plasmids at the inner cytosines of the sequences GGCC and CGNGC [*30*]. Using *E.coli* GM272, we showed that the expected HaeIII digest pattern of our donor plasmid was observed, but only when Φ3T I methyltransferase was also expressed from plasmids pAN1 or pAN3 (*Figure 2.4*). Using this purified plasmid DNA, a modified transformation protocol was established to transform *M. thermoacetica* with pMTL83151-*ΔpyrF::Km*^{R}. An analytical PCR amplification assay was used to show that a kanamycin resistance marker was successfully transformed and integrated into the *pyrF* genetic site (*Figure 2.5*).

Heterologous expression of ACS/CODH and candidate accessory factors in *E. coli*. In parallel with native acetogen studies, the top metallochaperone candidate, *AcsF*, was heterologously co-expressed with Strep-tagged ACS and CODH. After growth, cells were harvested and ACS/CODH was purified by affinity purification (*Figure 2.6*). The A cluster in active ACS is known to contain a binuclear nickel site in addition to a [4Fe-4S] cluster. Metal content analysis was performed using inductively-coupled plasma mass spectrometry (ICP-MS) to interrogate metal content of purified ACS. This analysis showed that Ni content in ACS expressed alone was no higher than buffer controls. However, iron content appeared to be 3.75 ± 0.23 Fe/ACS without AcsF and 3.89 ± 0.24 Fe/ACS in the presence of AcsF (*Figure 2.6*). One explanation is that iron loading into ACS can be performed by native iron loading factors in *E.coli*, as suggested in previous studies of recombinant ACS/CODH expression [*18*].

When AcsF was coexpressed with ACS/CODH, an increase in the nickel content of ACS was observed, but only to 0.30 ± 0.01 Ni/ACS molecule, although 2 Ni/ACS was expected (*Figure 2.6*). There are multiple possible explanations for this result. One possibility is that AcsF is insufficient for proper metal loading into ACS, which may suggest additional loading factors are necessary. However, the nickel present in the ACS active site has been reported to be labile [*23*], and it is also possible that the metal was properly loaded into the active site of ACS *in vivo*, but leached out during the purification process and was displaced by other metal contaminants, such as copper or zinc. To address this possibility, purified ACS could be reconstituted in the presence of a stoichiometric excess of $NiCl_2$, desalted and purified using the original chromatography protocol. By comparing the metal content of both of these samples, it may be possible to ascertain the effect of the affinity purification protocol on metal content. Alternatively, it may be possible to more directly interrogate the *in vivo* or crude lysate activity of *E.coli* expressing ACS without measuring metal content to circumvent the need to purify metallated ACS and risk the loss of these metals during purification.

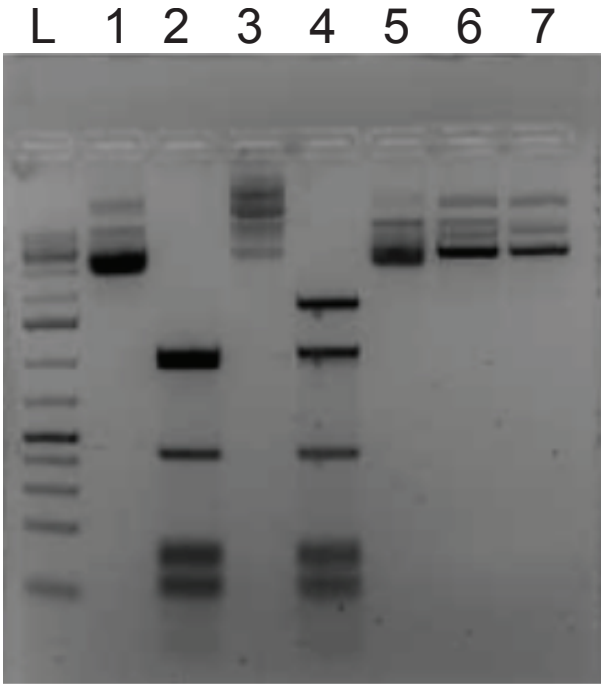


Figure 2.4 HaeIII digestion pattern of pre-methylated donor plasmids. *HaeIII digestion of pMTL83151-ΔpyrF::kan^r after co-transformation with pAN1 or pAN3 shows the absence of a pattern, suggesting that the pMTL83151 plasmid was successfully methylated by methyltransferases present on pAN1 or pAN3. L = ladder; 1 = pMTL83151 alone (E.coli DH10B strain); 2 = pMTL83151+ HaeIII (E.coli DH10B strain); 3 = pMTL83151 (E.coli GM272 strain); 4 = pMTL3151 + HaeIII (E.coli DH10B strain); 5 = pMTL83151 + pAN1 (E.coli GM272 strain); 6 = pMTL83151 + pAN1 + HaeIII (E.coli GM272 strain); 7 = pMTL3151 + pAN3 + HaeIII (E.coli GM272 strain).*

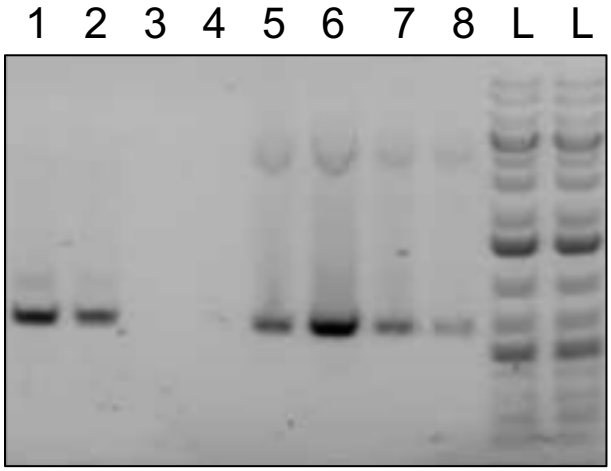


Figure 2.5 Confirmation of M. thermoacetica transformation with modified Clostridium protocol. *M. thermoacetica* transformants and control samples were prepared, and PCR was performed to show that the kanamycin resistance marker had successfully been recombined at the target *pyrF* site in *M. thermoacetica* transformants but not in wild type (*wt*) *M. thermoacetica*. Transformation was performed using a modified version of a general *Clostridia* transformation protocol. Expected PCR product size (652 bp), 1 = pMTL83151 (T_m 51°C); 2 = pMTL83151 (T_m 55°C); 3 = *wt M. thermoacetica* gDNA preparation protocol; 4 = *wt M. thermoacetica* DNA mini prep preparation protocol; 5 = *ΔpyrF* transformant 1; 6 = *ΔpyrF* transformant 2; 7 = *ΔpyrF* transformant 3; 8 = *ΔpyrF* transformant 4; L = 1 kb plus ladder.

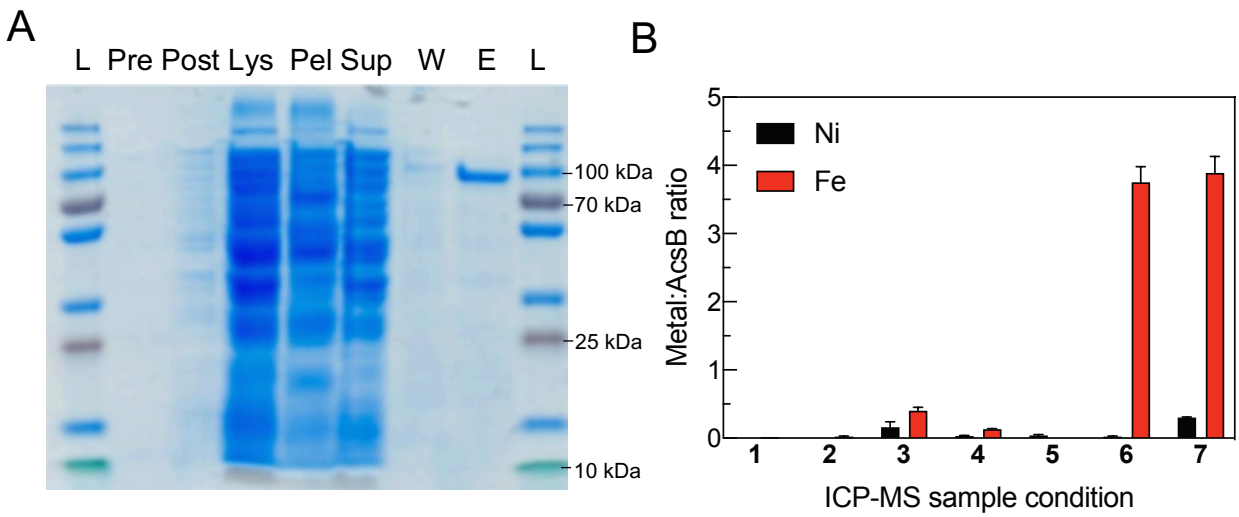


Figure 2.6 Strep-ACS purification and ICP-MS analysis. *Strep-tagged ACS was purified and metal content analysis was performed with and without AcsF, showing partial nickel loading and complete iron loading. A) Acetyl-CoA synthase (ACS) was Strep-tagged at the N-terminus and affinity purified. Coomassie stain shows a band eluted with the same size as ACS (83.4 kDa). L = Prestained PAGE ladder; Pre = pre-induction E.coli; Post = post-induction E.coli; Lys = lysate; Pel = pellet; Sup = supernatant; W = wash; E = elution. B) ICP-MS analysis of ACS and control samples shows that ACS has partial nickel loading (0.30 ± 0.01) in the presence of co-expressed AcsF. Iron loading appears to be at expected levels even in the absence of AcsF (3.75 ± 0.23), suggesting that native E.coli factors may be sufficient for the loading of the [4Fe-4S] cluster at the A site. 1 = MilliQ only; 2 = MilliQ + HNO_3 + 20 ppb Ga; 3 = Buffer (no desalting column); 4 = Buffer (after desalting); 5 = AcsF only; 6 = ACS only; 7 = ACS co-expressed with AcsF.*

Methods for measuring ACS/CODH activity. Two diagnostic isotopic exchange reactions have been used to measure the unique ability of the ACS/CODH complex to assembly acetyl groups from one-carbon units [*15,51,52*]: the carbonyl exchange assay and the CoA/acetyl-CoA exchange assay (*Figure 2.7*). In the case of the carbonyl exchange assay, ACS is used to catalyze the exchange of the carbonyl group acetyl-CoA with free CO [12]. Multiple steps are required for this carbonyl exchange to take place. Firstly, the enzyme must react with acetyl-CoA to generate an acetyl intermediate at the A cluster under reducing conditions. Then, the acetyl C-C bond must be broken. From this fragmented state, CO can dissociate and exchange with free CO in the headspace of the reaction vial. Conversely, unlabeled CO in the atmosphere may be converted to acetyl-CoA, which dilutes the molar fraction of labelled acetyl-CoA. If the ACS/CODH complex is used, it may be preferable to monitor CO_2 exchange since CO can be a poor substrate owing to the tight coupling between the A cluster of ACS and C cluster of CODH [*6*]. Another assay that could be performed is the CoA/acetyl-CoA exchange assay. In this assay, the exchange of labelled CoA and acetyl-CoA requires that enzyme is capable of carbon-sulfur bond cleavage [*52*]. Similar to the carbonyl exchange assay, the molar fraction of labelled compound will be diluted. The main difference between the two assays lies in the fact that we are only observing acetyl-intermediate formation and dissociation in the CoA/acetyl-CoA exchange assay, while the CO/acetyl-CoA exchange assay also requires C-C bond dissociation. Of these two, the CoA/acetyl-CoA exchange assay is simpler to set up and uses soluble reactants, so this assay was attempted first. A control experiment was performed using 3'-dephospho-CoA (instead of labelled CoA) and acetyl-CoA in the absence of enzyme. The reaction appeared to reach completion within ten minutes, suggesting a high background level of activity in the attempted conditions (*Figure 2.7*). In addition, there was an unexplained loss of 3'-dephospho-CoA signal during the course of the reaction. The reason for this is unclear but may be related to the buffer conditions of the media and unexpected mixed sulfides that may have formed. Further assay development will be required to reduce loss of 3'-dephospho-CoA, reduce background signal and measure enzymatic activity.

To probe the capability of the ACS to complete the catalytic cycle, the acetyl-CoA synthesis assay could be used (*Figure 2.7*). Using purified ACS/CODH that is reconstituted *in vitro* with the necessary accessory proteins and other factors, the entire catalytic cycle is required to produce acetyl-CoA. Instead of using CFeSP, MeTr and the methyl branch pathway, methylcobinamide can be used as a synthetic methyl donor [*53*]. This compound structurally resembles the corrinoid cofactor in CFeSP, bypassing the need to use purified CFeSP. Ti^{3+} NTA is used as a reducing agent to complete the catalytic cycle. Reverse phase HPLC analysis can be used to quantify CoA and acetyl-CoA over time, and initial rates could be obtained from first order fits. Finally, ACS/CODH can be expressed in *E. coli* and *in vivo* activity may be measured. The forward reaction may be difficult to measure because it requires a methyl donor source. Furthermore, the standard free energy of CO_2 reduction is +83.68 kJ/mol, even with reduced 2[4Fe-4S] ferredoxin. To increase the thermodynamic driving force for the forward direction, the ratio of substrates to products must be high. Typically, this would require high CO_2 pressure, but also a high ratio of reduced:oxidized ferredoxin. These may be challenging in a heterologous host where much of the metabolic infrastructure required for this may be missing. Conversely, the reverse reaction involving the decarbonylation of acetyl-CoA would only require acetyl-CoA as a substrate. Acetyl-CoA produced during glycolysis while fermenting labelled sugars can be decarbonylated to produce labelled CO_2, suggesting a possible method for *in vivo* characterization of ACS activity.

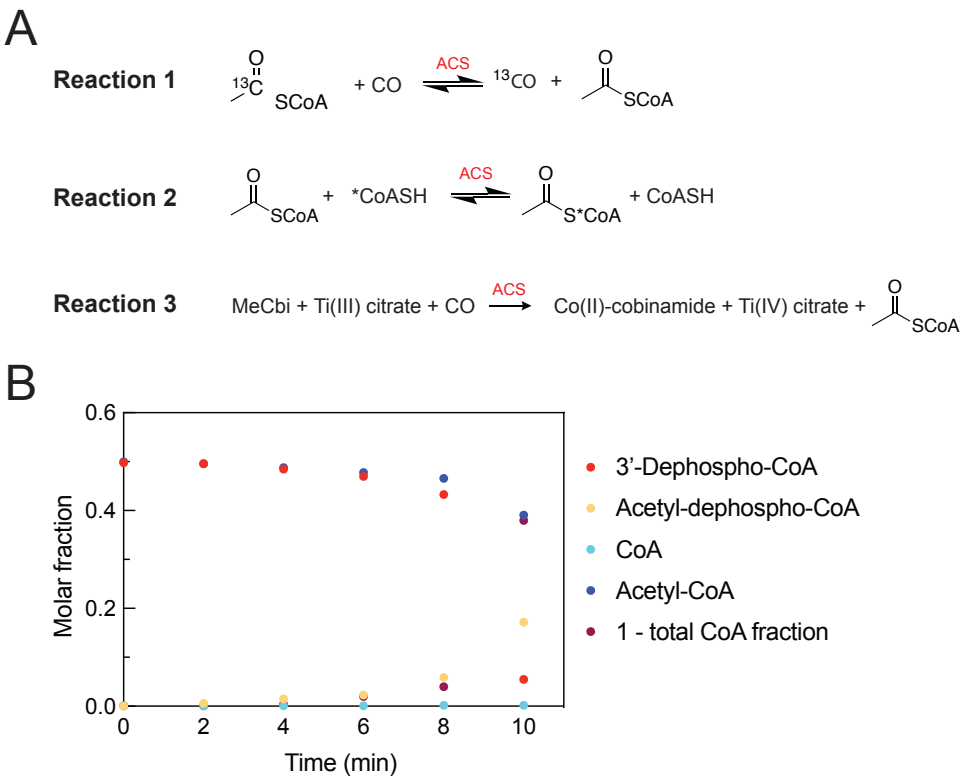


Figure 2.7 Acetyl-CoA synthase activity assays. *An overview of in vitro acetyl-CoA synthase activity and a no-enzyme control reaction for the acetyl-CoA/3'-dephospho-CoA exchange assay. (A) The three major assays used to measure in vitro activity of ACS. In the acetyl-CoA/CO exchange assay (reaction 1), labelled acetyl-CoA or CO is incubated in a reaction mixture in the presence of ACS enzyme, allowing for coordination of the acetyl group by the catalytic nickel, C-C bond fragmentation and CO release in the reverse direction, and acetyl-CoA formation in the forward direction. In the acetyl-CoA/CoA exchange assay (reaction 2), only the coordination of the acetyl-group and transfer to a free CoA are involved. The CoA group may be isotopically labelled. Alternatively, 3'-dephospho-CoA has been shown to be an acetyl group acceptor in this exchange assay, which circumvents the need to use labelled-CoA. In the acetyl-CoA synthesis reaction (reaction 3), methylcobinamide (MeCbi) is used as a methyl donor, and titanium citrate or titanium NTA can be used as a reductant. CO is added to the reaction and acetyl-CoA is produced in the presence of enzyme. (B) A no enzyme control for the acetyl-CoA/3'-dephospho-CoA assay was performed to observe the reaction rate in the absence of enzyme. The reaction appeared to near completion after ten minutes. However, a loss of 3'-dephospho-CoA was observed over time, suggesting possible degradation during the course of the reaction.*

2.4. Conclusions

ACS/CODH reconstitution in a genetically tractable recombinant host, such as *E. coli,* could open up new possibilities for studying ACS maturation and interactions with its partner metallochaperones. Furthermore, robust expression and activity of the complex could also pave the way towards engineering chemoautrotrophic metabolism into *E. coli*. The key knowledge gap that has thus far prevented progress towards these goals has been the unsuccessful effort to reconstitute the A cluster of ACS *in vivo* in a recombinant host. Similar to other metalloenzymes, ACS is known to use metallochaperones to facilitate nickel loading into its active site. To identify novel metallochaperones, *Moorella thermoacetica*, the model acetogen, was grown under autotrophic and heterotrophic conditions to identify genes that were differentially expressed. From a list of 251 gene candidates, bioinformatic analysis was used to identify genes that shared important biochemical features with known metallochaperones. From this analysis, it was concluded that AcsF and CooC are two accessory proteins required for ACS and CODH. In addition, another 9 accessory proteins were identified as potential accessory factors.

To investigate whether these candidates are essential, genetic knockouts can be made in *M. thermoacetica*. Method development was conducted to pre-methylate the pMTL83151 plasmid for transformation, and methylation was detected using HPLC-MS. After multiple transformation attempts of *M. thermoacetica*, genetic knockout was demonstrated with the *pyrF* gene, demonstrating that our protocol may be used for making knockout strains for metallochaperone candidate validation. In parallel, we attempted to reconstitute ACS/CODH in *E.coli* by co-expressing the top ranking metallochaperone candidate, AcsF. With the expression of these accessory proteins, ACS was shown to have increased amounts of nickel per active site, although iron remained unchanged. Further work was done to establish an assay to measure the activity of the ACS/CODH complex *in vitro*, with the CoA/3'-dephospho-CoA exchange assay performed in the absence of enzyme. The methods dem proposed in this section pave the path toward measuring the activity of purified ACS/CODH and *in vivo* activity of ACS/CODH in a recombinant host.

2.5. References

1. F. E. Fontaine, W. H. Peterson, E. McCoy, M. J. Johnson, and G. J. Ritter, "A new type of glucose fermentation by *Clostridium thermoaceticum"*, *Journal of Bacteriology.* **1942,** *43*, 701–715.
2. S. W. Ragsdale, "Life with carbon monoxide", *Critical Reviews in Biochemistry and Molecular Biology*. **2004**, *39*, 165–195.
3. S. W. Ragsdale E. Pierce, "Acetogenesis and the Wood-Ljungdahl pathway of CO_2 fixation, *Biochimica et Biophysica Acta - Proteins and Proteomics*. **2008**, *1784*, 1873–1898.
4. P. S. Adam, G. Borrel, S. Gribaldo, "An archaeal origin of the Wood–Ljungdahl H_4MPT branch and the emergence of bacterial methylotrophy", *Nature Microbiology*. **2019**, *4*, 2155–2163.
5. T. Shanmugasundarams, H. G. Woods, "Interaction of ferredoxin with carbon monoxide dehydrogenase from *Clostridium thermoaceticum"*, *The Journal of Biological Chemistry*. **1992**, *267*, 897-900.

6. P. H. Wang, M. Bruschi, L. De Gioia, J. Blumberger, "Uncovering a dynamically formed substrate access tunnel in carbon monoxide dehydrogenase/acetyl-CoA synthase", *Journal of the American Chemical Society.* **2013**. *135,* 9493–9502.
7. H. Alonso, P. L. Cummins, J. E. Gready, "Methyltetrahydrofolate:corrinoid/iron-sulfur Protein Methyltransferase (MeTr): Protonation state of the ligand and active-site residues", *Journal of Physical Chemistry B*. **2009**, *113*, 14787–14796.
8. J. Seravalli, S. Zhao, S. W. Ragsdale, "Mechanism of transfer of the methyl group from (6S)-methyltetrahydrofolate to the corrinoid/iron-sulfur protein catalyzed by the methyltransferase from *Clostridium thermoaceticum*: A key step in the Wood- Ljungdahl pathway of acetyl-CoA synthesis", *Biochemistry.* **1999**, *38,* 5728–5735.
9. E. L. Hegg, "Unraveling the structure and mechanism of acetyl-coenzyme A synthase", *Accounts of Chemical Research* **2004**, *37*, 775–783.
10. M. Can, F. A. Armstrong, S. W. Ragsdale, "Structure, function, and mechanism of the nickel metalloenzymes, CO dehydrogenase, and acetyl-CoA synthase", *Chemical Reviews*. **2014**, *114*, 4149–4174.
11. L. G. Ljungdhal, "The autotrophic pathway of acetate synthesis in acetogenic bacteria", *Annual Review of Microbiology.* **1986**, *40*, 415–450.
12. R. Burton, M. Can, D. Esckilsen, S. Wiley, S. W. Ragsdale, "Production and properties of enzymes that activate and produce carbon monoxide", *Methods in Enzymology*. **2018**, *613*, 297–324.
13. S. W. Ragsdale, "Nickel and the carbon cycle", *Journal of Inorganic Biochemistry*, **2007**, *101*, 1657–1666.
14. R. Burton, M. Can, D. Esckilsen, S. Wiley, S. W. Ragsdale, "Production and properties of enzymes that activate and produce carbon monoxide", *Methods in Enzymology*. **2018**, *613*, 297–324.
15. Z. Hu, N. J. Spangler, M. E. Anderson, J. Xia, P. W. Ludden, P. A. Lindahl, E. Münck, "Nature of the C-cluster in Ni-containing carbon monoxide dehydrogenases", *Journal of the American Chemical Society.* **1996***, 118*, 830–844.
16. S. W. Ragsdale, H. G. Wood, W. E. Antholine, "Evidence that an iron-nickel-carbon complex is formed by reaction of CO with the CO dehydrogenase from *Clostridium thermoaceticum"*, *Proceedings of the National Academy of Science U. S. A.* **1985**, *82*, 6811–6814.
17. C. L. Drennan, J. Heo, M. D. Sintchak, E. Schreiter, P. W. Ludden, "Life on carbon monoxide: X-ray structure of *Rhodospirillum rubrum* Ni-Fe-S carbon monoxide dehydrogenase", *Proceedings of the National Academy of Science U. S. A.* **2001**, *98*, 11973–11978.
18. H. K. Loke, G. N. Bennett, P. A. Lindahl, "Active acetyl-CoA synthase from *Clostridium thermoaceticum* obtained by cloning and heterologous expression of *acsAB* in *Escherichia coli*", *Proceedings of the National Academy of Science U. S. A.* **2000**, *97*, 12530–12535.
19. E. D. Carlson E. T. Papoutsakis, "Heterologous expression of the *Clostridium*

carboxidivorans CO dehydrogenase alone or together with the acetyl coenzyme A synthase enables both reduction of CO_2 and oxidation of CO by *Clostridium acetobutylicum"*, *Applied and Environmental Microbiology.* **2017**, *83*.

20. H. D. Urbina, J. J. Silberg, K. G. Hoff, L. E. Vickery, "Transfer of sulfur from IscS to IscU during Fe/S cluster assembly", *The Journal of Biological Chemistry.* **2001**, *276,* 44521–44526.
21. R. Gasper, A. Scrima, A. Wittinghofer, "Structural insights into HypB, a GTP-binding protein that regulates metal binding", *The Journal of Biological Chemistry.* **2006**, *281*, 27492–27502.
22. S. Gencic, D. A. Grahame, "Nickel in subunit β of the acetyl-CoA decarbonylase/synthase multienzyme complex in methanogens: Catalytic properties and evidence for a binuclear Ni-Ni site", *The Journal of Biological Chemistry.* **2003**, *278*, 6101–6110.
23. W. Shin, P. A. Lindahl, "Discovery of a labile nickel ion required for CO/acetyl-CoA exchange activity in the NiFe complex of carbon monoxide dehydrogenase from *Clostridium thermoaceticum"*, *Journal of the American Chemical Society*. **1992**, *114,* 9718–9719.
24. D. A. Capdevila, K. A. Edmonds, D. P. Giedroc, "Metallochaperones and metalloregulation in bacteria", *Essays in Biochemistry*. **2017**, *61*, 177–200.
25. C. M. Gregg, S. Goetzl, J. H. Jeoung, H. Dobbek, "AcsF catalyzes the ATP-dependent insertion of nickel into the Ni, Ni-[4Fe4S] cluster of acetyl-CoA synthase", *The Journal of Biological Chemistry.* **2016**, *291*, 18129–18138.
26. M. Merrouch, M. Benvenuti, M. Lorenzi, C. Léger, V. Fourmond, S. Dementin, "Maturation of the [Ni–4Fe4S] active site of carbon monoxide dehydrogenases", *Journal of Biological Inorganic Chemistry*. **2018**, *23*, 613–620.
27. A. G. Fast, E. T. Papoutsakis, "Functional expression of the *Clostridium ljungdahlii* acetyl-coenzyme A synthase in *Clostridium acetobutylicum* as demonstrated by a novel *in vivo* CO exchange activity en route to heterologous installation of a functional Wood-Ljungdahl pathway", *Applied and Environmental Microbiology.* **2018**, *84*.
28. R. S. Tanner, L. M. Miller, D. Yang, "*Clostridium ljungdahlii* sp. nov., an acetogenic species in clostridial rRNA homology group I", *International Journal of Systematic Bacteriology.* **1993**, *43*, 232–236.
29. D. G. Gibson, "Enzymatic assembly of overlapping DNA fragments", *Methods in Enzymology*. **2011**, *498*, 349–361.
30. B. Molitor, K. Kirchner, A. W. Henrich, S. Schmitz, M. A. Rosenbaum, "Expanding the molecular toolkit for the homoacetogen *Clostridium ljungdahlii"*, *Nature Scientific Reports.* **2016**, *6,* 1–10.
31. M. S. Lowenthal, E. Quittman, and K. W. Phinney, "Absolute quantification of RNA or DNA using acid hydrolysis and mass spectrometry", *Analytical Chemistry.* **2019**, *91*, 14569-14576.
32. B. Bhaskar, E. DeMoll, D. A. Grahame, "Redox-dependent acetyl transfer partial reaction of the acetyl-coa decarbonylase/synthase complex: Kinetics and mechanism", *Biochemistry*.

1998, *37*, 14491–14499.
33. A. F. Arendsen, M. Q. Soliman, S. W. Ragsdale, "Nitrate-dependent regulation of acetate biosynthesis and nitrate respiration by *Clostridium thermoaceticum"*, *Journal of Bacteriology.* **1999**, *181*.
34. Mann Hyung Lee, S. B. Mulrooney, M. J. Renner, Y. Markowicz, R. P. Hausinger, "*Klebsiella aerogenes* urease gene cluster: Sequence of ureD and demonstration that four accessory genes (ureD, ureE, ureF, and ureG) are involved in nickel metallocenter biosynthesis," *Journal of Bacteriology.* **1992**, *174*, 4324–4330.
35. M. B. C. Moncrief, R. P. Hausinger, "Characterization of UreG, identification of a UreD-UreF-UreG complex, and evidence suggesting that a nucleotide-binding site in UreG is required for *in vivo* metallocenter assembly of *Klebsiella aerogenes* urease", *Journal of Bacteriology.* **1997**, *179*, 4081–4086.
36. M. H. Yuen, Y. H. Fong, Y. S. Nim, P. H. Lau, K. B. Wong, R. J. Fletterick, "Structural insights into how GTP-dependent conformational changes in a metallochaperone UreG facilitate urease maturation", *Proceedings of the National Academy of Science U. S. A.* **2017**, *114,* 10890–10898.
37. W. B. Jeon, J. Cheng, P. W. Ludden, "Purification and characterization of membrane-associated CooC protein and its functional role in the insertion of nickel into carbon monoxide dehydrogenase from *Rhodospirillum rubrum"*, *The Journal of Biological Chemistry.* **2001**, *276*, 38602–38609.
38. P. A. Hubbard, D. Padovani, T. Labunska, S. A. Mahlstedt, R. Banerjee, C. L. Drennan, "Crystal structure and mutagenesis of the metallochaperone MeaB - insight into the causes of *Methylmalonic aciduria"*, *Journal of Biological Chemistry.* **2007**, *282*, 31308-31316.
39. S. Watanabe, T. Kawashima, Y. Nishitani, T. Kanai, T. Wada, K. Inaba, H. Atomi, T. Imanaka, K. Miki, "Structural basis of a Ni acquisition cycle for [NiFe] hydrogenase by Ni-metallochaperone HypA and its enhancer", *Proceedings of the National Academy of Science U. S. A.* **2015**, *112*, 7701–7706.
40. A. C. Rosenzweig, "Metallochaperones: Bind and deliver", *Chemistry and Biology*. **2002**, *9*, 673–677.
41. C. J. Hlynialuk, B. Ling, Z. Baker, P.A. Cobine, L. D. Yu, A. Boulet, T. Wai, A. Hossain, A.M. El Zawily, P. J. McFie, S. J. Stone, F. Diaz, C. T. Moraes, D. Viswanathan, M. J. Petris, S.C. Leary, "The mitochondrial metallochaperone SCO1 is required to sustain expression of the high-affinity copper transporter CTR1 and preserve copper homeostasis", *Cell Reports.* **2015**, *10*, 933–943.
42. W. L. Pang, A. Kaur, A.V. Ratushny, A. Cvetkovic, S. Kumar, M. Pan, A. Arkin, J. Aitchison, M. W. W. Adams, N. S. Baliga, "Metallochaperones regulate intracellular copper levels", *PLoS Compututational Biology.* **2013**, *9*, e1002880.
43. H. Huang, S. Wang, J. Moll, R. K. Thauer, "Electron bifurcation involved in the energy metabolism of the acetogenic bacterium *Moorella thermoacetica* growing on glucose or H_2 plus CO_2", *Journal of Bacteriology.* **2012**, *194*, 3689–3699.

44. S. C. Andrews, B. C. Berks, J. McClay, A. Ambler, M. A. Quail, P. Golby, J.R. Guest, "A 12-cistron *Escherichia coli* operon (*hyf*) encoding a putative proton-translocating formate hydrogen lyase system", *Microbiology*. **1997**, *143*, 3633–3647.
45. S. L. Benoit, J. L. McMurry, S. A. Hill, R. J. Maier, "*Helicobacter pylori* hydrogenase accessory protein HypA and urease accessory protein UreG compete with each other for UreE recognition", *Biochimica Biophysica Acta.* **2012**, *1820*, 1519–1525.
46. X. Yang, H. Li, T. P. Lai, H. Sun, "UreE-UreG complex facilitates nickel transfer and preactivates GTPase of UreG in *Helicobacter pylori"*, *The Journal of Biological Chemistry.* **2015**, *290,* 12474–12485.
47. N. J. Robinson, D. R. Winge, "Copper metallochaperones", *Annual Review of Biochemistry*. **2010**, *79*, 537–562.
48. N. P. Minton, M. Ehsaan, C. Humphreys, G. Little, J. Baker, A. Henstra, F. Liew, M. Kelly, L. Sheng, K. Schwarz, Y. Zhang, "A roadmap for gene system development in *Clostridium"*, *Anaerobe*. **2016**, *41*, 104–112.
49. Y. Iwasaki, A. Kita, K. Yoshida, T. Tajima, S. Yano, T. Shou, M. Saito, J. Kato, K. Murakami, Y. Nakashimada*,* "Homolactic acid fermentation by the genetically engineered thermophilic homoacetogen *Moorella thermoacetica* ATCC 39073", *Applied and Environmental Microbiology.* **2017**, *83*, 247–264.
50. A. Kita, Y. Iwasaki, S. Sakai, S. Okuto, K. Takaoka, T. Suzuki, S. Yano, S. Sawayama, T. Tajima, J. Kato, N. Nishio, K. Murakami, Y. Nakashimada, "Development of genetic transformation and heterologous expression system in carboxydotrophic thermophilic acetogen *Moorella thermoacetica"*, *Journal of Bioscience and Bioengeering.* **2013**, *115*, 347–352.
51. F. Rahayu, Y. Kawai, Y. Iwasaki, K. Yoshida, A. Kita, T. Tajima, J. Kato, K. Murakami, T. Hoshino, Y. Nakashimada*,* "Thermophilic ethanol fermentation from lignocellulose hydrolysate by genetically engineered *Moorella thermoacetica"*, *Bioresources Technology.* **2017**, *245*, 1393–1399.
52. J. Seravalli, K. L. Brown, S. W. Ragsdale, "Acetyl coenzyme a synthesis from unnatural methylated corrinoids: Requirement for 'base-off' coordination at cobalt-14", *Journal of the American Chemical Society*. **2001**, *123*, 1786–1787.
53. B. Bhaskar, E. DeMoll, D. A. Grahame, "Redox-dependent acetyl transfer partial reaction of the acetyl-coa decarbonylase/synthase complex: Kinetics and mechanism", *Biochemistry*, **1998.** *37*,14491–14499.

Chapter 3: *Conversion of carbon dioxide to bioplastic by Methanococcus maripaludis*

3.1. Introduction

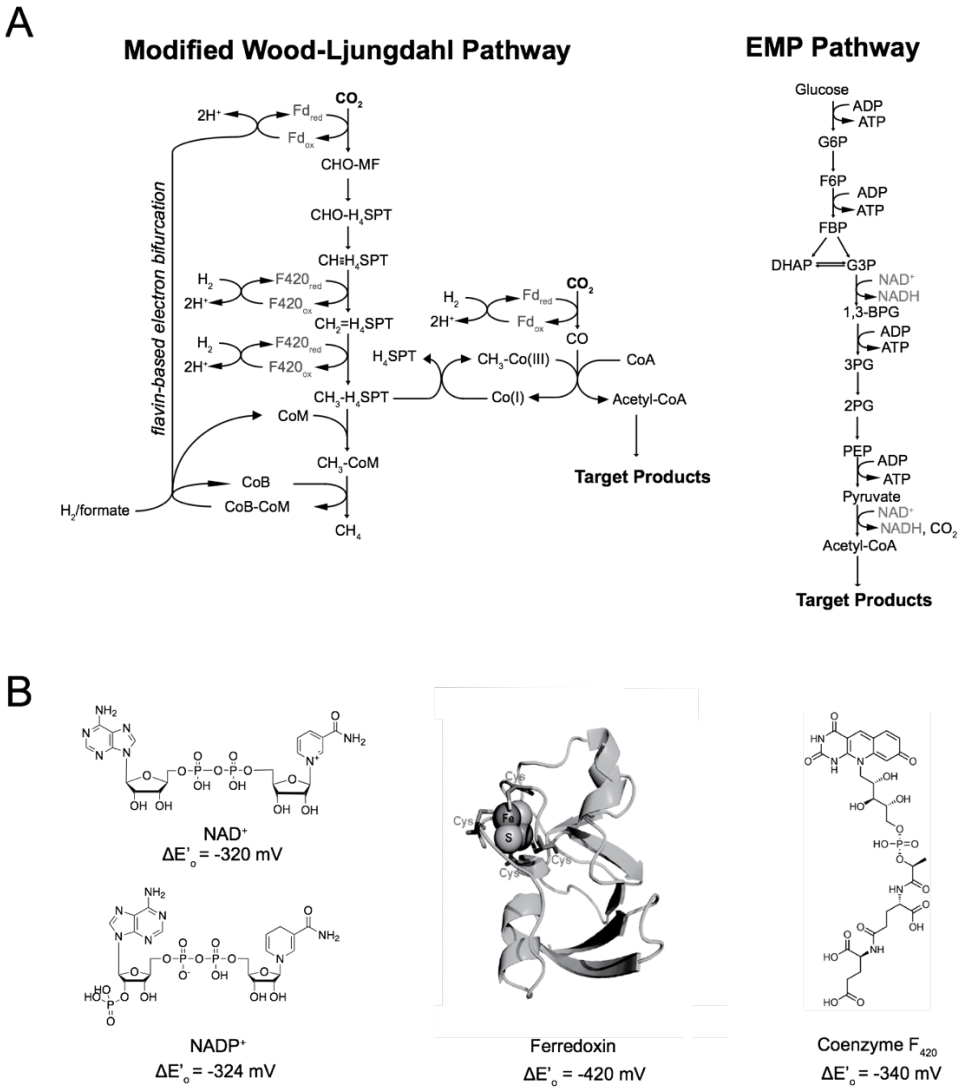
Chemical synthesis using petroleum-derived feedstocks has come under increasing scrutiny due to the steep environmental costs associated with its greenhouse gas emissions. In response, efforts have been made across scientific fields to capture, sequester and utilize CO_2 [*1*]. Nature itself has evolved diverse biological methods for CO_2 capture and incorporation into biomass via autotrophic metabolism. Due to their attractive metabolic features, photosynthetic autotrophs have garnered interest as a potential metabolic engineering host for environmentally sustainable chemical synthesis [*2*, *3*]. Compared to heterotrophic organisms, photoautotrophs offer the advantage of requiring only CO_2 as a carbon source to produce value-added chemicals. However, photoautotrophs also require light as an energy source, which creates a host of fermentation challenges associated with poor light penetration, low thermodynamic efficiency and solar intermittency [*4*, *5*].

In contrast, non-photosynthetic chemautotrophs can use hydrogen (H_2) and other inorganic compounds as sources of energy and electrons [*6*, *7*]. Furthermore, the hydrogen required for CO_2 fixation can be produced electrochemically using inorganic catalysts and renewable sources of energy, including but not limited to solar power [*8*]. Combining these complementary processes into a hybrid, versatile platform could provide a route to sustainable chemical synthesis from CO_2 that avoids key challenges associated with photosynthetic platforms. Amongst the chemoautotrophs, methanogens stand out for their industrially relevant metabolic features. Methanogens are a group of obligately anaerobic archaea that use a modified Wood-Ljungdahl pathway to incorporate CO_2 into biomass via acetyl-CoA. Their central metabolism is the most energy-efficient carbon fixation pathway in nature, as measured by the ATP requirement per mole of fixed CO_2. In addition, methanogens produce methane as a metabolic by-product, which could be used as a source of heating fuel, chemical precursors and energy [*9*]. Also, it may be possible to divert the acetyl-CoA pool in archaeal species towards the production of fatty acids, isoprenoids and other molecules, as is common for heterotrophic metabolic hosts [*10*, *11*].

However, very few studies have thus far reported the use of methanogens for chemical production [*12*, *13*]. Genetic elements used for robust, heterologous gene expression in methanogens have been determined empirically, and gene regulatory process have not been as fully elucidated as those of model bacterial species. Another feature of methanogen metabolism that is not yet fully elucidated is the role of NADH and NADPH. Unlike common fermentation hosts and other chemoautotrophs, methanogenic archaea have evolved to use 2[4Fe-4S] ferredoxin and coenzyme F_{420} instead of NAD(P)H in central metabolism [*14*,*15*] (*Scheme 3.1*). In nature, NAD(P)H is used in over 300 oxidation-reduction reactions, many of which are essential for energy generation and biosynthetic pathways [*16*,*17*] As such, there is a broad range of NAD(P)H-dependent pathways used for the production of value-added chemicals. However, further investigation and optimization are likely required to properly integrate NAD(P)H-dependent pathways into the unique metabolism of methanogens.

In this study, *Methanococcus maripaludis*, a methanogen with a fully sequenced genome and published genetic tools [*18*,*19*], was used as an engineering host to produce target chemicals from CO_2 and gain insights into NAD(P)H-dependent metabolism. Specifically, we aimed to engineer this species for the NADH-dependent biosynthesis of (S)-3-hydroxybutyrate ((S)-3HB), and the biodegradable plastic (R)-3-hydroxybutyrate ((R)-3HB) monomer and polyhydroxybutyrate (PHB) polymer [*20*]. In the course of our investigation, we studied the NADH-dependent metabolism of this species using proteomic, transcriptomic and metabolic modelling techniques.

We identified low NADH availability as a potential metabolic bottleneck in *M. maripaludis* and showed that alanine dehydrogenase-dependent growth and engineered strains with increased *in vivo* NADH pools and improved NAD^+ turnover can enhance bioplastic production. These model pathways show the potential of this species as a platform for the production of value-added chemicals from CO_2 and provide the first demonstration of an engineered NADH cofactor pool in an archaeal species.



Scheme 3.1. Comparison of glycolysis and the modified Wood-Ljungdahl pathway. *(A) The modified Wood-Ljungdahl pathway (top left) is a central metabolic pathway in methanogenic archaea that uses two molecules of CO_2, reduced ferredoxin and reduced coenzyme F_{420} to make acetyl-CoA and further downstream products. Hydrogen or formate can be used as the electron source in Methanococcus maripaludis, using flavin-based electron bifurcation to simultaneously reduce the sulfide bond in Coenzyme B and Coenzyme M (CoB-CoM) and ferredoxin* [21]*. This pathway has stark contrasts with EMP glycolysis (top right) commonly found in heterotrophic microbes, including different cofactor usage, different electron and carbon sources. (B) Methanogenesis requires coenzyme F_{420} and 2 [4Fe-4S] ferredoxin, while heterotrophic species typically use NADH and NADPH for catabolic and anabolic reactions* [22]*. The former two cofactors have lower standard reduction potentials compared to NAD(P)H.*

3.2. Materials and Methods

Commercial materials. Ultra-high purity gases purchased from Praxair (Danbury, CT) were used for all anaerobic manipulations. Distilled water (dH_2O) was deionized to a resistivity of 18.2 MΩ·cm using a Millipore Milli-Q UF Plus system (ddH_2O). Luria-Bertani (LB) Broth Miller, LB Agar Miller and yeast extract were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), chloramphenicol (Cm), sodium chloride, magnesium sulfate heptahydrate, magnesium chloride hexahydrate, ammonium chloride, potassium chloride, potassium phosphate monobasic, potassium phosphate dibasic, sodium phosphate monobasic, sodium phosphate dibasic, manganese (II) chloride tetrahydrate, dithiobis-[2-nitrobenzoic acid] (DTNB), methanol, and pressure release aluminum crimp seals (20 mm) were purchased from Fisher Scientific (Pittsburgh, PA). Calcium chloride dihydrate, adipic acid, iron(II) sulfate heptahydrate, sodium selenite, nickel(II) chloride hexahydrate, boric acid, cobalt(II) chloride hexahydrate, copper(II) chloride dihydrate, sodium molybdate dihydrate, sodium tungstate dihydrate, nitrilotriacetic acid, sodium sulfide, L-cysteine, biotin, folic acid, pyridoxine hydrochloride, thiamine hydrochloride, 3-morpholinopropane-1-sulfonic acid (MOPS), riboflavin, nicotinic acid, calcium D-(+)-pantothenate, vitamin B_{12}, p-aminobenzoic acid, nicotinamide, sodium bicarbonate, sodium acetate, magnesium formate dihydrate, thioctic acid, Tween® 20, nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), nicotinamide adenine dinucleotide 2′-phosphate reduced tetrasodium salt (NADPH), nicotinamide adenine dinucleotide hydrate (NAD^+), propionyl coenzyme A lithium salt, formic acid (LC-MS grade), polyethylene glycol 3350, dimethyl sulfoxide (DMSO), neomycin sulfate and puromycin dihydrochloride were purchased from Sigma Aldrich (St. Louis, MO). Bacto™ Casamino acids (acid hydrolysate of casein) was purchased from BD (Franklin Lakes, NJ). Zinc sulfate heptahydrate was purchased from Mallinckrodt Chemicals (St. Louis, MO). Resazurin was purchased from Eastman Kodak Company (Rochester, NY). Balch tubes (18 x 150 mm), anaerobic media bottles (250 mL) and butyl rubber stoppers (20 mm) were purchased from Chemglass (Vineland, NJ). Restriction enzymes and Phusion DNA polymerase were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), were purchased from Invitrogen (Carlsbad, California). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, Iowa), resuspended at a stock concentration of 100 µM in water and stored at -20°C. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, California). Parafilm M was purchased from Penchiney Plastic Packaging (Akron, Ohio). GC vials and glass inserts were purchased from National Scientific (Alexandria, Virginia). 0.2 µm syringe filter was purchased from Waters Corporation (Milford, Massachusetts).

Flux balance analysis. Flux balance analysis (FBA) was performed using the MATLAB COBRA toolbox and iMR539, a known metabolic model for *M. maripaludis* [*23*]. Reactions that oxidized ferredoxin (Fd), Coenzyme F_{420} (F420), reduced nicotinamide adenine dinucleotide (NADH) and reduced nicotinamide adenine dinucleotide phosphate (NADPH) were sorted by cofactor usage. The fluxes for reactions from each cofactor category were summed together to obtain total oxidative flux, measured in µmol/min/gram of dry cell weight (µmol/min/gDCW). Each cofactor's oxidative flux was divided by the total oxidative flux for all redox cofactors in order to obtain "%Total" values.

Bacterial strain details and culture conditions. *E. coli* DH10B-$T1^R$ was used for plasmid construction. To prepare chemically competent cells for plasmid transformations, *E. coli* was cultured from a single colony in 100 mL LB broth, then harvested at an OD_{600} of 0.4-0.6 and

resuspended in 10 mL of sterile transformation-storage solution (10% *w/v* PEG 3350, 5% *v/v* DMSO, 20 mM $MgCl_2$ in LB broth). 100 µL aliquots were flash frozen in liquid nitrogen and stored at -80°C for later use. Plasmid transformations were performed by adding the plasmid(s), 20 µL 5 x KCM solution (0.5 M KCl, 0.15 M $CaCl_2$, 0.25 M $MgCl_2$) and 175 µL of MilliQ-treated water to a thawed competent cell aliquot of the appropriate strain. After 30 min incubation on ice, cells were subjected to heat shock at 42°C for 90 seconds, diluted to 1 mL with LB broth, and incubated at 37°C for 1 hour before plating with sterile beads on LB agar containing appropriate antibiotics.

Gene and plasmid construction. Plasmids were constructed by Gibson assembly [*24*], except where otherwise noted. Strain and plasmid descriptions are listed in *Appendix 1.1.* All PCR amplifications were carried out with Phusion DNA polymerase using oligonucleotides listed in *Appendix 1.2*. Double stranded DNA fragments (gBlocks) were synthesized by Integrated DNA Technologies (*Appendix A1.4*) and resuspended at a stock concentration of 10 ng/µL in double-distilled water. Vectors were digested with appropriate restriction enzymes and mixed with inserts, either gel-purified PCR products or synthetic gBlocks. 1.5x Gibson mix was added into the mixture, and the Gibson reaction was carried out at 50ºC for 1 h in a thermocycler. The resulting mixture was transformed immediately into the chemically competent *E. coli* DH10B-$T1^R$ by heat shock. Constructs were verified by sequencing (Genewiz, San Francisco, CA). Details of individual plasmid constructs are listed below:

pTrc33-Δhpt::adhE-pac. pTrc33 [*25*] was digested with BamHI/XbaI. The upstream homology arm was PCR amplified from the genome of *Methanococcus maripaludis* using the *hpcE* f and r primer pair. The puromycin resistance cassette was PCR amplified from pWM321 [*26*] using the *pac* f and r primer pair. The downstream homology arm was PCR amplified from the genome of *M. maripaludis* using the *MMP0143* f and r primer pair. The P_{mcrB} promoter was PCR amplified from the gBLOCK P_{mcrB} using the P_{mcrB} f and r primer pair. *adhE* was amplified from the genome of *E.coli* using the *adhE* f and r primer pair. The P_{mcrB} promoter and *adhE* were joined by splicing by overlap extension (SOE) PCR with the P_{mcrB}-*adhE* SOE f and r primer pair. The cut vector and the PCR products were ligated by Gibson assembly.

pTrc33-Δhpt::phaA-phaB-tesB-pac. *pTrc33-Δhpt::adhE-pac* was digested with BamHI/XhoI. The gBLOCKs of *phaA*, *phaB*, and *tesB* were directly inserted into the cut vector by Gibson assembly.

pTrc33-Δhpt::phaA-Strep-phaB-Strep-tesB-Strep-pac. *pTrc33-Δhpt::adhE-pac* was digested with BamHI/XhoI. Inserts were generated by PCR from the *pTrc33-Δhpt::phaA-phaB-tesB-pac* plasmid using the *phaA-Strep* f and r primer pair, the *phaB-Strep* f and r primer pair, and the *tesB-Strep* f and r primer pair. The *phaA-Strep*, *phaB-Strep* and *tesB-Strep* inserts were joined by SOE PCR using the *phaA-phaB-tesB* SOE f and r primer pair. The inserts and vector were then ligated by Gibson assembly.

pTrc33-Δhpt::phaA-Strep-hbd-Strep-tesB-Strep-pac. *pTrc33-Δhpt::adhE-pac* was digested with BamHI/XhoI. Inserts were generated by PCR from the *pTrc33-Δhpt::phaA-phaB-tesB-pac* plasmid using the *phaA-Strep* f and r primer pair, the *hbd-Strep* f and r primer pair, and the *tesB-Strep* f and r primer pair. The *phaA-Strep*, *hbd-Strep* and *tesB-Strep* inserts were joined by SOE PCR using the *phaA-phaB-tesB* SOE f and r primer pair. The inserts and vector were then ligated by Gibson assembly.

pTrc33-Δhpt::phaA-Strep-phaB-Strep-tesB-Strep-(syntheticRBS)-pac. *pTrc33-Δhpt::adhE-pac* was digested with BamHI/XhoI. Inserts were generated by PCR from the *pTrc33-Δhpt::phaA-Strep-hbd-Strep-tesB-Strep-pac* plasmid using the *phaA-RBS* f and r primer pair, the *hbd-RBS* f and r primer pair, and the *tesB-RBS* f and r primer pair. The inserts and vector were then ligated by Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-phaA-P_{flaB1}-phaB-P_{hdrC1}-tesB-pac. *pTrc33-Δhpt::adhE-pac* was digested with BamHI/XhoI. Inserts were generated by PCR from the *pTrc33-Δhpt::phaA-phaB-tesB-pac* plasmid using the *phaA-P_{mcrB}* f and r primer pair, the *phaB-P_{flaB1}* f and r primer pair, and the *tesB-P_{hdrC1}* f and r primer pair. P_{flaB1} was amplified from the genome of *M. maripaludis* using the P_{flaB1} f and r primer pair, *P_{hdrC1}* was amplified from the genome of *M. maripaludis* using the *P_{hdrC1}* f and r primer pair. All inserts were joined by SOE PCR using the *phaA-phaB-tesB* SOE f and r primer pair. The inserts and vector were then ligated by Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-phaA-Strep-P_{flaB1}-phaB-Strep-P_{hdrC1}-tesB-Strep-pac.pTrc33-Δhpt::adhE-pac was digested with BamHI/XhoI. Inserts were generated by PCR from the *pTrc33-Δhpt::phaA-Strep-phaB-Strep-tesB-Strep-pac* plasmid using the *phaA-P_{mcrB}-Strep* f and r primer pair, the *phaB-P_{flaB1}-Strep* f and r primer pair, and the *tesB-P_{hdrC1}-Strep* f and r primer pair. P_{flaB1} was amplified from the genome of *M. maripaludis* using the P_{flaB1} f and r primer pair, P_{hdrC1} was amplified from the genome of *M. maripaludis* using the P_{hdrC1} f and r primer pair. All inserts were joined by SOE PCR using the *phaA-phaB-tesB* SOE primer pair. The inserts and vector were then ligated by Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-phaA-Strep-T-P_{flaB1}-hbd-Strep-T-pac. *pTrc33-Δhpt::P_{mcrB}-phaA-Strep-P_{flaB1}-hbd-Strep-P_{hdrC1}-tesB-Strep-pac* was digested with NotI. Inserts were generated by PCR from the genome of *Methanosarcina mazei* using the term73 f and r primer pair. Terminator fragments were joined by SOE PCR using the term73 f and term73 SOE r primer pair. These inserts and vector were joined using Gibson assembly to form a cloning intermediate. This cloning intermediate was then digested with BamHI/NotI. Another set of inserts were generated from a primer mix of term3050 f, term3050 r and hpcE-term3050 r using the term3050 f2 and r2 primer pair. The inserts were joined by SOE PCR using the term3050 SOE f and r primer pair. Another insert was generated from *pTrc33-Δhpt::P_{mcrB}-phaA-Strep-P_{flaB1}-hbd-Strep-P_{hdrC1}-tesB-Strep-pac* using LH-term3050 f and r primer pair. The inserts and vector were then joined using Gibson assembly.

pTrc33-Δhpt::P_{hdrC1}-phaA-Strep-T-P_{flaB1}-hbd-Strep-T-pac. *pTrc33-Δhpt::adhE-pac* was digested with BamHI/XhoI. Inserts were generated by PCR from the *pTrc33-Δhpt::P_{mcrB}-phaA-Strep-P_{flaB1}-hbd-Strep-P_{hdrC1}-tesB-Strep-pac* plasmid using the P_{hdrC1} f and r primer pair and the *pTrc33-Δhpt::pmcrB-phaA-Strep-T-p_{flaB1}-hbd-Strep-T-pac* using the *phaA-hbd* f and r primer pair. Inserts and vector were then joined using Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-hbd-pac. pTrc33-Δhpt::adhE-pac was digested with BamHI/XhoI. Inserts were generated from the *pTrc33-Δhpt::P_{mcrB}-phaA-P_{flaB1}-hbd-P_{hdrC1}-tesB-pac* using the *P_{mcrB}-hbd* f, r and r2 primer mix and the P_{mcrB}-hbd-LH f and r primer pair. The inserts and vector were then joined using Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-hbd-Strep-pac. pTrc33-Δhpt::adhE-pac was digested with BamHI/XhoI. Inserts were generated from the *pTrc33-Δhpt::P_{mcrB}-phaA-P_{flaB1}-hbd-P_{hdrC1}-tesB-pac* using the *P_{mcrB}-hbd-Strep* f, r and r2 primer mix and the P_{mcrB}-hbd-LH f and r primer pair. The inserts and vector were then joined using Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-tesB-pac. pTrc33-Δhpt::adhE-pac was digested with BamHI/XhoI. Inserts were generated from the *pTrc33-Δhpt::P_{mcrB}-phaA-P_{flaB1}-hbd-P_{hdrC1}-tesB-pac* using the *P_{mcrB}-tesB* f and r or r2 primer pair. The inserts and vector were then joined using Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-tesB-Strep-pac. pTrc33-Δhpt::adhE-pac was digested with BamHI/XhoI. Inserts were generated from the *pTrc33-Δhpt::P_{mcrB}-phaA-P_{flaB1}-hbd-P_{hdrC1}-tesB-pac* using the *P_{mcrB}-tesB-Strep* f and r or r2 primer pair. The inserts and vector were then joined using Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-phaA-P_{mcrB}–phaB-P_{mcrB}-tesB-pac. pTrc33-Δhpt::adhE-pac was digested with MfeI/XhoI to remove the *adhE* gene. gBlOCKs of *phaA, phaB* and *tesB* were directly inserted into the cut vector by the Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-phaA-P_{mcrB}–phaB-P_{mcrB}-phaC-pac. pTrc33-Δhpt::adhE-pac was digested with MfeI/XhoI to remove the *adhE* gene. gBlOCKs of *phaA, phaB* and *phaC* were directly inserted into the cut vector by the Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-phaA-P_{mcrB}–phaB-P_{mcrB}-phaC-P_{mcrB}-nadM-P_{mcrB}-nadV-pac. pTrc33-Δhpt::adhE-pac was digested with MfeI/XhoI to remove the *adhE* gene. gBlOCKs of *phaA, phaB, phaC, nadM and nadV* were directly inserted into the cut vector by the Gibson assembly.

pBAD24-Δupt::P_{hmv}-fdh1-neo^r. A *pBAD24* plasmid containing BbsI cut sites was digested with BbsI. Inserts generated from PCR of the *M. maripaludis* genome using the upt-LH f and r primer pair and the upt-RH f and r primer pair, PCR *of pRSF-hadh2* using the neo^r f and r primer pair, PCR of *pTrc33-Δhpt::P_{mcrB}-phaA-P_{flaB1}-hbd-P_{hdrC1}-tesB-pac* using the *P_{mcrB}* f and r primer pair. The cut vector, *fdh1* gBLOCK and PCR fragments were assembled using Golden Gate assembly [*27*].

pBAD24-Δupt::P_{hmv}-fdh1-P_{mcrB}-nadM-P_{mcrB}-nadV-neo^r. pTrc33-Δupt::P_{hmv}-fdh1-neo^r was digested with SacII/NcoI. Insert was generated from PCR of the *M. maripaludis* genome using the *pTrc33-Δhpt::P_{mcrB}-nadM-P_{mcrB}-nadV* plasmid and the *nadMV* f and r primer pair. The inserts and vector were then joined using Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-nadM-P_{mcrB}-nadV. *pTrc33-Δhpt::adhE-pac* was digested with MfeI/XhoI to remove the *adhE* gene. gBLOCKs of *nadM, nadV* were directly inserted into the cut vector by the Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-MMP0737. pTrc33-Δhpt::adhE-pac to remove the *adhE* gene. Insert was generated using *pTrc33-Δhpt::adhE-pac* with the *MMP0737* f and *MMP0737*r primer pair. Another insert was generated using PCR of the *M. maripaludis* genome with the *MMP0737* f2 and r2 primer pair. The inserts and vector were then joined using Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-nadA. pTrc33-Δhpt::adhE-pac to remove the *adhE* gene. Insert was generated using *pTrc33-Δhpt::adhE-pac* with the *MMP0737* f and *nadA* r primer pair. Another insert was generated using PCR of the *M. maripaludis* genome with the *nadA* f2 and r2 primer pair. The inserts and vector were then joined using Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-nadC. pTrc33-Δhpt::adhE-pac to remove the *adhE* gene. Insert was generated using *pTrc33-Δhpt::adhE-pac* with the *MMP0737* f and *nadC* r primer pair. Another insert was generated using PCR of the *M. maripaludis* genome with the *nadC* f2 and r2 primer pair. The inserts and vector were then joined using Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-MMP1578. pTrc33-Δhpt::adhE-pac to remove the *adhE* gene. Insert was generated using *pTrc33-Δhpt::adhE-pac* with the *MMP0737* f and *MMP1578* r primer pair. Another insert was generated using PCR of the *M. maripaludis* genome with the *MMP1578* f2 and r2 primer pair. The inserts and vector were then joined using Gibson assembly.

pTrc33-Δhpt::P_{mcrB}-MMP1349. pTrc33-Δhpt::adhE-pac to remove the *adhE* gene. Insert was generated using *pTrc33-Δhpt::adhE-pac* with the *MMP0737* f and *MMP1349* r primer pair. Another insert was generated using PCR of the *M. maripaludis* genome with the *MMP1349* f2 and r2 primer pair. The inserts and vector were then joined using Gibson assembly.

Methanogen strain details and culture conditions. *Methanococcus maripaludis S2* (ATCC BAA-2049) was purchased from the American Type Culture Collection (Manassas, VA). All culture manipulations were performed inside a Vacuum Atmospheres Nexus One glovebox with an atmosphere of 90% nitrogen and 10% hydrogen. Oxygen levels and humidity levels within the box were controlled using a STAK-PAK palladium catalyst and desiccant system in a Coy Laboratory Products unheated fan box. *M. maripaludis* cultures were propagated in 18 × 150 mm Balch tubes with butyl rubber stoppers and aluminum crimp seals using defined medium (McNA) or complex, undefined medium (McCas). In initial production conditions, 10 mL cultures were grown at 37°C without shaking. Under improved production conditions, 10 mL cultures were grown in McNA-Ala medium using 250 mL anaerobic glass bottles at 37°C with shaking at 250 rpm. Bottles were pressurized with H_2/CO_2 (80%, 20%) to 275 kPa every 12 hours for 5 days. Stock cultures were stored at room temperature in the dark after growth and were propagated by diluting 1:100 into fresh McCas or McNA medium at least every 2 weeks. The stock cultures were used until changes in growth patterns were observed. Glycerol stocks were kept at -80°C for long-term culture storage.

Methanogen growth media preparation. To prepare growth media for *M. maripaludis*, all components listed in the table below except the reducing agent were dissolved in the appropriate volume of MilliQ-treated water in an Airfree® Cajon storage flask (Chemglass, Vineland, NJ). The solution was made anaerobic using three freeze-pump-thaw cycles on a Schlenk line and transferred to serum bottles or tubes in an anaerobic chamber. The bottles or tubes were sealed with butyl rubber stoppers and aluminum crimps. Then, media was autoclaved and allowed to cool before reducing agent was added. If puromycin or neomycin were used, they would be added to the cooled media at a final concentration of 2.5 μg/mL or 500 μg/mL, respectively. A similar procedure was followed for the preparation of McCas agar bottle plates, except 2% (*w/v*) agar was added to 50 mL of media after it has been transferred to bottles in the anaerobic chamber.

A stock of 100x reducing agent was prepared by adding a 0.01% (*w/v*) resazurin solution (200 μL) to MilliQ-treated water (200 mL) and sparging the solution with N_2 for 20 minutes. Under continuous N_2 flow, 6 grams of L-cysteine was added, followed by 6 grams of sodium sulfide nonahydrate. Sparging with N_2 was continued until the sodium sulfide had dissolved completely and the solution was completely colorless. After preparation, the reducing agent was stored under an N_2 atmosphere in a 250 mL serum bottle and sealed with butyl rubber stoppers and aluminum crimps.

Per liter, the modified McCas medium contains:

ddH_2O	500 mL
N-free general salts solution	500 mL
$NaHCO_3$	5 g

NaCl	10.5 g
K_2HPO_4	140 mg
$FeSO_4 \cdot 7H_2O$	10 mg
N-free trace minerals solution (100X)	10 mL
Wolfe's vitamin solution	10 mL
Resazurin solution (0.01% (*w/v*))	1 mL
Sodium acetate	850 mg
MOPS (2 M, pH 7)	100 mL
Magnesium formate dihydrate	15 g
NH_4Cl	500 mg
Casamino acids	2 g

Per liter, the defined McNA medium contains:

ddH_2O	500 mL
N-free general salts solution	500 mL
$NaHCO_3$	5 g
NaCl	10.5 g
K_2HPO4	140 mg
$FeSO_4 \cdot 7H_2O$	10 mg
N-free trace minerals solution (100X)	10 mL
Resazurin solution (0.01% (*w/v*))	1 mL
Sodium acetate	850 mg
MOPS (2 M, pH 7)	100 mL
Magnesium formate dihydrate	15 g

And either of:

NH_4Cl	500 mg
L-Alanine	890 mg

Per liter, the N-free general salts solution contains:

KCl	670 mg
$MgCl_2 \cdot 6H_2O$	5.5 g
$MgSO_4 \cdot 7H_2O$	6.9 g
$CaCl_2 \cdot 2H_2O$	280 mg
ddH_2O	1000 mL

To prepare the N-free trace minerals solution, trisodium citrate was dissolved in water and the pH was adjusted to 6.5. Then the remaining salts (see below) were added. Per liter, the 100x N-free trace minerals solution contains:

Trisodium citrate dihydrate	2.1 g
$MnSO_4 \cdot H_2O$	500 mg

$CoCl_2 \cdot 6H_2O$	100 mg
$ZnSO_4 \cdot 7H_2O$	100 mg
$CuCl_2 \cdot 2H_2O$	100 mg
$FeSO_4 \cdot 7H_2O$	6.8 mg
$AlK(SO_4)_2 \cdot 12H_2O$	10 mg
H_3BO_3	10 mg
$Na_2MoO_4 \cdot 2H_2O$	100 mg
$NiCl_2 \cdot 6H_2O$	25 mg
Na_2SeO_3	200 mg
VCl_3	10 mg
$Na_2WO_4 \cdot 2H_2O$	3.3 mg

Methanogen genetic transformation. *M. maripaludis* was transformed using protocols previously described for the transformation of *Methanosarcina* species [*28*]. 10 mL of cultures were grown to late log phase for each transformation. The cells were harvested by centrifugation (2000 × *g* for 10 min) inside the anaerobic chamber. The media was removed, cells were washed with 5 mL of transformation buffer (see below). The cells were collected by centrifugation again. Supernatant was then discarded, and the cell pellet was resuspended in 1 mL transformation buffer per 10 mL of initial culture volume. In each transformation, 300 µL transformation buffer was aliquoted into a 6 mL serum vial. Vials were then closed with a butyl rubber stopper and removed from the anaerobic chamber. Using aseptic technique, 2-4 µg of plasmid and 30 µL DOTAP transfection agent (Roche, Basel, Switzerland). The vials were again stoppered and sealed, after which they were degassed by purging with argon for 5 min before the vials were returned to the anaerobic chamber. Then, 1 mL of the cell suspension was added to each vial. The vials were incubated at 37°C for 1 h. After the incubation, the suspension was transferred to 10 mL of fresh medium in a Balch tube. The cells were recovered at 37°C for at least 4 h. After recovery, the cells were harvested by centrifugation and resuspended in 200 µL modified McCas medium. They were then spread on the surface of an anaerobic bottle agar plate. Plates were incubated at room temperature with the surface of the plate horizontal for at least 30 minutes to allow cell suspension to leach into the agar. After incubation, bottles were stood upright. Headspace of the bottle plates was purged with a H_2/CO_2 (80%, 20%) mixture for 5 min before pressurizing to 275 kPa. Plates were incubated at 37°C for 3-4 days at which point colonies were picked into growth medium inside of the anaerobic chamber. After cultures reached saturation, presence of the desired knock-in was confirmed using colony PCR. Cultures were propagated in selective media containing puromycin and/or neomycin and stored as glycerol stocks in 30% glycerol (*v/v*) at -80°C.

To prepare transformation buffer, 100 µL of a 0.01% (*w/v*) resazurin solution was added to 100 mL of a 50 mM HEPES (pH 7.5) solution in an Airfree® Cajon storage flask (Chemglass, Vineland, NJ). The solution was made anaerobic using three freeze-pump-thaw cycles on a Schlenk line and transferred to a glass serum bottle in an anaerobic chamber. Then, the degassed HEPES buffer was combined with the remaining components of the transformation buffer (see below), and the solution was mixed until all solids were dissolved. The buffer was filtered into a sterile 250 mL serum bottle using a 0.2 µm syringe filter. The bottle was sealed with a sterile butyl rubber stopper and aluminum crimp. After approximately an hour of preparation, the buffer should become colorless. The buffer was stored at 4°C.

Per 100 mL, the transformation buffer contains:

HEPES (pH 7.5)	100 mL
Sucrose	12 g
NaCl	2.2 g
Cysteine	1 g
Dithiothreitol	771 mg

Measurement of intracellular NAD(H) pools. For quantification of the NAD^+ and NADH pools, overnight cultures of *M. maripaludis* or anaerobically grown *E. coli* were used. The assay was performed using an NAD^+/NADH-Glo Assay Kit (Promega, WI) according to the manufacture's protocol for measuring the NAD^+ and NADH pools individually with minor modifications. Cells were harvested by centrifugation for 1 min at 13,000 × *g* at room temperature. For *E.coli,* 500 µL of pelleted culture was used per sample. For *M. maripaludis*, 5 mL of pelleted culture was used per sample. After resuspending in 500 µL of lysis buffer, the samples were mixed with 300 µL 0.1 mm glass disruption beads in 2 mL O-ring sealed, screw cap tubes and were lysed using a BioSpec Products Mini-BeadBeater by beating for 45 s at 4°C. For cell dry weight measurements, 20 mL of culture from each sample was centrifuged and was collected into a pre-massed 2 mL tube by two rounds of centrifugation for 1 min at 13,000 × *g* at room temperature. As much residual media as possible was removed before samples were vacuum concentrated for 30 min at room temperature. The change in mass of tubes was determined using a microbalance.

Enzyme assays in clarified lysates. A 10 mL saturated, overnight *M. maripaludis* culture was harvested by centrifuging for 10 min at 10,000 × *g* at 4°C. The cell pellet was resuspended in 500 µL lysis buffer (50 mM KH_2PO_4, 1 mini-cOmplete EDTA-free protease inhibitor cocktail tablet (Roche) per 10 mL buffer, pH 7.4). 1 mM DTT was included in the lysis buffer, except when assaying TesB activity. The samples were mixed with 300 µL 0.1 mm glass disruption beads in 2 mL O-ring sealed, screw cap tubes and were lysed using a BioSpec Products Mini-BeadBeater by beating for 45 s at 4°C. Cell debris was pelleted by centrifuging for 10 min at 20,000 × *g* at 4°C. The supernatant was then removed to a fresh tube and total protein was quantified using Bradford reagent with a lysozyme standard curve.

Acetyl-CoA:acyl-CoA thiolase. Thiolase activity was assayed in the forward direction by monitoring NADH oxidation at 340 nm (ε = 6,220 M^{-1} cm^{-1}) at 30°C in a coupled assay with His_{10}-hbd. The assay mixture contained acetyl-CoA (1 mM), NADH (100 µM), and His_{10}-hbd (7.5 U/mL) in 100 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.5. The reaction was initiated by addition of acetyl-CoA.

Acetoacetyl-CoA reductase. The reduction of acetoacetyl-CoA was assayed by monitoring the oxidation of NADH at 340 nm (ε = 6,220 M^{-1} cm^{-1}) at 23°C. The assay mixture contained acetoacetyl-CoA (100 µM) and NADH (100 µM) in 100 mM Tris-HCl, 10 mM $MgCl_2$, pH 7.5. The reaction was initiated by addition of acetoacetyl-CoA.

Acyl-CoA thioesterase. Thioesterase activity was assayed by monitoring the reaction of free CoA with 5,5'-dithiobis-[2-nitrobenzoic acid] (DTNB) at 412 nm (ε = 41,150 M^{-1} cm^{-1}) at 30°C. The assay mixture contained propionyl-CoA (400 µM) and DTNB (300 µM) in PBS. The reaction was initiated by addition of cell lysate.

Formate dehydrogenase. The oxidation of formate was assayed by monitoring the reduction of NAD^+ at 340 nm ($\varepsilon = 6{,}220\ M^{-1}\ cm^{-1}$) at 23°C. The assay mixture contained magnesium formate dihydrate (10 mM) and NAD^+ (100 µM) in 100 mM Tris-HCl, 10 mM MgCl2, pH 7.5. The reaction was initiated by addition of magnesium formate dihydrate.

Quantitative real time PCR. *M. maripaludis* cultures (10 mL) were inoculated at 1:100 dilution and grown for 5.5 h at 37 °C without shaking. 6 mL of the culture were harvested by centrifugation for 2 min at 13,000 × *g* at room temperature. After centrifuging, the pellets were kept on ice, and RNA was extracted using an RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. For cell lysis, 600 µL buffer RLT with β-mercaptoethanol added was used, and the samples were homogenized using QIAshredder columns (Qiagen). The RNA samples were eluted into 50 µL RNase-free water and quantified using a Nanodrop spectrophotometer. cDNA synthesis was performed using the iScript gDNA Clear cDNA Synthesis Kit (Bio-Rad) according to the manufacturer's protocol, with 500 ng of input RNA per sample. Appropriate no-RT controls were included. The final cDNA synthesis reaction was diluted to 100 µL using DNase-free water. Quantitative real-time PCR was performed using 2X iQ SYBR Green Supermix (Bio-Rad) according to the manufacturer's protocol, using a 2-step cycling method with a 55°C annealing/extension temperature. Each well of the PCR plate was prepared by mixing a template mastermix (4 µL ddH_2O, 10 µL 2X iQ SYBR Green Supermix, 1 µL diluted cDNA synthesis reaction) with a primer mastermix (3.8 µL ddH_2O, 0.6 µL 10 µM forward primer, 0.6 µL 10 µM reverse primer). The PCR was performed using a Bio-Rad iQ5 Multicolor Real-Time PCR Detection System. The sequences of primers used for quantitative real-time PCR can be found in *Appendix Table A1.3*.

Expression analysis by Western blotting. A 10 mL saturated, overnight *M. maripaludis* culture was harvested by centrifuging for 10 min at 10,000 × *g* at 4°C. The cell pellet was resuspended in 500 µL lysis buffer (50 mM KH_2PO_4, 1 mini-cOmplete EDTA-free protease inhibitor cocktail tablet (Roche) per 10 mL buffer, pH 7.4). The samples were mixed with 300 µL 0.1 mm glass disruption beads (Research Products International) in 2 mL O-ring sealed, screw cap tubes and were lysed using a BioSpec Products Mini-BeadBeater by beating for 1 min at 4°C. Cell debris was pelleted by centrifuging for 10 min at 20,000 × *g* at 4°C. The supernatant was then removed to a fresh tube and total protein was quantified using Bradford reagent with a lysozyme standard curve. Gel samples were prepared by mixing the lysates with additional lysis buffer and Laemmli loading buffer (Sigma). The samples were boiled for 10 min at 98°C before being separated using SDS-PAGE gel electrophoresis. Once the gel run was complete, the contents of the gel were transferred to a PVDF membrane. After the transfer, the membrane was blocked with 5% non-fat milk in Tris-buffered saline with 0.1% TWEEN®20 detergent (TBS-T) overnight at 4°C. The membrane was blotted using Mouse anti Strep-tag Classic:HRP antibody (Bio-Rad) (1:4000 dilution in 5% non-fat milk in TBS-T) for at least 2 h at room temperature or overnight at 4°C. After staining, the membrane was washed for 10 min three times with fresh TBS-T. The blot was developed using Western Lightening Plus-ECL (PerkinElmer) and imaged using a Universal Hood II (Bio-Rad).

Isotopic labelling of 3-hydroxybutyrate (3HB). *Methanococcus maripaludis S2* was inoculated at 1:100 dilution in 10 mL McNA-Ala medium plus 10 mM ^{13}C-labelled and unlabelled alanine in 250 mL serum bottle with a butyl rubber stopper and aluminum crimp. The headspace of the bottles was pressurized to 275 kPa with an H_2/CO_2 (80%, 20%) gas mixture once daily. Cells were grown at 37°C with shaking at 250 rpm. When the cultures reached an OD_{600} of 0.60,

they were centrifuged at 10,000 × *g* for 30 min at 4°C. 100 µL of the cleared culture media was diluted with 100 µL of 50 µM adipic acid, which served as the internal standard. Samples were filtered through a 96-well MultiScreen$_{HTS}$ plate before injecting onto an Agilent 1290 HPLC equipped with an auto-sampler, a Phenomenex Rezex-ROA Organic Acid H^+ column (150 × 4.6 mm) and with a Carbo-H^+ Security Guard cartridge. 0.5% (*v/v*) formic acid was used as the mobile phase with a flow rate of 0.3 mL/min and the column temperature set to 65°C. The product was quantified using mass spectrometry on an Agilent 6460 triple quadrupole MS with ESI source, operating in precursor ion scan mode with the fragmentor voltage set at 0V. Between 5-8 min, the following m/z values were monitored: 103.1 (3-hydroxybutyric acid), 104.1, 105.1, 106.1 and 107.1. Samples were quantified relative to the extracted ion count (EIC) for unlabeled 3HB.

3HB quantification using HPLC-MS/MS. Production cultures were sampled 7 days post-inoculation. Cell culture samples (1 mL) were thawed, mixed with 300 µL 0.1 mm glass disruption beads in 2 mL O-ring sealed, screw cap tubes and were lysed using a BioSpec Products Mini-BeadBeater by beating for 45 s at 4°C. Lysates were centrifuged for 10 min at 20,000 × *g* at 4°C to pellet cell debris. 100 µL of the cleared culture media was diluted into 100 µL of 50 µM adipic acid, which served as the internal standard. Samples were filtered through a 96-well MultiScreen$_{HTS}$ plate before injecting onto an Agilent 1290 HPLC equipped with an auto-sampler, a Phenomenex Rezex-ROA Organic Acid H^+ column (150 × 4.6 mm) and with a Carbo-H^+ Security Guard cartridge. 0.5% (*v/v*) formic acid was used as the mobile phase with a flow rate of 0.3 mL/min and the column temperature set to 55°C. 3HB was quantified using mass spectrometry on an Agilent 6460 triple quadrupole MS with ESI source, operating in negative ion MRM transition mode with the fragmentor voltage set at 70V. Between 5-8 min, the following transitions and collision energies were monitored: *m/z* 145.1→83.1, 10V (adipic acid, internal standard); *m/z* 103.1→59.2, 5V (3-hydroxybutyric acid). Samples were quantified relative to a standard curve of 7.8125, 15.625, 31.25, 62.5, 125, 250, 500, and 1000 mg/L 3HB prepared in McCas medium.

PHB quantification using HPLC-UV. Production cultures were sampled 14 days post-inoculation. 1 mL culture samples were stored at -80°C until analysis. The culture samples were thawed, mixed with 300 µL 0.1 mm glass disruption beads in 2 mL O-ring sealed, screw cap tubes and were lysed using a BioSpec Products Mini-BeadBeater by beating for 45 s at 4°C. Lysates were centrifuged for 10 min at 20,000 × *g* at 4°C to pellet biomass. Samples were acid-digested using 500 µL of concentrated sulfuric acid at 90°C for 30 min. They were then diluted with another 500 µL of 1M KOH and filtered through a 96-well MultiScreen$_{HTS}$ plate. 100 µL of the filtered sample was diluted into 100 µL of 50 µM adipic acid, which served as the internal standard. Samples were injected onto an Agilent 1260 HPLC equipped with an Aminex HPX-87H ion-exclusion organic acid analysis column (300 x 7.8 mm) (Bio-Rad Laboratories, Richmond, CA) and an Aminex XPX-85X ion-exclusion guard column. 0.014N H_2SO_4 was used as a mobile phase with 0.6 mL/min and the column temperature set to room temperature. Absorbance of crotonic acid and adipic acid was measured at 235 nm and 425 nm, respectively, using an HPLC-UV-Vis detector. Samples were quantified relative to a standard curve of 31.25, 62.5, 125, 250, 500 and 1000 mg/L crotonic acid prepared in Milli-Q water.

3.3. Results and discussion

Design of a synthetic pathway for the (S)-3-hydroxybutyrate monomer. Using a model pathway for converting acetyl-CoA into (S)-3-hydroxybutyrate ((S)-3HB), we aimed to establish the fundamentals of pathway engineering in this host. The pathway consists of the well-characterized enzymes PhaA (*Ralstonia eutropha*), Hbd (*Clostridium acetobutylicum*) and TesB (*Escherichia coli*) [*29*] (*Figure 3.1*). In past work from our lab, a library of rationally designed polycistronic and monocistronic constructs were assembled and screened for mRNA and protein expression. Engineered strains expressing polycistronic constructs containing the *phaA-hbd-tesB* pathway had detectable mRNA for pathway genes in many cases (*Joe Gallagher's thesis*). However, protein levels from pathway genes were undetectable, suggesting a possible bottleneck in mRNA translation. When testing monocistronic constructs, we achieved robust protein expression of all enzymes in the pathway using the P_{mcrB} promoter from *Methanococcus voltae,* although (S)-3HB titers were below the limit of detection when engineered strains were grown in rich, undefined media (McCas). To troubleshoot the lack of production, we generated strains expressing each of the pathway enzymes alone and tested for enzyme activity in cell lysates supplemented with substrates and cofactors (*Figure 3.1*). This result suggests that the production problem is not due to an intrinsic lack of enzyme activity. Instead, it could be caused by a lack of substrate or appropriate redox cofactor.

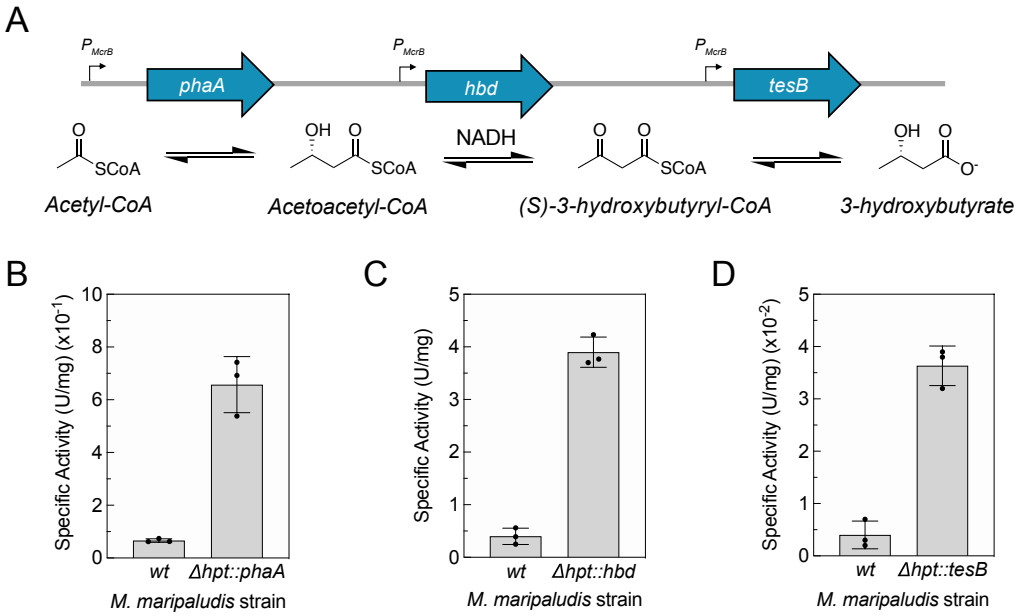
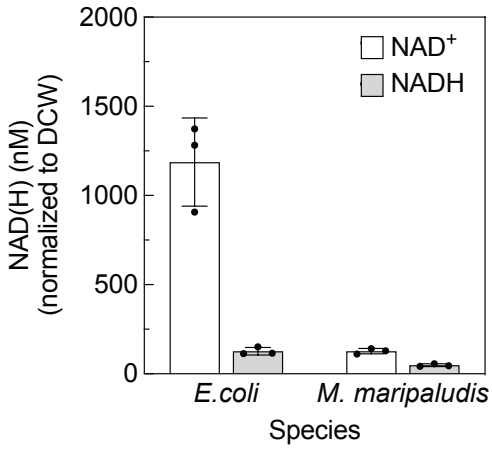


Figure 3.1. Targeting the acetyl-CoA pool of M. maripaludis for (S)-3-hydroxybutyrate ((S)-3HB) biosynthesis. *(A) Illustration of the phaA-hbd-tesB pathway. The pathway is composed of three enzymes that catalyze the synthesis of (S)-3HB from acetyl-CoA. Expression of each gene is driven by the P_{mcrB} promoter from Methanosarcina barkeri. Hbd is an NADH-dependent enzyme. Enzyme names are italicized in blue arrows. phaA, acetoacetyl-CoA thiolase/synthase from Ralstonia eutropha; hbd, 3-hydroxybutyryl-CoA dehydrogenase from Clostridium acetobutylicum; tesB, acyl-coA thioesterase from Escherichia coli. All constructs were targeted to the hypoxanthine phosphoribosyltransferase (hpt) site via homologous recombination. (B-D) Enzymatic activity of the (S)-3HB pathway enzymes were determined in M. maripaludis crude lysates supplemented with appropriate substrates and cofactors. (B) PhaA activity; (C) Hbd activity; (D) TesB activity. All three engineered strains show greater activity than wild type M. maripaludis. All data are mean ± SD (n = 3).*

Identifying metabolic bottlenecks for (S)-3HB production. Since methanogenesis has evolved to use coenzyme F_{420} and 2[4Fe-4S] ferredoxin as cofactors rather than NADH [*30*], we postulated that intracellular NADH levels would be low in methanogens. Furthermore, low intracellular NADH levels may be limiting Hbd activity, and therefore (S)-3HB biosynthesis, in this host. We measured NAD(H) levels in *M. maripaludis* grown in McCas media and compared this to NAD(H) levels in *E.coli* grown anaerobically in LB media. Remarkably, we found that the NAD(H) levels in *M. maripaludis* were approximately 5% of NAD(H) levels in *E. coli,* although the ratio of the two appeared similar in both species (*Figure 3.2*). In *E.coli,* NAD^+ and the ratio of NADH:NAD^+ have roles in metabolic processes through transcriptional regulation, allosteric modulation of enzyme activity, the overall flux of glycolysis and the molar ratios of excreted fermentation products [*33*]. In contrast, the role of NAD^+ and NADH in methanogens has not been clearly elucidated. Measuring expression levels of *de novo* NAD^+ biosynthesis pathway genes from *M. maripaludis* showed that they were expressed at less than 30% of the levels for a housekeeping gene (*Figure 3.3)*, suggesting that NAD^+ biosynthesis rates may also be low.

To determine the possible role of NAD(P)H in this species, we began by using a genome-scale metabolic model of *M. maripaludis* (iMR539) and performed flux balance analysis (FBA) assuming standard nutrient consumption rates in minimal medium with supplemented acetate. According to the genome-scale model, NAD(P)H appears to be predominantly associated with amino acid and coenzyme biosynthesis pathways. Despite the variety of reactions that are putatively associated with NAD(P)H, our flux balance analysis also suggests that the steady state turnover of these cofactors is less than 1% of the turnover of ferredoxin (Fd) and coenzyme F_{420} (*Figure 3.2*). This suggests that the majority of redox cofactor-dependent activity in *M. maripaludis* is confined to a small number of Fd- and coenzyme F_{420}-dependent reactions, most of which appear to be in central metabolism.

A



B

Cofactor	# of Reactions	Total Flux (μmol/min/gDCW)	% Total
NAD(H)	47	0.4	0.2%
NADP(H)	34	2.2	1.1%
Ferredoxin	12	106.0	51.7%
Coenzyme F420	11	96.4	47.0%
Total	104	205.0	100.0%

Figure 3.2. Total redox cofactor usage in M. maripaludis metabolic model. *(A) The concentration of NAD^+ (in white) and NADH (in gray) in M. maripaludis and E. coli were determined using a luciferase-based luminescence assay and normalized to milligrams of dry cell weight (DCW). The NAD(H) levels in E.coli are 30x higher than the levels in M. maripaludis. All data are mean ± SD (n = 3). (B) Flux balance analysis (FBA) was performed using the MATLAB COBRA toolbox and the most recent metabolic model for M. maripaludis, iMR539* [*23*]*. Reactions that oxidized ferredoxin (Fd), coenzyme F_{420}, reduced nicotinamide adenine dinucleotide (NADH) and reduced nicotinamide adenine dinucleotide phosphate (NADPH) were sorted by cofactor. The corresponding fluxes for reactions using each cofactor were added in order to obtain "Total oxidative flux" and "%Total" values. This analysis reveals that nicotinamide-based cofactors represent approximately 1% of total redox cofactor-dependent reactions*

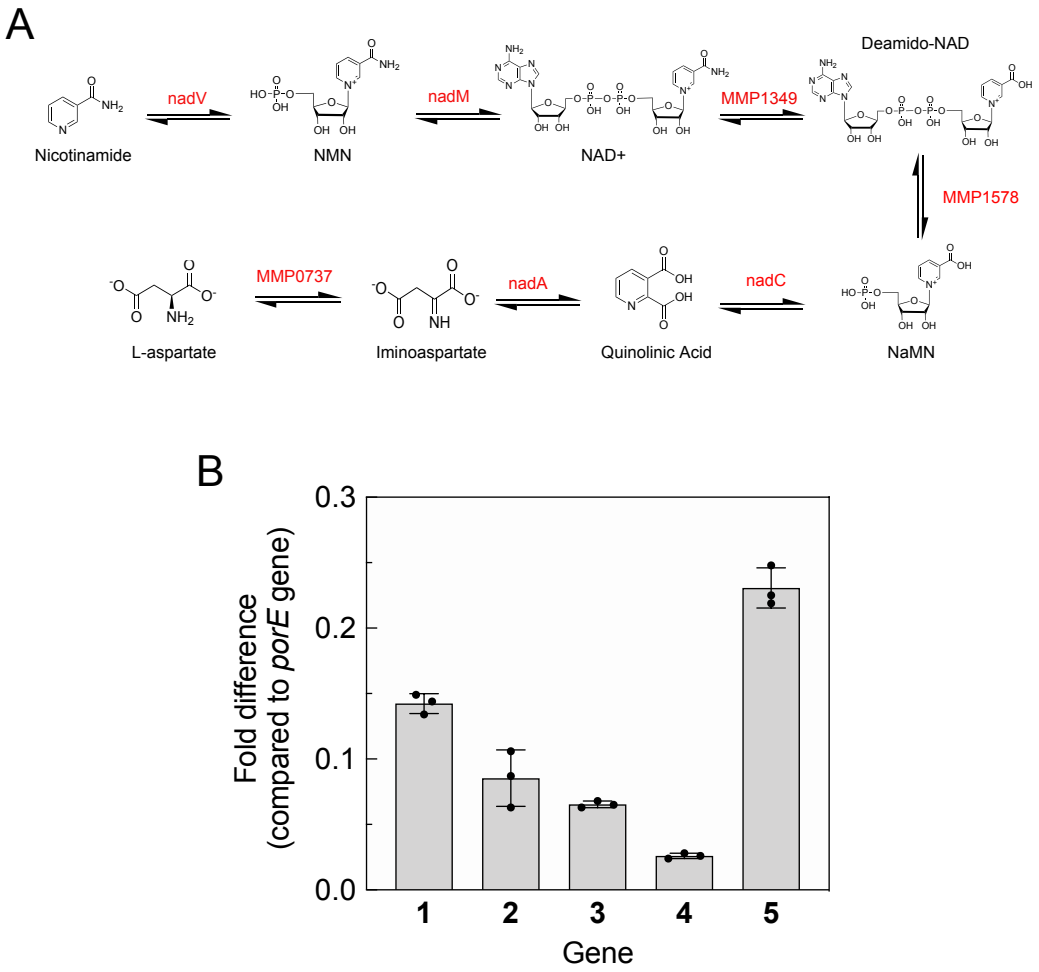


Figure 3.3. qRT-PCR analysis of de novo NAD⁺ biosynthesis pathway enzymes. *(A) The native de novo NAD⁺ biosynthesis pathway in M. maripaludis converts L-aspartate into NAD⁺ via five enzymatic steps. Enzyme names are displayed in red. MMP0737, aspartate dehydrogenase; nadA, quinolinate synthase; nadC, quinolinate phosphoribosyl transferase:nicotinate-nucleotide pyrophosphorylase; MMP1578, nicotinamide-nucleotide adenylyltransferase; MMP1349, NAD⁺ synthase. (B) The relative transcript abundance of the NAD⁺ biosynthetic pathway genes were compared to the abundance of pyruvate ferredoxin oxidoreductase subunit E (porE), an important housekeeping gene, using quantitative real-time PCR. 1: MMP0737, 2: nadA, 3: nadC, 4: MMP1578, 5: MMP1349. The level of expression of each gene of the NAD⁺ biosynthetic pathway is relatively low relative to porE. All data are mean ± SD (n = 3).*

Coupling NAD(H) production to nitrogen metabolism via alanine dehydrogenase. We next sought to gain more insight into NAD(H) use in this species by performing adenosine monophosphate (AMP)-agarose chromatography on crude lysates to enrich for proteins that bind to adenosine-containing cofactors like NAD^+, and then performed shotgun proteomics on this enriched pool (*Appendix Table A3.3*). Alanine dehydrogenase (*ald*) was the most abundant protein found in this proteomics approach. This enzyme catalyzes NAD^+-dependent oxidative deamination of L-alanine to liberate ammonium and pyruvate. AlaDH is also known to be essential for growth of *M. maripaludis* in mineral media with L-alanine as the sole nitrogen source (McNA-Ala) [*34*]. Because of these observations, we hypothesized that growth of *M. maripaludis* in McNA-Ala media may drive greater *in vivo* NAD^+ turnover than growth in alanine-independent conditions, leading to higher titers of (S)-3HB. To test this hypothesis, *phaA-hbd-tesB* containing strains of *M. maripaludis* were grown in McNA-Ala as well as a defined medium with 10 mM ammonium chloride as the sole nitrogen source (McNA) and McCas. Through these production experiments, we found that growth in McNA-Ala led to (S)-3HB titers of 14 ± 2.8 mg/L, the highest titer of the three media conditions (*Figure 3.4*).

Next, transcriptomic studies were performed to elucidate the effects of media choice on production titers (*Figure 3.5, Figure 3.6)*. During growth in either minimal media condition, most genes involved in the modified Wood-Ljungdahl and methanogenesis pathways were upregulated compared to McCas. Since amino acids are not supplied in minimal media (with the exception of alanine in the McNA-Ala condition), acetyl-CoA formation is necessary for cell growth. Conversely, acetyl-CoA formation may not be as relevant to growth in McCas, since most amino acids, vitamins and cofactors can be taken up from the medium. This suggests that acetyl-CoA production could be a metabolic bottleneck in McCas-dependent growth but can be mitigated by growth in defined mineral media. Pyruvate derived from supplemented alanine in McNA-Ala media can also be converted to acetyl-CoA for isoprenoid and (S)-3HB biosynthesis, as shown in isotopic labelling experiments of (S)-3HB in media supplemented with ^{13}C-alanine as the sole N-source (*Figure 3.7*).

Interestingly, a key gene involved in alanine metabolism was upregulated in both minimal media conditions: alanine dehydrogenase (*MMP1513*). The role of this enzyme in metabolism likely varies depending on the nitrogen source used for growth. During growth in McNA, *M. maripaludis* must use alanine dehydrogenase to produce alanine, an essential amino acid. Conversely, the reverse reaction is necessary for growth when alanine is the sole nitrogen source, suggesting that NAD^+ regeneration is a necessary feature of growth in McNA-Ala, in particular. During growth in McNA-Ala, NADH levels and the NADH:NAD^+ ratio were observed to be higher than during growth in McCas or McNA (*Figure 3.4*). In addition, the highest expression of the sodium-alanine symporter gene (*agcS*) was observed in McN-Ala growth conditions (*Figure 3.5*). These observations lead us to the model in which intracellular alanine flux could provide a thermodynamic driving force for the oxidative deamination and NADH regeneration activity of *ald* during growth in McNA-Ala, leading to improved NADH:NAD^+ ratios. Overall, our data suggests that the choice of media in our experiments has a profound influence on acetyl-CoA formation and NADH-dependent metabolism, particularly through the activity of alanine dehydrogenase.

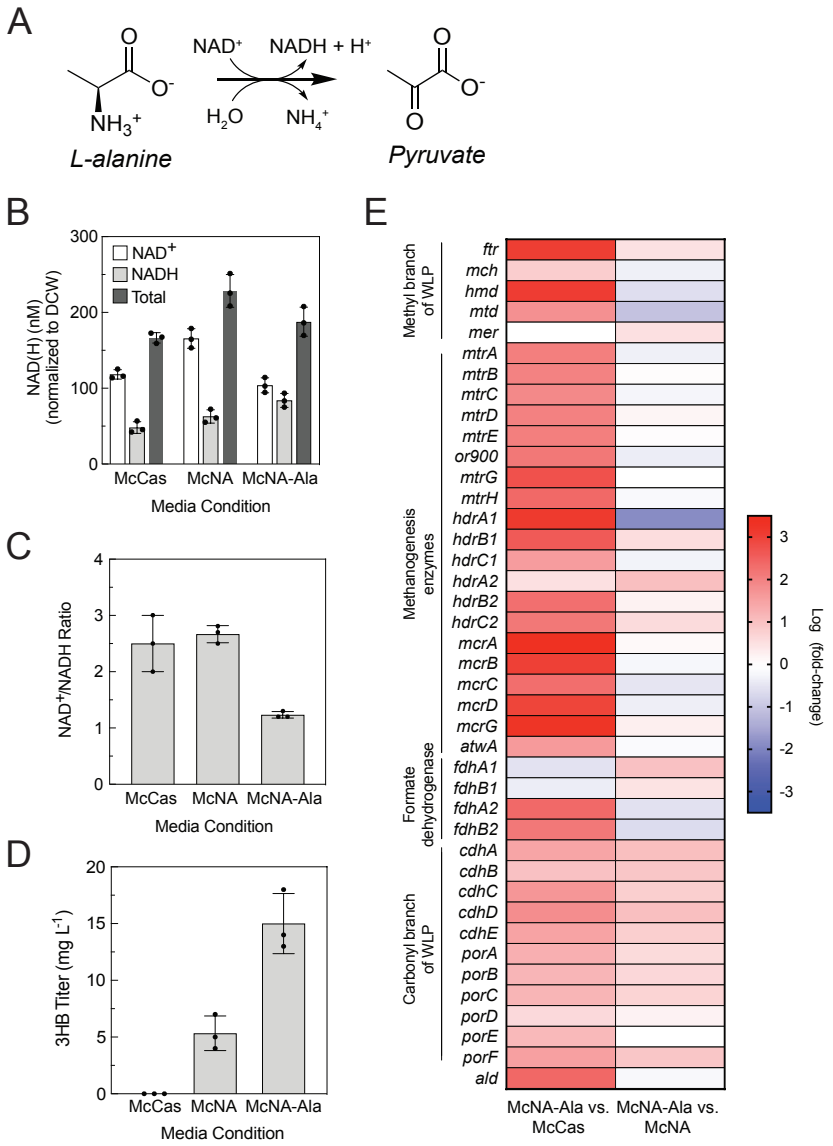


***Figure 3.4*. Physiological changes in *M. maripaludis* grown with L-alanine as the sole nitrogen source lead to greater (S)-3HB production titers.** *(A) Oxidative deamination of L-alanine to pyruvate is catalyzed by alanine dehydrogenase, using NAD^+ as an electron acceptor. Liberated ammonium ions can be incorporated into biomass, allowing alanine to be used as a sole nitrogen source for growth in M. maripaludis (McNA-Ala). Alanine formation via the activity of alanine dehydrogenase is also necessary for growth in defined medium with ammonium chloride as the sole nitrogen source (McNA), but it does not appear to be necessary during growth in rich, undefined media (McCas). (B) The concentration of NAD^+ (in white), NADH (in light gray) and both (in dark gray) in M. maripaludis grown with various nitrogen sources was determined using an enzyme-based luminescence assay. Values were normalized for comparison using milligrams (mg) of dry cell weight (DCW). McCas = Rich, undefined media with both ammonium chloride and alanine available as nitrogen sources; McNA = Minimal, defined media with 10 mM ammonium chloride; McNA-Ala = Minimal, defined media with 10 mM alanine. (C) Ratio of measured NAD^+ and NADH values from M. maripaludis grown in three media conditions with different nitrogen sources shows that strains grown in McNA-Ala have a lower NAD^+/NADH ratio than McNA and McCas. (D) (S)-3HB titers were measured for ⊿hpt::phaA-hbd-tesB strains grown in McCas, McNA or McNA-Ala media conditions after 7 days post-inoculation. Strains grown in McNA-Ala have 3HB product titers that are 3-fold higher than production in McNA, while product was not detected in McCas conditions. All data are mean ± SD (n = 3). (E) Heatmap showing genes of interest in carbon metabolism of M. maripaludis. Transcript abundance of three biological replicates from each growth condition was averaged and compared pairwise to determine the log_2(fold-change). Log_2(fold-change) of transcript levels greater than 0 is red and log_2(fold-change) lower than 0 is blue. Genes involved in alanine metabolism as well as the modified Wood-Ljungdahl and methanogenesis pathways were upregulated during growth in both defined, mineral media.*

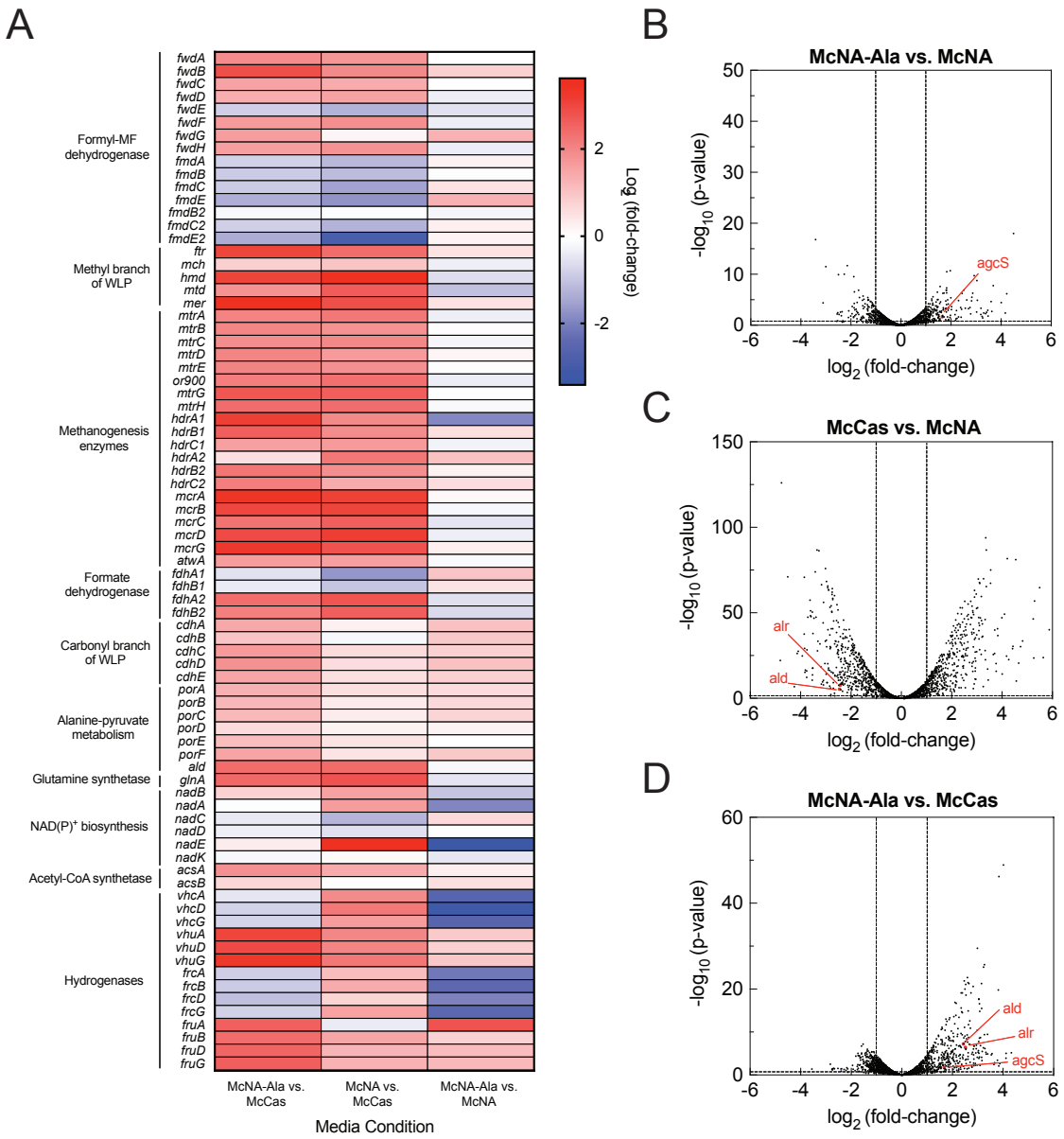


Figure 3.5. RNA-seq analysis of M. maripaludis grown in McCas, McNA and McNA-Ala. *(A) Heatmap showing genes of interest in the carbon metabolism and NAD⁺ biosynthesis pathway of M. maripaludis. Transcript abundance of three biological replicates from each growth condition was averaged and compared pairwise to determine the log_2(fold-change). Log_2(fold-change) of transcript levels greater than 0 is red and log_2(fold-change) lower than 0 is blue. Volcano plot representation of differential expression analysis in M. maripaludis grown in (B) McNA-Ala vs. McNA; (C) McCas vs. McNA; and (D) McNA-Ala vs. McCas. Dashed lines indicate boundaries of significantly increased or decreased expression ($p < 0.05$, log_2(fold change) is greater than 1 or less than -1), respectively. Broad shifts in methanogenesis, autotrophic acetyl-CoA synthesis and hydrogenase expression were observed in both minimal conditions, suggesting an increased requirement for carbon fixation and acetyl-CoA formation to provide the necessary precursors for biomass. Upregulation of alanine dehydrogenase (ald) and a sodium-alanine symporter (agcS) in McNA-Ala growth supports a model of greater alanine deamination and NADH regeneration compared to growth in McCas or McNA.*

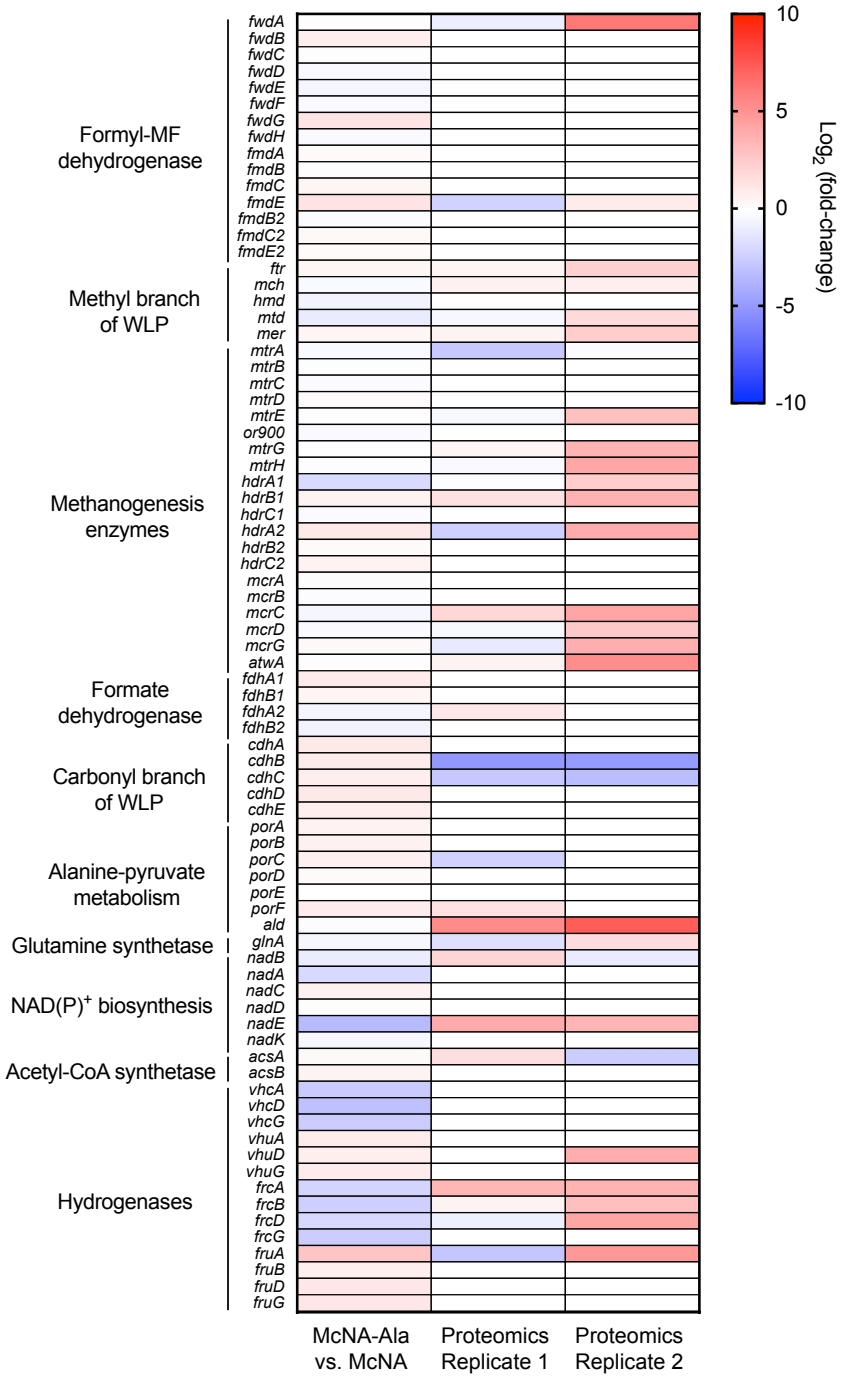


Figure 3.6. $^{14}N/^{15}N$-labeling quantitative proteomics of M. maripaludis grown in McNA-Ala vs. McNA. *Heatmap showing comparative proteomics analysis of M. maripaludis grown using $^{15}NH_4Cl$ or L-alanine as a sole nitrogen source in defined mineral media (Proteomics Replicate 1 and Proteomics Replicate 2). Data were compared to transcriptomic analysis for cell cultures grown in McNA-Ala vs. McNA. Two biological replicates were used for proteomics and three biological replicates were used for transcriptomic analysis. Mean protein or transcript abundance were calculated and log_2(fold-change) was used to compare differences in protein and transcript levels. Log_2(fold-change) of transcript levels greater than 0 is red and log_2(fold-change) lower than 0 is blue. Carbon metabolism and NAD^+ biosynthesis genes were isolated for comparisons, and key genes such as alanine dehydrogenase (ald) were upregulated at the protein level in cultures grown in McNA-Ala.*

A

L-Alanine

m/z	Abundance (%)	Formula
103.0401	95.2	$^{12}C_4H_8O_3$
104.0434	4.76	$^{12}C_3{}^{13}CH_8O_3$
105.0468	n.d.	$^{12}C_2{}^{13}C_2H_8O_3$
106.0501	n.d.	$^{12}C_1{}^{13}C_3H_8O_3$
107.0535	n.d.	$^{13}C_4H_8O_3$
$\%^{13}C$	1%	

B

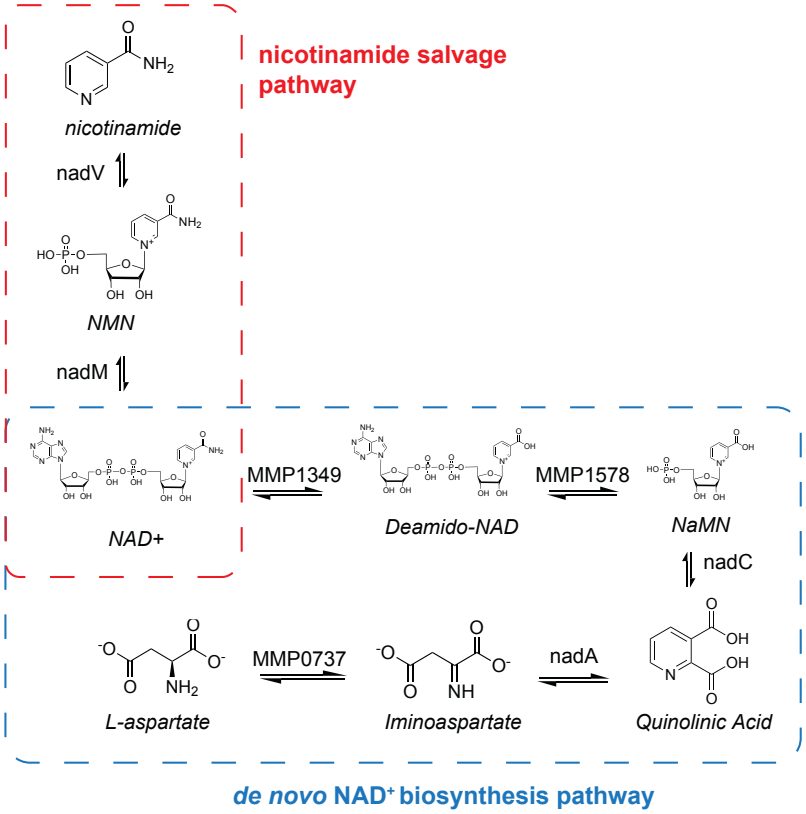
L-Alanine-3-^{13}C

m/z	Abundance (%)	Formula
103.0401	48.0	$^{12}C_4H_8O_3$
104.0434	35.0	$^{12}C_3{}^{13}CH_8O_3$
105.0468	17.0	$^{12}C_2{}^{13}C_2H_8O_3$
106.0501	n.d.	$^{12}C_1{}^{13}C_3H_8O_3$
107.0535	n.d.	$^{13}C_4H_8O_3$
$\%^{13}C$	17%	

Figure 3.7. ^{13}C-labelling of 3-hydroxybutyrate produced by M. maripaludis grown in McNA-Ala. *High-resolution mass spectrometry of extracted 3HB after 7-days post inoculation of M. maripaludis strain ΔhptA::phaA-hbd-tesB with either (A) L-alanine or (B) L-alanine-3-^{13}C, where alanine is the sole nitrogen source. Additional carbon sources provided in the media were acetate and formate. Tubes were pressurized to 275 kPa with H_2/CO_2 (80%, 20%). Labelled carbon from alanine is observed in final product with up to 17% isotopically labelled carbon, suggesting that alanine contributes a fraction of the carbon in the final 3HB titer. Experiments were performed in biological triplicate and mean was calculated for each expected 3HB isotope (n=3).*

Engineering NAD(H) pools in *M. maripaludis*. To more precisely test the hypothesis that NADH is a rate-limiting factor, we engineered *M. maripaludis* to increase NADH pools (*Scheme 3.2*). We first attempted to systematically overexpress each gene in the *de novo* NAD^+ biosynthesis pathway native to *M. maripaludis.* This strategy revealed that overexpression of quinolinate synthase (*nadA*), quinolinate phosphoribosyltransferase (*nadC*) and aspartate dehydrogenase (*MMP0737*) led to approximately 2-fold increases in total NAD(H) levels (*Figure 3.8*). *nadC* catalyzes an intermediate step of the pathway that converts iminoaspartate to quinolinate. Since iminoaspartate is a labile molecule, an increase in quinolinate synthase expression may be important for favoring quinolinate synthesis over iminoaspartate degradation. Alternatively, overexpression of the upstream *MMP0737* may increase iminoaspartate synthesis, which might compensate for its breakdown.

In addition to *de novo* NAD^+ biosynthesis, we also attempted to use a cofactor salvage approach to increase NADH pools. Tuning the concentration of supplemented NAD^+ precursor might allow for greater tunability and control of intracellular NAD(H) levels in an organism with relatively few tools for tuning gene expression. We chose nicotinamide as a precursor because it has no charge under physiological conditions. Unlike many bacterial and eukaryotic species, no nicotinamide salvage pathway appears to exist in this methanogen and therefore a novel pathway had to be designed. For this approach, we expressed nicotinamide phosphoribosyltransferase (NadV) from *H. ducreyi* [*35*], an enzyme that catalyzes the conversion of nicotinamide into nicotinamide mononucleotide (NMN). In addition, a second copy of the native nicotinamide mononucleotide adenylyltransferase (NadM) from *M. maripaludis* was also expressed. Combining these two genes allows for the constitution of a nicotinamide salvage pathway in *M. maripaludis.* Intracellular NAD(H) levels increased by almost 4-fold upon media supplementation of 50 mM nicotinamide (*Figure 3.9*). Comparing the two approaches for increased NAD(H) pools, the greatest improvement in NADH levels came from overexpression of NadMV.



Scheme 3.2. Expanding NAD(H) cofactor pools by salvage pathway or de novo pathway overexpression. *Synthetic nicotinamide salvage pathway (outlined in red dashed lines) involves expression of a second copy of nicotinamide mononucleotide adenylyltransferase (nadM) and nicotinamide phosphoribosyltransferase (nadV) from Haemophilus ducreyi. The de novo NAD⁺ biosynthesis approach (outlined in blue dashed lines) was used by overexpressing each individual gene from the pathway. Enzyme names are displayed above the equilibrium arrows of each step. MMP0737, aspartate dehydrogenase; nadA, quinolinate synthase; nadC, quinolinate phosphoribosyl transferase:nicotinate-nucleotide pyrophosphorylase; MMP1578, nicotinamide-nucleotide adenylyltransferase; MMP1349, NAD⁺ synthase.*

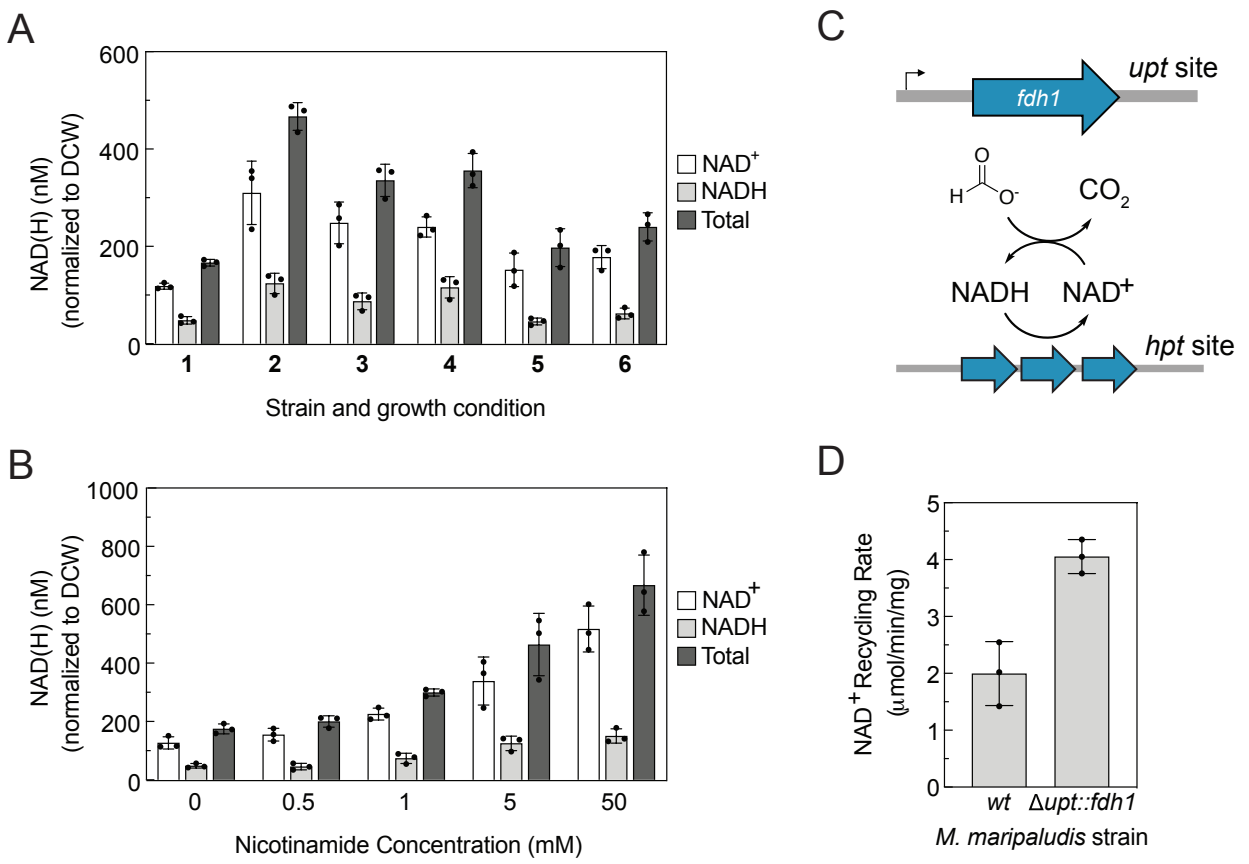


Figure 3.8. Engineered M. maripaludis shows improved NAD⁺ recycling rates and NAD(H) levels. *(A) The concentration of NAD⁺ (in white), NADH (in light gray) and both (in dark gray) in M. maripaludis engineered to overexpress genes from the de novo NAD⁺ biosynthetic pathway was determined using an enzyme-based luminescence assay. 1: Wild type M. maripaludis; 2: Δhpt::MMP0737; 3: Δhpt::nadA; 4: Δhpt::nadC; 5: Δhpt::MMP1578; 6: Δhpt::MMP1349. Overexpression of MMP0737, nadA and nadC increased the NAD(H) cofactor pool by approximately 2-fold compared to wild type M. maripaludis grown in McCas media. (B) The concentration of NAD⁺ (in white), NADH (in light gray) and both (in dark gray) was measured in engineered M. maripaludis expressing the nicotinamide salvage pathway (nadMV) with a supplementation of nicotinamide between 0 to 50 mM. Nicotinamide supplementation of 50 mM led to an approximately 3-fold improvement in total NAD(H) compared to wild type M. maripaludis in McCas media. (C) Genomic insertion of fdh1 at the upt genomic site allows for the use of formate as an energy source for NAD⁺ reduction and could balance the oxidation of NADH by a heterologous biosynthetic pathway inserted at the hpt site. CO_2 produced by formate oxidation can be used for the production of acetyl-CoA and downstream value-added chemicals. (D) NAD⁺ recycling rates were measured for clarified lysates of M. maripaludis engineered to express fdh1 and compared to wild type M. maripaludis. The NAD⁺ recycling rate was 2-fold higher in the engineered strain compared to wild type. All data are mean ± SD (n = 3).*

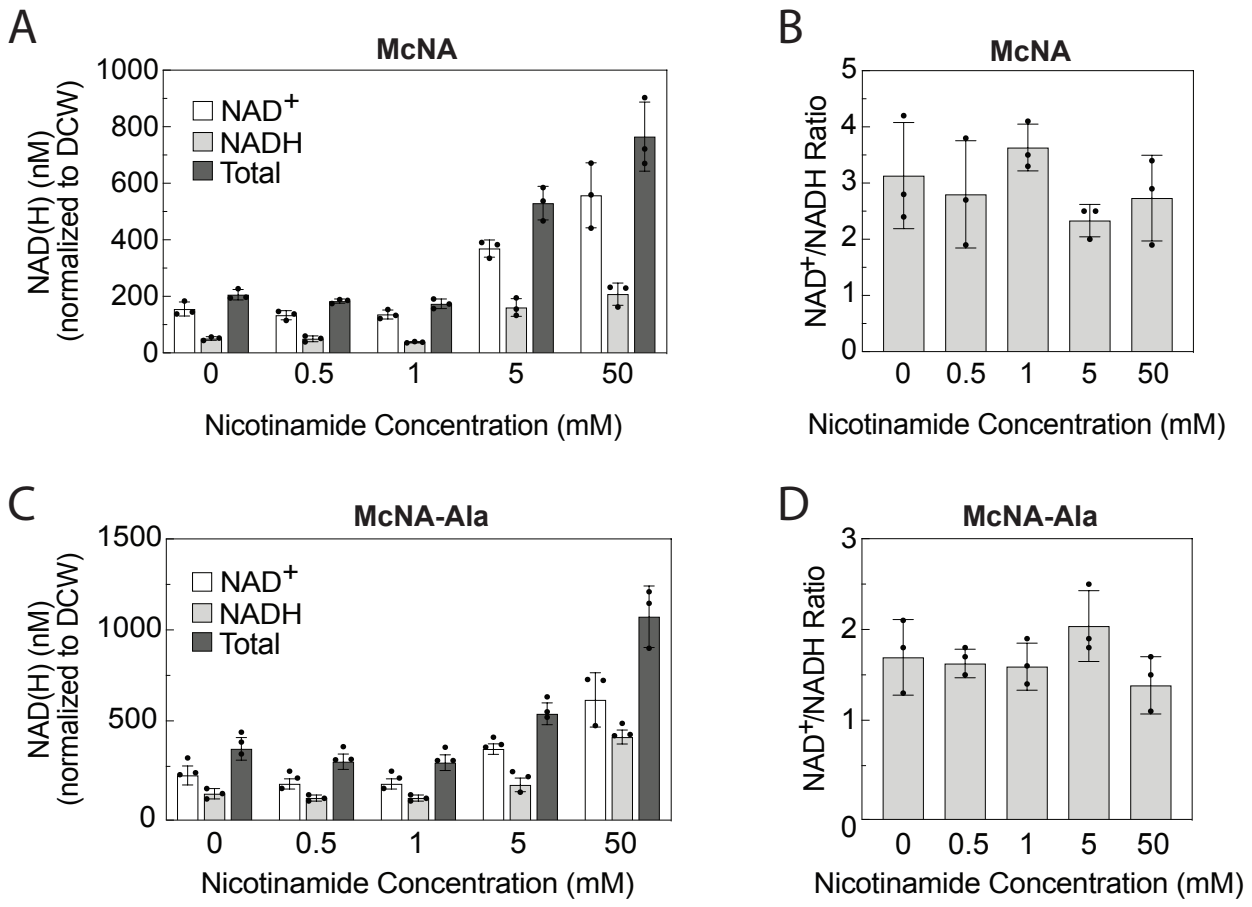


Figure 3.9. NAD(H) measurements for nicotinamide salvage pathway overexpression in McNA and McNA-Ala media. *The concentration of NAD^+ (in white) and NADH (in light gray) and total NAD(H) cofactor (in dark gray) in engineered Δhpt::nadMV strains grown in (A) McNA or (C) McNA-Ala with nicotinamide supplemented at the indicated concentrations. The corresponding NAD^+/NADH ratios for each strain grown in (B) McNA or (D) McNA-Ala. Total NAD(H) levels in engineered strains increased by 2- to 4-fold in the presence of 5 mM nicotinamide or greater. The NAD^+/NADH ratio did not vary across nicotinamide concentrations, but the overall ratio was lower in engineered strains grown in McNA-Ala compared to McNA. All data are mean ± SD (n = 3).*

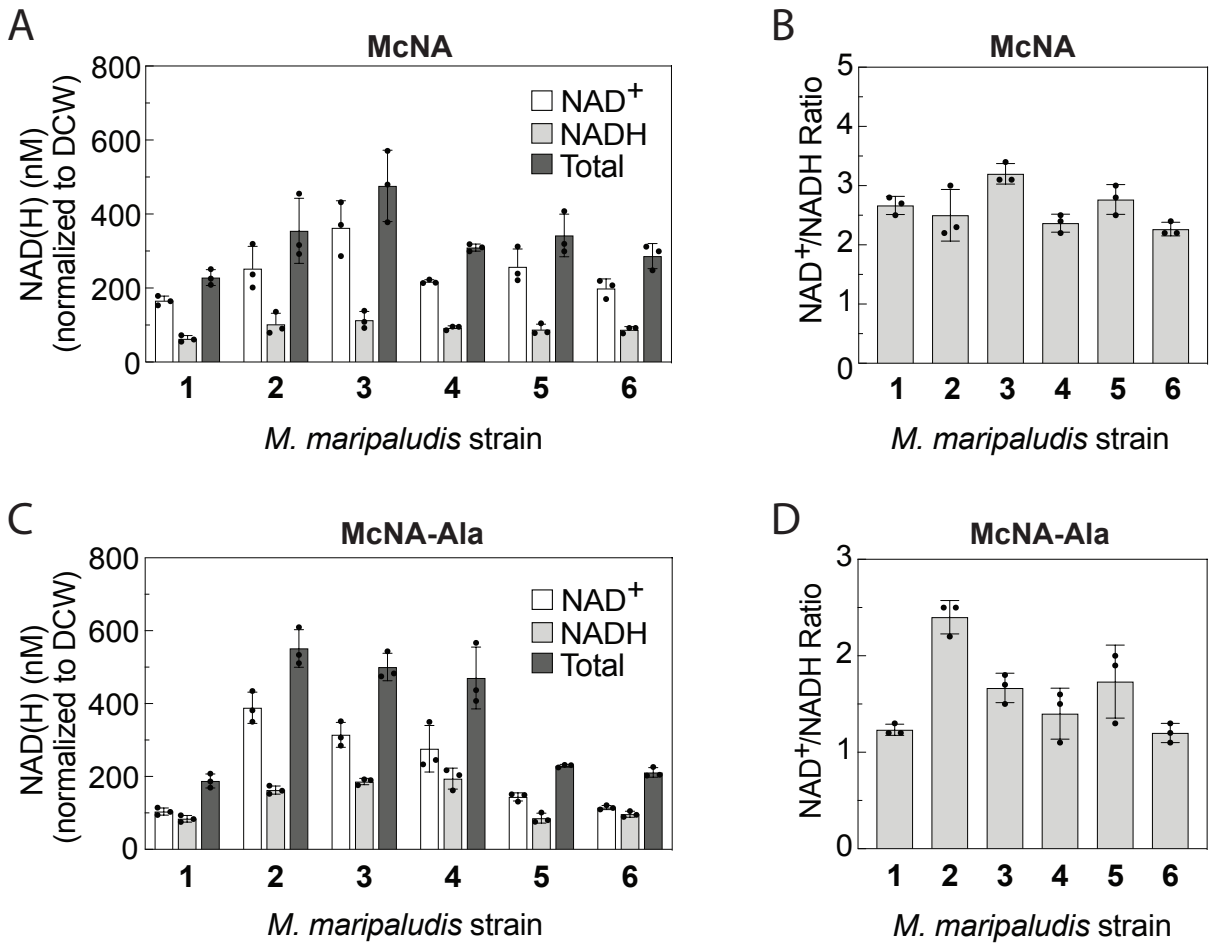
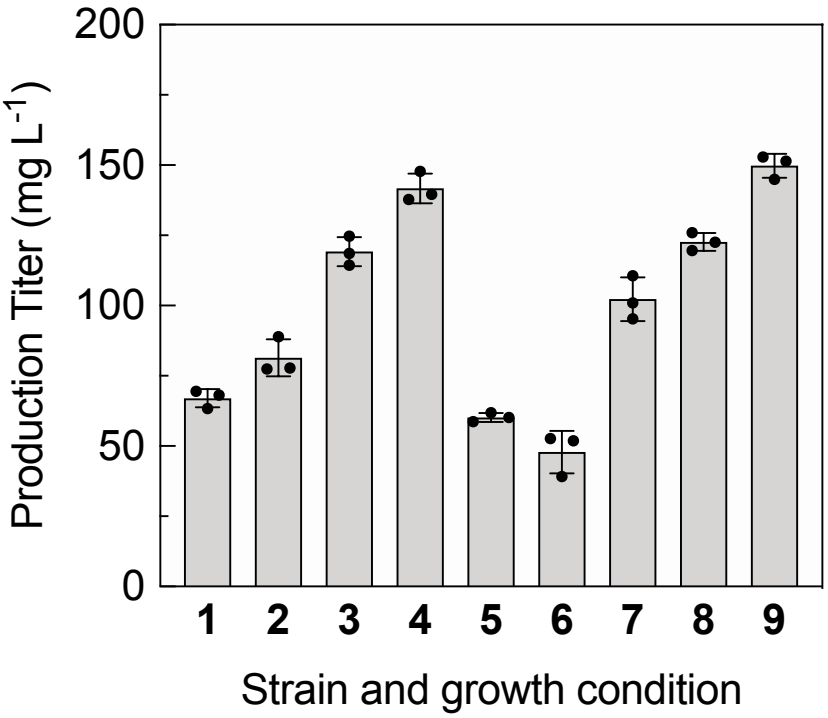


Figure 3.10. NAD(H) measurements for de novo NAD pathway overexpression in McNA and McNA-Ala media. *The concentration of NAD^+ (in white) and NADH (in light gray) and total NAD(H) cofactor (in dark gray) in engineered strains containing a second, overexpressed copy of a gene in the NAD^+ biosynthesis pathway, grown in (A) McNA or (C) McNA-Ala were determined using a luciferase-based luminescence assay. The corresponding NAD^+/NADH ratios for each engineered strain grown in (B) McNA or (D) McNA-Ala. 1: Wild type M. maripaludis; 2: Δhpt::MMP0737; 3: Δhpt::nadA; 4: Δhpt::nadC; 5: Δhpt::MMP1578; 6: Δhpt::MMP1349. Total NAD(H) levels in engineered strains with an overexpressed copy of MMP0737, nadA or nadC were approximately 2- to 3-fold higher than wild type strains, and the NAD^+/NADH ratio was not significantly different in any engineered strain. However, the NAD+/NADH ratio was lower in engineered strains grown in McNA-Ala compared to McNA. All data are mean ± SD (n = 3).*

In parallel, we also constructed (R)-3HB and polyhydroxybutyrate (PHB) pathways, which were chosen to demonstrate the production of value-added chemicals. Building on our investigation of NADH metabolism, we used a similar construct design as the *phaA-hbd-tesB* pathway but substituted Hbd with an NADH-dependent PhaB from *Halomonas bluephagenesis* [*36*]. Unlike Hbd, PhaB produces the (R)-stereoisomer of 3HB, which is required for the incorporation of monomers into a growing PHB polymer chain by PhaC. We then constructed a third biosynthetic pathway that used the PHB polymerase (PhaC) from *Cupriavadus necator* in combination with PhaA from *C. acetobutylicum* and PhaB from *H. bluephagenesis.* PhaC is known to incorporate (R)-3-hydroxybutyryl-CoA into an insoluble PHB polymer, which may also provide a thermodynamic driving force towards PHB biosynthesis from acetyl-CoA [*37*]. Initial measurements showed that PHB product titers were similar to (S)-3HB and (R)-3HB. However, combining the PhaABC pathway with NadMV and supplementing with 50 mM nicotinamide led to an approximately 32% increase in PHB titers (*Figure 3.11*).

To improve the titers further, we attempted to increase NAD^+ turnover by introducing a formate dehydrogenase (Fdh1) from *Candida boidini* [*38*]. This single subunit enzyme has been shown to improve NADH yield in engineered strains of heterotrophic hosts by catalyzing formate oxidation and NAD^+ reduction [*39*]. Since formate is already known to be a substrate for growth, we assumed that formate was transported into the cell using native transporters [*40*]. *M. maripaludis* strains containing Fdh1 had 4-fold higher formate-dependent NAD^+ turnover rates compared to wild type (*Figure 3.8*). This approach was combined with the NadMV and PhaABC pathway in *M. maripaludis,* which led to another 26% improvement in PHB titers (*Figure 3.11*). NadMV and Fdh1 were also combined with the PhaAB-TesB pathway, which showed an improvement of 26% to a final titer of 141 ± 5 mg/L. When combined with a formate dehydrogenase and nicotinamide salvage pathway, PHB product titers were improved by more than 2.5-fold compared to the PhaABC pathway-containing strain, with our highest titer of 150 ± 4 mg/L in McNA media (*Figure 3.11*). Titer improvements in the final strain were observed to be both formate- and nicotinamide-dependent. These improvements appeared to partly be growth-dependent and time-course monitoring of engineered strains showed that *NadMV-Fdh1* improved productivity in *M. maripaludis* over the course of growth (*Figure 3.12*).



Strain	*Pathway 1*	*Pathway 2*	*NADH regeneration*	*Supplements*
1	phaA-hbd-tesB			
2	phaAB-tesB			
3	phaAB-tesB	nadMV	fdh1	
4	phaAB-tesB	nadMV	fdh1	50 mM Nm
5	phaABC			
6	phaABC	nadMV		
7	phaABC	nadMV		50 mM Nm
8	phaABC	nadMV	fdh1	
9	phaABC	nadMV	fdh1	50 mM Nm

Figure 3.11. 3HB/PHB production titers were improved in a nicotinamide- and formate-dependent manner. *Production of 3HB or hydrolyzed PHB was measured 14 days post-inoculation of M. maripaludis strains containing pathway combinations. Samples varied by strain and supplemented precursors as indicated in the accompanying table in optimized production conditions. 3HB and PHB production were quantified by HPLC-UV/Vis. phaAB-tesB and phaABC containing strains both had improved production titers when NadMV or Fdh1 was used to increase the NADH pool and improve NAD^+ regeneration rates, respectively. All data are mean ± SD (n = 3).*

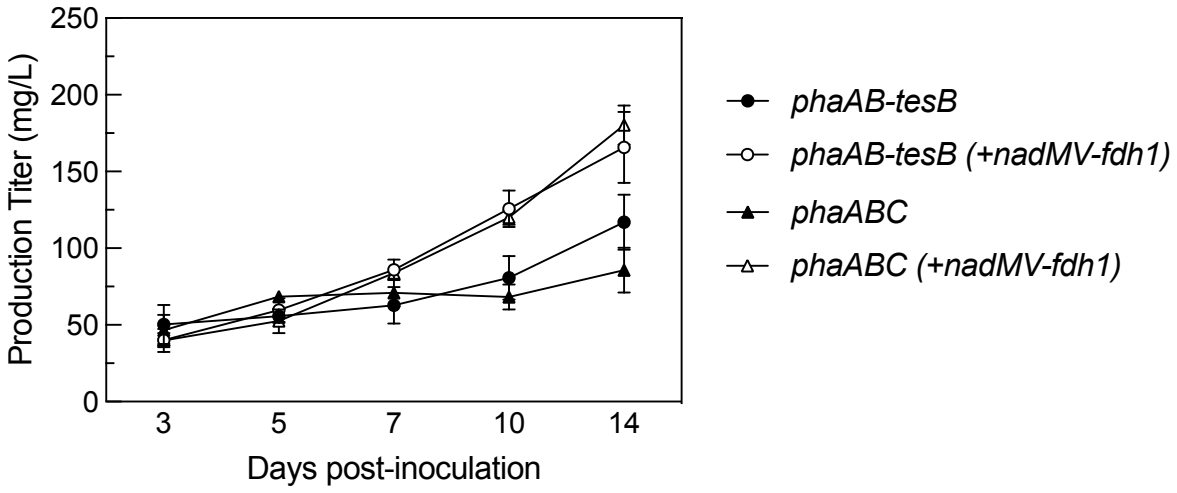


Figure 3.12. Time-course monitoring of (R)-3HB/PHB production titers *Production of both phaAB-tesB and phaABC was improved by the addition of the nadMV-fdh1 pathway, after an initial burst of production at day 3 post-inoculation. Production was conducted for 14 days post-inoculation. All data are mean ± SD (n = 3).*

3.4. Conclusions

Expanding the product scope of methanogenic archaea and gaining a better understanding of its metabolism improves the viability of this species as a host for metabolic engineering. Here, we show engineered strains of *M. maripaludis* capable of converting CO_2 and H_2 into (S)-3HB, (R)-3HB and PHB, while co-producing methane as a metabolic byproduct. The product titers achieved in this study greatly exceed those observed in previous reports of methanogen engineering and highlight effective strategies for alleviating metabolic bottlenecks in *M. maripaludis* [*41*]. Using *phaA-hbd-tesB* as a model pathway, it was shown that titers were higher in defined media (McNA and McNA-Ala) conditions compared to McCas. In defined media conditions, all essential components of the cell must be synthesized, and multiple gene clusters involved in acetyl-CoA synthesis were upregulated. This suggests the possibility that acetyl-CoA formation rates were higher in both conditions, which may have improved product titers. However, a key distinction between the McNA and McNA-Ala conditions is their alanine metabolism. When alanine is supplemented as a nitrogen source in McNA-Ala, it must undergo oxidative deamination by alanine dehydrogenase in order to liberate ammonium. Since this reaction uses NAD^+ as an electron acceptor, this reaction effectively couple nitrogen metabolism with NAD^+ turnover rates. Our results suggest that NAD^+/NADH ratios were lower in McNA-Ala conditions compared to McNA, which may also explain the improved (S)-3HB titers in McNA-Ala.

After targeted engineering of the NAD(H) cofactor pool and switching to (R)-3HB and PHB biosynthesis pathways, the highest product titers were observed in a strain containing the combination of a formate dehydrogenase from *C. boidinii* and NadMV, a synthetic nicotinamide salvage pathway that uses NadV from *H. ducreyi.* When NadMV was overexpressed in the absence of Fdh1, NAD(H) pools increased up to 4-fold higher than wild type *M. maripaludis* controls. Although titers improved, the NAD^+/NADH ratio did not appear to change greatly. On the other hand, formate dehydrogenase expressing strains have higher *in vivo* NAD^+ turnover rates. The combination of this enzyme with the NadMV pathway led to formate- and nicotinamide-dependent improvements in titer, resulting in PHB titers of approximately 150 ± 4 mg/L and (R)-3HB titers of 146 ± 15 mg/L. This suggests that the combination of higher NAD(H) pools from NadMV and

improved *in vivo* NAD^+ turnover had a synergistic benefit, leading to higher product titers than strains expressing either NadMV or Fdh1 alone.

Other Wood-Ljungdahl pathway containing organisms have also been engineered for 3HB or PHB production, including acetogenic bacteria [*42*, *43*]. These species also have Rnf complexes that can catalyze the oxidation of Fd^{2-} and reduction of NAD^+, unlike methanogens. Past efforts to produce (R)-3HB in engineered strains of the acetogen *Clostridium coskatii* led to a final (R)-3HB titer of 0.98 ± 0.12 g/L under autotrophic growth conditions [*42*]. In addition, a CRISPRi-based approach was used to divert carbon flux away from acetate formation towards 3HB production in *Clostridium ljungdahlii*, a model homoacetogen. Using this approach, approximately 46 mg/L of 3HB was produced using a heterologous *phaA-phaB-ptb-buk* pathway that also generated ATP for the cell, circumventing the need to produce ATP via acetate formation. Given the recent demonstration of CRISPR genetic tools in methanogenic species, a similar CRISPRi-based approach could be taken in *M. maripaludis* [*44*].

In contrast to the use of other engineered chemoautotrophs, there are unique benefits to the use of methanogens as engineering hosts. Methane is a byproduct of methanogen metabolism that can be continuously extracted during fermentation and can be used as a sustainable fuel or heating source. In addition, *M. maripaludis* and related methanogen species have been reported to use formate as an alternative energy source instead of hydrogen [*39*, *45*, *46*]. This may circumvent the growth limitations associated with poor hydrogen mass transfer kinetics in aqueous media and increase cellular productivity [*47*]. Furthermore, formate can be produced electrochemically, which allows for the use of a hybrid bio-inorganic platform that is independent of hydrogen [*48*]. Also, the modified Wood-Ljungdahl pathway is the most energy efficient carbon fixation pathway in nature, which may also help to minimize the energy requirements for chemical synthesis.

To further increase titers in *M. maripaludis,* target molecule biosynthesis rates could be coupled with growth to implement lab evolution strategies. Additionally, NADPH-dependent pathways may also be implemented for target chemical synthesis. Although NADPH levels were found to be low in this species, it may be improved by using NAD^+ kinase [*49*], supplemented by the approaches described in this study for increasing NAD(H) pools. In addition, Nfn oxidoreductase overexpression could be used to improve NADPH regeneration if necessary [*50*]. Overall, the favorable characteristics of this species enables it to become a promising host for future metabolic engineering efforts. Insights presented in this work provide a foundation for more extensive metabolic engineering efforts in *M. maripaludis* and may also be transferable to related archaeal species.

3.5. References

1. D. Y. C. Leung, G. Caramanna, M. M. Maroto-Valer, "An overview of current status of carbon dioxide capture and storage technologies". *Renewable and Sustainable Energy Reviews*. **2014**, *39*, 426–443.
2. D. C. Ducat, J. C. Way, P. A. Silver, "Engineering cyanobacteria to generate high-value products". *Trends in Biotechnology*. **2011**, *29*, 95–103.
3. J. Levering, J. Broddrick, K. Zengler, "Engineering of oleaginous organisms for lipid production". *Current Opinion in Biotechnology*. **2015**, *36*, 32–39.
4. X. G. Zhu, S. P. Long, D. R. Ort, "What is the maximum efficiency with which photosynthesis can convert solar energy into biomass?". *Current Opinion in Biotechnology*.

2008, *19*, 153–159.
5. J. Wang, J. Liu, T. Liu, "The difference in effective light penetration may explain the superiority in photosynthetic efficiency of attached cultivation over the conventional open pond for microalgae". *Biotechnology for Biofuels*. **2015**, *8*, 49.
6. K. C. Costa, S. H. Yoon, M. Pan, J. A. Burn, N. S. Baliga, J. A. Leigh, "Effects of H_2 and formate on growth yield and regulation of methanogenesis in *Methanococcus maripaludis*". *Journal of Bacteriology*. **2013**, *195*, 1456–1462.
7. M. Blaut, "Metabolism of methanogens". *Antonie van Leeuwenhoek*. **1994**, *66*, 187–208.
8. E. M. Nichols, J. J. Gallagher, C. Liu, Y. Su, J. Resasco, Y. Sun, P. Yang, M. C. Y. Chang, C. J. Chang, "Hybrid bioinorganic approach to solar-to-chemical conversion". *Proceedings of the National Academy of Sciences of the United States of America*. **2015**, *112*, 11461–11466.
9. D. E. Holmes, J. A. Smith, "Biologically produced methane as a renewable energy source". *Advances in Applied Microbiology*. **2016**, *97*, 1–61.
10. E. R. Marella, C. Holkenbrink, V. Siewers, I. Borodina, "Engineering microbial fatty acid metabolism for biofuels and biochemicals". *Current Opinion in Biotechnology*. **2018**, *50*, 39–46.
11. J. Nielsen, J. D. Keasling, "Engineering cellular metabolism". *Cell*. **2016**, *164*, 1185–1197.
12. M. J. McAnulty, V. G. Poosarla, J. Li, V. W. C. Soo, F. Zhu, T. K. Wood, "Metabolic engineering of *Methanosarcina acetivorans* for lactate production from methane". *Biotechnology and Bioengineering*. **2017**, *114*, 852–861.
13. Z. Lyu, R. Jain, P. Smith, T. Fetchko, Y. Yan, W. B. Whitman, "Engineering the autotroph *Methanococcus maripaludis* for geraniol production". *ACS Synthetic Biology*. **2016**, *5*, 577–581.
14. E. L. Hendrickson, J. A. Leigh, "Roles of coenzyme F_{420}-reducing hydrogenases and hydrogen- and F_{420}-dependent methylenetetrahydromethanopterin dehydrogenases in reduction of F_{420} and production of hydrogen during methanogenesis". *Journal of Bacteriology*. **2008**, *190*, 4818–4821.
15. K. AK, M. J, P. K, T. RK, "Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea.". *Proceedings of the National Academy of Sciences of the United States of America*. **2011**, *108*, 2981–2986.
16. C. Zhao, Q. Zhao, Y. Li, Y. Zhang, "Engineering redox homeostasis to develop efficient alcohol-producing microbial cell factories". *Microbial Cell Factories*. **2017**, *16*, 115.
17. L. Sellés Vidal, C. L. Kelly, P. M. Mordaka, J. T. Heap, "Review of NAD(P)H-dependent oxidoreductases: Properties, engineering and application". *Biochimica et Biophysica Acta - Proteins and Proteomics*. **2018**, *1866*, 327–347.
18. Z. Lyu, W. B. Whitman, "Transplanting the pathway engineering toolbox to methanogens". *Current Opinion in Biotechnology*. **2019**, *59*, 46–54.
19. E. L. Hendrickson, R. Kaul, Y. Zhou, D. Bovee, P. Chapman, J. Chung, E. Conway de Macario, J. A. Dodsworth, W. Gillett, D. E. Graham, M. Hackett, A. K. Haydock, A. Kang, M. L. Land, R. Levy, T. J. Lie, T. A. Major, B. C. Moore, I. Porat, A. Palmeiri, G. Rouse, C. Saenphimmachak, D. Söll, S. Van Dien, T. Wang, W. B. Whitman, Q. Xia, Y. Zhang, F. W. Larimer, M. V. Olson, J. A. Leigh, "Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*". *Journal of Bacteriology*. **2004**, *186*, 6956–6969.
20. R. A. J. Verlinden, D. J. Hill, M. A. Kenward, C. D. Williams, I. Radecka, "Bacterial

synthesis of biodegradable polyhydroxyalkanoates". *Journal of Applied Microbiology*. **2007**, *102*, 1437–1449.
21. R. K. Thauer, A. K. Kaster, H. Seedorf, W. Buckel, R. Hedderich, "Methanogenic archaea: Ecologically relevant differences in energy conservation". *Nature Reviews Microbiology*. **2008**, *6*, 579–591.
22. W. Xiao, R.-S. Wang, D. E. Handy, J. Loscalzo, "NAD(H) and NADP(H) redox couples and cellular energy metabolism", **2018**, *28*, 251-272.
23. M. A. Richards, T. J. Lie, J. Zhang, S. W. Ragsdale, J. A. Leigh, N.D. Price, "Exploring hydrogenotrophic methanogenesis: A genome scale metabolic reconstruction of *Methanococcus maripaludis*". *Journal of Bacteriology*. **2016**, *198*, 3379–3390.
24. D. G. Gibson, L. Young, R. Chuang, J. C. Venter, C. A. Hutchison III, H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases". *Nature Methods*. **2009**, *6*, 343–345.
25. M. E. Brown, A. Mukhopadhyay, J. D. Keasling, "Engineering bacteria to catabolize the carbonaceous component of sarin: Teaching *E. coli* to eat isopropanol". *ACS Synthetic Biology*. **2016**, *5*, 1485–1496.
26. W. W. Metcalf, J. K. Zhang, E. Apolinario, K. R. Sowers, R. S. Wolfe, "A genetic system for Archaea of the genus *Methanosarcina*: Liposome-mediated transformation and construction of shuttle vectors". *Proceedings of the National Academy of Sciences of the United States of America*. **1997**, *94*, 2626–2631.
27. E. Weber, C. Engler, R. Gruetzner, S. Werner, S. Marillonnet, "A modular cloning system for standardized assembly of multigene constructs". *PLoS ONE*. **2011**, *6*, e16765.
28. N. Buan, G. Kulkarni, W. Metcalf, "Genetic methods for *Methanosarcina* species" in *Methods in Enzymology*, (Academic Press Inc., 2011), pp. 23–42.
29. B. B. Bond-Watts, R. J. Bellerose, M. C. Y Chang, "Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways". **2011**, *7*, 222-227.
30. E. L. Hendrickson, A. K. Haydock, B. C. Moore, W. B. Whitman, J. A. Leigh, "Functionally distinct genes regulated by hydrogen limitation and growth rate in methanogenic archaea". *Proceedings of the National Academy of Sciences of the United States of America*. **2007**, *104*, 8930–8934.
31. K. C. Costa, T. J. Lie, M. A. Jacobs, J. A. Leigh, "H_2-independent growth of the hydrogenotrophic methanogen *Methanococcus maripaludis*". *mBio*. **2013**, *4*, e00062-13.
32. T. J. Lie, K. C. Costa, B. Lupa, S. Korpole, W. B. Whitman, J. A. Leigh, "Essential anaplerotic role for the energy-converting hydrogenase Eha in hydrogenotrophic methanogenesis". *Proceedings of the National Academy of Sciences of the United States of America*. **2012**, *109*, 15473–15478.
33. A. K. Holm, L. M. Blank, M. Oldiges, A. Schmid, C. Solem, P. R. Jensen, G. N. Vemuri, "Metabolic and transcriptional response to cofactor perturbations in *Escherichia coli*". *Journal of Biological Chemistry*. **2010**, *285*, 17498–17506.
34. B. C. Moore, J. A. Leigh, "Markerless mutagenesis in *Methanococcus maripaludis* demonstrates roles for alanine dehydrogenase, alanine racemase, and alanine permease". *Journal of Bacteriology*. **2005**, *187*, 972–979.
35. P. R. Martin, R. J. Shea, M. H. Mulks, "Identification of a plasmid-encoded gene from *Haemophilus ducreyi* which confers NAD^+ independence". *Journal of Bacteriology*. **2001**, *183*, 1168–1174.
36. C. Ling, G. Qiao, B. Shuai, K. Olavarria, J. Yin, R. Xiang, K. Song, Y. Shen, Y. Guo, G.

Chen., "Engineering NADH/NAD^+ ratio in *Halomonas bluephagenesis* for enhanced production of polyhydroxyalkanoates (PHA)". *Metabolic Engineering*. **2018**, *49*, 275–286.
37. J. Wodzinska, K. D. Snell, A. Rhomberg, A. J. Sinskey, K. Biemann, J. Stubbe, "Polyhydroxybutyrate synthase: Evidence for covalent catalysis". *Journal of the American Chemical Society*. **1996**, *118*, 6319–6320.
38. Q. Guo, L. Gakhar, K. Wickersham, K. Francis, A. Vardi-Kilshtain, D. T. Major, C. M. Cheatum, A. Kohen, "Structural and kinetic studies of formate dehydrogenase from *Candida boidinii*". *Biochemistry*. **2016**, *55*, 2760–2771.
39. S. J. Berríos-Rivera, G. N. Bennett, K. Y. San, "Metabolic engineering of *Escherichia coli*: Increase of NADH availability by overexpressing an NAD^+-dependent formate dehydrogenase". *Metabolic Engineering*. **2002**, *4*, 217–229.
40. G. E. Wood, A. K. Haydock, J. A. Leigh, "Function and regulation of the formate dehydrogenase genes of the methanogenic archaeon *Methanococcus maripaludis*". *Journal of Bacteriology*. **2003**, *185*, 2548–2554.
41. Z. Lyu, R. Jain, P. Smith, T. Fetchko, Y. Yan, W. B. Whitman, "Engineering the Autotroph *Methanococcus maripaludis* for geraniol production". *ACS Synthetic Biology*. **2016**, *5*, 577–581.
42. S. Flüchter, S. Follonier, B. Schiel-Bengelsdorf, F. R. Bengelsdorf, M. Zinn, P. Dürre, "Anaerobic production of poly(3-hydroxybutyrate) and its precursor 3-hydroxybutyrate from synthesis gas by autotrophic *Clostridia*". *Biomacromolecules*. **2019**, *20*, 3271–3282.
43. R. de Souza Pinto Lemgruber, K. Valgepea, R. Tappel, J. B. Behrendorff, R. W. Palfreyman, M. Plan, M. P. Hodson, S. D. Simpson, L. K. Nielsen, M. Köpke, E. Marcellin, "Systems-level engineering and characterisation of *Clostridium autoethanogenum* through heterologous production of poly-3-hydroxybutyrate (PHB)". *Metabolic Engineering*. **2019**, *53*, 14–23.
44. D. D. Nayak, W. W. Metcalf, "Cas9-mediated genome editing in the methanogenic archaeon *Methanosarcina acetivorans*". *Proceedings of the National Academy of Sciences of the United States of America*. **2017**, *114*, 2976–2981.
45. B. Lupa, E. L. Hendrickson, J. A. Leigh, W. B. Whitman, "Formate-dependent H_2 production by the mesophilic methanogen *Methanococcus maripaludis*". *Applied and Environmental Microbiology*. **2008**, *74*, 6584–6590.
46. "A Flexible System for Cultivation of *Methanococcus* and Other Formate-Utilizing Methanogens". *Archaea* **2017**, *2017*, 1-12.
47. A. Pauss, G. Andre, M. Perrier, S. R. Guiot, "Liquid-to-gas mass transfer in anaerobic processes: Inevitable transfer limitations of methane and hydrogen in the biomethanation process". *Applied and Environmental Microbiology*. **1990**, *56*, 1636–1644.
48. D. Kopljar, A. Inan, P. Vindayer, N. Wagner, E. Klemm, "Electrochemical reduction of CO 2 to formate at high current density using gas diffusion electrodes". **2014** *44,* 1107-1116.
49. S. Kawai, K. Murata, "Structure and function of NAD kinase and NADP phosphatase: Key enzymes that regulate the intracellular balance of NAD(H) and NADP(H). **2014**, *72,* 919.
50. S. Wang, H. Huang, J. Moll, R. K. Thauer, "$NADP^+$ reduction with reduced ferredoxin and $NADP^+$ reduction with NADH are coupled via an electron-bifurcating enzyme complex in *Clostridium kluyveri*". *Journal of Bacteriology*. **2010**, *192*, 5115–5123.

Chapter 4: *Investigating translation in Methanococcus maripaludis for improved heterologous gene expression*

4.1. Introduction

In all three domains of life, mRNA translation is an essential biological process with unique, domain-specific traits. This vital process is also important to the methanogen engineering efforts described in Chapter 3. In that chapter, an empirical approach was taken to design biosynthetic pathways for expression in *Methanococcus maripaludis*. As a result, we found that most of our promoters were capable of driving gene transcription, but the resulting transcripts were often not translated to observable protein products. Gaining a better insight into the translation of native transcripts could provide us a better insight into key design principles relevant to future metabolic engineering efforts in methanogenic Archaea.

Studies of archaeal translation have shown the presence of transcripts with 5′ UTR leader sequences of similar length to bacterial counterparts [*1-3*]. Similar to Bacteria, leadered transcripts in Archaea appear to use a Shine-Dalgarno (SD)-anti-SD duplex to facilitate the recruitment of the small 30S ribosomal subunit, followed by the formation of a ternary initiation complex with archaeal IF2 (aIF2):GTP:$tRNA^{fMet}$, aIF1 and aIF1A [*4*]. These three initiation factors also contribute to accurate selection of the start codon, similar to the role of bacterial IF3 [*5-9*]. To investigate the 5′ UTR of native mRNA transcripts in *M. maripaludis,* we generated an RNA-seq dataset and used *in silico* transcriptome assembly to identify both monocistronic and polycistronic transcripts with and without 5′ UTR sequences. Candidate transcript termini were confirmed using 3′ UTR sequencing. Transcripts were categorized to focus only on candidate transcripts with 5′ UTRs. Using this dataset, we identified a putative consensus SD sequence, which can be used to design pathways in the future. Furthermore, this process allows us to select specific native transcripts for further biochemical and genetic investigation into translation initiation (*Table 4.1*, *Table 4.2*).

Evidence for bacterial, eukaryal and archaeal transcripts with 5′ UTR of five nucleotides or less (so called “leaderless” transcripts) have also been found [*10*,*1*]. Leaderless transcripts do not contain an upstream SD sequence and it is unclear how translation initiation occurs with these transcripts. In addition, the relative abundance of leaderless transcripts varies greatly across Archaeal species [*11-13*] and its role in *M. maripaludis* is currently unknown. Multiple models of translation initiation for leaderless sequences have been proposed, but no mechanistic models have thus far been shown for *M. maripaludis*. A key distinction between these models is the precise positions at which the ribosome interacts with mRNA during translation initiation. One powerful method to gain a better understanding of these positions is through ribosome profiling. Ribosome profiling is an innovative technique that uses NGS of ribosome-protected mRNA fragments (“ribosome footprints”), allowing for the monitoring of *in vivo* translation [*14-16*]. These footprints can be generated by using nuclease activity to degrade RNA outside of the mRNA that is sequestered by ribosome [*17*, *18*]. Mapping these footprints back to the original mRNA can be used to identify the exact location of the translating ribosome, providing a precise record of the ribosome position when translation was halted. In addition, the fraction of protected footprints compared to the abundance of the corresponding transcript can also be used to measure translational efficiency, which can provide further insights into the principles for heterologous pathway design.

The 3’ UTR dataset was also further analyzed to identify common terminator sequences and the possible role of mRNA stability in determining heterologous gene expression outcomes. In this work, key determinants of gene translation were analyzed and new principles for gene expression in *M. maripaludis* are proposed.

4.2. Materials and Methods

Commercial materials. Ultra-high purity gases purchased from Praxair (Danbury, CT) were used for all anaerobic manipulations. Distilled water (dH_2O) was deionized to a resistivity of 18.2 MΩ·cm using a Millipore Milli-Q UF Plus system (ddH_2O). Glycerol, sodium bicarbonate, Tris base, kanamycin sulfate, sodium bicarbonate, sodium chloride, magnesium sulfate heptahydrate, magnesium chloride hexahydrate, ammonium chloride, potassium phosphate monobasic, potassium phosphate dibasic, sodium phosphate monobasic, sodium phosphate dibasic, potassium chloride, calcium chloride, manganese (II) chloride tetrahydrate, and pressure release aluminum crimp seals (20 mm) were purchased from Fisher Scientific (Pittsburgh, PA). Calcium chloride dihydrate, ammonium iron(II) sulfate hexahydrate, sodium selenite, nickel(II) chloride hexahydrate, boric acid, cobalt(II) chloride hexahydrate, copper(II) chloride dihydrate, sodium molybdate dihydrate, sodium sulfide, d-Desthiobiotin, L-cysteine, biotin, folic acid, pyridoxine hydrochloride, thiamine hydrochloride, riboflavin, nicotinic acid, calcium D-(+)-pantothenate, vitamin B12, p-aminobenzoic acid, PEG 3350, formic acid reagent grade, nitrilotriacetic acid, manganese(II) sulfate dihydrate, sodium tungstate dihydrate, dimethyl sulfoxide (DMSO), tris(2-carboxyethyl)phosphine (TCEP), cOmplete protease inhibitor cocktail tablets, acetyl coenzyme A sodium salt, 3'-dephosphocoenzyme A, sodium citrate dihydrate, nitrilotriacetic acid disodium salt, cyanocobalamin, MOPS and thioctic acid were purchased from Sigma Aldrich (St. Louis, MO). Zinc sulfate heptahydrate was purchased from Mallinckrodt Chemicals (St. Louis, MO). Resazurin was purchased from Eastman Kodak Company (Rochester, NY). Balch tubes (18 x 150 mm), anaerobic glass bottles (250 mL and 2L) and butyl rubber stoppers (20 mm) were purchased from Chemglass (Vineland, NJ).

Methanogen strain details and culture conditions. *Methanococcus maripaludis S2* (ATCC BAA-2049) was purchased from the American Type Culture Collection (Manassas, VA). All culture manipulations were performed inside a Vacuum Atmospheres Nexus One glovebox with an atmosphere of 90 N_2:10 H_2. Oxygen levels and humidity levels within the box were controlled using a STAK-PAK palladium catalyst and desiccant system in a Coy Laboratory Products unheated fan box. *M. maripaludis* cultures were propagated in 18 × 150 mm Balch tubes with butyl rubber stoppers using defined medium (McNA) or complex, undefined medium (McCas). In initial production conditions, 10 mL cultures were grown at 37°C without shaking. Stock cultures were stored at room temperature in the dark after growth and were propagated by diluting 1:100 into fresh McCas or McNA medium at least every 2 weeks. The stock cultures were used until changes in growth patterns were observed. Glycerol stocks were kept at -80°C for long-term culture storage.

Methanogen growth media preparation. To prepare growth media for *M. maripaludis*, all components listed in the table below except the reducing agent were dissolved in the appropriate volume of MilliQ-treated water in an Airfree® Cajon storage flask (Chemglass, Vineland, NJ). The solution was made anaerobic using three freeze-pump-thaw cycles on a Schlenk line and transferred to serum bottles or tubes in an anaerobic chamber. The bottles or tubes were sealed with butyl rubber stoppers and aluminum crimps. Then, media was autoclaved and allowed to cool before reducing agent was added. If chloramphenicol or cycloheximide were used, they would be added to the cooled media at a final concentration of between 1-500 µg/mL. A similar procedure was followed for the preparation of McCas agar bottle plates, except 2% (*w/v*) agar was added to 50 mL of media after it has been transferred to bottles in the anaerobic chamber.

A stock of 100x reducing agent was prepared by adding a 0.01% (*w/v*) resazurin solution (200 µL) to MilliQ-treated water (200 mL) and sparging the solution with N_2 for 20 minutes. Under continuous N_2 flow, L-cysteine (6 g) was added, followed by sodium sulfide nonahydrate (6 g). Sparging with N_2 was continued until the sodium sulfide had dissolved completely and the solution was completely colorless. After preparation, the reducing agent was stored under an N_2 atmosphere in a 250 mL serum bottle and sealed with butyl rubber stoppers and aluminum crimps.

Per L, the modified McCas medium contains:

ddH_2O	500 mL
N-free general salts solution	500 mL
$NaHCO_3$	5 g
NaCl	10.5 g
K_2HPO_4	140 mg
$FeSO_4 \cdot 7H_2O$	10 mg
N-free trace minerals solution (100X)	10 mL
Wolfe's vitamin solution	10 mL
Resazurin solution (0.01% (*w/v*))	1 mL
Sodium acetate	850 mg
MOPS (2 M, pH 7)	100 mL
Magnesium formate dihydrate	15 g
NH_4Cl	500 mg
Casamino acids	2 g

Per L, the defined McNA medium contains:

ddH_2O	500 mL
N-free general salts solution	500 mL
$NaHCO_3$	5 g
NaCl	10.5 g
K_2HPO4	140 mg
$FeSO_4 \cdot 7H_2O$	10 mg
N-free trace minerals solution (100X)	10 mL
Resazurin solution (0.01% (*w/v*))	1 mL
Sodium acetate	850 mg
MOPS (2 M, pH 7)	100 mL
Magnesium formate dihydrate	15 g
NH_4Cl	500 mg

Per L, the N-free general salts solution contains:

KCl	670 mg
$MgCl_2 \cdot 6H_2O$	5.5 g
$MgSO_4 \cdot 7H_2O$	6.9 g
$CaCl_2 \cdot 2H_2O$	280 mg
ddH_2O	1000 mL

To prepare the N-free trace minerals solution, trisodium citrate was dissolved in water and the pH was adjusted to 6.5. Then the remaining salts (see below) were added. Per L, the 100x N-free trace minerals solution contains:

Trisodium citrate dihydrate	2.1 g
$MnSO_4 \cdot H_2O$	500 mg
$CoCl_2 \cdot 6H_2O$	100 mg
$ZnSO_4 \cdot 7H_2O$	100 mg
$CuCl_2 \cdot 2H_2O$	100 mg
$FeSO_4 \cdot 7H_2O$	6.8 mg
$AlK(SO_4)_2 \cdot 12H_2O$	10 mg
H_3BO_3	10 mg
$Na_2MoO_4 \cdot 2H_2O$	100 mg
$NiCl_2 \cdot 6H_2O$	25 mg
Na_2SeO_3	200 mg
VCl_3	10 mg
$Na_2WO_4 \cdot 2H_2O$	3.3 mg

RNA seq analysis. Cells were grown as described previously in 10 mL of McCas medium. All measurements were collected in biological triplicates. When the OD_{600} reached 0.5-0.6 (after approximately 18 hours), the cultures were rapidly cooled in an ice-water bath. The cells were collected via centrifugation at 8,000 × *g* for 30 min at 4°C and rapidly frozen in liquid nitrogen before storage at −80°C. The cells were then resuspended in 200 µL of their respective growth medium and lysed using 250 µL of RLT buffer (Qiagen, Valencia, CA). 100% ethanol (350 µL) was added, and the RNA was purified on RNeasy columns (Qiagen). Aliquots of RNA were treated on-column with RNAse-free DNase (Norgen, Thorold, ON, Canada) according to manufacturer's protocol. Quality of total RNA samples was verified using an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Palo Alto, CA). rRNA was removed using a Ribo-Zero rRNA removal kit for bacteria (Illumina, San Diego, CA) according to the manufacturer protocol and a modified RNeasy MinElute Cleanup Kit (Qiagen) was used to recover mRNA. Sample quality was verified using an Agilent 2100 Bioanalyzer (Agilent). Library preparation was performed by the Genome Center at the University of California, Davis, using the TruSeq RNA library prep kit (Illumina) according

to the manufacturer's protocols. cDNA was sequenced using Illumina HiSeq 4000 SE50 at the UC Davis Genomic Center. Transcriptome was assembled using the *M. maripaludis* S2 NCBI reference genome (accession ID: NC_005791), RNA-seq data from three biological replicates, and Rockhopper software [*19*].

3′-terminus RNA sequencing. Cells were grown as described previously in 10 mL of McCas medium. All measurements were collected in biological triplicates. When the OD_{600} reached 0.5-0.6 (after approximately 18 hours), the cultures were rapidly cooled in an ice-water bath. The cells were collected via centrifugation at 8,000 × *g* for 30 min at 4°C and rapidly frozen in liquid nitrogen before storage at −80°C. The cells were then resuspended in 200 µL of their respective growth medium and lysed using 250 µL of RLT buffer (Qiagen, Valencia, CA). 100% ethanol (350 µL) was added, and the RNA was purified on RNeasy columns (Qiagen). Aliquots of RNA were treated on-column with RNAse-free DNase (Norgen, Thorold, ON, Canada) according to manufacturer's protocol. Quality of total RNA samples was verified using an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Palo Alto, CA). rRNA was removed using a Ribo-Zero rRNA removal kit for bacteria (Illumina, San Diego, CA) according to the manufacturer protocol and a modified RNeasy MinElute Cleanup Kit (Qiagen) was used to recover mRNA. Sample quality was verified using an Agilent 2100 Bioanalyzer (Agilent).

Library preparation was continued as described previously [*20*]. Briefly, RNA was polyadenylated with poly(A) polymerase (New England Biolabs, Ipswich, MA). RNA was polyadenylated in 1x PAP buffer containing 10 units recombinant RNase inhibitor (Takara Bio, Mountain View, CA) with 0.1 mM ATP at 37°C for 10 min. Reaction was stopped by incubating at 65°C for 20 min. Then, a cDNA strand synthesis is performed in the presence of a poly(dT) oligonucleotide containing a custom 3′-adaptor sequence and the Moloney murine leukemia virus reverse transcriptase (MMLV-RT). When the reverse trancriptase reaches the 5′ terminus of the RNA template, the enzyme adds additional nucleotides (predominantly dC) that are not encoded by the template. Template switching oligonucleotide (TSO) containing three 3′-terminal rG nucleotides and a custom 5′-adaptor sequence are added to the RT reaction product. The rG nucleotides at the 3′-end of the TSO and the dC-rich extended sequence are expected to promote template switching by MMLV-RT. Reverse and forward primers are used for PCR amplification of the cDNA and standard Illumina library preparation is performed or the cDNA is analyzed by gel electrophoresis.

The following primers were used:

Name	Sequence
Illumina poly(dT)	AGACGTGTGCTCTTCCGATCT(T)x30
Template switching oligonucleotide (TSO)	GTTCAGAGTTCTACAGTCCGACGATCrGrGrG
cDNA f	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
cDNA r	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA

Ribosome profiling. Single colonies of *M. maripaludis* were picked from anaerobic agar plates and grown overnight at 37°C with shaking at 200 rpm in 500 mL McCas medium until late log phase (OD_{600} of 0.6 – 0.8). Cells were harvested by centrifugation, rapidly resuspended in lysis buffer (1% (*v/v*) Triton X-100 and 25 U/mL Turbo DNase I in polysome buffer) and dripped into in liquid nitrogen contained in a 50 mL conical tube with several holes punched into the cap.

Chilled metal spatulas were used to scrape and transfer cell pellets. The tube was capped and placed upright in a -80°C freezer until the liquid nitrogen evaporates.

The frozen pellets were added to the grinding jar, a grinding ball was inserted and sealed tightly. The assembled chamber was kept in liquid nitrogen. *M. maripaludis* was ground for 10 s for 10 cycles with 1 min of rest in between. An open conical tube was filled with liquid nitrogen and placed upright in a liquid nitrogen bath. The grinding chamber was opened and the recovered powder was transferred to liquid nitrogen using the chilled metal spatulas. The tube was capped and again placed in a -80°C freezer until the liquid nitrogen evaporates.

M. maripaludis lysate was centrifuged at 3000 × *g* at 4°C for 5 min and supernatant was recovered into multiple non-stick RNase-free microcentrifuge tubes. The supernatant was further clarified by spinning at 20,000 × *g* at 4°C for 10 min. The supernatant was recovered and 50 µL aliquots were made and flash frozen in liquid nitrogen. 30 to100 µg of total RNA was diluted to 200 µL with polysome buffer (20 mM Tris, pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT with either 1 or 100 µg/mL cycloheximide or 5, 50 and 500 µg/mL chloramphenicol). 1.5 µL RNase I or MNase (10 U/µL) were added and the reaction was incubated for 45 min at room temperature with gentle agitation. 10 µL SUPERase*In RNase inhibitor to pause nuclease digestion and the digestion was transferred to ice. The remainder of the protocol was completed as described previously [*21*]. Data analysis was performed using Plastid and Cutadapt [*22*]. Offset was determined by setting the distance between the 5′ end and the start codon peak, which was approximately 12-14 nucleotides, depending on the fragment length.

Polysome profiling. The ribosome profiling described above was used to generate polysomes, which were either digested with 1.5 µL of RNase I (10 U/µL) or 1.5 µL ddH_2O during a 45 min incubation at room temperature with gentle agitation. 10 µL SUPERase*In RNase inhibitor was added to pause nuclease digestion and the digestion was transferred to ice.

10% and 50% (*w/v*) sucrose solutions were prepared and 1.0 µL SUPERase*In was added per 1.0 mL gradient buffer. The sucrose solutions were nutated at 4°C for 1 h. The sucrose solutions were spun down to remove air bubbles. 6 mL of 10% sucrose solution was transferred to 14 mm × 89 mm ultracentrifuge tubes and 8 mL of 50% (*w/v*) sucrose solution was transferred to the bottom of the labelled tube by a cannula. The tube was capped with a short cap and gradients were prepared using a BioComp GradientMaster 108. The tube was removed and 200 µL of the polysome reaction mixture was transferred to the top of the sucrose mixture. Tubes were balanced and spun in an ultracentrifuge at 4°C for 3 h at 36,000 rpm using a Beckman SW41 rotor and Beckman L8-M ultracentrifuge. Polysome profiling was performed using the BioComp GradientMaster 108 and UV was monitored using Logger Lite. The plunger was washed with MilliQ water between samples.

3′ UTR terminus analysis. An algorithm was developed to map WIG file data on to the *M. maripaludis* S2 reference genome. By denoting the last nucleotide in the terminator with reads observed in at least two biological replicates, we identified the 3′ terminus. This allowed us to also calculate the distance of the terminator from the closest start and stop codons, helping to determine terminator lengths and intragenic terminators. Nucleotide frequency analysis was also conducted up to -11 bp of the 3′ terminus.

4.3. Results and discussion

Ribosome profiling development. Ribosome profiling is a powerful technique for identifying ribosome position on mRNA transcripts using NGS analysis of protected ribosome footprints [*23*]. Since translation initiation is usually the rate-limiting step of the translation process, gaining a better understanding of the ribosome position could shed light on the mRNA-ribosome interactions during translational initiation. Although ribosome profiles have been obtained for eukaryotic and bacterial species, profiles for *M. maripaludis* have not yet been published and only one archaeal profile has been published so far [*23-25*]. To conduct a ribosome profile with an archaeal, specific modifications to the protocol were required.

In the first step, translation can be inhibited by rapidly freezing ribosomes during translation. In this step, it is also possible to use antibiotics to assist with the arrest of ribosomes during translation. Several translation inhibitors have been used to halt ribosomes during translation initiation [*26-30*]. Of these, cycloheximide is often the antibiotic of choice for eukaryotic translation arrest before ribosome profiling [*31*, *32*], In bacterial species such as *E. coli*, chloramphenicol has been used [*33*]. In both cases, however, doubts have been raised about the fidelity of ribosome positions after antibiotic treatment [*34-36*]. Despite the challenges, these inhibitors can be useful for the initial development of a ribosome profiling method because it can improve the intact monosome fraction used to generate footprint libraries. Therefore, antibiotics were used for the remainder of our method development.

In our growth experiments, we attempted to determine which antibiotic would be best for ribosome arrest. Since Archaeal translation shows similarities to both eukaryotic and bacterial translation methods, we screened both chloramphenicol and cycloheximide. In our growth experiments, *M. maripaludis* showed defects in growth with chloramphenicol concentrations of 5 μg/mL or greater, while cycloheximide concentrations at even 100 μg/mL had no effect on growth (*Figure 4.1*). Therefore, we used chloramphenicol in the lysis buffer during cell lysis.

After frozen lysates were pulverized, allowed to thaw and centrifuged, nuclease digestion was performed on the supernatant. RNase I and micrococcal nuclease (MNase) were used for this step. RNase I was used because of its common use for RNA digest in yeast and other eukaryotes [*37*]. However, RNase I has also been shown to digest rRNA in bacterial ribosomes, causing the ribosome complex to dissociate [*38*]. Because of this, MNase is commonly used for bacterial ribosome profiling and was also used in our nuclease screening experiments [*39*]. According to our nuclease digest gel electrophoresis results, it appeared that RNase I was effective for generating the expected gel band in the range of 17-34 nucleotides at the concentration of 100 μg/mL (*Figure 4.1*). To confirm that polysomes were appropriately digested into monosomes using this technique, we also used polysome profiling of clarified lysates before or after nuclease digestion. RNA gel electrophoresis showed that the large peak at 2.5 min and the absorbance after 3.5 min contained both 23S and 16S rRNA, suggesting that these fractions could be associated with monosomes and polysomes, respectively. After digestion with RNase I, our profile of the reaction mixture indicated a loss of absorbance after 3.5 min. Furthermore, the peak at 2.5 min appeared larger, suggesting that polysome digestion may have occurred. Using these results, we proceeded with the ribosome profiling methodology as described in the literature, allowing us to generate our first ribosome profile of *M. maripaludis.*

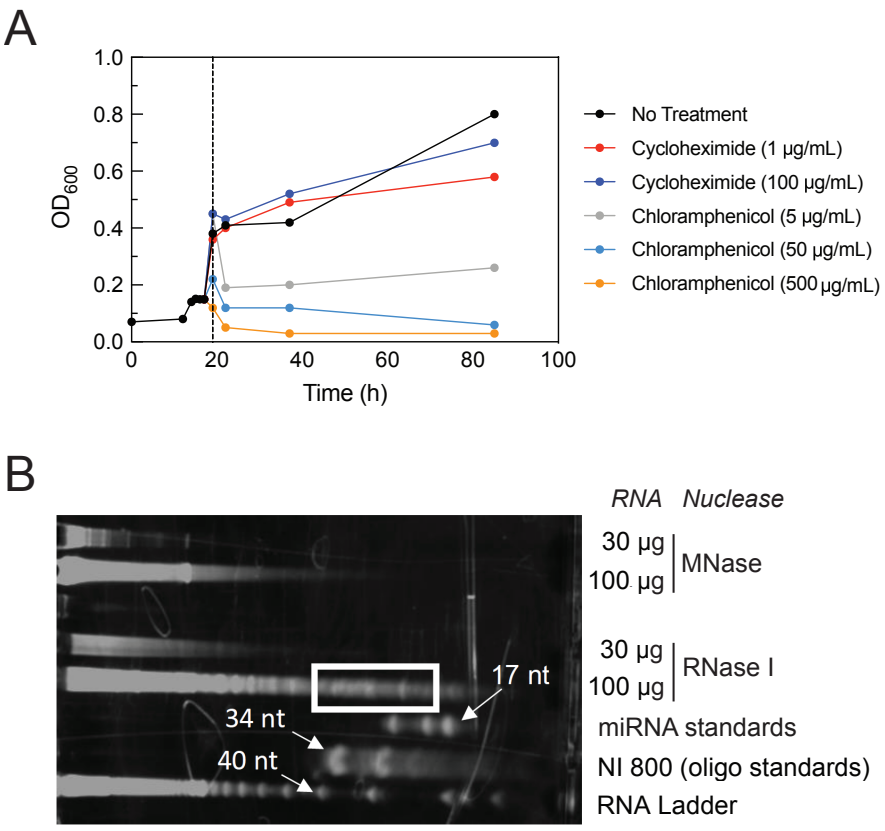


Figure 4.1 Ribosome profiling method development for M. maripaludis. *Antibiotics and nucleases were screened to determine their efficacy for ribosome profiling in M. maripaludis. (A) Growth of M. maripaludis in McCas media was measured when chloramphenicol and cycloheximide, known translation inhibitors, were added to the media at 18 h post inoculation (highlighted with dashed line). Growth was not affected by the addition of 1 to 100 µg/mL cycloheximide, but chloramphenicol had a dose-dependent inhibitive effect on growth. Data are from single samples of each condition (n = 1). (B) Micrococcal ribonuclease (MNase) and RNase I (10 U/µL) were screened for digestion of polysomes in clarified cell extract with either 100 or 30 µg of RNA. RNase I treatment of 100 µg/mL of RNA showed the typical nuclease digest pattern, as compared to the miRNA standards and NI-800 oligonucleotide standards. The expected ribosome footprints were between 17 to 34 nucleotides [21, 40].*

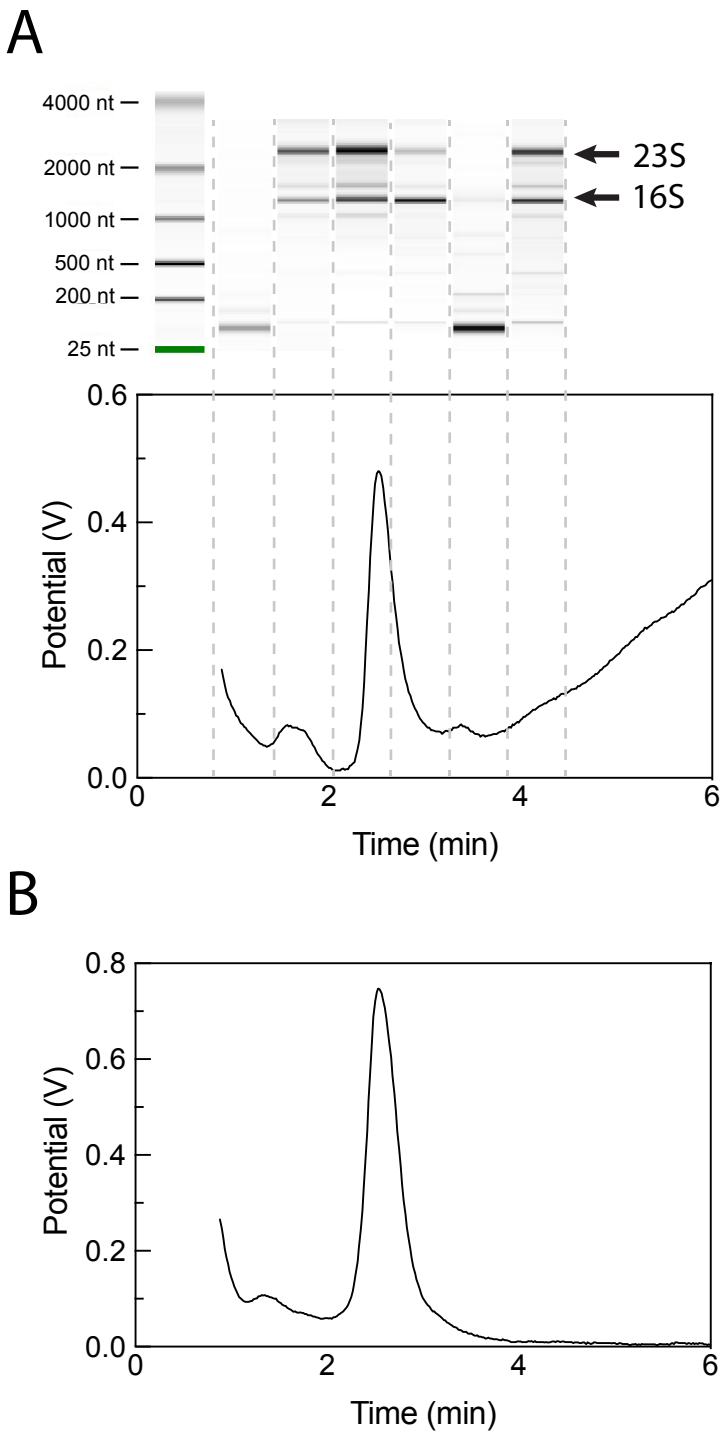


Figure 4.2 Polysome profiling of ribosomes from M. maripaludis suggests RNase I is active. *Clarified cell lysate used for polysome extraction showed evidence of RNase digestion. (A) Six fractions were collected from the ultracentrifugation and RNA was extracted for electrophoretic analysis. 16S and 23S rRNA bands were observed in lanes 2-4 and 6, while only a faint 16S band was seen in lane 2. Absorbance profile show likely 50S ribosome peak (with 23S rRNA) at 1.3 min, a monosome peak at 2.5 min and an increasing background signal, which is also associated with both 23S and 16S rRNA bands. (B) RNase I was added to clarified cell lysate to digest polysomes into individual monosomes. A reduction of absorbance after 3.3 min and an increase in the monosome peaks at 2.5 min suggest that the polysomes were digested, leading to an increase in monosome fraction.*

Analyzing the ribosome profile of *M. maripaludis*. Using computational tools, raw sequencing reads from ribosome footprints were analyzed to measure the quality of the profiling data and to identify the fraction of total reads that map to mRNA. Some important quality metrics for a ribosome profile are ribosome fragment length distributions and codon periodicity in the P-site of the ribosome (Figure 4.3). In yeast, ribosome footprint fragment length distributions typically have a mean value of 28 nt, although a set of smaller size footprints have also been identified [36]. In comparison, the fragment length distribution of *M. maripaludis* appeared to have a mean value of 31 nt. There could be many factors that explain this larger mean fragment size, including a larger pocket that is protected by the archaeal ribosome or incomplete RNase digestion. Increasing the digestion time may allow us to better elucidate the cause of the skewed footprint lengths. The footprints are reasonably similar in size, providing some evidence towards the quality of the ribosome profile.

Another metric for ribosome profile quality is the codon periodicity at the P-site, which is analyzed using the ribosome metagene position during translation. The position of the P-site on the transcript is estimated using a P-site offset. By analyzing codon periodicity behavior, one can determine whether the profile resembles typical ribosome behavior observed in profiles seen in other species [16]. Codon periodicity is ascribed to the ratcheting movement of ribosomes from codon to codon, with the 5′ end of the transcript typically in register at the first nucleotide of every translated codon (after applying an offset to the footprint reads) [41]. The metagene position of the ribosome for all translated genes is determined by mapping the 5′ terminus of the ribosome footprint to its gene position. Using this technique, our analysis showed that the codon periodicity of our preliminary dataset resembled that of typical profiles (Figure 4.3), since the majority of reads had 5′ ends that mapped to the first base in each codon across most of the footprint fragment lengths.

Our attention then turned to an analysis of rRNA fragments seen in our library, since these fragments represent greater than 90% of our reads, limiting our ability to obtain good depth with the sequencing of our mRNA footprints. This large rRNA fraction is within a range that may suggestspossible rRNA degradation by RNase I. To reduce the rRNA fraction, depletion has been a commonly recommended method(cite). In our initial dataset, 100 unique rRNA fragments represented over 50% of all rRNA-associated reads (Figure 4.4). Specific anti-sense oligonucleotides can be made to the ten most abundant rRNA molecules to remove approximately 30% of all reads. Selective subtraction of ribosomal fragments is highly effective, and often useful for removing rRNA that are overly abundant [42, 43]. To conduct an initial analysis of read density, rRNA sequences were removed from our dataset and mRNA footprints were mapped back to their respective coding gene by mapping the 5′ terminus or 3’ terminus each read to the corresponding nucleotides in the transcript. By dividing these transcript values by the total number of footprints, one arrives at a footprint density for every position of the transcript, where 0 is the start codon and negative values represent nucleotides upstream of the start codon. After mapping each footprint in this way, an offset was applied. Since the exact distance is not known, offsets were determined by the distance between the 5′ terminus and the start codon, with a shift of 12-14 nucleotides made for reads of each fragment size, similar to other published ribosome profiles [44]. The resulting read density plot can provide insights into the approximate position of the P-site and other protected elements in the ribosome profile.

We observed that the majority of footprint density clustered within a range of 100 nucleotides from position 0 (Figure 4.5). The upstream sequence was not represented in our dataset, which

may be due to insufficient sampling. Another reason could be the abundance of leaderless transcripts in this species. Based on our current model of translation initiation using leaderless transcripts, the 70S ribosome or 30S ribosome:$tRNA^{fMet}$ complex binds to the start codon through base pairing between the start codon and the $tRNA^{fMet}$ anticodon. The footprint distribution predicted by this model fits with our ribosome density plot, although further refinement of our ribosome profiling method, greater sequencing depth and grouping of known, leaderless transcripts from the *M. maripaludis* transcriptome will be required to appropriately determine whether the patterns observed in this preliminary density plot are artefacts or not. In the next iteration, rRNA depletion and greater sequencing depth will be used with and without antibiotics.

In 2020, an article that presented the first archaeal ribosome profile of Haloferax volcanii [43]. In this paper, ribosome profiling was performed with and without translation inhibitors, including cycloheximide and thiostrepton. By identifying different transcript lengths and correlating these lengths to putative leaderless transcripts, the paper also inferred that the 5′ terminus of leaderless transcripts is bound by the P-site. Although this is circumstantial evidence, it lends support to the notion that binding of the start codon to the $tRNA^{fMet}$ in complex with either the 30S subunit or 70S complex may be necessary and/or sufficient for translation initiation of these transcripts.

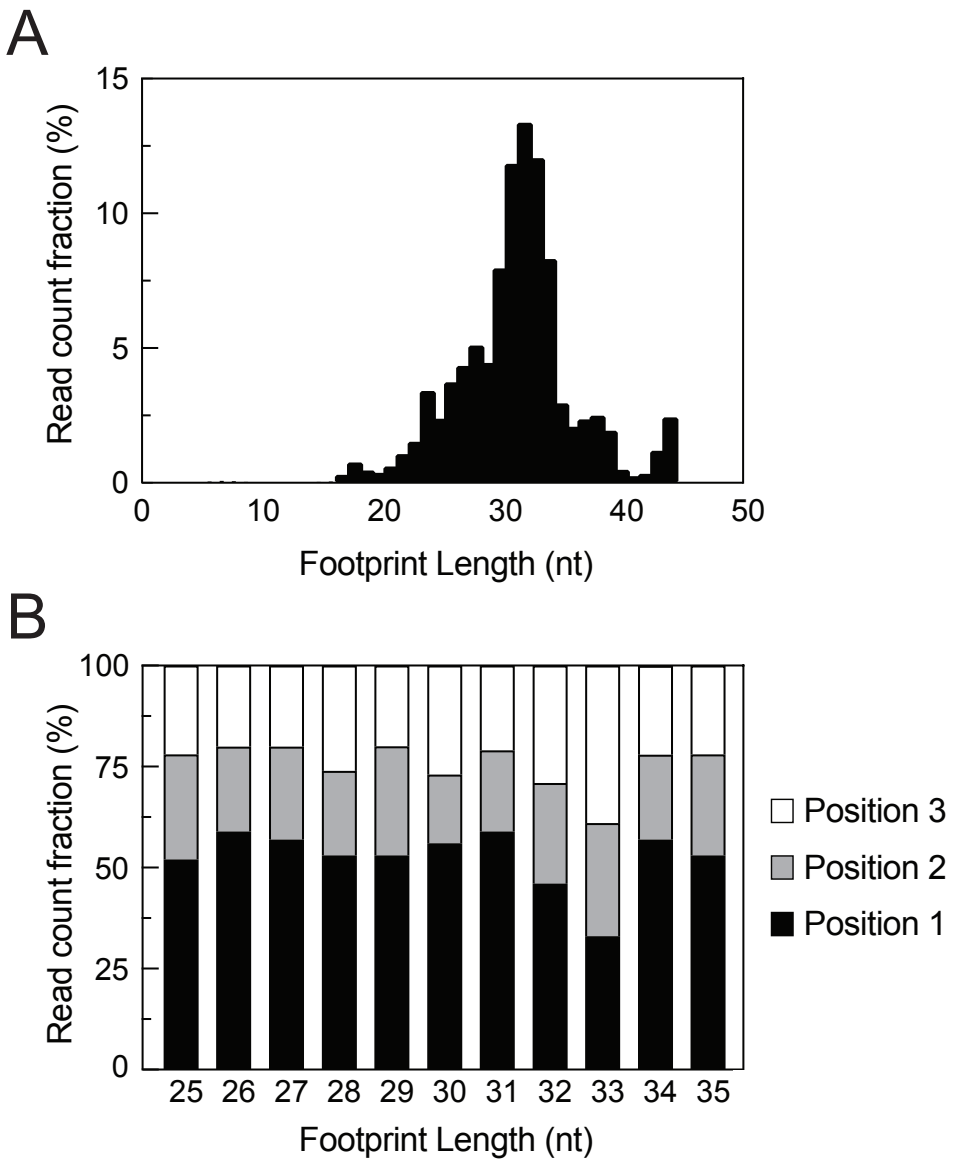


Figure 4.3 Preliminary ribosome profile quality control measurements. *Our ribosome footprint library was sequenced using NGS and its profile quality was analyzed. Reads were mapped back to genes and derived footprint lengths were measured. (A) Footprint read size distribution has a mean value of 31, which is larger than the typical nucleotide length in S. cerevisiae, possibly due to a larger ribosome-protected region of mRNA [21]. (B) After applying a P-site offset of 10-12 nucleotides, a meta-gene analysis of fragments of all lengths was conducted by mapping the 5' terminus of reads to genes. The nucleotide at that position was classified based on the position of the 5' terminus and the position of that nucleotide in a codon. This data illustrates the ribosome ratcheting effect seen in typical ribosome profiles in Bacteria and Eukarya [41].*

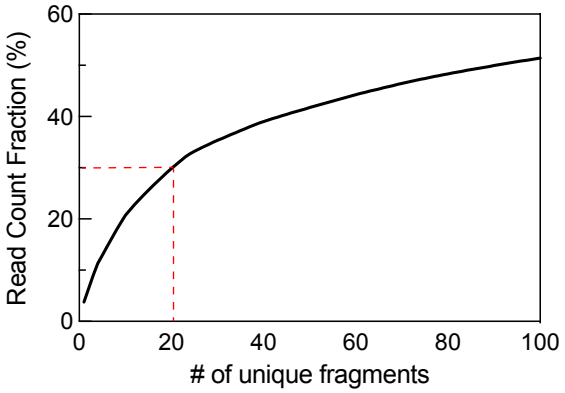


Figure 4.4 rRNA contaminants represent the majority of unique ribosome footprints. *Unique ribosome-protected footprints were counted and unique rRNA fragments were plotted based on the % of the rRNA read count fraction that they represent. Over 100 rRNA fragments were found, but the top ten fragments represented over 30% of all contaminants, suggesting that the targeted depletion of these sequences in particular could meaningfully reduce contamination.*

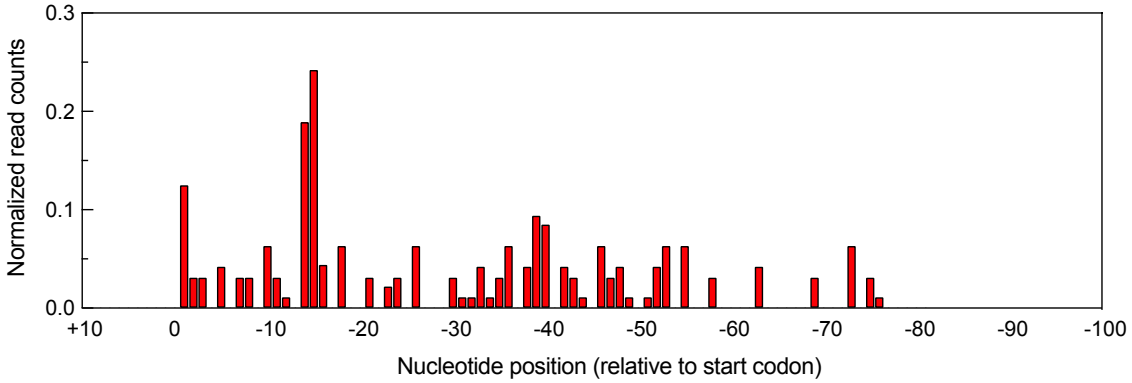


Figure 4.5 Ribosome density profile from preliminary ribosome footprint library. *A meta-gene profile of ribosome footprints from our preliminary dataset, in which center-weighted reads were assigned to a nucleotide position in each of their respective genes. Normalized read counts were obtained by normalizing read counts at every nucleotide position by the total number of reads. P-site offset of 12-14 nucleotides was applied to meta-gene profile to account for the distance of the P-site and the 5' terminus. Reads were mapped in the region from 0 to -100 and none were found in the region upstream of the open reading frame. A peak was seen between position -10 and -20, suggesting a possible stalling of the ribosome at this position during translation initiation*.

Whole transcript sequencing and genome reannotation. To complement our ribosome profiling studies, we also attempted to use transcriptome assembly of RNA-seq reads to sort transcripts with 5′ UTR sequences and polycistronic transcripts into different categories, allowing us to probe translation initiation in leadered and leaderless transcripts as well as polycistronic or monocistronic transcripts in a targeted fashion. To perform this assembly, we first prepared an RNA library from wild type *M. maripaludis* grown in rich media and performed RNA-seq. Rockhopper software was then used to assemble NGS reads into whole transcripts.

After in silico assembly of these putative transcripts, we verified the 3′ termini using the "Capture and Amplification by Tailing and Switching (CATS)" sequencing method [20] (Figure 4.6). In this method, the 3′ UTR is tailed with a poly(A) sequence using a poly(A) polymerase. This tail serves as the primer region for the synthesis of the complementary strand, after which both strands can be amplified by PCR. After library preparation, amplicons were sequenced, and the resulting reads were then mapped back to the genome. The predicted 3′ termini allowed us to further filter our Rockhopper transcript dataset to focus further analysis on 3′ terminus-verified transcripts only. Doing this, we generated a table of putative operons (Appendix Table A4.1) and 5′ UTR sequences (Table 4.2). The operons identified varied greatly in predicted 5′ UTR length, number of genes, and 3′ UTR length. Most genes did not appear to have a 5′ UTR, with leaderless, monocistronic genes representing roughly 40% of total genes.

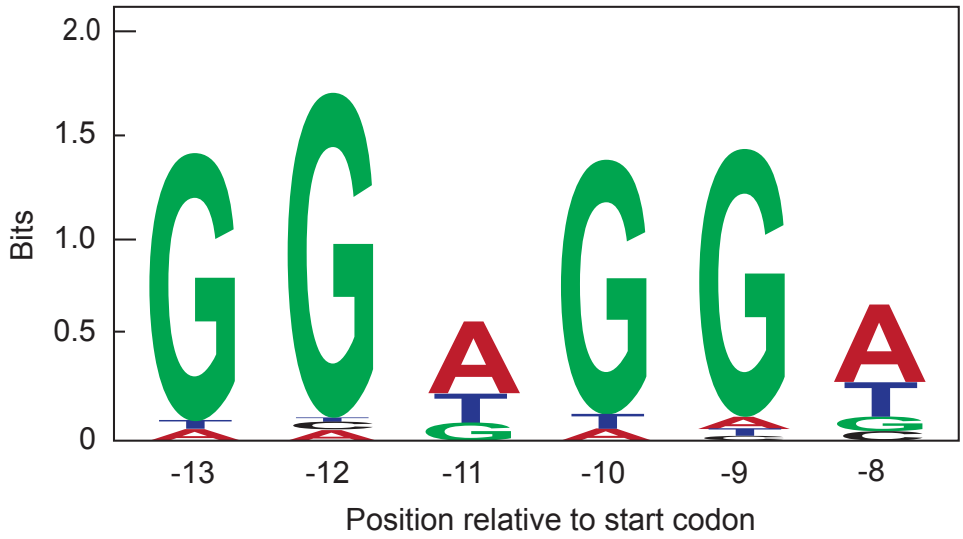
With these categories, we aimed to identify the consensus ribosome binding site and compare this to our promoters. Using the RBS finder software, we identified a series of putative RBS sequences that were almost precisely 8 nucleotides away from the start site. After determining the relative frequency of each nucleotide in the region of -8 to -13, we concluded that the consensus RBS from our dataset was GGAGGA (Figure 4.6). When comparing this sequence to our expression promoters, it was noted that this exact SD-sequence was found in the McrB promoter sequence that led to the highest protein expression. In accordance with this observation, it was observed in M. barkeri, a related methanogenic species, that the consensus sequence was GGAGG [45].

Past studies of multi-cistronic transcripts with and without 5′ UTR sequences indicates that the SD sequence is relevant for the translation of distal genes [11], although a similar analysis conducted on our operon database did not reveal obvious patterns of a SD sequence precisely 8 nucleotides upstream of the start codon. Further biochemical analysis of intergenic regions in operons and 5' UTR-containing transcripts can use this transcript dataset as a starting point (Table 4.1).

Gene Name	UTR Length	Part of operon?	Nt Position	Gene Alias
Methylated-DNA--protein-cysteine methyltransferase	8	No	80649	MMP0069
DNA/RNA-binding protein AlbA	8	No	1559489	MMP1613
Phosphoribosylaminoimidazole carboxylase catalytic subunit	10	No	283401	MMP0282
Metal dependent phosphohydrolase	12	Yes	54361	MMP0037
Aspartate/glutamate/uridylate kinase	13	Yes	1585939	MMP1645
Malate dehydrogenase	14	Yes	636877	MMP0645
Precorrin-8X methylmutase	17	Yes	327278	MMP0330
ATPase RIL	17	Yes	379591	MMP0382
Rhodanese domain-containing protein	17	Yes	54361	MMP0900
Hexapeptide repeat-containing transferase	20	No	1334880	MMP1355
ADP-specific phosphofructokinase	22	Yes	1281207	MMP1296
SAM-binding motif-containing protein	43	No	1356373	MMP1375
Sulfopyruvate decarboxylase subunit alpha	54	No	409867	MMP0411
Glutamate-binding protein	84	No	1575110	MMP1632
Preflagellin peptidase	116	Yes	558325	MMP0555
Transcriptional regulator	191	Yes	107207	MMP0097
Amino acid-binding ACT domain-containing protein	282	Yes	709877	MMP0716
Phosphate transporter PhoU	383	No	1185544	MMP1199
TIM-barrel protein	385	Yes	405344	MMP0406
Arsenite-activated ATPase ArsA	541	No	172067	MMP0163
Multiple resistance and pH regulation protein F	579	No	1041634	MMP1049
Major facilitator transporter	658	No	1577524	MMP1636
GCN5-like N-acetyltransferase	739	Yes	1264147	MMP1281
Phosphoribosyl-ATP pyrophosphatase	788	Yes	66822	MMP0051
ATPase	1092	No	281507	MMP0281
Microsomal signal peptidase 21 kDa subunit	1340	Yes	385673	MMP0387
AIR synthase-like domain-containing protein	1579	Yes	875462	MMP0882
Prephenate dehydrogenase	1691	No	1473136	MMP1514
Iron ABC transporter ATPase subunit	1770	Yes	1166627	MMP1183
RNA-processing protein	2039	Yes	600046	MMP0605
Molybdopterin-guanine dinucleotide biosynthesis protein A	2184	Yes	572518	MMP0573

Table 4.1 Predicted 5' UTR sequences from Rockhopper transcriptome assembly. *5' untranslated regions of predicted transcripts from Rockhopper assembly software [19]. UTR length (in nucleotides) suggest a large variety of 5' UTR lengths were detected. Some, but not all of the 5' UTR-containing transcripts have polycistronic operons. 5' UTR sequences could play a role in translation initiation through direct interaction with the initiation complex or through its effect on mRNA stability, a known phenomenon in bacterial species*.

A



B

McrB promoter region (Methanosarcina barkeri): 5' -TAAATTAA**GGAGGA**AATTCAT- 3'

FlaB1 promoter region (Methanosarcina barkeri): 5' -CAAACC**AAATAC**AGCACCA- 3'

Ribosome protein L31E promoter region (M. maripaludis): 5' -CATGC**GGAGGT**GCATAAT- 3'

McrB promoter region (Methanococcus voltae): 5' -AAAAAAATA**GGAGTG**ATTCTC- 3'

Figure 4.6 Comparing promoter elements and consensus ribosome binding sites in M. maripaludis. *The GGAGGA consensus sequence was identified in our study. (A) The consensus binding site was obtained by using the sequence of the start codon region for all leadered genes in the genome using our Rockhopper assembly and a reference M. maripaludis NCBI file (accession ID: NC_005791). Two GG pairs were most dominant in position -13, -12, -10, and -9, suggesting their importance for SD-anti-SD interactions. (B) When comparing this sequence to the putative ribosome binding sites in a panel of promoters (highlighted in red), the presence of two GG pairs is apparent in all except for the flaB1$_{Mb}$ promoter, which produced no heterologous translation product. In contrast the two mcrB promoters that were effective at producing translated product had a GGAG in their putative RBS.*

Validation of 5′ UTR-containing and leaderless transcripts. To validate transcripts before further translation studies, candidates will be sequenced again to confirm the upstream sequence in leadered and leaderless transcripts. To identify 5′ UTR, 5′ rapid analysis of cDNA ends (RACE) may be performed [46]. In this method, a complementary cDNA strand is made using the strand of interest as a template (Figure 4.7). The 3′ end of the new cDNA strand can then be adenylated, serving as an annealing region for a DNA primer, which can replicate the original strand by using the complementary poly(A) strand as a template. After PCR amplification, the resulting amplicons can be sequenced to reveal the sequence of the 5′ UTR, if any exists. Alternatively, a variety of whole transcript sequencing methods have become available, allowing for the sequencing of intact transcripts from the 5′ to the 3′ end [47, 48]. Examples of this method include Nanopore and PacBio sequencing, both of which allow for transcript sequencing without requiring the fragmentation of the transcripts in their workflow. This allows for improved fidelity during the read mapping process, reducing the probability that reads will be incorrectly matched to upstream or downstream transcripts and providing greater accuracy for the prediction of 5′ UTR sequences, operons, and 3′ UTR sequences. Although this strategy has a more complex workflow and issues with 5′ terminus sequencing accuracy compared to 5′ RACE, it allows for a more comprehensive analysis of regulon structure.

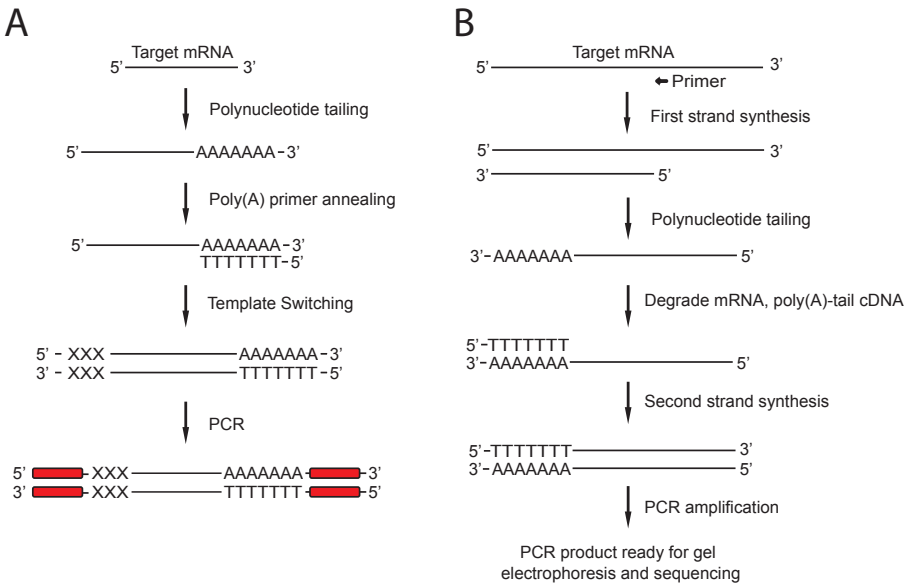


Figure 4.7 Sequencing strategies for the 5' and 3' mRNA termini. *Distinct methods are required for the sequencing of the 5' and 3' termini of mRNA transcripts. Capture and amplification by tailing and switching (CATS) sequencing (left) is a method for adding 3' poly-A tails to mRNA, Reverse transcribing this extended 3' end allows for first strand cDNA synthesis. From that point, standard Illumina library preparation methods can be used. In the 5′ RACE method (right), A gene specific primer is used for targeted cDNA synthesis of the RNA of interest. This first cDNA strand is extended by poly-A tailing of the 3' end, creating a complementary region for a small, poly(T) oligonucleotide. Reverse transcription is used for the second strand synthesis, after which PCR products can be amplified for gel electrophoresis or further processed to make Illumina libraries for NGS.*

Sequencing the 3′ UTR to characterize 3′-terminus boundaries. Similar to the 5′ UTR, the 3′ UTR can also play a role in translation regulation [49]. Long 3′ UTRs are commonly used for regulation in eukaryotes and typically contain binding sites for non-coding RNA [50]. In eukaryotes, the 3′ UTR can interact with binding proteins that circularize the transcript and target the 5′ mRNA cap [51, 52, 2]. Evidence has also been found in bacterial species for the presence of antisense RNA (asRNA), which can be formed by long 3′ UTR sequences and effect levels of transcription, mRNA stability or translation [53]. Since 3′ UTR regulation of translation has been shown to be relevant to both bacterial and eukaryotic species, we aimed to gain a better understanding of translation regulation by sequencing the 3′ UTR in a model archaeal species.

Using CATS-seq data and a custom algorithm, we analyzed the resulting 3′ UTR sequence predictions. From these results, we found that 30% of the terminators were intragenic, while the remainder were found in the intergenic region (Figure 4.8). Many of the intragenic terminators were in convergent genes, suggesting the possibility that 3′ UTR sequences may bind to adjacent transcripts, potentially leading to a change in mRNA stability and regulating gene expression level. Indeed, several cases of overlapping transcriptional units have been observed to lead to mRNA destabilization or reduced translation efficiency [53-55]. This suggests the importance of proper gene termination in designing pathway expression vectors. To better elucidate possible terminators, we analyzed the nucleotide frequency of 3′ ends of identified terminators. This analysis showed that polyuridine tracts are common in these reads, possible forming stable RNA secondary structures (Figure 4.8). Mapping of genes from other Euryarchaea showed that in vivo RNA 3′ ends are uridine-rich and have the potential to form stable RNA secondary structures [56]. In vivo data using reporter assays also confirmed that uridine rich tracts were necessary for termination [57]. Interestingly, we found that there were several instances of more than one putative 3′ UTR sequence for a transcript (Figure 4.8). These terminator isoforms varied in length and may have similar properties as those proposed for the 3′ UTR of multicistronic transcripts in M. mazei, in which leaky termination was proposed to have a role in tuning gene stoichiometry in operons [55].

Overall, our analysis of 3′ UTR sequences reveals possible terminator sequences for 3′-termini, which can be used for further characterization of transcripts. Further study of our 3’ UTR dataset revealed possible strategies for tuning gene expression using intragene transcript termination or multiple terminator isoforms. With our sequencing results, a pattern of AT rich motifs at the 3’ terminus of transcripts emerges. These results also provide a starting point towards designing terminators for heterologous gene expression in *M. maripaludis*, with possible applications to other archaeal species.

Gene Name	UTR Length	Part of operon?	Nt Position	Gene Alias
Integral membrane protein	2	No	716325	MMP0725
Homoserine kinase	3	No	295253	MMP0295
Amino acid-binding ACT domain-containing protein	5	Yes	709877	MMP0716
Aspartate kinase	5	No	1004396	MMP1017
Bifunctional RNAse	5	Yes	1327842	MMP1348
Cytidylate kinase	10	Yes	619232	MMP0626
TOBE domain-containing protein	11	No	620371	MMP0628
RNA-processing protein	12	No	600046	MMP0605
Zinc finger protein	22	No	1416669	MMP1446
Carbonic anhydrase	70	Yes	65867	MMP0049
Sulfopyruvate decarboxylase subunit alpha	73	No	409867	MMP0411
Multiple resistance and pH regulation protein F	78	Yes	1041634	MMP1049
Mevalonate kinase	258	Yes	1315756	MMP1335
Hydrogenase assembly chaperone (HypC/HupF)	275	No	1310965	MMP1330
Rubredoxin-type Fe(Cys)4 protein	303	No	301604	MMP0303
Endonuclease III-like protein	398	Yes	583252	MMP0586
Holliday junction resolvase	429	No	332791	MMP0336
XRE family transcriptional regulator	496	Yes	981206	MMP0993
Imidazole glycerol phosphate synthase (HisH)	741	No	258316	MMP0256
50S ribosomal protein L14e	1032	No	618916	MMP0625
Xylose isomerase domain-containing protein	2245	No	122675	MMP0114
Radical SAM domain-containing protein	3901	Yes	64389	MMP0047

Table 4.2 Predicted 3'-UTR sequences from Rockhopper transcriptome assembly. *3' untranslated regions of predicted transcripts from Rockhopper assembly software. UTR length (in nucleotides) suggest a large variety of 3' UTR lengths were detected, with some lengths approaching or exceeding the length of 3' UTR from eukaryotic species. The role of the 3' UTR has not yet been clearly elucidated in M. maripaludis***.**

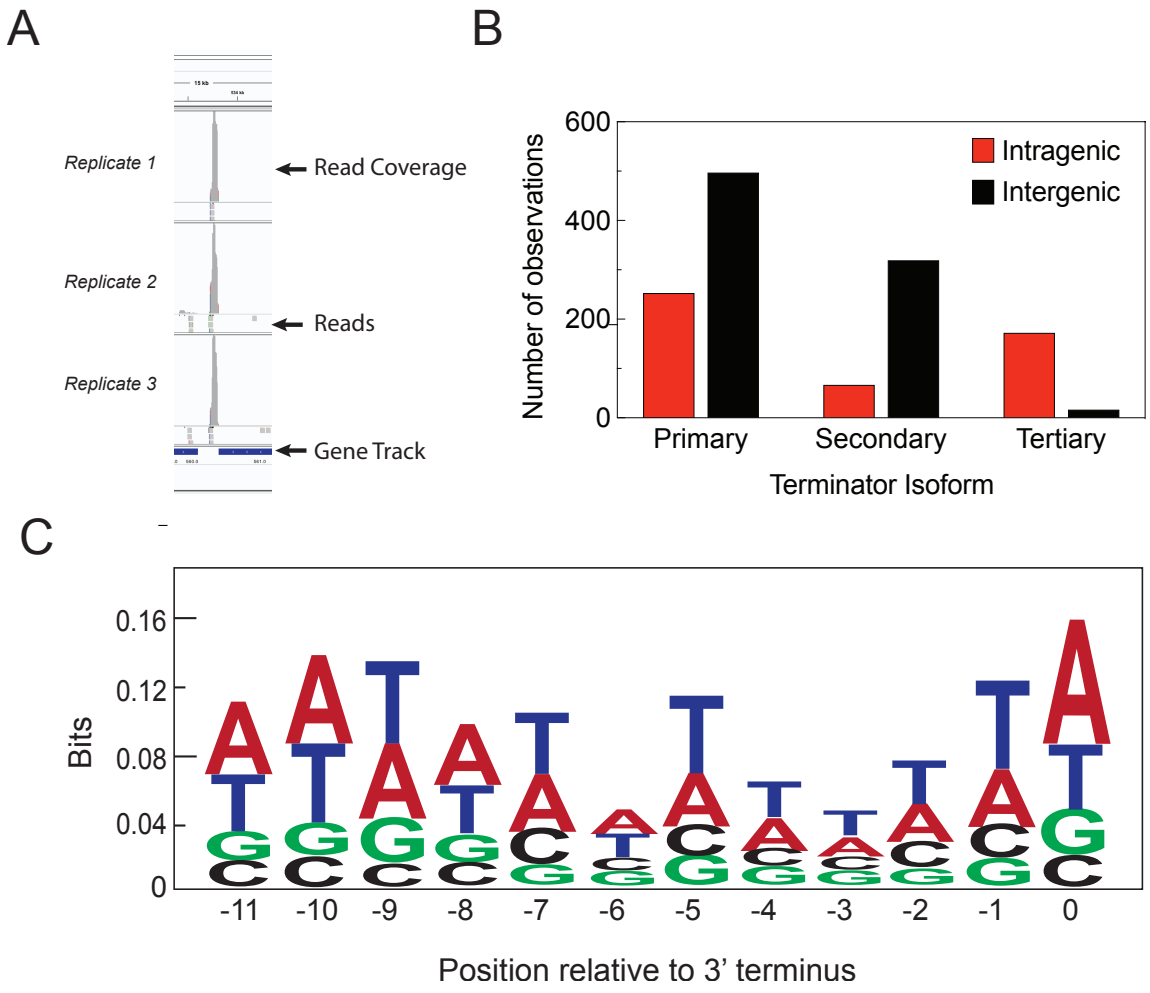


Figure 4.8 3' termini have poly(A) tracts that can stretch into intragenic regions. *Sequencing data suggests terminators can be intragenic, an occurrence that can cause mRNA instability in converging gene pairs. (A) CATS-seq reads were mapped to a visual map of the genome in the Integrated Genome Viewer (IGV) [58], showing the presence of several primary, secondary and tertiary terminator isoforms in downstream genes, Data is from three independent biological replicates. (B) Frequency counts of intragene and intergene terminator isoforms. Intragenic terminators represent about one-third of all primary terminators identified by our algorithm. Multiple terminators suggest a possible role for "leaky" genes to control gene stoichiometries in polycistronic transcripts. (C) Sequence logo plot of the 3' ends of CATS-seq identified terminators. Position 0 is the 3' end of a terminator, while -11 represents the 11th nucleotide away from the 3' end of the transcript. Clear patterns of A/T rich regions spaced a few nucleotides away suggests a model of RNA hairpins, which could be involved in mRNA stability in a similar way to bacterial mRNA hairpins [59].*

4.4. Conclusions

The translation of heterologous transcripts in *M. maripaludis* has been built on the use of a small toolbox of known promoters. Many of these genetic elements were used to express pathways for 3-hydroxybutyrate (3HB) and polyhydroxybutyrate (PHB) in Chapter 3, but extension of the toolbox frequently did not lead to translation, despite the fact that almost all of the pathways were successfully transcribed. The regulation of translation and mRNA stability in archaea is an area of continuing research, but key knowledge gaps in this domain of life challenge our ability to robustly predict gene expression from the promoters and terminators cloned with our pathway.

To better understand the translation process, we focused on translation initiation because this step is typically rate limiting [60]. Among transcripts with 5′ UTR sequences, SD-anti-SD interactions are predicted to be necessary for translation initiation [61]. According to our analysis, the consensus SD sequence GGAGGA should be used for translation initiation. Besides leadered transcripts, our analysis suggests that the majority of transcripts in *M. maripaludis* may be leaderless, potentially adding new challenges in translation of heterologous genes. The method of translation initiation for leaderless transcripts is not fully elucidated, although models have been proposed (e.g. one hypothesis is that the $tRNA^{fMet}$ base pairs with the start codon while complexed to the 30S subunit or with the 70S complex, allowing for the stable formation of a translation initiation complex) [13]. To investigate the models of translation initiation for both leadered and leaderless transcripts, we developed a method to determine ribosome pausing sites using ribosome profiling in the model methanogenic Archaea, *M. maripaludis*. After determining that chloramphenicol and RNase I were an effective translation inhibitor and nuclease, respectively, we moved forward with the standard ribosome profiling method.

Performing ribosome profiling allowed for the initial identification of common rRNA fragments contaminants and common quality assessment metrics validated the ribosome profile. The ribosome density in the non-rRNA contaminants suggested that footprints were found within the open reading frame, with almost no footprints seen in the upstream region before the start codon. Although this is preliminary evidence, it suggests that ribosome pausing appears to happen within the reading frame, which resembles the expected results from a predominantly leaderless mRNA pool. Elucidating the transcriptome further would allow for the appropriate identification of putative leaderless transcripts, operon structures and 3′ UTR sequences. This assembly process revealed a large set of putative operons and several sequences without 5′ UTR, which sets the stage for further biochemical and genetic analysis of leadered and leaderless sequences. These transcripts were confirmed with 3′-termini sequencing, which also provided more details on 3′ UTR sequences. Further analysis of the 3′-termini dataset revealed evidence for intragenic terminators and multiple terminator isoforms. These features may also have important roles in mRNA stability, operon gene stoichiometry and possible direct roles in translation initiation as well. Overall, this work lays some of the foundation for more predictable design of heterologous biosynthetic pathways for robust gene expression in *M. maripaludis* and possibly other archaea.

4.5. References

1. M. Kozak, "Initiation of translation in prokaryotes and eukaryotes". Gene. **1999**, 234, 187–208.
2. M. Brenneis, J. Soppa, "Regulation of Translation in Haloarchaea: 5′- and 3′-UTRs Are Essential and Have to Functionally Interact In Vivo". PLoS ONE. **2009**, 4, e4484.
3. H. Jian, L. Xiong, G. Xu, X. Xiao, F. Wang, "Long 5′ untranslated regions regulate the RNA stability of the deep-sea filamentous phage SW1". Scientific Reports. **2016**, 6, 1–9.
4. M. Kozak, "Comparison of initiation of protein synthesis in prokaryotes, eukaryotes, and organelles". Microbiological Reviews. **1983**, 47, 1–45.
5. N. Pedullà, R. Palermo, D. Hasenöhrl, U. Blasi, P. Cammarano, P. Londei, "The archaeal eIF2 homologue: Functional properties of an ancient translation initiation factor". Nucleic Acids Research. **2005**, 33, 1804–1812.
6. D. Hasenöhrl, D. Benelli, A. Barbazza, P. Londei, U. Bläsi, "Sulfolobus solfataricus translation initiation factor 1 stimulates translation initiation complex formation". RNA. **2006**, 12, 674–682.
7. K. Gäbel, J. Schmitt, S. Schulz, D. J. Näther, J. Soppa, "A comprehensive analysis of the importance of translation initiation factors for Haloferax volcanii applying deletion and conditional depletion mutants". PLoS ONE. **2013**, 8, e77188.
8. P. D. Coureux, C. Lazennec-Schurdevin, S. Bourcier, Y. Mechulam, E. Schmitt, "Cryo-EM study of start codon selection during archaeal translation initiation". Nature Communications. **2016**, 7, 13366.
9. A. Monestier, C. Lazennec-Schurdevin, P. D. Coureux, Y. Mechulam, E. Schmitt, "Role of aIF1 in Pyrococcus abyssi translation initiation". Nucleic Acids Research. **2018**, 46, 11061–11074.
10. E. Torarinsson, H. P. Klenk, R. A. Garrett, "Divergent transcriptional and translational signals in archaea". Environmental Microbiology. **2005**, 7, 47–54.
11. J. Ma, A. Campbell, S. Karlin, "Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures". Journal of Bacteriology. **2002**, 184, 5733–5745.
12. A. Lomsadze, K. Gemayel, S. Tang, M. Borodovsky, "Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes". Genome Research. **2018**, 28, 1079–1089.
13. E. Schmitt, P. Coureux, R. Kazan, G. Bourgeois, C. Lazennec-Schurdevin ,Y. Mechulam, "Recent advances in archaeal translation initiation". Frontiers in Microbiology. **2020**, 11, 2259.
14. L. Guillon, E. Schmitt, S. Blanquet, Y. Mechulam, "Initiator tRNA binding by e/aIF5B, the eukaryotic/archaeal homologue of bacterial initiation factor IF2". Biochemistry. **2005**, 44, 15594–15601.
15. N. T. Ingolia, "Ribosome profiling: New views of translation, from single codons to genome scale". Nature Reviews Genetics. **2014**, 15, 205–213.
16. G. A. Brar, J. S. Weissman, "Ribosome profiling reveals the what, when, where and how of protein synthesis". Nature Reviews Molecular Cell Biology. **2015**, 16, 651–664.
17. S. L. Wolin, P. Walter, "Ribosome pausing and stacking during translation of a eukaryotic mRNA". The EMBO journal. **1988**, 7, 3559–3569.
18. J. A. Steitz, "Polypeptide chain initiation: Nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA". Nature. **1969**, 224, 957–964.

19. R. McClure, D. Balasubramaniam, Y. Sun, M. Bobrovskyy, P. Sumby, C. A. Genco, C. K. Vanderpool, B. Tjaden, "Computational analysis of bacterial RNA-Seq data". Nucleic Acids Research. **2013**, 41, e140.
20. A. Turchinovich, H. Surowy, A. Serva, M. Zapatka, P. Lichter, B. Burwinkel, "Capture and amplification by tailing and switching (CATS), Anultrasensitive ligation-independent method for generation of DNA libraries for deep sequencing from picogram amounts of DNA and RNA". RNA Biology. **2014**, 11, 817–828.
21. N. J. McGlincy, N. T. Ingolia, "Transcriptome-wide measurement of translation by ribosome profiling". Methods. **2017**, 126, 112–129.
22. J. G. Dunn, J. S. Weissman, "Plastid: Nucleotide-resolution analysis of next-generation sequencing and genomics data". BMC Genomics. **2016**, 17, 958.
23. N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, "Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling". Science. **2009**, 324, 218–223.
24. N. T. Ingolia, L. F. Lareau, J. S. Weissman, "Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes". Cell. **2011**, 147, 789–802.
25. F. Mohammad, C. J. Woolstenhulme, R. Green, A. R. Buskirk, "Clarifying the translational pausing landscape in bacteria by ribosome profiling". Cell Reports. **2016**, 14, 686–694.
26. I. Chopra, M. Roberts, "Tetracycline antibiotics: Mode of action, applications, molecular biology, and epidemiology of bacterial resistance". Microbiology and Molecular Biology Reviews. **2001**, 65, 232–260.
27. A. S. Weisberger, S. Wessler, L. V. Avioli, "Mechanisms of action of chloramphenicol". JAMA: The Journal of the American Medical Association. **1969**, 209, 97–103.
28. M. Gaynor, A. Mankin, "Macrolide antibiotics: Binding site, mechanism of action, resistance". Current Topics in Medicinal Chemistry. **2005**, 3, 949–960.
29. J. Spížek, T. Řezanka, "Lincosamides: Chemical structure, biosynthesis, mechanism of action, resistance, and applications". Biochemical Pharmacology. **2017**, 133, 20–28.
30. B. D. Davis, "Mechanism of bactericidal action of aminoglycosides". Microbiological Reviews. **1987**, 51, 341–350.
31. N. Garreau De Loubresse, I. Prokhorova, W. Holtkamp, M. V. Rodnina, G. Yusupova, M. Yusupova, "Structural basis for the inhibition of the eukaryotic ribosome". Nature. **2014**, 513, 517–522.
32. G. Yusupova, L. Jenner, B. Rees, D. Moras, M. Yusupov, "Structural basis for messenger RNA movement on the ribosome". Nature. **2006**, 444, 391–394.
33. K. S. Long, B. T. Porse, "A conserved chloramphenicol binding site at the entrance to the ribosomal peptide exit tunnel". Nucleic Acids Research. **2003**, 31, 7208–7215.
34. M. V. Gerashchenko, V. N. Gladyshev, "Translation inhibitors cause abnormalities in ribosome profiling experiments". Nucleic Acids Research. **2014**, 42.
35. J. A. Hussmann, S. Patchett, A. Johnson, S. Sawyer, W. H. Press, "Understanding Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in Yeast". PLoS Genetics. **2015**, 11.
36. L. F. Lareau, D. H. Hite, G. J. Hogan, P. O. Brown, "Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments". eLife. **2014**, 2014.
37. G. W. Rushizky, H. A. Sober, “Characterization of the major compounds found in

ribonuclease T1 digests of ribonucleic acid" Tetranucleotides. **1962**, 2883-2888.
38. M. V. Gerashchenko, V. N. Gladyshev, "Ribonuclease selection for ribosome profiling". Nucleic Acids Research. **2017**, 45, e6.
39. E. Oh, A. H. Becker, A. Sandikci, D. Huber, R. Chaba, F. Gloge, R. J. Nichols, A. Typas, C. A. Gross, G. Kramer, J. S. Weissman, B. Bukau, "Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo". Cell. **2011**, 147, 1295–1308.
40. A. La Teana, A. Brandi, M. O'Connor, S. Freddi, C. L. Pon, "Translation during cold adaptation does not involve mRNA-rRNA base pairing through the downstream box". RNA. **2000**, 6, 1393–1402.
41. W. Zhang, J. A. Dunkle, J. H. D. Cate, "Structures of the ribosome in intermediate states of ratcheting". Science. **2009**, 325, 1014–1017.
42. F. Cholet, U. Z. Ijaz, C. J. Smith, "Differential ratio amplicons (R amp) for the evaluation of RNA integrity extracted from complex environmental samples". Environmental Microbiology. **2019**, 21, 827–844.
43. D. R. Gelsinger, E. Dallon, R. Reddy, F. Mohammad, A. R. Buskirk, J. DiRuggiero, "Ribosome profiling in archaea reveals leaderless translation, novel translational initiation sites, and ribosome pausing at single codon resolution". Nucleic acids research. **2020**, 48, 5201–5216.
44. C. Legrand, F. Tuorto, "RiboVIEW: A computational framework for visualization, quality control and statistical analysis of ribosome profiling data". Nucleic Acids Research. **2020**, 48, 7.
45. K. Sauer, U. Harms, R. K. Thauer, "Methanol: Coenzyme M methyltransferase from Methanosarcina barkeri. Purification, properties and encoding genes of the corrinoid protein MT1". European Journal of Biochemistry. **1997**, 243, 670–677.
46. G. Schramm, I. Bruchhaus, T. Roeder, "A simple and reliable 5'-RACE approach.". Nucleic acids research. **2000**, 28, 96.
47. D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. Sean Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Peshin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, J. Schloss, "The potential and challenges of nanopore sequencing" in Nanoscience and Technology: A Collection of Reviews from Nature Journals, (World Scientific Publishing Co., 2009), pp. 261–268.
48. A. Rhoads, K. F. Au, "PacBio sequencing and its applications". Genomics, Proteomics and Bioinformatics. **2015**, 13, 278–289.
49. E. Szostak, F. Gebauer, "Translational control by 3'-UTR-binding proteins". Briefings in Functional Genomics. **2013**, 12, 58–65.
50. D. C. Di Giammartino, K. Nishida, J. L. Manley, "Mechanisms and consequences of alternative polyadenylation". Molecular Cell. **2011**, 43, 853–866.
51. N. Sonenberg, A. G. Hinnebusch, "Regulation of translation initiation in eukaryotes: Mechanisms and biological targets". Cell. **2009**, 136, 731–745.
52. R. J. Jackson, C. U. T. Hellen, T. V. Pestova, "The mechanism of eukaryotic translation initiation and principles of its regulation". Nature Reviews Molecular Cell Biology. **2010**, 11, 113–127.
53. N. Sesto, O. Wurtzel, C. Archambaud, R. Sorek, P. Cossart, "The excludon: A new concept in bacterial antisense RNA-mediated gene regulation". Nature Reviews Microbiology. **2013**, 11, 75–82.

54. O. Wurtzel, N. Sesto, J. R. Mellin, I. Karunker, S. Edelheit, C. Bécavin, C. Archambaud, P. Cossart, R. Sorek, "Comparative transcriptomics of pathogenic and non-pathogenic Listeria species". Molecular Systems Biology. **2012**, 8, 583.
55. D. Dar, D. Prasse, R. A. Schmitz, R. Sorek, "Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea". Nature Microbiology. **2016**, 1.
56. M. Brenneis, O. Hering, C. Lange, J. Soppa, "Experimental characterization of cis-acting elements important for translation and transcription in halophilic archaea". PLoS Genetics. **2007**, 3, e229.
57. T. J. Santangelo, L. Cubonová, K. M. Skinner, J. N. Reeve, "Archaeal intrinsic transcription termination in vivo". Journal of Bacteriology. **2009**, 191, 7102–7108.
58. H. Thorvaldsdottir, J. T. Robinson, J. P. Mesirov, "Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration". Briefings in Bioinformatics. **2013**, 14, 178–192.
59. S. A. Emory, P. Bouvet, J. G. Belasco, "A 5'-terminal stem-loop structure can stabilize mRNA in Escherichia coli". Genes and Development. **1992**, 6, 135–148.
60. B. Reeve, T. Hargest, C. Gilbert, T. Ellis, "Predicting translation initiation rates for designing synthetic biology". Frontiers in Bioengineering and Biotechnology. **2014**, 2.
61. D. Benelli, E. Maone, P. Londei, "Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation". Molecular Microbiology. **2003**, 50, 635–643.

Appendix 1: Plasmids and oligonucleotides

Table A1.1. Plasmid constructs.

Plasmid	Description	Source
pETDuet-AcsB-Strep	acsB-Strep (T7), lacO, ColE1, $lacI^q$, Amp^r	Ch. 2
pETDuet-AcsB-Strep-CODH	acsB-Strep (T7), CODH (T7), lacO, ColE1, $lacI^q$, Amp^r	Ch. 2
pACYCDuet-AcsF	acsF (T7), lacO, P15A, $lacI^q$, Cm^r	Ch. 2
pMTL83151-ΔpyrF::Kan^r	P_{G3PD}-kan^r – kanamycin resistance cassette, catP, ColE1, pCB102, 1 kb homology arms flanking the pyrF gene	Ch. 2
pTrc33-Δhpt::adhE	P_{mcrB} – adhE, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::hbd	P_{mcrB} – hbd, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::hbd-Strep	P_{mcrB} – hbd tagged with a C-terminus Strep-tag, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::tesB	P_{mcrB} – tesB, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::tesB-Strep	P_{mcrB} – tesB tagged with a C-terminus Strep-tag, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::P_{mcrB}-phaA-P_{flaB1}-phaB-P_{hdrC1}-tesB	P_{mcrB} – phaA, P_{flaB1} – phaB, P_{hdrC1} – tesB, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::P_{mcrB}-phaA-P_{flaB1}-phaB-P_{hdrC1}-tesB-Strep (all)	P_{mcrB} – phaA, P_{flaB1} – phaB, P_{hdrC1} – tesB, all tagged with a C-terminus Strep-tag, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::P_{hdrC1}-phaA-P_{flaB1}-hbd-T (all)	P_{hdrC1} – phaA, P_{flaB1} – hbd with terminators at the C-terminus, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::P_{hdrC1}-phaA-P_{flaB1}-hbd-Strep-T (all)	P_{hdrC1} – phaA, P_{flaB1} – hbd with terminators and Strep-tags at the C-terminus, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::phaA-hbd-tesB-P_{mcrB} (all)	P_{mcrB} – phaA-hbd-tesB, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::phaA-hbd-tesB-Strep(all)	P_{mcrB} – phaA-hbd-tesB, each tagged with a C-terminus Strep-tag, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb	Ch. 3
pTrc33-Δhpt::phaA-phaB-tesB-Strep(all)-syntheticRBS	P_{mcrB} – phaA-hbd-tesB, each tagged with a C-terminus Strep-tag, intergenic regions contain synthetic RBS, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb	Ch. 3
pTrc33-Δhpt::MMP0737	P_{mcrB} – MMP0737, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::nadA	P_{mcrB} – nadA, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3

pTrc33-Δhpt::nadC	P_{mcrB} – nadC, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::MMP1578	P_{mcrB} – MMP1578, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::MMP1349	P_{mcrB} – MMP1349, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::nadMV	P_{mcrB} – nadM and nadV, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pBad24-Δupt::fdh1	P_{hmv} – fdh1, (neo^r) – neomycin resistance cassette, $lacI^q$, Cb^r, M13, 1 kb homology arms flanking the upt gene	Ch. 3
pBad24-Δupt::fdh1-nadMV	P_{hmv} – fdh1, P_{mcrB} – nadM and nadV, (neo^r) – neomycin resistance cassette, $lacI^q$, Cb^r, M13, 1 kb homology arms flanking the upt gene	Ch. 3
pTrc33-Δhpt::phaAB-tesB	P_{mcrB} – phaA-phaB-tesB, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::phaABC	P_{mcrB} – phaA-phaB-phaC, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3
pTrc33-Δhpt::phaABC-nadMV	P_{mcrB} – phaA-phaB-phaC, nadM-nadV, (pac^r) – puromycin resistance cassette, $lacI^q$, Cm^r, M13, 1 kb homology arms flanking the hpt gene	Ch. 3

Table A1.2. Primers used in the construction of plasmids in Table A1.1.

Name	Sequence
pyrF LH f	ggatgattgttatggattataagcggccggcctgtcctcaacaccctcaccagg
pyrF LH r	cttggcaaccgtccagttattatttcacctggcaagcaggccagaagc
pyrF RH f	gagaaaataaaatattatattttactggatgaattgttttagtaacttcggcctgctttc
pyrF RH r	ggtgcataagcttggacagtttaaactccttttttgataatctcatgaccaaa
hpcE f	gcggataacaatttcacacgagctcggtacccgggatccgatggaaccagcttttctgaaatatcg
hpcE r	aaaaaatccgcttaaattataaattaaatttaaattggtgaattatttttcaggtgaaattttggatataaaacc
pac f	aaatatttatgatacacaaaaagaaacagctaaaaaataagagtattcaaattcaaattattgtg
pac r	tcttttgtgttctccgataaatgaagcatgcgaattgccccagtgaattaaaaatatat
MMP0143 f	aataacacaataatttgaatttgaatactcttatttttttagctgtttctttttgtgtatcataaa
MMP0143 r	caagcttgcatgcctgcaggtcgactctagaggatcgaagataaatcctttgaaatcaactcttgtaaa
P_{mcrB} f	ctttttttatatattttttaattcactgggggcaattcgcatgcttcatttatcggagaac
P_{mcrB} r	cgctctacgagtgcgttaag
adhE f	cctgaaaaataattcaccaatttaaatttaatttataatttaagcggatttttttcgcttttttc
adhE r	ctttttttatatattttttaattcactgggggcaattcgcatgcttcatttatcggagaac
P_{mcrB}-adhE SOE f	cctgaaaaataattcaccaatttaaatttaatttataatttaagcggatttttttcgcttttttc
P_{mcrB}-adhE SOE r	ctttttttatatattttttaattcactgggggcaattcgcatgcttcatttatcggagaac
phaA-Strep f	accttctaaacgaatgagatttcatt
phaA-Strep r	cgtttattaccttacttctcgaactgcggatggctccatttttctttcaactgcgagagc
phaB-Strep f	gtgacataaaatcacttacttctcgaactgcggatggctccaacccatgtgcaatccacc
phaB-Strep r	ctttgtgtcatcgtttattaccttacttctcgaactgcggatggctccatttttctttcaa
tesB-Strep f	ttaatttataatttacttctcgaactgcggatggctccagttgtggtttctcatcactcc
tesB-Strep r	aatgcttgtgacataaaatcacttacttctcgaactgcggatggctccaacccatgtgc
hbd-Strep f	ctacagtaaatggagccatccgcagttcgagaagtaagcatgcttcatttatcggagaac
hbd-Strep r	ctccgataacacaaacttttttcatatgaatttcctccttaatttattaaaatcattttg
phaA-phaB-tesB SOE f	accttctaaacgaatgagatttcatt
phaA-phaB-tesB SOE r	gcggataacaatttcacacgagctcggtacccgggatccgatggaaccagcttttctgaaatatcg
phaA-RBS f	accttctaaacgaatgagatttcatt
phaA-RBS r	gtgtcattataattctcctaataaatctccgttaatatccttacttctcgaactgcggat
phaB-RBS f	aaggatattaacggagatttattaggagaattataatgacacaaagaattgcttacgtaa
phaB-RBS r	atgcttgtgacatgcgtgcttcctatatgattaaggggagttacttctcgaactgcggat
tesB-RBS f	agaagtaactccccttaatcatataggaagcacgcatgtcacaagcattaaaaaacctct
tesB-RBS r	gcggataacaatttcacacgagctcggtacccgggatccgatggaaccagcttttctgaaatatcg
phaA-P_{mcrB} f	accttctaaacgaatgagatttcatt
phaA-P_{mcrB} r	cagattttcgggatgttctttctgttgggcggccgcttattttctttcaactgcgagagc
P_{flaB1} f	tgggcgttgctctcgcagttgaaagaaaataagcggccgcccaacagaaagaacatcccg

P_{flaB1} r	tacgtaagcaattctttgtgtcatatggtgctgtatttggtttgg
phaB-P_{flaB1} f	cataccaaaccaaatacagcaccatatgacacaaagaattgcttacgtaa
phaB-P_{flaB1} r	atgtacgaaggagaaattgcaatcgttaacccatgtgcaatccac
P_{hdrC1} f	ataccggggaggtgaaatcgaatgtcacaagcattaaaaaacctct
P_{hdrC1} r	aatggtggattgcacatgggttaacgattgcaatttctccttcgt
tesB-P_{hdrC1} f	gcggataacaatttcacacgagctcggtacccgggatccgatggaaccagcttttctgaaatatcg
tesB-P_{hdrC1} r	ataccggggaggtgaaatcgaatgtcacaagcattaaaaaacctct
phaA-P_{mcrB}-Strep f	accttctaaacgaatgagatttcatt
phaA-P_{mcrB}-Strep r	cagattttcgggatgttctttctgttgggcggccgcttattttctttcaactgcgagagc
phaB-P_{flaB1}-Strep f	cataccaaaccaaatacagcaccatatgacacaaagaattgcttacgtaa
phaB-P_{flaB1}-Strep r	atgtacgaaggagaaattgcaatcgttaacccatgtgcaatccac
tesB-P_{hdrC1}-Strep f	ataccggggaggtgaaatcgaatgtcacaagcattaaaaaacctct
tesB-P_{hdrC1}-Strep r	gcggataacaatttcacacgagctcggtacccgggatccgatggaaccagcttttctgaaatatcg
term73 f	ttttcgggatgttctttctgttgggcggccgcccctcccgaggcgggact
term73 r	gggactcgggaaagaaagaaaaagcaggcccgaaggcctgctctgttttaagttgtttaa
phaA-term73 f	cctgctctgttttaagttgtttaatttattgagtttacttctcgaactgcggatggctcc
phaA-term73 r	tgaatttcctccttaatttattaaaatcattttgggactggtcacctactcgagtgtccactattcccaatgaaatc
P_{flaB1}-term73 SOE r	cctgctctgttttaagttgtttaatttattgagtttacttctcgaactgcggatggctcc
term3050 f	attccatctcaatctctgttttttttacttctcgaactgcggatggctccatttactgtag
term3050 f2	tgataaaagagaaaaagaaaagaatggaataaaaagttattccatctcaatctctgtttt
hpcE-term3050 r	tctttttctctttttatcaaatcatttctattataaattaaatttaaattggtgaattatt
term3050 f2	aataattcaccaatttaaatttaatttataatag
term3050 r2	ctacagtaaatggagccatccgcag
LH-term3050 f	tctttttctctttttatcaaatcatttctattataaattaaatttaaattggtgaattatt
LH-term3050 r	gcggataacaatttcacacgagctcggtacccgggatccgatggaaccagcttttctgaaatatcg
term3050 SOE f	ttctttctttcccgagtcccgcctcgggaggggcggccgcccaacagaaagaacatcccg
term3050 SOE r	gcggataacaatttcacacgagctcggtacccgggatccgatggaaccagcttttctgaaatatcg
P_{hdrC1} f	ttccttttttttatatatttttaattcactgggggcaattccgattgcaatttctccttcg
P_{hdrC1} r	ctactgcagttcttgcagcagaaacgataacaacgtcagtcattcgatttcacctccccg
phaA-hbd f	aaaacaaagattaaattataccggggaggtgaaatcgaatgactgacgttgttatcgttt
phaA-hbd r	gcggataacaatttcacacgagctcggtacccgggatccgatggaaccagcttttctgaaatatcg
P_{mcrB}-hbd f	caaaatgattttaataaattaaggaggaaattcatatgaaaaaagtttgtgttatcggag
P_{mcrB}-hbd r	aataattcaccaatttaaatttaatttataatttatttactgtagtcgtaaaatcctttg
P_{mcrB}-hbd r2	gatttcattgggaatagtggacactcgagtaggtgaccagtcccaaaatgattttaataaattaaggaggaaattca
P_{mcrB}-hbd-LH f	gcggataacaatttcacacgagctcggtacccgggatccgatggaaccagcttttctgaaatatcg
P_{mcrB}-hbd-LH r	ttttacgactacagtaaataaattataaattaaatttaaattggtgaattatttttcagg
P_{mcrB}-hbd-Strep f	caaaatgattttaataaattaaggaggaaattcatatgaaaaaagtttgtgttatcggag
P_{mcrB}-hbd-Strep r	acctgaaaaataattcaccaatttaaatttaatttataatttacttctcgaactgcggat
P_{mcrB}-hbd-Strep r2	gatttcattgggaatagtggacactcgagtaggtgaccagtcccaaaatgattttaataaattaaggaggaaattca

P_{mcrB}-tesB f	aaaatgattttaataaattaaggaggaaattcatatgtcacaagcattaaaaaacctc
P_{mcrB}-tesB r	gcggataacaatttcacacgagctcggtacccgggatccgatggaaccagcttttctgaaatatcg
P_{mcrB}-tesB r2	gatttcattgggaatagtggacactcgagtaggtgaccagtcccaaaatgattttaataaattaaggaggaaattca
P_{mcrB}-tesB-Strep f	aaaatgattttaataaattaaggaggaaattcatatgtcacaagcattaaaaaacctc
P_{mcrB}-tesB-Strep r	gcggataacaatttcacacgagctcggtacccgggatccgatggaaccagcttttctgaaatatcg
P_{mcrB}-tesB-Strep r2	gatttcattgggaatagtggacactcgagtaggtgaccagtcccaaaatgattttaataaattaaggaggaaattca
upt-LH f	agacgtgaagacaggcagctttcattggaccaaaatcc
upt-LH r	agacgtgaagacagtcgcatttatatgtttctatttctttttttaagt
neo^r f	agacgtgaagactccgttagaaaaactcatcgagca
neo^r r	agacgtgaagactctcatgagccatattcaacgg
upt -RH f	agacgtgaagacgtcaatttaccctctaaatccttaattt
upt -RH r	agacgtgaagacgtggtacacatttataacgacaatactcg
nadMV f	cgtgttcagcagggtcggcgtgttcggtgtgtcccttattcttcatcgtagtctttttcg
nadMV r	gacccgcagcgcccgaccgaaaggagcgcacgaccattcactgggggcaattc
MMP0737 f	gcatcgattccagacactatttttgggccaattggggaaaaagtatcaaaagatttagct
MMP0737 r	tctgaacctgtaagaatcggaacttaaggatataaaaccttcaaaaatagtatgtgctgg
MMP0737 f2	ccagcacatactatttttgaaggttttatatccttaagttccgattcttacaggttcaga
MMP0737 r2	cccaaaatgattttaataaattaaggaggaaattatgttaaaaataggtctagtaggctg
nadA r	gcaattgaaagaatgctttcaatttaaggatataaaaccttcaaaaatagtatgtgctgg
nadA f2	actatttttgaaggttttatatccttaaattgaaagcattctttcaattgcttttcttgc
nadA r2	ttttcaatttggttattctctcaattatatccataatttcctccttaatttattaaaatc
nadC r	atatcggacttgatttactttaaggatataaaaccttcaaaaatagtatgtgctggattg
nadC f2	ccagcacatactatttttgaaggttttatatccttaaagtaaatcaagtccgatatcaac
nadC r2	cccaaaatgattttaataaattaaggaggaaattatgttgaaaaatcatgctttaaggat
MMP1578 r	agcgaaaaagactacgatgaagaataaggatataaaaccttcaaaaatagtatgtgctgg
MMP1578 f2	tactatttttgaaggttttatatccttattcttcatcgtagtctttttcgcttaatcttc
MMP1578 r2	ccaaaatgattttaataaattaaggaggaaattatgagagcttttctaatcggcaggtgg
MMP1349 r	cacttccattaccaatgccttaaggatataaaaccttcaaaaatagtatgtgctggattg
MMP1349 f2	cacatactatttttgaaggttttatatccttaaggcattggtaatggaagtgatttgtgc
MMP1349 r2	caaaatgattttaataaattaaggaggaaattatgatgagaacaaacgaacagttagaag
nadMV r	gcgaaaaagactacgatgaagaataaggatataaaaccttcaaaaatagtatgtgctgg
nadMV f2	acatactatttttgaaggttttatatccttattcttcatcgtagtctttttcg
nadMV r2	aaaatgattttaataaattaaggaggaaattatggataacctattaaattatagtagtcg

Table A1.3. Primers used in qRT-PCR.

Name	Sequence
atpD QF	gaagagttaacgctcttgaatacgttg
atpD QR	ctttagcgtccattctcgctttaatg
hbd QF	ctatgggaccactcgaattagg
hbd QR	gaatctccggtttcgctgta
MMP0737 QF	gaaggaacggtttttgatgc
MMP0737 QR	tgcatttggatctgcaatga
MMP1349 QF	tcccacttttcaggaggaa
MMP1349 QR	caaactgacgaagacgaact
MMP1578 QF	actggaaattcgctcgttag
MMP1578 QR	gagatgttcccaagcactac
nadA QF	gcagattgcgtcagttctta
nadA QR	ttatcggaacagaatgcgac
nadC QF	ccgatatcaacagaagtggc
nadC QR	agcggcggaataaatgaaaa
pac QF	tctcacccgaccaccag
pac QR	gcggaggtctccaggaa
phaA QF	cacttgcagtacaccaacaaat
phaA QR	ctacatcctgaagcaccgatag
phaB QF	gaaaggacaattcggtcagacta
phaB QR	cggtgtttacggttactcctt
porE QF	gcttgttcgtctcttccaaa
porE QR	cattggatgtggactctgtg
tesB QF	catcagctagaggtttcgttagag
tesB QR	

Table A1.4. gBlock sequences used in the construction of plasmids in Table A1.1.

Name	Sequence	Source
$acsB_{Mt}$	gtttaactttaagaaggagatataccatggatgacggattttgacaaaatcttcgaaggtgcgattccggagggcaaagagc cggtcgcgctgttccgtgaagtttatcacggtgccattacggccacctcctatgcggagattctgctgaatcaagcgattcgta cctatggcccggatcacccggtgggttacccggacacggcatattacctgccggttattcgttgctttagcggcgaagaagtt aagaaactgggcgatctgccgccgattttgaaccgtaagcgtgcacaggtgagcccggtcttgaattttgaaaacgctcgtct ggcaggcgaggcaacctggtacgctgcggagatcattgaggcgctgcgttacctgaaatataaaccggacgagccgttgc tgccgccgccgtggaccggtttcatcggtgacccggttgtgcgtcgtttcggtatcaaaatggtggattggacgattccgggtg aggcgatcatcctgggtcgtgcgaaggacagcaaagcattggcgaagatcgtgaaggagctgatgggcatgggtttcatg ctgttcatttgcgatgaagccgttgagcaactgctggaggagaacgttaaactgggtattgactatattgcgtatccgctgggc aattttacccaaattgtccacgcagcaaactacgcgctgcgtgccggcatgatgtttggtggtgtgaccccgggtgcccgcg aagagcagcgtgactatcaacgtcgtcgtatccgcgcatttgttctgtacctgggcgagcacgatatggtcaaaaccgcagc ggcgttcggtgcaatttttaccggcttcccggttatcaccgaccaaccgctgccggaagataaacagattccggactggttttt tccgttgaagactatgacaagattgtgcagatcgcaatggaaacgcgtggtatcaagctgaccaaaattaagttggatctgc cgattaacttcggcccggcttttgaaggcgaaagcatccgtaaaggcgatatgtacgtggaaatgggtggtaaccgtaccc cggcgtttgaactggttcgtacggtttccgagtctgaaatcaccgacggtaagatcgaggtcatcggcccggacatcgatca	Ch. 2

	aatcccggaaggcagcaagctgccgctgggtattctggttgacatctatggtcgcaaaatgcaagcggatttcgagggtgtc ctggagcgccgtatccacgattttattaactatggcgagggtctgtggcataccggccagcgtaatattaactggctgcgtgtg agcaaagacgcggttgcgaaaggttttcgttttaagaactatggtgagattctggtggctaaaatgaaagaggaattcccgg caattgtcgatcgcgtccaggtgaccatctttaccgacgaagctaaagtcaaggagtacatggaagttgcgcgtgagaaat acaaggaacgtgacgatcgtatgcgtggcctgaccgatgagacggtggacaccttctacagctgcgtcctgtgccaatcctt cgctccgaatcatgtgtgtatcgttaccccggaacgtgtgggtctgtgcggcgcagtgtcctggctggatgcaaaggcaagct acgagatcaaccacgcaggtccgaaccaaccgattccgaaagaaggtgaaattgacccgattaagggtatctggaaga gcgtgaacgattacctgtacacggcaagcaatcgtaacctggaacaggtctgtctgtataccctgatggaaaatccgatgac gagctgcggctgcttcgaggctatcatggcgatcctgccggaatgcaatggtatcatgattaccacccgcgatcacgcaggt atgaccccgagcggtatgacgttcagcaccttggcgggtatgatcggtggtggcacccagacgccgggcttcatgggtattg gccgcacctacattgtgtctaagaaattcatcagcgcggatggtggtatcgctcgcatcgtttggatgccgaaatccctgaag gactttctgcatgacgagtttgttcgtcgtagcgttgaagaaggtctgggtgaggactttattgataagattgctgacgaaacca tcggcacgacggtcgatgaaatcctgccgtacctggaggaaaagggtcatccggcactgaccatggacccgatcatggg cagcggtggtagcgcgtggagccacccgcaatttgagaaaggtgcgggttaagatccgaattcgagctcgg	
	cgatcgctgacgtcggtaccctcgagatgccgcgtttccgtgacctgagccacaactgccgtccgagcgaggcaccgcgt gtcatggagccgaaaaatcgtgaccgcaccgtcgatccggcagtgctggaaatgctggttaaaagcaaggacgacaag gtcattaccgcgttcgaccgtttcgttgctcagcaaccgcaatgcaagattggctacgagggcatctgttgtcgcttctgcatgg caggcccgtgccgtatcaaagcgaccgacggtccgggcagccgcggcatttgtggtgcgtccgcgtggacgattgtggcg cgtaacgtgggcctgatgatcttgacgggcgcggcggcgcactgtgaacacggcaatcatatcgcgcacgctctggtcga gatggctgaaggtaaagcaccggattacagcgtcaaagacgaggcgaaactgaaggaagtttgtcgccgtgttggtatcg aagtggagggtaaaagcgttctggagctggcccaggaagttggtgaaaaggcgctggaggattttcgccgcctgaagggt gaaggtgaagcgacgtggctgatgacgaccatcaacgagggtcgtaaggaaaaatttcgtacgcataatgtcgtcccgttt ggcatccatgcgagcatctccgagctggttaaccaggcgcatatgggtatggataacgacccggtcaatctggtcttcagcg caattcgtgtggcattggcagattacaccggtgagcatattgcgaccgacttcagcgacattctgttcggtaccccgcagccg gttgtgtctgaggcgaacatgggtgtgctggacccggaccaagtgaactttgtcctgcacggccataatccgctgctgtccga aattattgtgcaagcggctcgtgagatggaaggtgaggccaaagcggcgggcgctaaaggtattaacctggttggtatctg ctgcaccggcaatgaggtgctgatgcgccaaggcattccgctggtcacgagctttgcgtctcaggagctggcgatctgtacg ggtgcaatcgatgcaatgtgtggacgtgcagtgtattatgccgagcattagcgcagtggcagaatgttaccacacccgca ttatcaccaccgcagataatgcgaaaattccgggtgcgtaccatatcgattaccagaccgctaccgcgattgaaagcgcga aaacggcaatccgtatggcgattgaggcgttcaaagagcgtaaagagtccaatcgcccggtgtacattccgcagattaaa aatcgcgtggtggctggctggtctctggaggcactgaccaagttgctggcgacccaaaatgcgcaaaacccgatccgtgtt ctgaaccaagcaattttggacggcgagctggcgggcgttgcgctgatctgtggttgcaataatctgaagggctttcaggataa cagccacttgaccgtcatgaaggaactgctgaaaaacaacgtgtttgtggtggccacgggttgttctgcacaagcagcagg caagctgggtttgctggatccggcgaacgttgagacctactgtggcgacggtctgaaaggtttcttgaagcgtctgggcgag ggtgcaaatattgagattggcctgccgccggtttttcacatgggtagctgcgtcgacaatagccgtgcggttgatctgctgatg gcaatggcgaacgacctgggtgtcgatacccccgaaagtgccgttcgtggccagcgctccggaggcaatgagcggtaaag ccgcggcgattggcacgtggtgggtttctctgggcgtgccgacccatgttggcaccatgccgccggttgaaggtagcgacct gatctatagcattctgacgcaaattgccagcgatgtttacggtggctactttatttttgaaatggacccgcaggtggcggcccgc aagattctggatgcgctggagtaccgcacgtggaaactgggtgtgcacaaggaagtcgcggaacgttacgaaaccaaac tgtgtcaaggctactaattaattaacctaggctgctgccaccg	Ch. 2
$acsA_{Mt}$		
	aattttgtttaactttaataaggagatatacatggctcgtcacatcgctgtcgcaggcaaaggtggtacgggcaaaacgacctt tgccgcgctgatgatccgttacttgattgaaggccaaaagggtagcatcctggccgttgacgcggacccgaatgcgaatctg aatgaagccctgggcgtccagattgacacggttattgccgacatcctggacgcaaccaagaacccgaagtctatcccgga gggcatgagcaaagaaatttttgtccaataccagttggcgcaggcactggtcgaaaccaaagatttcgacctgctgacgat gggtcgcccgcagggtccgggctgttattgttatccgaacgacttgctgcgtaaacacctggaaaccctgagcgacaattac gactacatgattatcgacagcgaggcgggcctggagcacatttcccgtcgtatcatccaaaacgtcagcgatctgtttgttatc agcgacgcttccgcgcgtggtattcgtagcgcaggtcgtgtccgtgagctggttcaggagctgcaactgccgatcaataatct gtatctgattgtcacgaagaccaccggcgatattgcgccgctgcaagaggagattgagcgtacgggtatcccgctgacggg tgtcatcccgtacgatgaacaaattgttgattacgatattcatagcaagccgttgtttgacttgccggcgacgagcgttagcgtc caagcggtcaaagcgatcctggcacgttgtcaattctaaagcttgcggccgcataatgcttaag	Ch. 2
$acsF_{Mt}$		
	cgctctacgagtgcgttaagttcagcgacattagtaacagccatatgaatttcctccttaatttattaaaatcattttgggactggt cacctactcgagtgtccactattcccaatgaaatctcattcgtttagaaggtacttaaatcttttgtgttctccgataaatgaagcat gcgaattgcccccagtgaattaaaaatatataaaaaaag	Ch. 3
P_{mcrB}		
	gcccattcctcctccgatgcaaagtgaagctaaacctttttagcgtctcttctttttcatttcgtgaagtaatgtaactaaaattctac atcctgaagcaccgatagggtgtccaattgcaattgccccaccgtttacattaactttgcttgtgtcccaacccatttgttggtgta ctgcaagtgcctgagcagcgaatgcttcgttaatttccattaaatcgagatcttgaggcgtccattcagctcttgataatgctctttt tgaagcaggtacaggtcccattcccataaccttggatcaacacctgcgtttgcgtagcttttgatggttgctaatggtgtgaga cccaattctttagcttttgcagctgacattactacaactgcagcagcgccatcgttaagtcctgatgcgtttgctgcagtaacggt	Ch. 3
phaA		

	tcctgctttgtcgaaagctggttttagtccactcatggaatctaaggttgctccttgtcttacaaattcatctgttttgaatgcaacag ggtcacctttctctgaggaattaatactggaacaatttcttcatcaaatttaccagctttctgtgctgcttctgctttatttttgtgatcct actgcgaattcatcctgagcttctcttgtgattccgtattcttttgctacgttttctgctgtaatacccatgtggtactggttgtaaacat cccataatccatccacaatcattgtgtctaccaatttagcatctcccattctaaatccatctcttgatcctggtaaaacgtgtggag ctgcactcatgttttcttgaccacctgcaacaacaatttctgcatctcctgccataattgcgtttgcagctaacattactgctttgagt cctgagccacataccttgttaatagtcattgctggtaccatagctggcaaaccagctttgatagcagcttgtcttgctgggttttgg cctgatcctgctgtaagaacctgtcccatgattacttctgatacctgttctggttttactccagctctttctaaagctgctttaataact actgcaccgagttctggagcaggtatctttgctaaactacctccaaatttacctactgcagttcttgcagcagaaacgataaca acgtcagtcatatgaatttcctccttaatttattaaaatcattttgggactggtcacctactcgagtgtccactattcccaatgaaat ctcattcgtttagaaggt	
hbd	tctgattgacctctaaagagaccttcctcaatttttttctaaattcaacaatgttaagaggttttttaatgcttgtgacataaaatcactt atttactgtagtcgtaaaatcctttgccagattttctgcccaaccatccagctctaacgtatttttttaataaggtgtgaggtctgtatt ttgaatctccggtttcgctgtacaaaacatccattatagcaagacaaatatctaaaccgatgaagtcacctaattcgagtggtc ccatagggtgattagcacctaatttcattgctttgtcgatgtcttctactgaagcaattccttcggctaagattccaacagcttcatt gatcattggaataaggattctgtttactacgaatccaggagcttctgcaacttcgacagggtcttttccgattgcaattgatgtttct ttaacagcatcgaatgtttcctggcttgttgcaattcctctgataacttctacgagtttcataacaggtgcagggttaaagaagtg catgccaataactttatcaggtcttttagttgcagatgctacttcagtgattgataagcttgaagtgttggatgctaatattgtttcag gtttacagatgttatctaagtctgcaaaaatttgctttttgatgtccattctttctacagctgcttcgattactaagtcacagtcagctg ccatgtttaagtctactgtacctgagattcttgttaaaatctctacttttgtagcttcttcgattttaccttttttaactaattttgagaggttt ttgttgataaaatctaaacctctgtctacgaattcgtctttgatatctcttaagacaacttcaaatcctttagctgcgaatgcttgag ctattcctgaacccatggtaccagctccgataacacaaactttttttcatcgtttattaccttatttttctttcaactgcgagagcaacg cccattcctcctccgatgcaaagt	Ch. 3
tesB	cttttggtaacttcattttaagttcttctgcatggtctttgtaattcaatccagcacatactatttttgaaggttttatatccaaaatttcac ctgaaaaataattcaccaatttaaatttaatttataatttagttgtggtttctcatcactccttcttgtacagttgatgctacaagaact ccatcttgagtgtaaaattcacctctaacgaaacctctagctgatgatgctgaggtgctttctactgaataaagtaaccattcgtt aagattgaaaggtctgtggaaccacattgagtggtcgattgttgcaatctgaattccaggttctaagaaaccaattccgtgag gctgtaaagcaacaggtaaaaagttcaagtctgaagcgtatcctaataagtattggtgtactcttaagtcatcaggtacgcttc cgtttgctctaatccaaacttgtctgtgtggttcagcaacgtgacctttaagtgggttgtggaattcaactggtcttacttctaaagg tctgtcgcagataaatttgtcttttaatacaggtggtaataagtgtgctaagctctgagcaatttgagtttcacttggaagtccatct ggtgctggtgctgaaggcattgtttttttggtgttcaaaaccagcttctggagcttggaatgaagcggtcatgtagaaaataggttt accgttttgtattgcagctactcttcttgctgaaaagctgttaccgtctctcaatgtttccacgtcgtagataattggtttttggagtct ccaggtcttaaaaagtagctgtgaaagctgtgtacgagtctttcttccggaactgtttctttagcagcgtataatgcttggcctact acttgtccaccgaaaacctgtctgagacctaaatcttctgattgacctctaaagagaccttcctcaatttttctaaattcaacaat gttaagagg	Ch. 3
phaB	gctaagtacgattggagcagtgagaaatatgcgggattttcctgccacgcggtatccgaaaaccgacgatcacgcgggtc gactggcatgggagcatcaggatcgtttgcattgaacacgcgcgtagctgctgcggggattttcgccaggcggccggcata ctcggcggtcgctgctgtgaccgcttgcggattccgcactgcgcgaccggctacggaagtaacagctggccccccagccaa tagggtccagtgttgacttcagttttttatgcacgtgatccagcatatgaatttcctccttaatttattaaaatcattttgggactggtc acctactcgagtgtccactattcccaatgaaatctcattcgtttagaaggtacttaaatcttttgtgttctccgataaatgaagcat gcttaacccataaactgtcccccgttaatgtcgatatttgcccctgtaataaagccactttccttatcggcaagaaaggttacaa gcctgcctatctcttctggtgtgccgtacctcttgactggtatggtttctctaatagcctctctcacctgctccgggatcttcataatc atgtcggtagcaatgtatccaggggagacggtgtttactgtgatacctttttgtggcggtctcctgggcaagactcatggtgagg ccgtgcattccggctttggcagctgcgtagttcacctgaccgaactgacccttcctaccatttatgctgctgatgttaataattctg ccataaccctgctccagcatcatctcgataacgctcctgcatgtattgaatactgagttgaggttggtatcgataacctgatgcc actgctcgtatgacattttcttcatggttccatctcttgtgataccggcacagtttacgaggactgaaaccggaccatatttctcct gtatttccctggcaccttcaagacatgcgttgtgatcggagaggtctactcctgacagagcaatattgtcatatcctgctgcctg ctgtgtttcgagccaggttttggccttctctgggttatggtatccagcgacgacaaggtacccggcatcggccaggctatggca aattgatgttccaattccacctgttccgccggtaacccaggccaccggagcctgattggccatatgaatttcctccttaatttatta aaatcattttgggactggtcacctactcgagtgtccactattcccaatgaaatctcattcgtttagaaggtacttaaatcttttgtgtt ctccgataaatgaagcatgc	Ch. 3
phaC	gcatcgattccagacactatttttgggccaattggggaaaaagtatcaaaagatttagctcttgtccactgcccgtctttttcctga aggtctcttgcagtaacgtcgtttaaaattgtaaatccttttatgtaatttttaacatcttctttttttacagatttacacgtttttccaatta cgattgcaagttcagcttcgtaatcgactctttcagatgcttttggtattttgattgcttcttcgttatatattattgaagaagtaggcttt aaaaatataactggctcttttggtaacttcattttaagttcttctgcatggtctttgtaattcaatccagcacatactatttttgaaggttt tatatccaaaatttcacctgaaaattagctaaaaacgtatgtaccaggggcatcctcgagcggagggtgggccgtgctaccc atggccggaggcgctactaattcgccggcacgtttgttagaccacgccgtccaatcttcccaccaactatgttcgtaacgggt cgctgcctcgtcccatacttgtggatcggcaggtaacggcgtgccactctcggcgtctggcgcgccaaccgctttgaaccac acgcgtttacccgggggattcacagcaccggcgacgtgacccccgttggtcagaacataacgtacatcgccccccagaa	Ch. 3

	gatttactgcctgatacgacgaggtccaaggcacgatgtggtcattaatcgcacccacaacatacgtatcacaagcgatgtc gtgaagattcagaggctgcccatccagaacgtacagaccttcagccagttcattgcggccatacaaactgcgcagataatg tgaatgcatctctgcgggcatactcgtagaatcttcattccatgccagaatgtcgaacgctgccggcttttccccttcatccagc gagatacccaatagttaaacacaagatccttggcacgaatcatgtcaaaggaaccggccatctctttgcctgaaaggaagc cctgctggcgcatccggaactctaccagatcaagggtagccgggtcagtcagtaagcccagctcccccacttggctgtagt ccaggagcgtattcagcatcgtaaatgcgctaacacgtttatctccgacggcgaaagcacgtgcggctgccatcgccgcca tcgcacccccccaggcaaatactcagcacttcaatcttcgggctgccggtaatctcttcgactacatccagggcggttgcaatc ccgtccacatagtaatcatccatggtaatatgccgcatactttcatcagggttacggtacgaaatcataaatacggtccgacc atgctgcacagcccattcggcgaggctgcgccccggcgccaggtccagaatgtagtatttgttgatccagggtggagctgc cagaatcggcaccgcgtgcacctgctccgtctggggagcgtattggatgagctcaatcagatcattgcggaatactactttgc ccggcgtcgcggctaagttctcacctaccgtaaacgcgtcagaatccaccttcagaggcaggccaccgcggttcaaaatat catgtgcggcgtagcgagctccacgcagcagcgacgcacctccggtttcaaatgcgcgcgtcaatacacccggattccac aaaaaattggatggtgcgagggcatcgaacatcaggttagcgaactgccgagccttgccatcctgtaacgggtcgccgctg ccagcttcggtcagttcttcaacataggcgcgagtggctaagtacgattggagcagtgagaaatatgcgggattttcctgcca cgcggtatc	
nadM	gcatcgattccagacactatttttgggccaattggggaaaaagtatcaaaagatttagctcttgtccactgcccgtcttttcctga aggtctcttgcagtaacgtcgtttaaaattgtaaatccttttatgtaatttttaacatcttctttttttacagatttacacgtttttccaatta cgattgcaagttcagcttcgtaatcgactctttcagatgcttttggtattttgattgcttcttcgttatatattattgaagaagtaggctttt aaaaatataactggctcttttggtaacttcattttaagttcttctgcatggtctttgtaattcaatccagcacatactatttttgaaggttt tatatccttattcttcatcgtagtctttttcgcttaatcttctaatcctgttgattccatcgatttcttcaattacttttacgacttcttcagga acgagatgttcccaagcactaccttctaacatttttttccttatttttgttcctgaataatctgttctattgtataattctggttttttaacca cataatttctttcagagaataattctctaacgagcgaatttccagtgtaaatggtagtaaatggcggggtcaatgactctacag atgaaacccaaactgcattataatctatatcgataattggaattgcgtaatagtttatatcgtaattttcaagggtttttgtaatcatc atcattctttctcccgcggtgaacgggtctgttaaagtatggctttttgacagcttccaatacccacaattatctcatcaacttccg cagagatttttttaattatttctaaatgccctttgtgaaatggctgccacctgccgattagaaaagctctcat	Ch. 3
nadV	ctttgtgaaatggctgccacctgccgattagaaaagctctcatatgaatttcctccttaatttattaaaatcattttgggactggtca cctactcgagtgtccactattcccaatgaaatctcattcgtttagaaggtacttaaatcttttgtgttctccgataaatgaagcatgc tcatagcgtagtgcgactaactaacaagttttgccgaatttcatcaaagtctgtttggcgtaataacttaccatcttcaaataaca gctctaataaatcatcactaaaatcatccgctgaagttaaaccatcaacgtaagtatcacgagaaagcactttaacacgacc tttttgcgatttttaaaaccatcatcggttttaggatttttgaaaatagctttttcttcgccattaatagtgatagatgtcgctttaagcg caaagcccaacgtatcacgtgtattacgttgataggtttgtgcaccaacgccaaacacaatattacttgaggcaaatcctttgg cttgtaatccttctaagatcttaaacatttttttcataagtcacgccatcaccataaattgccccaatatgtggattgatcactttataa cctttctgattaactgtaccaccaaaaatatcccataaacattcaatgagtcctttacgttcatgctcagtatcagcggttggatc accacaaataatcgcaaaaaagttaccgctatctggacgaatgactaaacgggcattttctggccttgctataatttcttgcttt aataacggcaaattaacggtaatgttatgccaaaagtctgtagtatctgacacaattgacaacatattatgcggaaattttgcc attaaataacgaaatgttgataattcatcgacaccatgtgaactcattactgaatgctcagaagcgggtatagacgtgccaatt agactgcttgaaccataatacgcttcaacaaaagagagtgcaggaatagtgtctgtacctaaaaaagaagttaaatggcc agcacctgaagtttctgcggattctaaagaactcataccacgcattgaaaaatcatgcgattgaaatggcacatgttcttgatt atcacaagtttcattagcaaatttaattaatgctgtccgataagcaaaagcaatcgaggcagaagtcatcggctgccaaagt gatacattaattaatgtttctaaatagttagtaagccaaaagaaatcagaatgcgtattttcaatcgtcatcaccggaactttaat tgccaccgtttttccttcaggaatagctttaatccggataggcaaataacccaactcatgtaattttgcaatgtgttcacccgtatc ctctaactgtaaggttttctcaataaatgcagagtattcagtcacaacatcatgtttatctcgagaaaaaaagttatcattaaaat aatgaattaaatatttaatgataaaggcttgaaaaccaaatgacacaacttgtgttaaataaggcgcttgttcattgctacgag gtgtaaatgtactataaataatttgtgaacattcgggatacattattcgatgagatgttttgtaaaaatcgcataataatgatggta tagcactagcacgactactataatttaataggttatccataatttcctccttaatttattaaaatcattttgggactggtcacctactc gagtgtccactattcccaatgaaatctcattcgtttagaaggtacttaaatcttttgtgttctccgataaatgaagcatgcgaattg cccccagtgaattaaaaatatat	Ch. 3
fdh1	gaattgcccccagtgaattaaaaatatataaaaaaaggaaaccatattaggtttcctttagttttctttttactattttacccaatcta ataataaattaataataaattagtaattaacttagattttagcaggtaagatagggtctctttcaccagcagcttatttttttatcatgtt tgccatacgctttggtcacatattcgccgttcagcagaataatatcctgcgggcgataatcaaatttgccggtaaaaaagctttc cagaatgtttttggtgccttccgcatagcgggtctgcgcatccagggtggtgccgctataatgcggggtcatcgcgttgcccgc gccatatttgttgcgcatatcgcgccacggatgatctttcggcgccggctgcggaaaccacacatcgccgccatagccgcg cagctggccgctttccagcgccgccgccacatcttccgccacgcaaatcgcgccgcgcgcggtgttcaccagccacgcgc ctttttttaaatttgctcagcagttctttgttaatcaggcctttggtgcccgcatgcagcggcgcgttcacggtcacaatatccgcct gcgccaccagttcttcaatgttttccacgcggcgcgcgcccactttttcttccgcttctttcggcagcgcctgataatcataataca gcagttctttcgggttaaacggcagcaggcgttccagcacgcgatagccaatgcggcccgcgccaatggtcgcaatggtttt gccttcaatatcatacgcatctttcgcaatcgccgccacttcccaatcatggttaataatctgttcatgcgccggcacaaagttg cgcaccagcaccagcatggtcatcaccacatgttccgccacgctcaccacgttgctgccggtcacttccagcacgctaatttt tttgccggtctggttaatataatccagatcaatatgatcgctgcccacgcccgccaccaccaccagtttcaggtttttcgctttatc	Ch. 3

caggcgttctttggtaatatacgccggatgaaacggggtggtaataataatatccgcatccggaatatgtttatccagttcgctg
gtttcgccttctttatcgctggtggtaatcagttcatggccctgatctttcagccagttcgcaatgcccagtttgttttcggtgcagcc
atacagtttttcttcatccgccgcatgtttgcccgcatcatacagcaccagcacaattttcatatgttccacctttcataagccaat
attttacattgtttatataaaatatgcctttaaattatttaaacctattaattcccgatccgctcctgggcaccgaactgcgccgcgt
gttcagcagggtcggcgtgttcggtgtgtcccccgcggtgggcctcgggggcgggtgcggggtcggcggggccgccccg
ggtggcttcggtcggagccatggggtcgtgcgctcctttcggtcgggcgctgcgggtcgtggggcgggcgtttcgtcttcacgt
ct

Table A1.5. Gene accession IDs

Protein	Source	Gene accession no.
PyrF	Moorella thermoacetica	WP_025774316.1
AcsA	Moorella thermoacetica	WP_011392716.1
AcsB	Moorella thermoacetica	WP_053104380.1
AcsF	Moorella thermoacetica	AF502245
PhaA	Cupriavadus necator	WP_010810132.1
Hbd	Clostridium acetobutylicum	WP_010965995.1
PhaB	Halomonas bluephagenesis	EGP20949
TesB	Escherichia coli	WP_000075876.1
PhaC1	Cupriavadus necator	WP_011615085.1
NadM	Methanococcus maripaludis	WP_011171522.1
NadV	Haemophilus ducreyi	WP_064081888.1
Fdh1	Candida boidinii	AJ245934

Appendix 2: Supplementary materials for Chapter 2

Table of Contents

Supplementary methods

RNA isolation and sequencing. Cells were grown in 10 mL of modified ATCC 1754 PETC medium with either 60 mM fructose and 10 mM sodium nitrate or 200 kPa H_2/CO_2 for heterotrophic or autotrophic growth, respectively. All measurements were collected in biological triplicates. When the OD_{600} reached mid- to late-log phase (after approximately 48 hours), the cultures were rapidly cooled in an ice-water bath. The cells were collected via centrifugation at 8,000 × g for 30 min at 4°C and rapidly frozen in liquid nitrogen before storage at −80°C.

RNA isolation was performed using a NucleoSpin® RNA isolation kit (Macherey-Nagel, Duren, Germany) according to the manufacturer protocol. Quality of total RNA samples was verified using an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Palo Alto, CA). rRNA was removed using a Ribo-Zero rRNA removal kit for bacteria (Illumina, San Diego, CA) according to the manufacturer protocol and a modified RNeasy MinElute Cleanup Kit (Qiagen, Hilden, Germany) was used to recover mRNA. Sample quality was again verified using an Agilent 2100 Bioanalyzer (Agilent).

Library preparation was performed by the Functional Genomics lab at the University of California, Berkeley, using the TruSeq RNA library prep kit (Illumina) according to the manufacturer's protocols. cDNA was sequenced using Illumina HiSeq 4000 PE100 at the UC Berkeley Vincent J. Coates Genomic Sequencing Laboratory. Gene-level quantification and differential expression analysis was performed using the CLC genomics workbench (Qiagen).

Supplementary results

Table A2.1. Transcriptional response of M. thermoacetica to growth in 60 mM fructose or H_2/CO_2. Comparison of the log_2-fold ratio of the top 100-ranked transcripts from cells grown in the presence of 60 mM fructose with sodium nitrate or H_2/CO_2 (80%, 20%), including p-value for each log_2(fold-change). (A) Upregulated genes. (B) Downregulated genes.

A

Gene ID	Log_2 (fold-change)	p-value
Moth_1884	8.010	0.000
Moth_1883	7.677	0.000
Moth_1888	7.348	0.000
Moth_1885	7.079	0.000
Moth_2266	6.942	0.000
Moth_1887	6.724	0.000
Moth_1635	6.583	0.000
Moth_1413	5.607	0.000
Moth_1411	5.573	0.000
Moth_1410	5.436	0.000
Moth_1409	5.297	0.000
Moth_1412	5.273	0.000
Moth_1886	5.065	0.000
Moth_1408	4.468	0.000
Moth_1889	4.310	0.000
Moth_1197	3.758	0.000
Moth_0109	3.711	0.000
Moth_1336	3.708	0.000
Moth_1190	3.527	0.000
Moth_1415	3.511	0.000
Moth_1191	3.456	0.000
Moth_0831	3.439	0.000
Moth_1198	3.430	0.000
Moth_1196	3.388	0.000
Moth_1203	3.348	0.000
Moth_1269	3.347	0.000
Moth_1202	3.346	0.000
Moth_1195	3.340	0.000
Moth_R0083	3.287	0.208
Moth_1193	3.257	0.000
Moth_1199	3.231	0.000
Moth_1194	3.231	0.000

Moth_1200	3.225	0.000
Moth_1427	3.207	0.000
Moth_1204	3.206	0.000
Moth_1201	3.140	0.000
Moth_1426	3.133	0.000
Moth_1192	3.090	0.000
Moth_1339	3.078	0.000
Moth_0088	2.983	0.000
Moth_1338	2.959	0.000
Moth_1428	2.915	0.000
Moth_1589	2.854	0.000
Moth_R0016	2.839	0.058
Moth_R0069	2.784	0.300
Moth_2050	2.777	0.000
Moth_2049	2.771	0.000
Moth_1337	2.771	0.000
Moth_1340	2.739	0.000
Moth_1719	2.707	0.000
Moth_1341	2.689	0.000
Moth_2051	2.678	0.000
Moth_1240	2.574	0.000
Moth_0396	2.536	0.001
Moth_1429	2.497	0.000
Moth_1416	2.474	0.000
Moth_1750	2.462	0.000
Moth_1430	2.406	0.000
Moth_1713	2.383	0.000
Moth_0089	2.356	0.000
Moth_1514	2.335	0.000
Moth_1513	2.310	0.000
Moth_1718	2.301	0.000
Moth_0698	2.296	0.000
Moth_2048	2.262	0.000
Moth_2053	2.230	0.000
Moth_1835	2.222	0.000
Moth_1712	2.209	0.000
Moth_2045	2.195	0.000
Moth_1515	2.187	0.000
Moth_2290	2.119	0.000
Moth_1239	2.117	0.000

Moth_1601	2.067	0.000
Moth_2052	2.056	0.000
Moth_1720	2.055	0.000
Moth_R0038	2.000	0.479
Moth_R0063	2.000	0.477
Moth_0699	1.975	0.000
Moth_1600	1.956	0.000
Moth_1437	1.936	0.000
Moth_1312	1.924	0.000
Moth_0934	1.914	0.000
Moth_1721	1.901	0.000
Moth_2212	1.888	0.000
Moth_1791	1.885	0.000
Moth_2047	1.878	0.000
Moth_1626	1.877	0.000
Moth_1311	1.872	0.000
Moth_2211	1.859	0.000
Moth_2046	1.842	0.000
Moth_1453	1.834	0.000
Moth_2412	1.826	0.000
Moth_0189	1.821	0.000
Moth_1722	1.820	0.000
Moth_0940	1.807	0.000
Moth_2019	1.803	0.000
Moth_1516	1.802	0.000
Moth_2044	1.779	0.000
Moth_1851	1.768	0.000
Moth_2015	1.756	0.000

B

Gene ID	**Log_2 (fold-change)**	**p-value**
Moth_1898	-6.310	0.000
Moth_0339	-6.187	0.000
Moth_1318	-6.150	0.000
Moth_0137	-6.131	0.000
Moth_1905	-5.991	0.000
Moth_1897	-5.971	0.000
Moth_2413	-5.744	0.000
Moth_0728	-5.628	0.000

Moth_0727	-5.563	0.000
Moth_0305	-5.465	0.000
Moth_0328	-5.408	0.000
Moth_0329	-5.389	0.000
Moth_R0090	-5.330	0.000
Moth_0334	-5.314	0.000
Moth_0333	-5.293	0.000
Moth_0325	-5.233	0.000
Moth_0324	-5.212	0.000
Moth_0331	-5.178	0.000
Moth_0330	-5.178	0.000
Moth_0345	-5.177	0.000
Moth_R0055	-5.149	0.034
Moth_0726	-5.128	0.000
Moth_0306	-5.121	0.000
Moth_0327	-5.103	0.000
Moth_0326	-5.093	0.000
Moth_0340	-5.075	0.000
Moth_0308	-5.054	0.000
Moth_0341	-5.031	0.000
Moth_0347	-5.019	0.000
Moth_0344	-5.008	0.000
Moth_0307	-4.965	0.000
Moth_0346	-4.914	0.000
Moth_0332	-4.908	0.000
Moth_0336	-4.904	0.000
Moth_0312	-4.846	0.000
Moth_0310	-4.821	0.000
Moth_0337	-4.787	0.000
Moth_0316	-4.779	0.000
Moth_0343	-4.772	0.000
Moth_0311	-4.769	0.000
Moth_0279	-4.767	0.000
Moth_0338	-4.729	0.000
Moth_0348	-4.728	0.000
Moth_0335	-4.725	0.000
Moth_1317	-4.695	0.000
Moth_1902	-4.653	0.000
Moth_0302	-4.635	0.000
Moth_0285	-4.536	0.000

Moth_1899	-4.513	0.000
Moth_0420	-4.495	0.000
Moth_0286	-4.489	0.000
Moth_2095	-4.488	0.000
Moth_0304	-4.453	0.000
Moth_0301	-4.447	0.000
Moth_1901	-4.375	0.000
Moth_0352	-4.373	0.000
Moth_0419	-4.364	0.000
Moth_1904	-4.358	0.000
Moth_0342	-4.341	0.000
Moth_0315	-4.335	0.000
Moth_0313	-4.331	0.000
Moth_0280	-4.296	0.000
Moth_0309	-4.262	0.000
Moth_0283	-4.225	0.000
Moth_1900	-4.218	0.000
Moth_0288	-4.217	0.000
Moth_1945	-4.130	0.000
Moth_0303	-4.110	0.000
Moth_1316	-4.088	0.000
Moth_0314	-4.071	0.000
Moth_0359	-4.064	0.000
Moth_0729	-4.020	0.000
Moth_1908	-4.018	0.000
Moth_0457	-4.008	0.000
Moth_0349	-3.975	0.000
Moth_R0066	-3.974	0.107
Moth_R0025	-3.974	0.109
Moth_0360	-3.952	0.000
Moth_2096	-3.943	0.000
Moth_1907	-3.906	0.000
Moth_0358	-3.905	0.000
Moth_0231	-3.873	0.000
Moth_0282	-3.841	0.000
Moth_1906	-3.820	0.000
Moth_0410	-3.817	0.008
Moth_1971	-3.807	0.000
Moth_1972	-3.767	0.000
Moth_0289	-3.746	0.000

Moth_0691	-3.717	0.139
Moth_1890	-3.678	0.000
Moth_0287	-3.650	0.000
Moth_1447	-3.642	0.000
Moth_0015	-3.639	0.000
Moth_0281	-3.597	0.000
Moth_1446	-3.574	0.000
Moth_0284	-3.558	0.000
Moth_0357	-3.539	0.000
Moth_0012	-3.531	0.000
Moth_0320	-3.506	0.000
Moth_1315	-3.501	0.000

Appendix 3: Supplementary materials for Chapter 3

Table of Contents

Supplementary methods

AMP-agarose affinity chromatography. Methanococcus maripaludis S2 was inoculated at 1:100 dilution for 18 h in 50 mL McNA medium with 10 mM $^{15}NH_4Cl$ or 10 mM $^{14}NH_4Cl$ in 250 mL serum bottles sealed with a butyl rubber stoppers and aluminum crimps. The headspace of the bottles was pressurized to 275 kPa with an H_2/CO_2 (80%, 20%) gas mixture. Cells were grown at 37°C with shaking at 250 rpm. When the cultures reached an OD_{600} of 0.60, they were centrifuged at 10,000 × g for 30 min at 4°C. After resuspending the cells in 5 mL of buffer A (50 mM Tris-HCl, 50 mM NaCl, 0.1% (v/v) Triton X-100, pH 7.5) supplemented with 1 mM DTT and one cOmplete EDTA-free protease inhibitor tablet, cells were lysed using three freeze-thaw cycles. After thawing, 10 units of DNase I were added, and the suspension was incubated for 15 min at 37°C. Lysate was centrifuged at 6,000 × g for 30 min at 4°C to separate soluble and insoluble fractions. Soluble lysate was then mixed with either a slurry of AMP-agarose resin (for ^{15}N-labelled samples) or agarose resin (for ^{14}N-labelled samples) and batch binding was performed at 4°C. The slurry was centrifuged at 6,000 × g for 1 minute and supernatant was discarded. The resins were resuspended in buffer B (lysis buffer with 0.4 mM TCEP) and incubated at 4°C again. This process was repeated twice before proteins were eluted with Buffer C (Buffer B with 20 mM NAD^+). Isotopically labelled and non-labelled samples were combined pairwise so that equal weights of heavy (^{15}N) and light (^{14}N) total protein were used for the two conditions being compared.

Sample preparation for $^{14}N/^{15}N$-labeling quantitative proteomics. Methanococcus maripaludis S2 was were inoculated at 1:100 dilution for 18 h in 10 mL McNA medium with 10 mM $^{15}NH_4Cl$ or 10 mM L-alanine in 250 mL serum bottles sealed with a butyl rubber stoppers and aluminum crimps. The headspace of the bottles was pressurized to 275 kPa with an H_2/CO_2 (80%, 20%) gas mixture. Cells were grown at 37°C with shaking at 250 rpm. When the cultures reached an OD_{600} of 0.60, they were centrifuged at 10,000 × g for 30 min at 4°C. After resuspending the cells in 5 mL of buffer A (50 mM Tris-HCl, 50 mM NaCl, 0.1% (v/v) Triton X-100, pH 7.5) supplemented with 1 mM DTT and one cOmplete EDTA-free protease inhibitor tablet, cells were lysed using three freeze-thaw cycles. After thawing, 10 units of DNase I (10 U/µL) were added, and the suspension was incubated for 15 min at 37°C. Lysate was centrifuged at 6,000 × g for 30 min at 4°C to separate soluble and insoluble fractions. The soluble fraction was used for proteomics analysis. Isotopically labelled and non-labelled samples were combined pairwise so that equal weights of heavy (^{15}N) and light (^{14}N) total protein were used for the two conditions being compared.

Proteome extraction and digestion. In preparation for tryptic digest, 100% (w/v) TCA was added to the sample to bring the TCA concentration to 20%, after which the sample was kept at -80°C to precipitate proteins. Samples were then centrifuged at 12,000 × g for 10 min at 4°C. Pellets were washed with a solution of ice cold 0.01 M HCl/90% acetone. The pellets were air dried and then resuspended in 30 µL of 8M urea in PBS. 30 µL of 0.2%/1X ProteaseMax surfactant from Promega (Madison, WI) was added to the solution followed by 40 µL of ammonium bicarbonate, resulting in a 100 µL total volume. 10 µL of 110 mM TCEP was then added, followed by a 30-minute incubation at 60°C. Then, 2.5 µL of 500 mM iodoacetamide was added. Samples were wrapped in foil and left in the dark at room temperature for 30 minutes. 120 µL of PBS was added followed by 1.2 µL of 5x ProteaseMax. Details of trypsin digestion, solution volumes, and 2-D capillary HPLC/tandem mass spectrometry, AKA MudPIT, was described previously [2]. Briefly,

protein was extracted from three biological replicates of either ^{14}N or ^{15}N-labelled proteomes. The six protein extracts were digested with Trypsin Gold (1:50 enzyme:substrate) at 37°C overnight and then acidified with a final concentration of 5% formic acid (v/v). After a final centrifugation step at 13,200 rpm for 30 minutes, the supernatant was transferred to a new tube and stored at -80°C.

2-D capillary HPLC/tandem mass spectrometry analysis. Each pooled proteomics sample was pre-fractionated and analysed by 2-D capillary HPLC/tandem mass spectrometry. Data from three biological replicates from AMP-agarose affinity purifications were pooled, yielding a single light/heavy mixture dataset containing abundance ratios from the 3 replicates. A p-value cut-off of 0.05 was used for all significance. In terms of qualitative peptide identifications, the estimated FDR for the work reported here was 0.01, based on matches with reversed protein sequences in the decoy portion of the database. Peptides from all proteomic experiments were pressure-loaded onto a 250 μm inner diameter fused silica capillary tubing packed with 4 cm of Aqua C_{18} reverse-phase resin (Phenomenex no. 04A-4299), which was previously equilibrated on an Agilent 600 series high-performance liquid chromatography using the gradient from 100% buffer A to 100% buffer B over 10 min, followed by a 5 min wash with 100% buffer B and a 5 min wash with 100% buffer A. The samples were then attached using a MicroTee PEEK 360 μm fitting (Thermo Fisher Scientific no. p-888) to a 13 cm laser pulled column packed with 10 cm Aqua C_{18} reverse-phase resin and 3 cm of strong-cation exchange resin for shotgun proteomics studies. Samples were analyzed using an Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) using a five-step Multidimensional Protein Identification Technology (MudPIT) program, using 0, 25, 50, 80 and 100% salt bumps of 500 mM aqueous ammonium acetate and using a gradient of 5–55% buffer B in buffer A (buffer A: 95:5 water:acetonitrile, 0.1% formic acid; buffer B: 80%:20% acetonitrile:water, 0.1% formic acid). Data were collected in data-dependent acquisition mode with dynamic exclusion enabled (60 s). One full mass spectrometry (MS1) scan (400–1,800 mass-to-charge ratio (m/z)) was followed by 15 MS2 scans of the nth most abundant ions. Heated capillary temperature was set to 200°C and the nanospray voltage was set to 2.75 kV. MS spectra were analyzed by Byonic v3.9-32 and Byologic v3.9.6s software (Protein Metrics, CA).

RNA seq analysis. Cells were grown as described previously in 10 mL of McCas medium, McNA or McNA-Ala. All measurements were collected in biological triplicates. When the OD_{600} reached 0.5-0.6 (after approximately 18 hours), the cultures were rapidly cooled in an ice-water bath. The cells were collected via centrifugation at 8,000 × g for 30 min at 4°C and rapidly frozen in liquid nitrogen before storage at −80°C. The cells were then resuspended in 200 μL of their respective growth medium and lysed using 250 μL of RLT buffer (Qiagen, Valencia, CA). 350 μL of 100% ethanol was added, and the RNA was purified on RNeasy columns (Qiagen). Aliquots of RNA were treated on-column with RNAse-free DNase (Norgen, Thorold, ON, Canada) according to manufacturer's protocol. Quality of total RNA samples was verified using an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Palo Alto, CA). rRNA was removed using a Ribo-Zero rRNA removal kit for bacteria (Illumina, San Diego, CA) according to the manufacturer protocol and a modified RNeasy MinElute Cleanup Kit (Qiagen) was used to recover mRNA. Sample quality was verified using an Agilent 2100 Bioanalyzer (Agilent). Library preparation was performed by the Genome Center at the University of California, Davis, using the TruSeq RNA library prep kit (Illumina) according to the manufacturer's protocols. cDNA was sequenced using Illumina HiSeq 4000 SE50 at the UC Davis Genomic Center. Gene-level quantification and differential expression analysis was performed using kallisto [3] and sleuth [4], respectively.

Anaerobic reagent preparation. To prepare anaerobic reagents and transformation buffer, the reagent was first prepared in standard aerobic laboratory conditions in Pyrex glass beakers. After thoroughly mixing components with a stir bar and magnetic stand, the solution is transferred to an Airfree® Cajon storage flask (Chemglass, Vineland, NJ). 100x resazurin solution (0.01% (w/v)) was added to the reagent and flask was stoppered. The solution was made anaerobic using three freeze-pump-thaw cycles on a Schlenk line, using liquid nitrogen to freeze reagent solutions. Solutions are swiftly transferred to a glass serum bottle in an anaerobic chamber. Reagents that need to be steriled are filtered into a sterile 250 mL serum bottle or 18 × 150 mm Balch tubes using a 0.2 μm syringe filter. The bottle was sealed with a sterile butyl rubber stopper and aluminum crimp. Reagents are stored at +4°C or -80°C.

Media preparation. After combining necessary media components in standard aerobic laboratory conditions, media is transferred to an Airfree® Cajon storage flask (Chemglass, Vineland, NJ). After transferring the media into serum bottles in an anaerobic chamber, 200x reducing agent is added. The media is then autoclaved. After approximately an hour of preparation, the reagent should become colorless. If reagents need to be added to media, it will be thawed and added in the anaerobic chamber before the media is autoclaved.

For 2% (w/v) agar plates, agar was brought into the glovebox separately from the media, in Pyrex bottles without a lid. Securing a kimwipe with a rubber band on the bottle top can be used to cap the bottles for the vacuum cycles without hermetically sealing the contents of the bottle. Agar is added to the media in a 250 mL serum bottle in the anaerobic chamber. The mixture is sealed with a butyl rubber stopper and crimped before autoclaving the media. While autoclaving, any antibiotics that are required for agar plate preparation will be thawed. After autoclaving, the bottle is allowed to cool inside the anaerobic chamber. During cooling, antibiotic solutions are added to the mixture. Before media is used in experiments that require dry biomass measurements, media is filtered again to remove metal sulfides.

A stock of 100x reducing agent was prepared by adding a 0.01% (w/v) resazurin solution (200 μL) to MilliQ-treated water (200 mL) and sparging the solution with N_2 for 20 minutes. Under continuous N_2 flow, 6 grams of L-cysteine was added, followed by 6 grams of sodium sulfide nonahydrate. Sparging with N_2 was continued until the sodium sulfide had dissolved completely and the solution was completely colorless. After preparation, the reducing agent was stored under an N_2 atmosphere in a 250 mL serum bottle and sealed with butyl rubber stoppers and aluminum crimps.

Methanogen strain details and culture conditions. Methanogen growth Methanococcus maripaludis S2 (ATCC BAA-2049) was purchased from the American Type Culture Collection (Manassas, VA). All culture manipulations were performed inside a Vacuum Atmospheres Nexus One glovebox with an atmosphere of 90% nitrogen and 10% hydrogen. Oxygen levels and humidity levels within the box were controlled using a STAK-PAK palladium catalyst and desiccant system in a Coy Laboratory Products unheated fan box. *M. maripaludis* cultures were propagated in 18 × 150 mm Balch tubes with butyl rubber stoppers and aluminum crimp seals using defined medium (McNA) or complex, undefined medium (McCas). Before beginning experiments, *M. maripaludis* is passaged at least once after inoculation from glycerol stock.

In initial production conditions, 10 mL cultures were grown at 37°C without shaking. Under improved production conditions, 10 mL cultures were grown in McNA-Ala or McNA media using 250 mL anaerobic glass bottles at 37°C with shaking at 250 rpm. Bottles were pressurized with

H_2/CO_2 (80%, 20%) to 275 kPa every 12 hours for up to 14 days. Stock cultures were stored at room temperature in the dark after growth and were propagated by diluting 1:100 into fresh McCas, McNA-Ala or McNA medium at least every 2 weeks. The stock cultures were used until changes in growth patterns were observed. Glycerol stocks were kept at -80°C for long-term culture storage.

Methanogen transformation. 10 mL of cultures were grown to late log phase for each transformation. The cells were harvested by centrifugation (2000 × g for 10 min) inside the anaerobic chamber. The media was removed, cells were washed with 5 mL of transformation buffer. The cells were collected by centrifugation again. Supernatant was then discarded, and the cell pellet was resuspended in 1 mL transformation buffer per 10 mL of initial culture volume. In each transformation, 300 µL transformation buffer was aliquoted into a 6 mL serum vial. Vials were then closed with a butyl rubber stopper and removed from the anaerobic chamber. Using aseptic technique, 2-4 µg of plasmid and 30 µL DOTAP transfection agent (Roche, Basel, Switzerland). The vials were again stoppered and sealed, after which they were degassed by purging with argon for 5 min before the vials were returned to the anaerobic chamber. Then, 1 mL of the cell suspension was added to each vial. The vials were incubated at 37°C for 1 h. After the incubation, the suspension was transferred to 10 mL of fresh medium in a Balch tube. The cells were recovered at 37°C for at least 4 h. After recovery, the cells were harvested by centrifugation and resuspended in 200 µL modified McCas medium. They were then spread on the surface of an anaerobic bottle agar plate. Plates were incubated at room temperature with the surface of the plate horizontal for at least 30 minutes to allow cell suspension to leach into the agar. After incubation, bottles were stood upright. Headspace of the bottle plates was purged with a H_2/CO_2 (80%, 20%) mixture for 5 min before pressurizing to 275 kPa. Plates were incubated at 37°C for 3-4 days at which point colonies were picked into growth medium inside of the anaerobic chamber. After cultures reached saturation, presence of the desired knock-in was confirmed using colony PCR. Cultures were propagated in selective media containing puromycin and/or neomycin and stored as glycerol stocks in 30% glycerol (v/v) at -80°C.

Supplementary results

Table A3.1. Transcriptional response of M. thermoacetica to growth in 60 mM fructose or H_2/CO_2. Comparison of the log_2(fold-change) ratio of the top 100-ranked transcripts from cells grown in the presence of McNA-Ala, McNA, or McCas media, including q-value for each log_2(fold-change). (A) Upregulated genes. (B) Downregulated genes.

A

		McNA-Ala vs. McCas		McCas vs. McNA		McNA vs. McNA-Ala	
Gene ID	Protein names	Log_2 (fold-change)	q-value	Log_2 (fold-change)	q-value	Log_2 (fold-change)	q-value
MMP0858	Nitrogenase iron-molybdenum cofactor biosynthesis protein NifE	4.33	0.00	-0.73	0.18	0.01	-3.47
MMP0383	S-layer protein	4.12	0.00	-4.25	0.00	0.86	0.26
MMP0856	Nitrogenase molybdenum-iron protein	4.12	0.00	-0.35	0.53	0.02	-3.63
MMP1347	Archaeal histone B	4.02	0.00	-4.76	0.00	0.17	0.87
MMP0853	Nitrogenase iron	3.87	0.01	-0.26	0.74	0.09	-3.48

MMP1213	Uncharacterized protein	3.84	0.00	-2.12	0.00	0.00	-1.58
MMP0855	Nitrogen fixation nifHD region glnB-like protein 2	3.83	0.02	-0.43	0.55	0.15	-3.27
MMP0601	Uncharacterized protein	3.82	0.00	-4.51	0.00	0.35	0.83
MMP0857	Nitrogenase molybdenum-iron protein beta chain	3.78	0.00	-0.45	0.48	0.05	-3.19
MMP0860	Nitrogen fixation-related protein	3.67	0.01	-0.23	0.34	0.03	-3.31
MMP0854	Nitrogen fixation nifHD region glnB-like protein 1	3.64	0.01	-0.27	0.70	0.10	-3.24
MMP0058	5,10-methylenetetrahydromethanopterin reductase	3.57	0.00	-2.92	0.00	0.75	-0.51
MMP0859	Iron-molybdenum cofactor biosynthesis protein, subunit N	3.54	0.00	-0.21	0.64	0.01	-3.20
MMP1559	Methyl-coenzyme M reductase subunit alpha	3.40	0.00	-3.13	0.00	0.92	-0.13
MMP1685	UPF0333 protein MMP1685	3.39	0.00	-2.30	0.00	0.21	-0.95
MMP1695	F_{420}-non-reducing hydrogenase subunit	3.33	0.00	-2.24	0.00	0.38	-0.96
MMP0065	Ammonium transporter	3.32	0.00	-1.53	0.00	0.10	-1.66
MMP1558	Methyl-coenzyme M reductase subunit gamma	3.28	0.00	-2.85	0.00	0.78	-0.30
MMP1407	Ribonuclease P protein component 1	3.27	0.00	-2.13	0.00	0.03	-1.01
MMP1212	Uncharacterized protein	3.27	0.00	-1.52	0.00	0.03	-1.61
MMP1302	Uncharacterized protein	3.23	0.00	-4.14	0.00	0.15	1.04
MMP0457	Probable ATP dependent RNA helicase	3.22	0.00	-2.41	0.00	0.64	-0.68
MMP0259	50S ribosomal protein L10	3.18	0.00	-3.52	0.00	0.64	0.48
MMP0260	50S ribosomal protein L1	3.17	0.00	-3.54	0.00	0.70	0.51
MMP1414	30S ribosomal protein S8	3.16	0.00	-2.38	0.00	0.36	-0.64
MMP1697	CoB-CoM heterodisulfide reductase iron-sulfur subunit A	3.16	0.00	-1.97	0.00	0.32	-1.05
MMP0164	Sirohydrochlorin nickel chelatase	3.15	0.00	-2.72	0.00	0.73	-0.30
MMP1413	30S ribosomal protein S14 type Z	3.11	0.00	1.50	0.02	0.00	-4.50
MMP1016	Uncharacterized protein	3.11	0.00	-3.84	0.00	0.34	0.87
MMP0127	5,10-methenyltetrahydromethanopterin hydrogenase	3.09	0.02	-3.52	0.00	0.80	0.57
MMP1158	Uncharacterized protein	3.08	0.00	-2.97	0.00	0.98	0.02
MMP1415	50S ribosomal protein L6	3.07	0.00	-2.95	0.00	0.99	0.01
MMP1609	H4MPT formyltransferase	3.07	0.00	-2.43	0.00	0.50	-0.50
MMP1694	F_{420}-non-reducing hydrogenase subunit alpha	3.07	0.00	-2.01	0.00	0.40	-0.92
MMP1412	50S ribosomal protein L5	3.05	0.00	-2.63	0.00	0.75	-0.29
MMP1555	Methyl-CoM reductase subunit beta	3.04	0.00	-3.09	0.00	0.88	0.18
MMP1418	50S ribosomal protein L18	3.04	0.00	-2.71	0.00	0.83	-0.20
MMP1406	Protein translation factor SUI1 homolog	3.04	0.00	-2.32	0.00	0.68	-0.59
MMP0888	Putative cobalt transport protein	3.04	0.00	-2.97	0.00	0.95	0.07
MMP1692	Polyferredoxin	3.03	0.00	-1.86	0.00	0.22	-1.03

MMP1370	Elongation factor 1-alpha	3.02	0.00	-3.53	0.00	0.53	0.64
MMP0684	Heat shock protein Hsp20	2.99	0.00	-4.80	0.00	0.00	1.95
MMP1556	Methyl-coenzyme M reductase I, protein D	2.98	0.00	-3.17	0.00	0.75	0.32
MMP1696	F_{420}-non-reducing hydrogenase Vhu	2.98	0.02	-2.05	0.00	0.65	-0.80
MMP0695	Proteasome subunit beta	2.98	0.00	-3.75	0.00	0.18	0.90
MMP1482	Cobalt transport protein CbiN	2.95	0.00	-3.09	0.00	0.80	0.27
MMP1299	Carbonic anhydrase	2.94	0.00	-2.80	0.01	0.99	-0.01
MMP0889	Cobalamin (Vitamin B12) biosynthesis CbiM	2.92	0.00	-2.64	0.00	0.92	-0.15
MMP0064	Nitrogen regulatory protein P-II	2.92	0.02	-1.00	0.00	0.27	-1.79
MMP1691	Tungsten-containing formylmethanofuran dehydrogenase 2 subunit B	2.91	0.01	-1.97	0.00	0.59	-0.81
MMP1544	50S ribosomal protein L4	2.85	0.00	-2.82	0.00	0.92	0.10
MMP1329	Ferredoxin	2.85	0.00	-2.18	0.00	0.73	-0.54
MMP0258	50S ribosomal protein L12	2.85	0.02	-3.81	0.00	0.49	1.09
MMP1421	50S ribosomal protein L15	2.85	0.00	-3.33	0.00	0.48	0.62
MMP1321	30S ribosomal protein S11	2.84	0.01	-4.09	0.00	0.27	1.39
MMP1405	50S ribosomal protein L29	2.84	0.00	-2.33	0.00	0.52	-0.38
MMP1419	30S ribosomal protein S5	2.83	0.00	-2.89	0.00	0.87	0.20
MMP1420	50S ribosomal protein L30	2.82	0.00	-2.84	0.00	0.86	0.15
MMP1408	30S ribosomal protein S17	2.80	0.00	-2.43	0.00	0.86	-0.23
MMP1709	50S ribosomal protein L44e	2.78	0.04	-3.83	0.00	0.51	1.18
MMP1566	Tetrahydromethanopterin S-methyltransferase subunit G	2.77	0.03	-2.65	0.00	0.99	0.01
MMP1417	50S ribosomal protein L19e	2.77	0.00	-2.13	0.00	0.66	-0.51
MMP0386	Archaeal histone A	2.73	0.00	-2.25	0.00	0.67	-0.35
MMP1411	30S ribosomal protein S4e	2.71	0.00	-2.72	0.00	0.88	0.14
MMP0575	Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase subunit C	2.70	0.00	-3.10	0.00	0.57	0.53
MMP1416	50S ribosomal protein L32e	2.70	0.00	-2.70	0.00	0.94	0.14
MMP1211	UPF0219 protein	2.69	0.00	-1.42	0.00	0.04	-1.14
MMP0585	Universal stress protein (Usp)	2.66	0.00	-3.47	0.00	0.17	0.94
MMP1323	50S ribosomal protein L18e	2.65	0.00	-3.85	0.00	0.01	1.33
MMP0144	5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase	2.65	0.06	-3.60	0.00	0.61	1.09
MMP1155	Ferredoxin reductase subunit B1	2.64	0.00	-1.93	0.00	0.43	-0.57
MMP1384	Coenzyme F_{420}-reducing hydrogenase subunit gamma	2.64	0.00	-1.29	0.00	0.05	-1.21
MMP0298	50S ribosomal protein L15e	2.63	0.00	-3.86	0.00	0.04	1.36
MMP0176	CDC48 cell division cycle protein family member	2.62	0.00	-3.76	0.00	0.14	1.27
MMP1546	50S ribosomal protein L2	2.62	0.00	-2.81	0.00	0.81	0.32
MMP0692	Uncharacterized protein	2.62	0.00	-2.93	0.00	0.54	0.44
MMP1382	Coenzyme F_{420}-reducing hydrogenase subunit alpha	2.62	0.00	0.35	0.30	0.00	-2.83

MMP0166	MatE family member	2.60	0.00	-2.10	0.00	0.68	-0.37
MMP1100	Putative transcriptional regulator	2.60	0.00	-2.97	0.00	0.46	0.51
MMP1693	F_{420}-non-reducing hydrogenase Vhu subunit U	2.58	0.08	5.87	0.00	0.00	-8.34
MMP0961	Uncharacterized protein	2.58	0.00	-3.51	0.00	0.17	1.06
MMP0624	Putative_PNPOx domain-containing protein	2.56	0.00	-2.41	0.00	0.98	-0.01
MMP1369	Elongation factor 2 (EF-2)	2.55	0.00	-2.51	0.00	0.94	0.09
MMP1383	Coenzyme F_{420}-reducing hydrogenase delta subunit	2.54	0.01	-1.24	0.00	0.38	-1.16
MMP1543	50S ribosomal protein L3	2.54	0.00	-2.58	0.00	0.87	0.18
MMP0235	Uncharacterized protein	2.54	0.00	-2.37	0.00	0.97	-0.03
MMP0157	DNA-binding protein	2.53	0.00	-2.50	0.00	0.92	0.11
MMP1157	Desulfoferrodoxin, ferrous iron-binding site	2.52	0.00	-2.76	0.00	0.50	0.37
MMP1512	Alanine racemase	2.51	0.00	-2.43	0.00	0.97	0.05
MMP0686	2-amino-3,7-dideoxy-D-threo-hept-6-ulosonate synthase	2.51	0.00	-2.57	0.00	0.89	0.20
MMP1206	Glutamine synthetase	2.51	0.00	-2.83	0.00	0.62	0.46
MMP0916	PRC domain-containing protein	2.50	0.00	-3.17	0.00	0.09	0.80
MMP1401	Elongation factor 1-beta	2.49	0.00	-2.10	0.00	0.67	-0.26
MMP1320	30S ribosomal protein S4	2.48	0.00	-3.39	0.00	0.15	1.04
MMP0626	Cytidylate kinase	2.48	0.07	-3.09	0.00	0.70	0.75
MMP1224	ABC-type amino acid transport/signal transduction systems periplasmic component-related	2.47	0.00	-2.79	0.00	0.50	0.45
MMP1319	30S ribosomal protein S13	2.45	0.00	-3.61	0.00	0.05	1.30
MMP1576	Uncharacterized protein	2.45	0.00	5.50	0.00	0.00	-7.82
MMP1325	30S ribosomal protein S9	2.44	0.03	-3.93	0.00	0.23	1.62

B

		McNA-Ala vs. McCas		**McCas vs. McNA**		**McNA vs. McNA-Ala**	
Gene ID	**Protein names**	**Log_2 (fold-change)**	**q-value**	**Log_2 (fold-change)**	**q-value**	**Log_2 (fold-change)**	**q-value**
MMP0145	Hypoxanthine/guanine phosphoribosyltransferase	-3.55	0.09	6.07	0.00	0.00	0.00
MMP0424	Uncharacterized protein	-2.83	0.01	4.01	0.00	0.00	0.00
MMP0594	Conserved hypothetical archaeal protein	-2.76	0.09	2.94	0.00	0.95	0.19
MMP1539	DUF3343 domain-containing protein	-2.74	0.13	0.58	0.00	0.35	2.53
MMP0892	Acetyltransferase related protein	-2.73	0.09	2.57	0.00	0.87	0.53
MMP1125	Uncharacterized protein	-2.70	0.07	1.73	0.00	0.59	1.35
MMP0752	HTH_34 domain-containing protein	-2.67	0.08	2.07	0.00	0.73	0.97
MMP1572	Uncharacterized protein	-2.52	0.07	2.94	0.00	0.98	-0.04

MMP1269	YCII domain-containing protein	-2.42	0.10	2.10	0.00	0.81	0.69
MMP0016	Uncharacterized protein	-2.40	0.21	0.48	0.01	0.43	2.29
MMP0726	Uncharacterized protein	-2.38	0.08	1.78	0.00	0.70	0.97
MMP0828	Uncharacterized protein	-2.28	0.10	2.04	0.00	0.83	0.62
MMP0745	Uncharacterized protein	-2.21	0.15	0.00	0.00	0.00	0.00
MMP0389	Ferredoxin	-2.17	0.05	2.86	0.00	0.93	-0.31
MMP0746	Uncharacterized protein	-2.14	0.08	4.18	0.00	0.00	0.00
MMP0987	Probable ribosomal RNA small subunit methyltransferase A	-2.13	0.34	0.15	0.41	0.48	2.35
MMP0790	Uncharacterized protein	-1.95	0.02	2.89	0.00	0.63	-0.79
MMP0978	Uncharacterized protein	-1.92	0.10	1.41	0.00	0.69	0.89
MMP0816	Bacterial seryl-tRNA synthetase related	-1.85	0.18	1.40	0.00	0.75	0.82
MMP0756	Uncharacterized protein	-1.82	0.00	1.89	0.00	0.94	0.07
MMP0757	Uncharacterized protein	-1.76	0.03	2.80	0.00	0.55	-0.89
MMP1378	Uncharacterized protein	-1.76	0.12	2.96	0.00	0.00	0.00
MMP0747	Uncharacterized protein	-1.74	0.01	5.26	0.00	0.04	-3.38
MMP0365	Uncharacterized protein	-1.73	0.00	3.34	0.00	0.03	-1.48
MMP0195	Uncharacterized protein	-1.73	0.00	4.57	0.00	0.01	-2.70
MMP0367	Uncharacterized protein	-1.72	0.07	3.71	0.00	0.35	-1.83
MMP1357	ThiamineS	-1.70	0.22	0.55	0.30	0.53	1.52
MMP0740	Possible purine NTPase	-1.69	0.00	2.18	0.00	0.62	-0.36
MMP0248	DNA-directed RNA polymerase related protein	-1.68	0.24	0.00	0.00	0.00	0.00
MMP0881	Uncharacterized protein	-1.68	0.02	3.23	0.00	0.29	-1.41
MMP0032	ArsR family member	-1.67	0.00	2.08	0.00	0.83	-0.26
MMP1339	SAM (And some other nucleotide) binding motif	-1.65	0.00	2.55	0.00	0.34	-0.76
MMP1037	Putative ribosomal RNA large subunit methyltransferase H	-1.64	0.01	1.92	0.00	0.93	-0.13
MMP0484	Possible sodium/proton antiporter	-1.64	0.00	2.54	0.00	0.48	-0.76
MMP0199	Conserved hypothetical archaeal protein	-1.64	0.00	2.84	0.00	0.20	-1.06
MMP0837	RNase HI related protein	-1.63	0.00	2.60	0.00	0.31	-0.83
MMP0741	Uncharacterized protein	-1.60	0.04	3.65	0.00	0.25	-1.90
MMP0362	DUF58 domain-containing protein	-1.59	0.00	2.62	0.00	0.10	-0.88
MMP0364	DUF4350 domain-containing protein	-1.59	0.00	2.91	0.00	0.02	-1.18
MMP0673	TPR repeat	-1.59	0.00	2.87	0.00	0.07	-1.14
MMP1078	Conserved Hypothetical Archael Protein	-1.59	0.03	3.08	0.00	0.30	-1.35
MMP1055	UPF0235 protein MMP1055	-1.59	0.06	2.90	0.00	0.43	-1.16
MMP1338	Uncharacterized protein	-1.55	0.13	4.82	0.00	0.00	0.00
MMP0485	CAAX prenyl protease 2	-1.55	0.00	2.50	0.00	0.22	-0.81
MMP0307	Conserved hypothetical archaeal protein	-1.55	0.00	2.21	0.00	0.57	-0.52
MMP0771	Uncharacterized protein	-1.54	0.04	2.99	0.00	0.35	-1.31

MMP0047	B12-binding domain-containing protein	-1.53	0.00	2.29	0.00	0.23	-0.62
MMP0436	Putative SSU ribosomal protein S6 modification protein	-1.53	0.00	2.21	0.00	0.54	-0.54
MMP0750	Uncharacterized protein	-1.53	0.09	0.00	0.00	0.00	0.00
MMP0461	Uncharacterized protein	-1.52	0.01	3.82	0.00	0.05	-2.16
MMP1276	Uncharacterized protein	-1.52	0.00	2.50	0.00	0.38	-0.84
MMP0022	Uncharacterized protein	-1.51	0.37	-0.45	0.08	0.35	2.32
MMP0299	Uncharacterized protein	-1.50	0.01	2.75	0.00	0.25	-1.11
MMP0778	Uncharacterized protein	-1.50	0.00	5.24	0.00	0.00	-3.60
MMP0513	Molybdopterin biosynthesis moeA protein	-1.45	0.00	2.50	0.00	0.08	-0.91
MMP0473	Uncharacterized protein	-1.45	0.00	3.38	0.00	0.00	-1.79
MMP1650	Probable amino acid ABC transporter, periplasmic component	-1.44	0.01	3.11	0.00	0.09	-1.52
MMP0842	Uncharacterized protein	-1.44	0.12	2.82	0.00	0.48	-1.21
MMP0193	Uncharacterized protein	-1.44	0.12	3.17	0.00	0.34	-1.58
MMP0515	Molybdenum ABC transporter, permease protein	-1.43	0.02	3.08	0.00	0.08	-1.50
MMP1110	Uncharacterized protein	-1.43	0.00	3.06	0.00	0.00	-1.49
MMP0217	Transcriptional repressor related protein	-1.43	0.15	3.76	0.00	0.31	-2.18
MMP1534	Uncharacterized protein	-1.42	0.50	-0.78	0.00	0.40	2.57
MMP0196	ABC-type iron(III) transport system, periplasmic binding protein	-1.41	0.00	3.54	0.00	0.00	-1.99
MMP1677	Digeranylgeranylglyceryl phosphate synthase	-1.41	0.00	2.81	0.00	0.01	-1.26
MMP0769	Conserved hypothetical archaeal protein	-1.40	0.00	3.35	0.00	0.00	-1.81
MMP0832	2Fe-2S ferredoxin-type domain-containing protein	-1.38	0.03	1.45	0.00	0.95	0.07
MMP1179	SAM (And some other nucleotide) binding motif:Generic methyltransferase	-1.38	0.01	2.38	0.00	0.39	-0.85
MMP1639	Acyl-CoA dehydrogenase	-1.38	0.01	2.94	0.00	0.04	-1.42
MMP0435	RLAN domain-containing protein	-1.38	0.00	3.11	0.00	0.02	-1.59
MMP0197	ABC-type iron(III) transport system, permease component	-1.38	0.00	4.55	0.00	0.00	-3.03
MMP0376	Uncharacterized protein	-1.37	0.00	1.71	0.00	0.80	-0.20
MMP1231	Putative HAD family hydrolase, a	-1.37	0.01	1.44	0.00	0.95	0.07
MMP0754	Uncharacterized protein	-1.37	0.00	2.43	0.00	0.05	-0.92
MMP1631	PA-phosphatase related phosphoesterase	-1.37	0.02	2.16	0.00	0.54	-0.65
MMP0652	Uncharacterized protein	-1.36	0.02	1.36	0.00	0.89	0.15
MMP0957	DsbD domain-containing protein	-1.36	0.01	2.96	0.00	0.05	-1.45
MMP1284	DEAD/DEAH box helicase:Helicase, C-terminal	-1.36	0.00	2.24	0.00	0.06	-0.74
MMP1203	Cobalt-precorrin-5B C(1)-methyltransferase	-1.35	0.03	1.41	0.00	0.94	0.09
MMP1057	Uncharacterized protein	-1.35	0.00	2.01	0.00	0.42	-0.52

MMP0200	Molybdenum containing formylmethanofuran dehydrogenase, subunit E	-1.35	0.00	2.80	0.00	0.02	-1.31
MMP0843	Uncharacterized protein	-1.34	0.18	2.81	0.00	0.49	-1.31
MMP0829	Coenzyme B12-binding:Cobalamin-dependent methionine synthase, B12-binding	-1.34	0.00	0.81	0.00	0.15	0.67
MMP1423	Putative glycosyl transferase, family 4:Aldehyde dehydrogenase	-1.33	0.00	2.66	0.00	0.01	-1.19
MMP0517	Lactamase_B domain-containing protein	-1.33	0.01	1.96	0.00	0.55	-0.49
MMP1526	Ribonuclease III family:Double-stranded RNA binding (DsRBD) domain	-1.32	0.02	2.60	0.00	0.09	-1.14
MMP0292	Uncharacterized protein	-1.32	0.30	0.69	0.25	0.63	0.99
MMP0329	Uncharacterized protein	-1.31	0.54	0.98	0.00	0.86	0.70
MMP0834	Uroporphyrinogen decarboxylase	-1.31	0.02	1.37	0.00	0.94	0.07
MMP0508	Molybdenum containing formylmethanofuran dehydrogenase, subunit E	-1.31	0.03	1.69	0.00	0.82	-0.23
MMP1079	Glycosyl transferase, family 2	-1.31	0.04	0.32	0.08	0.20	1.13
MMP0896	Polysaccharide biosynthesis protein	-1.30	0.00	2.47	0.00	0.09	-1.03
MMP0288	Conserved hypothetical archeal protein	-1.30	0.00	3.26	0.00	0.00	-1.83
MMP0831	Uroporphyrinogen decarboxylase	-1.29	0.02	0.86	0.00	0.48	0.57
MMP1529	Conserved hypothetical archaeal protein	-1.28	0.11	2.58	0.00	0.39	-1.16
MMP0804	Cyclophil_like-2 domain-containing protein	-1.28	0.02	2.47	0.00	0.28	-1.05
MMP0483	Uncharacterized protein	-1.28	0.17	1.18	0.00	0.89	0.25
MMP0354	Putative oligosaccharide transporter	-1.27	0.00	1.62	0.00	0.73	-0.21
MMP0833	CobW C-terminal domain-containing protein	-1.27	0.01	1.73	0.00	0.76	-0.31
MMP1341	DNA double-strand break repair Rad50 ATPase	-1.27	0.00	2.34	0.00	0.03	-0.94

Table A3.2. List of proteins from intracellular protein samples at 48 h by MUDPIT analysis. Comparison of the top 100 extracted light/heavy ion count ratios for *M. maripaludis* grown in the presence of McNA-Ala or McNA media. (A) upregulated genes. (B) downregulated genes.

A

Gene ID	Light Ion XIC (10^5)	Heavy Ion XIC (10^5)	Ratio
MmarC6_0825	31.627	0.018	1785.4132
MMP0961	11.046	0.013	825.492534
GYY_04835	0.916	0.006	141.278396
MMJJ_14480	8.53	0.132	64.4695172
MMJJ_06260	1.04	0.017	63.4154699
MMP0874	35.686	0.699	51.049868
MMP1513	1006.277	23.041	43.6726178

WP_048064121	0.897	0.023	38.5964171
MMJJ_11620	44.202	1.17	37.6279257
MmarC7_0908	32.184	0.888	36.2393798
MMJJ_18200	0.508	0.015	35.109197
MMJJ_09400	7.84	0.257	30.4603399
MMJJ_07450	12.378	0.418	29.6390577
MmarC7_1637	2.47	0.09	27.49232
MMJJ_15250	1.38	0.051	27.0112728
MMP0181	2.97	0.117	25.3545386
MMJJ_17140	0.957	0.043	22.4477337
GYY_01685	2.39	0.107	22.2786067
MMJJ_17310	2.21	0.1	21.9572971
MmarC5_1739	11.423	0.53	21.5328805
MMP0451	0.673	0.035	19.1413126
MMJJ_15710	1.56	0.082	19.1102697
MMP0785	2.63	0.149	17.6331176
MMP1071	2.49	0.142	17.4835331
MMJJ_14900	0.257	0.016	16.3847235
MMP1668	101.243	7.46	13.5691667
MMJJ_12670	2.05	0.156	13.1104159
MMP1404	5.28	0.411	12.8488913
MMJJ_06230	6.97	0.593	11.7606423
MMJJ_10550	2.65	0.235	11.2748583
MMJJ_01770	36.728	3.26	11.2712477
MMP0001	1.18	0.111	10.5656886
MMJJ_03020	0.307	0.03	10.273163
MMJJ_15210	14.192	1.47	9.68060636
MMP0064	7.26	0.751	9.66543844
MMJJ_11100	0.374	0.039	9.59159397
MMP1492	0.271	0.029	9.39573339
MMP1248	12.066	1.3	9.26664922
GYY_05995	6.59	0.783	8.40995663
MMJJ_11000	2.68	0.334	8.00599172
MmarC6_1494	1.43	0.181	7.87502133
MMJJ_02610	1.67	0.218	7.62499944
MmarC7_0857	1.36	0.179	7.60752509
MMJJ_17320	0.031	0.004	7.4005891
MMJJ_12820	4.43	0.6	7.38178694
MMP0457	1.7	0.239	7.09214688
MMJJ_02810	4.09	0.586	6.98001351

MMJJ_16270	1.14	0.174	6.56878486
MMP0736	1.69	0.264	6.40512678
MMJJ_14740	1.68	0.275	6.12714181
MMJJ_10320	13.28	2.26	5.88474014
MMJJ_17250	1.53	0.26	5.88025377
MmarC7_1360	1.74	0.3	5.81868618
MMJJ_03580	11.61	2.02	5.75284734
MMJJ_08060	3.43	0.614	5.58858787
MMP0447	8.43	1.52	5.55545485
GYY_06190	1.08	0.2	5.39951515
WP_048064123	6.67	1.28	5.20214917
MMP0625	6.83	1.32	5.16319929
MMJJ_11650	0.8	0.156	5.12492142
GYY_07485	4.34	0.868	4.99758796
MMP1635	13.545	2.86	4.74364341
MMP1259	7.23	1.56	4.62436522
MMJJ_14350	5.68	1.23	4.6071455
MMJJ_03180	1.27	0.285	4.471837
GYY_05105	17.121	3.86	4.43119501
MMP0975	5.96	1.42	4.2074179
MMJJ_18090	15.063	3.64	4.1389137
MMJJ_02380	0.355	0.088	4.05339534
MMP0341	0.265	0.066	4.00126085
MMJJ_11130	66.846	16.971	3.93885081
MMP1671	33.019	8.39	3.93472311
GYY_02040	0.312	0.08	3.90463918
MMP1557	0.513	0.134	3.84386444
MMJJ_12790	0.257	0.067	3.84347839
MMP0092	1.2	0.32	3.74268287
MMJJ_15090	2.03	0.543	3.73168747
MMP1330	0.675	0.181	3.73168747
GYY_07160	1.23	0.333	3.70058385
MMJJ_07730	3	0.833	3.60296907
MMP0155	2.72	0.774	3.51321174
MmarC5_1205	0.575	0.164	3.51174778
WP_048064099	0.196	0.056	3.49235541
MMP1053	115.332	33.268	3.46680819
MMP0284	1.94	0.56	3.45343307
MMJJ_07770	2.92	0.906	3.21913721
MMJJ_06300	26.539	8.27	3.20853587

MMJJ_02390	2.54	0.793	3.20256339
GYY_00635	6.35	2.01	3.1650047
MMJJ_16600	1.94	0.624	3.10364366
GYY_01190	0.938	0.31	3.02421327
MMJJ_07410	68.451	22.662	3.02047216
GYY_08185	0.813	0.275	2.95339421
MMP1369	161.173	54.579	2.95304025
MMP1714	19.605	6.68	2.93614051
MMJJ_17720	0.148	0.051	2.93085003
MMJJ_08090	27.106	9.34	2.90269797
MMJJ_11610	45.565	16.07	2.8354548
MMJJ_16340	38.896	13.838	2.81081978
MMP1178	5.48	1.95	2.80925318

B

Gene ID	Light Ion XIC (10^5)	Heavy Ion XIC (10^5)	XIC Ratio
GYY_02865	0.013	1.39	0.0010249
MmarC6_0571	0.005	0.434	0.00109798
MmarC5_0652	0.005	0.381	0.00110994
MMJJ_10300	0.049	3.53	0.00160973
MMP0603	0.007	0.422	0.00347609
MMJJ_11890	0.007	0.399	0.00364901
MmarC5_0187	0.063	3	0.00449799
MMJJ_07860	0.034	1.41	0.006133
MMJJ_11530	0.104	4.12	0.00617518
MMP1554	0.097	3.82	0.00705455
MMJJ_04000	0.356	12.755	0.00739781
MMJJ_15970	0.253	8.73	0.00845683
MMP1398	0.158	5.3	0.00857403
MMJJ_10460	0.228	7.58	0.00913246
MMP0983	1.35	43.27	0.00913246
GYY_06675	0.03	0.905	0.01040932
MMJJ_04320	0.029	0.743	0.01273165
MMP1564	0.428	10.945	0.01377071
MMJJ_15520	0.019	0.487	0.01710263
MMP0357	0.105	2.63	0.0181059

MMJJ_07680	0.014	0.328	0.02095887
MMP0358	0.091	2.16	0.02401431
MmarC6_1358	0.173	4	0.02525565
MMJJ_18480	0.038	0.867	0.02535479
MMJJ_04150	0.047	1.01	0.02792465
GYY_06605	0.022	0.477	0.02894022
GYY_07390	0.021	0.438	0.02983361
MMJJ_18370	0.14	2.87	0.0300506
MMJJ_02730	0.228	4.44	0.03122565
GYY_02050	1.01	19.288	0.03296963
MMJJ_02950	0.088	1.64	0.03844171
GYY_03630	0.105	1.92	0.03909022
MmarC6_1304	4.46	80.716	0.03946851
MMP0568	0.092	1.63	0.03984868
MMP0055	0.304	5.26	0.04106479
MMP1509	0.103	1.64	0.04221149
GYY_06670	0.068	1.04	0.04330826
MMJJ_15480	0.017	0.224	0.04353688
MMJJ_14120	0.044	0.578	0.04626746
MMJJ_02690	0.214	2.76	0.04668819
MMJJ_14870	0.256	3.2	0.04694644
MMJJ_00770	1.75	21.191	0.0489535
MMJJ_12200	1.13	12.817	0.0513054
MmarC6_1678	0.391	4.41	0.05222918
GYY_03740	1.03	11.547	0.05350992
MmarC5_1548	0.039	0.42	0.05482243
WP_048064119	2.04	21.778	0.05526172
MMJJ_12880	0.378	3.78	0.05631479
MMJJ_13470	0.046	0.445	0.05774425
MMJJ_06040	0.01	0.09	0.06286145
MMJJ_16000	0.312	2.83	0.06469166
MMJJ_18470	0.94	8.53	0.07431226
MMJJ_07290	0.308	2.76	0.07638895
GYY_08825	0.042	0.35	0.07745821
MMP0331	0.281	2.35	0.08006858
MMJJ_18520	1.05	8.44	0.12470465
MMJJ_07310	0.292	2.32	0.12571956
MMJJ_00600	2.61	20.474	0.1273662
MMJJ_15470	0.091	0.7	0.13013271
MMJJ_09240	0.059	0.448	0.13127879

MMJJ_11270	0.022	0.17	0.13131954
MMJJ_12230	0.461	3.47	0.13280527
MMJJ_07580	0.398	2.85	0.13969158
MMJJ_14070	1.1	7.77	0.14167549
MMP0820	4.28	29.786	0.14367251
MMP0959	0.264	1.84	0.14378801
MMJJ_02070	0.211	1.45	0.14563551
MMP1565	0.62	4.24	0.1461267
MMJJ_13290	0.309	2.08	0.14838578
MMP1500	13.674	91.93	0.14874778
MMP1069	0.118	0.789	0.14903631
GYY_00030	0.357	2.37	0.15060322
MmarC6_1653	0.159	1.04	0.1525446
MMP0873	0.13	0.826	0.15789333
MMP1031	8.25	51.89	0.15904568
MMJJ_06450	0.129	0.807	0.15975121
MMP0397	1.48	8.88	0.16642018
GYY_00445	0.471	2.82	0.16675648
MMP0003	0.078	0.465	0.16750729
MMJJ_02800	0.229	1.37	0.16795712
MMJJ_13660	0.036	0.209	0.1707739
MMJJ_06000	1.45	8.47	0.17171406
MMJJ_03470	2.32	13.325	0.17395871
MMP0931	0.406	2.33	0.17453546
MMJJ_17300	0.203	1.14	0.17778026
MMJJ_07270	0.844	4.73	0.1783975
MMJJ_09540	1.48	8.27	0.17874109
MMP0141	0.227	1.27	0.17925131
MMP0370	0.341	1.82	0.18754225
MMP0704	0.456	2.43	0.18755427
MMJJ_13630	0.568	3.01	0.18830844
MMP1721	0.795	4.13	0.19252582
GYY_06110	1.63	8.34	0.19520635
MMP0684	9.69	49.493	0.19569475
MmarC7_1572	0.348	1.78	0.1959952
MMJJ_16780	0.014	0.072	0.19707504
MMP1409	0.988	5	0.19738751
MMJJ_03770	0.877	4.35	0.20171314
MMP1250	0.344	1.68	0.2051431
MMP0500	0.428	2.08	0.20555876

Table A3.3. List of upregulated proteins from enriched protein samples after AMP-agarose affinity purification and MUDPIT analysis. Comparison of the ion count in AMP-agarose enriched samples compared to agarose control for *M. maripaludis* grown in the presence of $^{15}NH_4Cl$ or $^{14}NH_4Cl$-containing media.

		Ion Count (10^{10})	
Gene ID	**Protein Name**	**Agarose**	**AMP Agarose**
MMP1513	Alanine dehydrogenase	0	201
MMP1630	Arginine transport ATP-binding protein ArtM	0	51.4
MMJJ_17340	Hypothetical protein	0	10.1
MMP1015	Histone-like transcription factor (CBF/NF-Y) and archaeal histone	0	6.39
MMP1437	DNA topoisomerase VI subunit A	0	5.9
MMP0959	Pyridine nucleotide-disulfide oxidoreductase	0	4.9
MMJJ_17340	Hypothetical protein	0	4.54
MMP1559	Coenzyme-B sulfoethylthiotransferase subunit alpha	0	4.49
MMP0922	4-hydroxy-tetrahydrodipicolinate reductase	0	4.43
MMP1515	Thermosome subunit	0	2.52
MmarC5_0070	Phthiodiolone/phenolphthiodiolone dimycocerosates ketoreductase	0	1.64
MMJJ_12000	Hypothetical protein	0	1.58
MMP0352	Oxidoreductase	0	1.15
MMP1431	2-phosphoglycerate kinase	0	1.04
MMP0650	Acetolactate synthase large subunit	0	0.948
MMP0383	S-layer protein	0	0.707
MMP0228	tRNA (adenine(58)-N(1))-methyltransferase TrmI	0	0.644
MMJJ_07500	Hypothetical protein	0	0.602
MMP0822	Coenzyme F_{420}-non-reducing hydrogenase subunit gamma	0	0.521
MMP0572	Peptidylprolyl isomerase	0	0.443
MMP1155	CoB--CoM heterodisulfide reductase subunit B	0	0.437
MMP1599	Trk system potassium uptake protein TrkA	0	0.409
MMP1658	30S ribosomal protein S8e	0	0.364
MMJJ_14690	Bifunctional enzyme CysN/CysC	0	0.349
MMP1191	Methenyltetrahydromethanopterin cyclohydrolase	0	0.327
GYY_05620	Thioredoxin-1	0	0.327
MMP0087	Hydroxymethylglutaryl-CoA reductase (NADPH)	0	0.325
MMP0346	2-hydroxyacyl-CoA dehydratase	0	0.306
MMP1412	50S ribosomal protein L5	0	0.302
MMJJ_03290	Putative acetolactate synthase small subunit	0	0.298
MMP0509	Formylmethanofuran dehydrogenase subunit A	0	0.291
MMJJ_07240	Hypothetical protein	0	0.282
MMP0051	Phosphoribosyl-ATP pyrophosphatase	0	0.281
MMP0372	F_{420}-dependent methylenetetrahydromethanopterin dehydrogenase	0	0.281
MMP1249	Formyltransferase/hydrolase complex Fhc subunit C	0	0.263

MMP0640	30S ribosomal protein S28e	0	0.262
MMJJ_01430	Pyridoxamine 5'-phosphate oxidase	0	0.261
MmarC6_1042	NADH-quinone oxidoreductase subunit I	0	0.248
MMP0654	Ketol-acid reductoisomerase	0	0.247
MMJJ_04000	Putative universal stress protein	0	0.243
MMP0633	Reverse rubrerythrin-2	0	0.237
MMP0667	30S ribosomal protein S2	0	0.219
MMP1212	Putative acetyl-CoA acyltransferase	0	0.216
MMP1558	Coenzyme-B sulfoethylthiotransferase subunit gamma	0	0.215
MMJJ_07490	Sulfur transfer complex subunit TusD	0	0.214
MMP0353	Bucleotide sugar dehydrogenase	0	0.211
MMJJ_09530	Inosine-5'-monophosphate dehydrogenase	0	0.195
MMJJ_08580	Coenzyme A biosynthesis bifunctional protein CoaBC	0	0.189
MMJJ_12270	Formyltransferase/hydrolase complex subunit D	0	0.181
MMJJ_12930	50S ribosomal protein L4P	0	0.18
MMP1208	Translation initiation factor IF-2 subunit beta	0	0.156
MMP1419	30S ribosomal protein S5	0	0.154
MMJJ_17170	Peptidyl-prolyl cis-trans isomerase B	0	0.148
MMJJ_07270	HIT-like protein	0	0.134
MMP_RS06480	Nitrate ABC transporter substrate-binding protein	0	0.133
MMJJ_10180	23S rRNA (uracil-C(5))-methyltransferase RlmCD	0	0.118
MMP1169	FeS cluster assembly protein SufB	0	0.115
MMJJ_11220	General stress protein 13	0	0.113
MMP1411	30S ribosomal protein S4e	0	0.107
MMOS7_14080	Formate/nitrite transporter family protein	0	0.104
MMP1505	Pyruvate synthase subunit PorA	0	0.1
MMP1031	Adenylate kinase	0	0.092
MMJJ_10520	Ribonuclease J 2	0	0.0851
MMJJ_16380	30S ribosomal protein S6e	0	0.08
MMJJ_18030	Hypothetical protein	0	0.079
MmarC6_1258	50S ribosomal protein L6	0	0.0755
MMP1567	Tetrahydromethanopterin S-methyltransferase subunit H	0	0.074
MMP1514	Prephenate dehydrogenase	0	0.0733
MMP0817	Coenzyme F_{420}-reducing hydrogenase subunit beta	0	0.0726
MMP0539	3-isopropylmalate dehydrogenase	0	0.0721
MMP1555	Methyl-coenzyme M reductase beta subunit, C-terminal domain	0	0.0703
MMP1368	30S ribosomal protein S7	0	0.0679
MMJJ_11940	Hypothetical protein	0	0.0663
MMP0854	Nitrogen regulatory protein P-II	0	0.0643
MMP1067	Succinate dehydrogenase/fumarate reductase iron-sulfur subunit	0	0.0588

MMP0720	Hydroxylamine reductase	0	0.0583
MMP1208	Translation initiation factor IF-2 subunit gamma	0	0.0578
MMP1289	50S ribosomal protein L10e	0	0.0575
MMJJ_13050	Hypothetical protein	0	0.0561
MMP0933	Response regulator	0	0.0553
MMP0127	H_2-dependent methylenetetrahydromethanopterin dehydrogenase	0	0.0528
MMJJ_00300	Hypothetical protein	0	0.0466
MMJJ_15910	Phosphoribosylformylglycinamidine cyclo-ligase	0	0.0451
MMJJ_17550	UDP-glucose 4-epimerase	0	0.044
MMJJ_15400	Selenium binding protein	0	0.0439
MMP0062	50S ribosomal protein L31e	0	0.0438
MMP1436	Cell division protein FtsZ	0	0.0436
MmarC5_0070	NADH-dependent phenylglyoxylate dehydrogenase subunit gamma	0	0.0413
MmarC6_1311	DNA-directed RNA polymerase subunit beta	0	0.0406
MMJJ_16220	Hypothetical protein	0	0.0405
MMP0081	Glutamate synthase	0	0.0396
MMP1206	Glutamine synthetase	0	0.0364
MMJJ_16320	Hypothetical protein	0	0.0356
MMP1391	Aspartate-semialdehyde dehydrogenase 2	0	0.0354
MMP0561	Carboxymuconolactone decarboxylase family protein	0	0.0348
GYY_02010	Digeranylgeranylglycerophospholipid reductase	0	0.0327
MMP1401	Elongation factor 1-beta	0	0.0288
MMP0156	30S ribosomal protein S19e	0	0.0252
MJJ_18380	Arabinose 5-phosphate isomerase KdsD	0	0.0237
MMJJ_10370	Hypothetical protein	0	0.0233
MMJJ_09270	Hypothetical protein	0	0.0232
MMP1666	Flagellin	0	0.0216
MMP_RS08875	Class III signal peptide-containing protein	0	0.0202
MMP1094	Phosphoenolpyruvate synthase	0	0.0197
MMJJ_18110	V-type ATP synthase subunit I	0	0.0184
MMP1409	50S ribosomal protein L14	0	0.0173
MMJJ_06440	Hypothetical protein	0	0.0137
MMP1265	Glutamyl-tRNA(Gln) amidotransferase subunit A	0	0.0126
MMP0259	Acidic ribosomal protein P0	0	0.00906
MMP0947	ATP phosphoribosyltransferase	0	0.00709
MMP1148	Small nuclear ribonucleoprotein	0	0.00705
MMP1564	Tetrahydromethanopterin S-methyltransferase subunit A	0	0.007
MMJJ_15840	Double zinc ribbon	0	0.00642
MMJJ_01770	Coenzyme F_{420}-dependent hydrogenase subunit alpha	0	0.0062

MMP0049	Beta-carbonic anhydrase 1	0	0.0053
MMP1225	Amino acid ABC transporter substrate-binding protein	0	0.00512
MMJJ_16340	3-oxoacyl-[acyl-carrier-protein] synthase 3	0	0.00409
MMP1613	DNA-binding protein Alba	0	0.0035
MMP0024	ATP-dependent zinc metalloprotease FtsH	0	0.00336
MMP1222	DNA repair and recombination protein RadA	0	0.00296
MMJJ_03630	Glycine betaine/carnitine/choline transport ATP-binding protein OpuCA	0	0.00106
MMP0641	Ribosome-associated protein L7Ae-like protein	0	0.000355

References

1. D. G. Gibson, "Enzymatic assembly of overlapping DNA fragments", Methods in Enzymology. **2011**, 498, 349–361
2. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, "Near-optimal probabilistic RNA-seq quantification". Nature Biotechnology. **2016**, 34, 525–527.
3. H. Pimentel, N. L. Bray, S. Puente, P. Melsted, L. Pachter, "Differential analysis of RNA-seq incorporating quantification uncertainty". Nature Methods. **2017**, 14, 687–690.

Appendix 4: Supplementary materials for Chapter 4

Table of Contents

Supplementary methods

Rockhopper assembly. Rockhopper software for microbial transcriptome assembly was used as described previously [1] using default parameters. A reference-based analysis was performed using the *Methanococcus maripaludis* S2 genome (accession ID: NC_005791). Rockhopper text and .wig file outputs were analyzed and show to contain 343 multi-gene operons with 974 genes, 26 monocistronic genes with 5' UTR (Appendix Table A4.1, Table 4.1), representing the majority of protein coding genes. However, many other transcripts were shown to be leaderless, indicating that it could be a common phenomenon. RNA-seq datasets for three biological replicates was used.

3′ UTR terminus analysis. An algorithm was developed to map WIG file data on to the *M. maripaludis* S2 reference genome. By denoting the last nucleotide in the terminator with reads observed in at least two biological replicates, we identified the 3′ terminus. This allowed us to also calculate the distance of the terminator from the closest start and stop codons, helping to determine terminator lengths and intragenic terminators. Nucleotide frequency analysis was also conducted up to -11 bp of the 3′ terminus.

Supplementary Results

Table A4.1. Table of putative operons and their key features. Operons predicted by Rockhopper assembly software including the number of genes in the operon, genes in the operon, genomic start and stop positions, strand polarity, 5' UTR (in nt) and 3' UTR (in nt).

Gene IDs	Genes	Start	Stop	Strand	5'UTR	3'UTR
iorA2, iorB2, MMP0715, MMP0716	4	706252	710302	+	5	282
MMP1348, MMP1349, MMP1350, MMP1351	4	1327842	1331550	-	5	-
ehbC, ehbD	2	1064872	1065378	+	6	-
cmk, rpl34e	2	619232	620047	-	10	-
leuS, MMP0698	2	686587	689799	+	10	10
MMP0979, cdh, cdhD, MMP0982, cdhB, cdh, cdhA, thyA, ksgA, MMP0988	10	964340	972377	-	10	-
MMP0432, MMP0433	2	433202	434476	+	38	20
MMP1299, MMP1300, fdhC	3	1286147	1288015	-	70	-
MMP1568, MMP1569, MMP1570, MMP1571	4	1520508	1523103	-	81	-
MMP1335, selB	2	1315756	1318126	-	258	-
MMP0586, napA	2	582554	584447	-	398	-
MMP0993, MMP0994	2	981206	981744	-	496	-
MMP1633, MMP1634	2	1575947	1577150	-	1586	-

MMP0001, MMP0002	2	10104	11480	+	-	-
korA, MMP0004	2	11517	13543	-	-	-
DP1, MMP0009, MMP0010	3	17602	20931	-	-	-
truD, MMP0015, MMP0016	3	24061	26552	-	-	-
MMP0022, MMP0023	2	30411	31567	-	-	-
MMP0030, MMP0031	2	45639	48755	+	-	-
MMP0033, MMP0034, MMP0035, tfe, MMP0037	5	50520	54948	+	-	12
MMP0038, MMP0039, MMP0040, tfb, MMP0042	5	54975	59924	-	-	-
MMP0044, idsA	2	61568	63881	+	-	-
MMP0048, MMP0049, ribH, hisE	4	65556	67112	+	-	788
MMP0053, MMP0054, MMP0055	3	68156	69994	-	-	-
rplX, aIF6, MMP0062	3	74160	75380	-	-	-
glnK1, amtB	2	76751	78342	+	-	-
glnK2, amtB	2	78986	80588	+	-	-
MMP0080, gltB, MMP0082, MMP0083	4	90509	95063	+	-	-
MMP0084, MMP0085, tfx, hmgA, hemA, MMP0089, MMP0090, MMP0091	8	95381	102422	-	-	-
MMP0100, MMP0101	2	109297	110975	-	-	-
MMP0106, MMP0107, MMP0108, afuB_1, MMP0110	5	115117	119991	+	-	-
MMP0111, MMP0112, MMP0113, MMP0114	4	119995	123316	-	-	-
MMP0118, birA, MMP0120, MMP0121, MMP0122	5	125924	128889	-	-	-
flpA, bioB	2	130979	133565	-	-	-
MMP0129, fumA	2	135643	137160	+	-	-
thrC, leuD	2	141612	143347	+	-	-
fdhA, fdhB	2	145038	148219	+	-	-
MMP0141, MMP0142, MMP0143, hpcE, hpt, MMP0146, MMP0147	7	150629	157773	-	-	-
MMP0150, rpl40e	2	161087	161750	+	-	-
MMP0156, MMP0157, MMP0158, rpl39e	4	167490	169072	+	-	-
MMP0165, MMP0166, MMP0167, MMP0168	4	173677	177259	-	-	-
MMP0169, cofF, MMP0171	3	177372	180128	+	-	-

purQ, purL	2	186678	190466	-	-	-
mtaP, MMP0186	2	193763	195159	+	-	-
MMP0196, MMP0197, MMP0198, MMP0199, fmdE, moeA	6	204032	209379	+	-	-
modA, modB, modC	3	212528	215158	+	-	-
MMP0212, MMP0213, MMP0214	3	218916	222096	-		1814
MMP0216, MMP0217	2	222645	225067	-	-	-
MMP0222, MMP0223, hemL	3	230588	232408	+	-	-
gldA, MMP0226, MMP0231, MMP0232, MMP0233	5	232456	241361	-	-	-
MMP0234, MMP0235, MMP0236, MMP0237	4	241487	244393	+	-	-
MMP0238, MMP0239, MMP0240, MMP0241	4	244445	248100	+	-	4
MMP0242, MMP0243, MMP0244, pfdB	4	248102	250146	-	-	-
MMP0246, MMP0247, MMP0248	3	250204	251097	-	-	-
rpl37ae, MMP0250	2	251152	252176	-	-	-
MMP0254, MMP0255	2	255837	258280	+	-	-
hisH, tbp	2	258316	259484	-	-	-
rplP0, rpl1P	2	260003	261665	-	-	-
MMP0261, MMP0262	2	261844	262748	-	-	-
tyrS, MscMJ, MMP0265	3	263868	266379	+	-	-
MMP0266, MMP0267, truA	3	266419	268598	-	-	-
MMP0270, MMP0271, MMP0272, comA	4	270082	272984	-	-	2543
MMP0278, mptG, hisI	3	279585	281446	+	-	-
MMP0285, MMP0286, MMP0287, MMP0288	4	286576	289852	+	-	-
MMP0292, MMP0293	2	292999	294122	-	-	-
MMP0300, hypA	2	300053	300894	+	-	-
MMP0305, MMP0306	2	302475	305053	-	-	-
MMP0308, MMP0309	2	305972	306538	+	-	-
MMP0311, MMP0312, MMP0313	3	307302	308610	+	-	-
MMP0314, iorB1, iorA1	3	308925	312596	-	-	-
cbiL, MMP0320	2	315937	317494	-	-	-

MMP0321, rfcB	2	317627	319633	+	-	-
MMP0328, MMP0329, cbiC	3	326152	327910	+	-	17
hjc, MMP0337, MMP0338	3	332791	334742	-	-	-
MMP0350, aspB-like1, MMP0352, MMP0353, MMP0354	5	345820	351034	+	-	-
MMP0355, MMP0356, MMP0357, MMP0358, MMP0359	5	351015	356598	-	-	-
MMP0361, MMP0362, MMP0363, MMP0364, MMP0365	5	357702	363307	-	-	-
MMP0366, MMP0367	2	363355	364338	+	-	-
MMP0376, MMP0377, MMP0378, MMP0379, dnaB	5	373010	378844	-	-	-
MMP0388, MMP0389	2	386328	387667	-	-	1340
hemD, MMP0395	2	391835	392863	-	-	-
MMP0398, htpX, ehbQ, metE, MMP0402, MMP0403, cofD, gch3, MMP0406	9	397319	406114	+	-	385
hisA, MMP0418	2	418670	420323	+	-	-
MMP0420, MMP0421, MMP0422	3	422287	424667	+	-	-
MMP0425, MMP0426	2	426937	428161	+	-	-
MMP0435, rimK, MMP0437	3	435021	437676	+	-	-
rpoE2, MMP0442, rps24e, MMP0444	4	440245	441449	+	-	-
MMP0447, MMP0448	2	442530	443207	-	-	-
MMP0449, MMP0450, MMP0451	3	443364	445432	-	-	-
MMP0461, MMP0462	2	458318	459142	-	-	-
MMP0467, MMP0468, MMP0469, MMP0470, MMP0471, MMP0472, MMP0473, MMP0474	8	464479	471008	+	-	-
MMP0476, MMP0477	2	473717	474264	-	-	-
MMP0482, MMP0483	2	478571	479240	-	-	-
MMP0490, MMP0491, MMP0492	3	487189	488881	-	-	-
MMP0494, MMP0495	2	496024	499391	-	-	-
modC, MMP0505, modB, modA	4	508346	511690	-	-	-
fmdA, fmdC, fmdB	3	512896	517241	+	-	-
modA, modB	2	520134	521586	+	-	-
MMP0519, MMP0520, MMP0521, MMP0522, MMP0523, MMP0524, MMP0525	7	523566	531122	-	-	-

MMP0526, MMP0527	2	531626	532985	+	-	-
MMP0530, MMP0531	2	535522	536635	+	-	-
MMP0534, MMP0535, MMP0536, nth1	4	539194	542571	-	-	-
purC, serB, MMP0542, MMP0543, MMP0544, MMP0545	6	545004	550601	+	-	-
hisB, MMP0549, MMP0550, MMP0551	4	552862	555651	+	-	-
MMP0556, MMP0557, MMP0558	3	559014	560577	-	-	116
MMP0564, MMP0565, MMP0566, MMP0567, MMP0568	5	563804	568318	-	-	-
MMP0570, moaA	2	570187	571589	-	-	-
MMP0574, gatC, dapA, rps17E	4	573568	575031	-	-	2184
MMP0579, act	2	575501	577033	-	-	-
MMP0582, MMP0583	2	577959	578976	+	-	-
MMP0590, MMP0591, MMP0592, MMP0593, MMP0594, MMP0595	6	587336	591674	-	-	-
MMP0596, flpA	2	591868	594008	+	-	-
cbiJ, MMP0600	2	595456	596920	-	-	-
aIF-1A, MMP0604	2	598821	599965	+	-	-
MMP0606, nrpR, MMP0608	3	600654	604337	-	-	2039
MMP0618, MMP0619, atwA	3	613463	615816	+	-	-
MMP0634, MMP0635, MMP0636, MMP0637	4	626477	630426	+	-	-
MMP0639, rps28e, rpl7ae	3	632868	633699	-	-	-
MMP0644, mdh	2	635985	637821	-	-	1818
MMP0647, rtcA, MMP0649	3	638236	641215	-	-	117
ilvB, ilvH	2	641377	643663	+	-	-
MMP0661, MMP0662, MMP0663, MMP0664	4	651185	654544	+	-	-
rps2P, MMP0668, MMP0669, MMP0670, ywbE	5	657133	660303	+	-	-
MMP0674, MMP0675	2	662813	663413	-	-	-
upp, MMP0681	2	668393	670366	-	-	-
MMP0690, MMP0691	2	678832	680084	-	-	-
MMP0699, MMP0700	2	689869	691787	-	-	-
MMP0701, MMP0702	2	691982	693340	+	-	-

wbpI, MMP0706, MMP0707	3	695370	699024	-	-	-
MMP0710, corA	2	701140	704988	+	-	-
MMP0718, MMP0719	2	711235	712030	+	-	-
MMP0722, MMP0723	2	714749	715743	+	-	-
uvrB, uvrC, uvrA, MMP0730, MMP0731, xseA	6	717504	723887	-	-	-
MMP0736, MMP0737	2	729123	730205	+	-	-
MMP0743, MMP0744, MMP0745, MMP0746, MMP0747, MMP0748, MMP0749	7	737294	738867	-	-	260
MMP0755, MMP0756	2	746442	749538	+	-	-
MMP0757, MMP0758, MMP0759	3	749567	752053	-	-	-
MMP0767, MMP0768, MMP0769	3	757172	761437	-	-	-
MMP0771, MMP0772, MMP0773	3	762446	765096	-	-	-
MMP0775, MMP0776, MMP0777, MMP0778, MMP0779, MMP0780, MMP0781, MMP0782	8	765796	767464	-	-	-
MMP0783, MMP0784	2	773799	775686	+	-	-
MMP0792, MMP0793, MMP0780, MMP0781, MMP0782	4	784694	786652	+	-	-
MMP0802, MMP0803	2	794946	797239	+	-	-
MMP0805, MMP0806	2	798048	799346	+	-	-
MMP0808, MMP0809	2	800375	802365	+	-	-
MMP0814, MMP0815	2	807432	810987	-	-	-
frcB, frcG, frcD, frcA	4	811639	814994	-	-	-
vhcD, vhcG, vhcA, vhcB	4	815353	819333	+	-	-
MMP0829, mtaA, mtbA, MMP0832	4	823118	827824	+	-	-
MMP0851, MMP0852	2	842609	843745	-	-	-
NifI1, nifI2, nifD, nifK, nifE, nifN, nifX, kamA, yodP	9	844960	856164	+	-	-
proX, MMP0867, proV	3	857796	860715	-	-	-
MMP0869, MMP0870	2	861171	862584	+	-	-
cofG, nadC	2	869522	871442	-	-	-
MMP0883, MMP0884	2	876371	877864	-	-	1579
MMP0888, MMP0889	2	880636	881593	-	-	-
pyrG, guaA, MMP0895, MMP0896	4	885559	890125	+	-	-

MMP0899, MMP0900	2	894927	896249	+	-	17
MMP0902, MMP0903	2	897383	898014	-	-	-
MMP0905, MMP0906, MMP0907	3	899384	901325	-	-	-
MMP0908, MMP0909	2	901511	902388	+	-	-
MMP0913, MMP0914	2	903792	904895	-	-	-
cobY/cofC, MMP0916	2	904980	905833	+	-	-
MMP0919, ahcY	2	908415	910461	-	-	-
MMP0922, dapB, MMP0924, cheW, cheB, cheA, cheD, MMP0929	8	911058	919413	+	-	-
MMP0931, MMP0932, cheY	3	920322	922027	-	-	-
MMP0935, aroE, cofE	3	923434	925536	+	-	-
cobS, MMP0939, MMP0940	3	925590	927366	-	-	-
MMP0942, taqD, MMP0944	3	929000	931311	-	-	-
MMP0949, MMP0950, cobD	3	936440	940148	+	-	-
dsbD, MMP0958, trxB	3	946238	948177	-	-	-
MMP0961, MMP0962, MMP0963	3	949611	952077	+	-	-
MMP0967, hisD, MMP0969, lig, purB	5	954150	959445	+	-	-
MMP0973, MMP0974	2	959986	961489	-	-	-
MMP0976, CooC	2	962612	964051	+	-	-
MMP0999, MMP1000	2	984278	985110	-	-	-
trpA, trpB, trpF, trpG, trpE, trpD, trpC, pyrC, MMP1010	9	986571	995216	-	-	6383
MMP1021, MMP1022	2	1009427	1012157	-	-	-
MMP1024, MMP1025	2	1012955	1015603	-	-	-
MMP1033, tmk, MMP1035, MMP1036	4	1025552	1030432	+	-	-
atpH, atpI	2	1031411	1033813	+	-	-
atpE, atpC, atpF, atpA, atpB, atpD, MMP1047	7	1034595	1039912	+	-	-
hdrB2, hdrC2	2	1044633	1046104	-	-	-
MMP1055, MMP1056	2	1046236	1047766	+	-	-
MMP1059, cysS, MMP1061, MMP1062	4	1050094	1055482	+	-	-
MMP1064, MMP1065	2	1057240	1058234	-	-	-

MMP1068, MMP1069, MMP1070, MMP1071, MMP1072	5	1060440	1064672	-	-	-
dut, MMP1076, MMP1077	3	1065449	1068508	-	-	-
MMP1078, MMP1079	2	1068603	1070360	+	-	-
wbpG, hisH	2	1072342	1074063	+	-	-
MMP1084, MMP1085, MMP1086	3	1074958	1078429	-	-	-
MMP1089, MMP1090	2	1081000	1083453	-	-	-
MMP1092, coaD	2	1084457	1085814	-	-	-
MMP1096, MMP1097, pstB	3	1089326	1091855	+	-	-
MMP1102, MMP1103	2	1094532	1095817	+	-	-
sucC, MMP1106	2	1096371	1097783	+	-	-
MMP1107, corA, MMP1109, MMP1110, modA, MMP1112	6	1097784	1103162	-	-	-
rpe, MMP1115	2	1104218	1105718	-	-	-
MMP1125, MMP1126	2	1115557	1116881	+	-	-
endA, comC	2	1122281	1123875	+	-	-
thiM, thiE	2	1128004	1129477	+	-	-
MMP1141, MMP1142	2	1130105	1133442	+	-	-
MMP1143, MMP1144	2	1133478	1134623	-	-	-
mtaA, mtbC	2	1139143	1140848	-	-	-
hdrC1, hdrB1, MMP1156, rbo, MMP1158, MMP1159, MMP1160, MMP1161, MMP1162, MMP1163, MMP1164, MMP1165, MMP1166, MMP1167, MMP1168, MMP1169, MMP1170, pssA	18	1142799	1155227	+	-	-
MMP1176, MMP1177	2	1158991	1160114	+	-	-
MMP1179, MMP1180	2	1161713	1164126	+	-	-
MMP1182, MMP1183	2	1165563	1167388	+	-	1770
MMP1185, lon, fucA	3	1168035	1171238	+	-	-
MMP1201, MMP1202, cbiD, MMP1204	4	1188095	1191209	-	-	-
aIF2_gamma, MMP1209	2	1194834	1197167	+	-	-
MMP1211, MMP1212, MMP1213	3	1198055	1200732	+	-	-
MMP1217, MMP1218	2	1204315	1206239	-	-	-
cbiE, MMP1228	2	1216502	1217903	+	-	-

MMP1234, moaE, MMP1236, MMP1237, bioB, MMP1239, MMP1240, MMP1241	8	1221654	1227174	+	-	-
fwdH, fwdF, fwdG, fwdD, fwdA, fwdC	6	1228973	1233772	+	-	-
MMP1251, MMP1252	2	1234459	1235881	+	-	-
MMP1257, hemB	2	1240285	1241949	-	-	-
MMP1261, MMP1262, MMP1263, MMP1264	4	1244911	1248324	-	-	-
MMP1268, MMP1269	2	1252548	1253871	+	-	-
vorA, vorB, vorC	3	1255312	1258058	-	-	-
sdhA, MMP1278, MMP1279	3	1261146	1263490	-	-	-
MMP1282, MMP1283	2	1264697	1265489	-	-	739
MMP1286, MMP1287, MMP1288	3	1269354	1272308	+	-	-
MMP1290, MMP1291	2	1273018	1274950	-	-	-
glgA, MMP1295, pfkC	3	1278266	1282598	+	-	22
fdhB, fdhA	2	1282622	1285845	-	-	-
MMP1303, MMP1304	2	1288715	1291021	-	-	-
MMP1309, purO, MMP1311, MMP1312, fen-1, MMP1314, korG, korB	8	1295407	1300796	-	-	-
rps13p, rps4p, rps11p	3	1305905	1307298	+	-	-
MMP1323, rpl13p, rps9p, rpoN	4	1307955	1309414	+	-	-
rpoK, MMP1328	2	1309563	1310703	+	-	-
MMP1331, MMP1332	2	1311328	1312994	-	-	-
mre11, rad50, MMP1342	3	1319955	1324472	+	-	-
MMP1343, MMP1344, MMP1345, MMP1346	4	1324489	1327322	-	-	-
MMP1354, MMP1355	2	1333713	1335560	+	-	17
MMP1356, MMP1357, MMP1358, MMP1359	4	1335578	1336783	-	-	-
MMP1365, nusA, rps12P, MMP1368	4	1346873	1348766	+	-	-
rnhB, MMP1375, MMP1376, MMP1377	4	1355645	1358683	+	-	-
fruA, fruD, fruG, fruB	4	1362470	1365924	+	-	-
aroD, MMP1395	2	1372973	1375897	-	-	-
smc1, dapE, MMP1399, MMP1400, ef1B, MMP1402, rpl22p, rps3p, rpmC	9	1377736	1385363	+	-	-
rps17p, rpl14p, rpl24p, MMP1411, rpl5p, rps14P, rps8p, rpl6p, rpl32e, rpl19e, rpl18p, rps5p, rpl30p, rpl15p	14	1386287	1393014	+	-	-

MMP1425, dcd, MMP1427	3	1398277	1400840	+	-	-
MMP1430, 2pgk	2	1401518	1403914	-	-	-
secE, ftsZ1	2	1406294	1407646	-	-	-
MMP1443, MMP1444, guaA	3	1414018	1416629	+	-	-
MMP1447, ehaA, ehaB, ehaC, ehaD, ehaE, ehaF, ehaG, ehaH, ehaI, ehaJ, ehaK, ehaL, ehaM, ehaN, ehaO, ehaP, ehaQ, ehaR, ehaS, ehaT	21	1417426	1428736	+	-	-
MMP1471, MMP1472, MMP1473, ileS	4	1431231	1432967	+	-	-
cbiA, cbiA, MMP1479	3	1438364	1440662	+	-	-
cbiM, MMP1482, cbiQ, cbiO, moaB	5	1442205	1445351	+	-	-
pheS, MMP1497, MMP1498	3	1457002	1459787	+	-	-
porF, porE, porB, porA, porD, porC	6	1462535	1466397	-	-	-
MMP1508, MMP1509	2	1466665	1467332	+	-	-
MMP1516, MMP1517	2	1476368	1477657	-	-	-
MMP1518, MMP1519	2	1477790	1479510	-	-	-
MMP1521, MMP1522	2	1480378	1481829	-	-	-
secD, secF	2	1481881	1483910	-	-	-
pheA, MMP1529, MMP1530, MMP1531	4	1487425	1490206	-	-	-
MMP1535, MMP1536	2	1492980	1494788	-	-	-
MMP1537, MMP1538, MMP1539, MMP1540	4	1494996	1497385	+	-	-
rpl3p, rpl4lp, rplW, rpl2p, rps19p	5	1498876	1502238	+	-	-
MMP1549, MMP1550	2	1502975	1504474	+	-	-
ffh, MMP1552	2	1504831	1506767	-	-	-
mcrB, mcrD, mcrC, mcrG	4	1509697	1512916	+	-	-
mtrE, mtrD, mtrC, mtrB, mtrA, or900, mtrG, mtrH	8	1514893	1520421	+	-	-
bioF, bioW	2	1524069	1525924	-	-	-
MMP1583, MMP1584, MMP1585	3	1533600	1535691	+	-	-
MMP1587, serA	2	1536058	1538075	-	-	-
cbiG, trpS, MMP1593, MMP1594	4	1540191	1543764	-	-	-
MMP1598, trkA	2	1546728	1547963	+	-	-
MMP1600, MMP1601	2	1547994	1549610	-	-	-

MMP1605, MMP1606, MMP1607	3	1552396	1555523	+	-	-
MMP1611, MMP1612	2	1557924	1559377	-	-	301
MMP1617, aIF-2B_I	2	1563021	1564260	-	-	-
MMP1619, MMP1620	2	1564401	1565915	+	-	-
ehbO, ehbM, ehbL, ehbK, ehbJ, MMP1626, ehbG, ehbF, ehbE	9	1565956	1572593	-	-	-
MMP1638, MMP1639	2	1579796	1581871	+	-	-
MMP1642, MMP1643, MMP1644, MMP1645	4	1584091	1586712	+	-	13
modC, MMP1650, modA	3	1589977	1592611	-	-	-
MMP1653, MMP1654, MMP1655	3	1594017	1600218	+	-	-
MMP1660, MMP1661	2	1604456	1605002	+	-	-
cbiF, MMP1663	2	1604999	1606998	-	-	-
flaB1, flaB2, flaB3, flaC, flaD, flaE, flaF, flaG, flaH, flaI, flaJ	11	1609807	1618850	+	-	-
MMP1684, MMP1685	2	1626448	1627375	-	-	-
cbf5, korD, MMP1688	3	1627999	1629918	+	-	-
comE, MMP1690, fwdB, vhuB, vhuU, vhuA	6	1629931	1635388	-	-	-
MMP1699, MMP1700, MMP1701, hom	4	1640528	1643999	+	-	-
MMP1704, MMP1705, MMP1706, aIF2_alpha, rps27e, rpl44e, MMP1710, pcnA	8	1645101	1649130	-	-	512
MMP1714, MMP1715, hmdII, MMP1717, MMP1718	5	1651340	1654681	-	-	-
MMP1720, MMP1721, hisF	3	1657683	1661129	+	-	-

References

1. R. McClure, D. Balasubramanian, Y. Sun, M. Bobrovskyy, P. Sumby, C. A. Genco, C. K. Vanderpool, B. Tjaden, "Computational analysis of bacterial RNA-Seq data". Nucleic Acids Research. **2013**, 41, e140.