Title
Adaptive cellular strategies to improve commodity chemical production in Escherichia coli

Author
Koleski, Edward Jon

Publication Date
2023

Peer reviewed|Thesis/dissertation

Adaptive cellular strategies to improve commodity chemical production in *Escherichia coli*

By

Edward J Koleski

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michelle C. Y. Chang, Chair
Professor David Savage
Professor Michael Marletta

Spring 2023

Adaptive cellular strategies to improve commodity chemical production in *Escherichia coli*

by Edward J Koleski

Abstract

Adaptive cellular strategies to improve commodity chemical production in *Escherichia coli*

by

Edward J Koleski

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Michelle C. Y. Chang, Chair

Biology holds an amazing propensity for chemistry. Living systems continuously carry out a vast plethora of chemical reactions within a complex network, known as metabolism, to sustain growth and improve evolutionary fitness. Metabolic engineers seek to utilize this aptitude for chemistry by creating biological catalysts for chemical production from renewable feedstocks. Biological catalysts offer an eco-friendly, and in some cases, superior, alternative to petrochemical-based chemical production.

In this work, we examine biological catalysts designed for the production of two C_4 commodity chemicals, *n*-butanol and (*R*)-1,3-butanediol. These catalysts are strains of *Escherichia coli* containing constructed biosynthetic pathways. Leveraging an anaerobic growth selection, laboratory adaptive evolution identified several mutant strains with improved phenotypes. We set out to understand the mechanism by which these adaptive mutations confer improved production. Through detailed analysis of *n*-butanol fermentation, we discovered that the parent strain for our evolution was unable to support sustained anaerobic growth via *n*-butanol fermentation, potentially due to metabolic burden associated with overexpression of the pathway enzymes. Further experimentation suggested that the mutations arose as a strategy to relieve metabolic burden through decreased expression of our biosynthetic pathway. The results of this study highlight the importance of balanced pathway expression when designing biological catalysts.

We then shifted our focus to design a microbial catalyst for production of polyhydroxyalkanoates (PHAs) containing unsaturated monomers. Sites of unsaturation provide functional handles for downstream chemical modification. We devised a metabolic strategy to convert two non-canonical amino acids with unsaturated functional groups to their respective 2-hydroxy acids and activate these acids as coenzyme A thioesters for polymerization within *E. coli*. We identified and tested candidate enzymes for the appropriate activities *in vitro* and successfully showed that our identified enzymes can form a functional biosynthetic pathway. These experiments lay the groundwork for creation of a microbial catalyst capable of generating PHAs with unsaturated functional groups using glucose as a carbon source.

Table of Contents

Chapter 4: Towards *in vivo* production of PHAs with amino-acid derived monomers

Appendices

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations

^{13}C-MFA	^{13}C metabolic flux analysis
2H4Pe	2-hydroxy-4-pentenoate
2H4Py	2-hydroxy-4-pentynoate
2HIC	2-hydroxyisocaproate
2O4Pe	2-oxo-4-pentenoate
2O4Py	2-oxo-4-pentynoate
3HB	3-hydroxybutyrate
3HV	3-hydroxyvalerate
4HB	4-hydroxybutyrate
4HBCS	4-hydroxy butyryl-CoA synthetase
AlaDH	alanine dehydrogenase
Alg	allylglycine
ATP	adenosine triphosphate
ADP	adenosine diphosphate
BDO	1,3-butanediol
Cb	carbenicillin
Cm	chloramphenicol
CO_2	carbon dioxide
CoA	coenzyme A
CRISPR	clustered regularly interspaced short palindromic repeats
cryo-EM	cryogenic electron microscopy
DEG	differentially expressed gene
DNA	deoxyribonucleic acid
DNC	Dansyl chloride
dNTP	deoxynucleotides
DTNB	5,5´-dithiobis-(2-nitrobenzoic acid), Ellman's reagent
DTT	dithiolthreitol
ESI	Electrospray ionization
EtOH	ethanol
FADH	flavin adenine dinucleotide reduced

GC	gas chromatography
GFP	green fluorescent protein
GO	gene ontology
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high-performance liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Km	kanamycin
LB	lysogeny (Luria-Bertani) broth
LC-MS	liquid chromatography – mass spectrometry
LC-QQQ	liquid chromatography – triple quadrupole mass spectrometer
LeuDH	leucine dehydrogenase
L-HicDH	L-2-hydroxyisocaproate dehydrogenase
mRNA	messenger RNA
MPB	maltose binding protein
MRM	multiple reaction monitoring
NAD^+	nicotinamide adenine dinucleotide
NADH	nicotinamide adenine dinucleotide reduced
NMR	Nuclear magnetic resonance
OD	optical density
P3HB	poly 3-hydroxybutyrate
P3HV	poly 3-hydroxyvalerate
PAM	protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	polymerase chain reaction
PDHc	pyruvate dehydrogenase complex
PE	polyethylene
PHA	polyhydroxyalkanoate
PFL	pyruvate formate lyase
PLA	polylactic acid
PMSF	phenylmethanesulfonyl fluoride
PP	polypropylene
Pra	propargylglycine

PROSS	protein repair one-stop shop
qPCR	quantitative PCR
RBS	Ribosome binding site
RNA	ribonucleic acid
RNA-seq	RNA sequencing
SOE-PCR	splicing by overlap extension PCR
TB	terrific broth
TCA	tricarboxylic acid cycle
TMAO	trimethylamine N-oxide
TNB	2-nitro-5-thiobenzoate anion
tRNA	transfer RNA
UPLC	ultra-performance liquid chromatography

Acknowledgments

I had the great fortune of being able to complete my PhD within the Michelle Chang lab, which through generations continues to be a friendly community of bright scientists. I thank Michelle for welcoming me into the lab and for her mentorship through my scientific pursuits. I know that she has my best interests in mind and never lets intellectual differences come in the way of our relationship. Michelle's supportive nature has fostered a great community within the lab. The Chang lab that I joined was full of great mentors. I owe my early training as an independent scientist to the graduate students who preceeded me: Vivian, Jorge, Monica, Sasilada, Kersh, and Jason. These students all served as role models for a work ethic that I strive for. Specifically, I have a vast and growing appreciation for the time that Kersh devoted to long conversations about my research project and the intellectual and friendly support that he provided.

In the year after I joined, the lab welcomed in Eli, Doug, and Max. I have become increasingly fond of each of these students and have had the opportunity to see them flourish as scientists. Max, Jason, and I became especially close during the pandemic when lab occupancy was restricted. The three of us formed the afternoon lab shift and never missed an opportunity to get a meal from one of the many fine restaurants in Berkeley. The joining of the new era of lab members, Andrew, Elizabeth, Veronica, Gabby, and Sunnie, marked an increase in group activities outside of lab that has brought us all even closer together. Weekly lunches at our favorite restaurant, MoMo Masalas, have now become ingrained in tradition. I know that the lab's culture will be in good hands with the newest additions of Yanni, Nicole, and Helena.

During my time in the Chang Lab, I had the great opportunity to mentor an undergraduate student, Albert Qiang. During the two years that I worked with Albert, he developed into a competent scientist and an invaluable resource. I wish him the best of luck with his future endeavors and am confident that he will be a great addition to any team.

A special part of this acknowledgment is reserved for Hongjun Dong, who as a veteran post-doc, taught me everything I know about microbiology. It was hard to say goodbye to Hongjun when he left the lab, but I was encouraged that he had realized his dream to become a professor in China. With a tragic stroke of fate, Hongjun lost his life to brain cancer shortly after assuming his new research faculty position. I wish the very best for his wife and two children who now have to carry on in his absence. I want them to know that he was a great scientist, mentor, and friend, who was loved by all in the lab.

Outside of the lab, I am grateful for weekly burrito lunches with my two closest friends in my cohort, Ryan and Derek. The lunches provided a great escape from the toils of research and were often the highlight of my week. I am also thankful for the unwavering support of my girlfriend, Kincso, which was especially needed during the writing of this thesis. Lastly, I dedicate this work to my parents, Jon and Deb who nourished my scientific curiosity from a young age and provided me with a firm foundation to stand on.

Chapter 1: *Introduction*

Portions of this work were performed in collaboration with the following persons:

Kersh Thevasundaram assisted with research, and figure making.

1.1. Historical use of whole-cell catalysis

Although they may not be traditionally thought of as catalysts, organisms have evolved over millions of years to perform chemistry on a plethora of substrates. This chemistry is necessary to construct a functioning metabolic network to support cellular growth and organism fitness. Throughout history, human society has leveraged the synthetic capacity of biology. Due to their relative simplicity compared to higher organisms, microbes have often been the preferred type of whole-cell catalysts.

In fact, the earliest known evidence of anthropomorphic use of microbes to perform chemistry can be traced back to as early as 6000 BC (*Figure 1.1*) [1]. Fungal and bacterial species were used to digest complex carbohydrates and proteins in grains, vegetables and milk through a process known as fermentation, in order to preserve food. Brewing beer through fungal fermentation of malt has been a part of human society for thousands of years, with records of brewing dating back to ancient Egypt [2]. By the end of the 19th century, microbial cultures were being used as an approach to waste management. Biogas resulting from anaerobic fermentation of wastewater and solid agricultural waste was collected and used as a fuel for lighting [3]. In the early 20th century, industrial fermentation came to the forefront to meet the increased demand for acetone by the British Army in World War I, who were struggling from an unstable supply chain. The solventogenic species, *Clostridium acetobutylicum* isolated in 1916 was found to produce acetone via fermentation. Using existing ethanol plant infrastructure, acetone production at industrially relevant levels could be achieved by microbial fermentation using molasses as a main feedstock [4]. Industrial citric acid fermentation was also developed in the early 20th century [5]. By 1919, Pfizer, Inc. had opened a pilot plant for citric acid production via *Aspergillus niger* fermentation of molasses. After citric acid, Pfizer commercialized a number of other fermentation products including gluconic acid, itaconic acid and fumaric acid. Given previous experience with fermentation, Pfizer was well equipped for the fermentation of *Penicillium* species to produce the much needed, newly discovered penicillin antibiotics during World War II [6]. Production of penicillin marked only the first foray of microbial fermentation for production of medicinally relevant compounds. In addition to penicillin, a number of other antibiotics are now produced through microbial fermentation, including β-lactams and macrolides [7–9].

In the latter half of the 20th century, emergence of the central dogma of molecular biology and genome sequencing tools spurred an understanding of gene-dependent activity [10]. The development of tools for introducing genomic manipulations and transplanting genes into new organisms opened the flood gates for “metabolic engineers” to begin tinkering with genomic material of organisms to develop catalysts for new applications [11]. One of the earliest demonstrations of such metabolic engineering was the recombinant expression of naphthalene dioxygenase from *Pseudomonas putida* in *Escherichia coli*, allowing for the conversion of indole to indoxyl leading to the formation of indigo dye [12]. Recombinant gene expression has also been

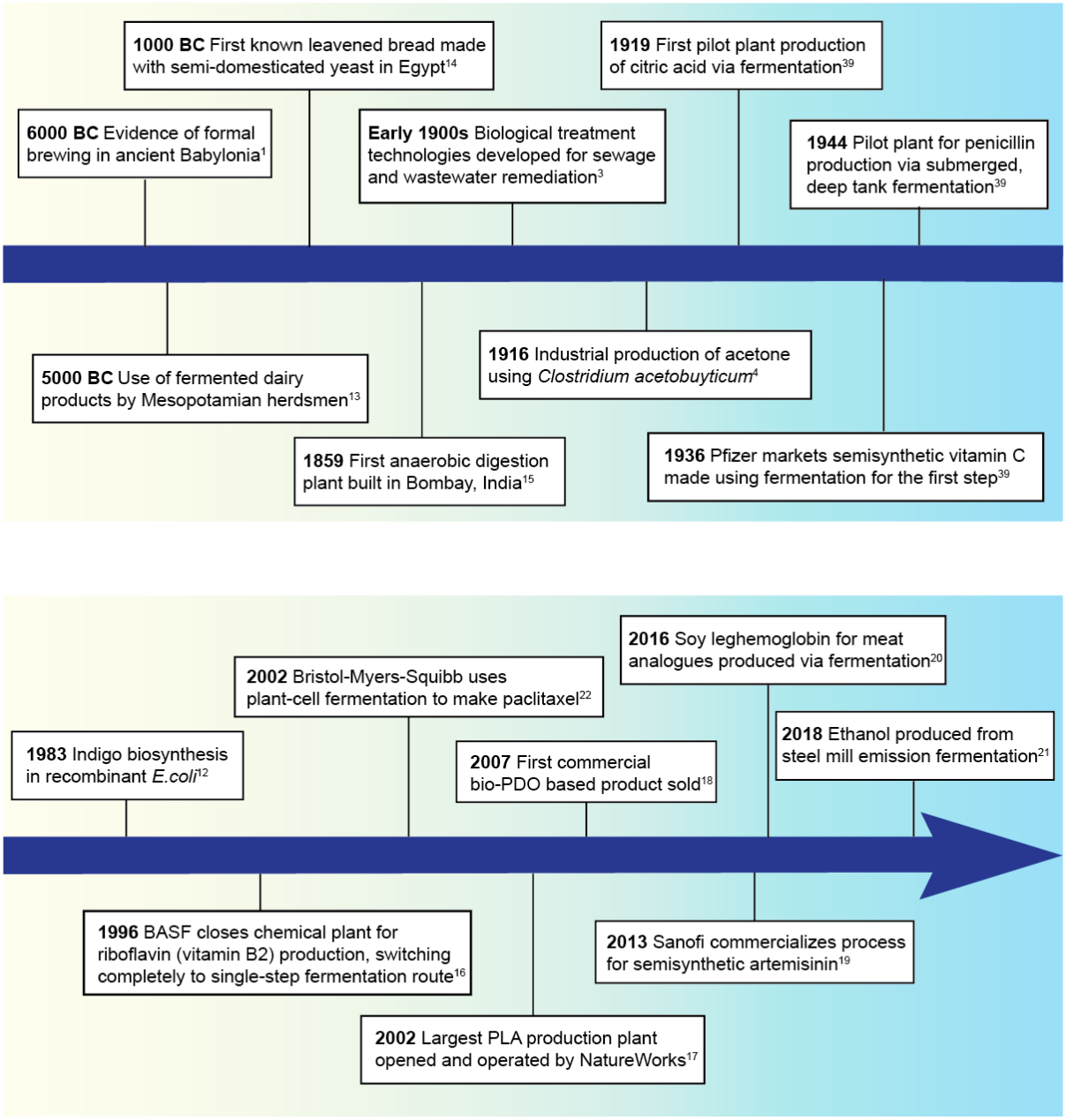


Figure 1.1. *Timeline of the historic use of biological catalysts [1,3,4,6,12–22].*

used to streamline production of 2-ketogluconic acid (2-KLG), the final precursor in commercial processes to produce ascorbic acid (vitamin C). One route to 2-KLG involved two successive fermentations by separate organisms [23]. By transferring a gene from one of the organisms to the other, 2-KG could be produced by a single fermentation with only one organism [24]. These studies served as the foundation for industrial metabolic engineering.

1.2. Advantage of whole-cell catalysis

Whole-cell catalysis offers a number of benefits over traditional synthesis. First, a variety of feedstocks can be used to support microbial growth and production, ranging from fermentable sugars to other small molecules. Second, biology excels at chemistry that can be synthetically challenging. Metabolic pathways use enzymes to catalyze reactions within a carefully designed environment, allowing for the stereo- and regioselective control required to form complicated specialty metabolites. Last, whole-cell catalysts are often able to achieve these reactions in an environmentally friendly way. Fermentations are carried out in aqueous solutions at ambient temperatures and pressures and generally don't produce toxic byproducts.

Microbial metabolism has evolved a so called "Bow-Tie" structure to support growth on a wide variety of substrates [25]. Various carbon sources are assimilated and processed to generate a smaller group of common metabolites which can then be recombined to form an abundance of chemically complex and diverse molecules. As an example, the common laboratory bacteria, *E. coli*, can support growth on numerous carbohydrates such as mannose, xylose, glycerol, maltose, galactose, fructose, lactose, and glucose [26,27]. These sugars can be generated from mechanical processing of a variety of feedstocks. The highest yielding feedstocks include edible food crops such as rice, wheat, barley, potato, corn, and sugarcane. However, competition with food supply and land use has spurred interest into other alternatives such as the widely abundant lignocellulosic biomass [28]. Although this carbon source can grow on poor quality land with less water and fertilizer than food crops, more sophisticated processing is required to release fermentable sugars [29]. Aside from carbohydrates, other small molecules can support *E. coli* growth including succinate, pyruvate, oxaloacetate, proteogenic amino acids, and fatty acids [26,30]. Autotrophic organisms have the unique metabolic potential to utilize CO_2 as a carbon source, which could play a significant role in decarbonization of the atmosphere. In fact, organisms from the *Clostridia* genus have been commercially used to convert industrial waste gas streams from steel mills to value added alcohols like ethanol and isopropanol [21]. Microbial utilization of other C1 feedstocks such as methane, methanol, CO and formate has also been explored [31]. The scope of utilizable substrates is expanded further when considering microbial communities. Studies have shown that co-cultures of lactic acid bacteria in combination with the common laboratory yeast, *Saccharomyces cerevisiae,* can use macroalgae as a feedstock for lactic acid fermentation [32]. There have even been preliminary reports on the microbial degradation of polystyrene, suggesting that plastic waste streams could potentially support microbial chemical production [33].

In addition, microbes are experts of stereochemistry. Whole-cell catalysis has found a place in the commercial production of poly(lactic acid) (PLA), a common bioplastic. The precursor to PLA, lactate can be generated through chemical synthesis or fermentation. Chemical synthesis generates a racemic mixture of the two lactate enantiomers, whereas optically pure lactate can be achieved through microbial fermentation [34]. Thus, whole-cell catalysis is the preferred source of lactate for PLA production [17]. Additionally, whole-cell catalysis has been employed in the production

of paclitaxel, an antitumor agent with 8 chiral centers. Although paclitaxel can be obtained from extraction of the Pacific yew tree, *Taxus brevifolia,* harvesting trees was considered unsustainable [22,35,36]. Conversely, published chemical synthesis strategies required approximately 40 steps to achieve the complex stereochemistry with an overall yield of 2% [37,38]. Bristol-Myers-Squib now solely employs plant cell fermentation for production of paclitaxel, supporting cell growth with only sugars, amino acids, vitamins and trace elements [39]. Whole-cell catalysis has also been utilized for the production of the antimalarial drug, artemisinin. Artemisinin can be extracted from Chinese sweet wormwood (*Artemisia annua*), but seasonal supply and consistent demand for this drug generated interest in other sources. A total chemical synthesis was developed for artemisinin, addressing the challenging regio- and stereochemical requirements [40]. However, chemical synthesis alone is not economically feasible [19]. To address this need, *Saccharomyces cerevisiae* was metabolically engineered to generate artemisinic acid with the correct stereochemistry that could be further chemically modified to yield artemisinin [41]. More recently, production of artemisinin has also been achieved with transgenic plant whole-cell catalysts [42]. These examples highlight the utility of whole-cell catalysts for complex regio- and stereochemistry that is difficult to achieve by traditional synthetic approaches.

Whole-cell catalysts are also environmentally friendly alternatives to traditional chemical production methods. As discussed earlier, plant cell fermentation is now used for paclitaxel production. This process has replaced the extraction from plant material, which required the use of multiple hazardous, often toxic organic solvents not easily amenable to recycling, and energy-intensive processing. Additionally, the six intermediates typically required in semisynthesis as well as ten solvents and six drying steps were eliminated [39]. As another example, for almost five decades starting from the mid-1930s, synthetic routes were the only commercial method for riboflavin production. This synthesis involved six to eight chemical steps from glucose, several of which used toxic agents and produced waste products which required specialized effluent treatment and environmental control. Now, riboflavin is commercially produced via microbial fermentation with either the fungal species, *Ashbya gossypii*, or *Bacillus subtilis* in a more economically and environmentally conscious manner [16]. Fermentation-based methods in *E. coli* for the production of 1,4-butanediol have been shown to reduce total CO_2-equivalent emissions/kg BDO by 68% and fossil energy usage by 67% relative to petrochemical processes [43]. Similarly, a bio-based approach to PDO production results in 40% less energy consumption and 20% reduction in greenhouse gas emissions than production of petrochemical-derived PDO [44]. Whole-cell catalysts have also been used to generate fertilizer for food crops. This represents a unique approach where the catalyst, an engineered strain of *Klebsiella varicola,* is introduced at the site of end use. This microbe was shown to colonize the rhizosphere of crops and generate ammonia from atmospheric nitrogen resulting in its slow and sustained release. This technology is predicted to offset the estimated $200 billion in environmental damages caused by nitrogen fertilizer run off in addition to the large carbon footprint of nitrogen generation via the Haber-Bosch process. As was shown here, whole-cell catalysts can greatly reduce the environmental impact of chemical production.

1.3. Technologies assisting development of whole-cell catalysts

Thus far we have given both historical and contemporary uses of whole-cell catalysts to achieve chemical transformations. A number of the catalysts above were developed through metabolic engineering. The Design-Build-Test-Learn cycle is often borrowed from classical

engineering and used as a thought framework to guide metabolic engineering efforts (*Figure 1.2*) [45–47]. We will now examine some technology developments that have assisted each arm of the cycle.

The cycle begins with design. One facet of attempts to hijack living organisms to synthesize the molecules that we desire, is to fully understand the microbial host. Despite their lack of complexity compared to higher organisms, we still do not fully understand the intricacies of microbes. *E. coli* is the most studied organism, yet we do not know the function of 30% to 40% of protein coding sequences in the genome [48]. Accurate models describing cellular metabolism would allow metabolic engineers to optimize host metabolism for chemical synthesis *in silico*, prior to performing experiments. Metabolic models generally fall into two catagories, constraint-based or kinetic. Constraint-based models are built from reaction stoichiometry and thermodynamic information, while kinetic models take into consideration biochemical network stoichiometry, kinetic parameters and enzyme concentrations. Given their theoretical foundation, kinetic models hold the potential to offer an even more detailed look into microbial metabolism and provide the opportunity to integrate more types of omics data than can be handled using constraint-based models [53]. However, kinetic models of metabolism are still in their nascent stage and computationally intensive. As computing power improves, we will likely see kinetic models play a more significant role in the design of microbial catalysts. However, constraint-based models have proven to provide metabolic engineers with metabolic insight. These models have been successfully applied to guide genomic manipulations to improve production of several whole-cell catalysts including dodecanedioic acid production in *Yarrowia lipolytica*, production of aromatic polyesters in *E. coli,* and itaconic acid in *E. coli* among others [49–52]. Using constraint-based models, the algorithm, OptKnock, was developed to predict sets of gene deletions to optimize production of specific metabolites and has been used to establish growth coupling necessary for adaptive laboratory evolution of 1,4-butanediol production in *E. coli* [43,54]. Aside from metabolism, new tools have been added to the metabolic engineer's toolbox for pathway construction [55]. RetroPath was designed to help create metabolic retrosynthetic approaches while Selenzyme provides a tool to mine for pathway candidate enzymes [56,57]. As these tools develop, they will undoubtably become essential to the design process of metabolic engineering.

With respect to the build step of the cycle, advances in DNA synthesis technology have been impactful. The scope and goals of synthetic biologists has expanded beyond the need for short oligonucleotides used for priming and detection. There is now a necessity to create whole genes, including those from metagenomic data for which no template for replication exists. Recently, there has been significant interest, and funding, in companies pursuing template-free enzymatic DNA synthesis using technology built around the enzyme terminal deoxynucleotidyl transferase [58]. This technology has the potential to synthesize oligonucleotides with improved accuracy over existing phosphoramidite-based chemical DNA synthesis, allowing for accessible creation of longer, hi-fidelity DNA sequences. New developments in DNA synthesis will further increase the throughput of the build step and allow synthetic biologists and metabolic engineers access to a higher capacity of DNA constructs to test and learn from.

Another aspect of the build step is performing genetic manipulations. Twenty years ago, genetic elements of the lambda bacteriophage were leveraged for genomic manipulations to perform every single non-essential gene deletion in *E. coli* [59,60]. Since then, homologous

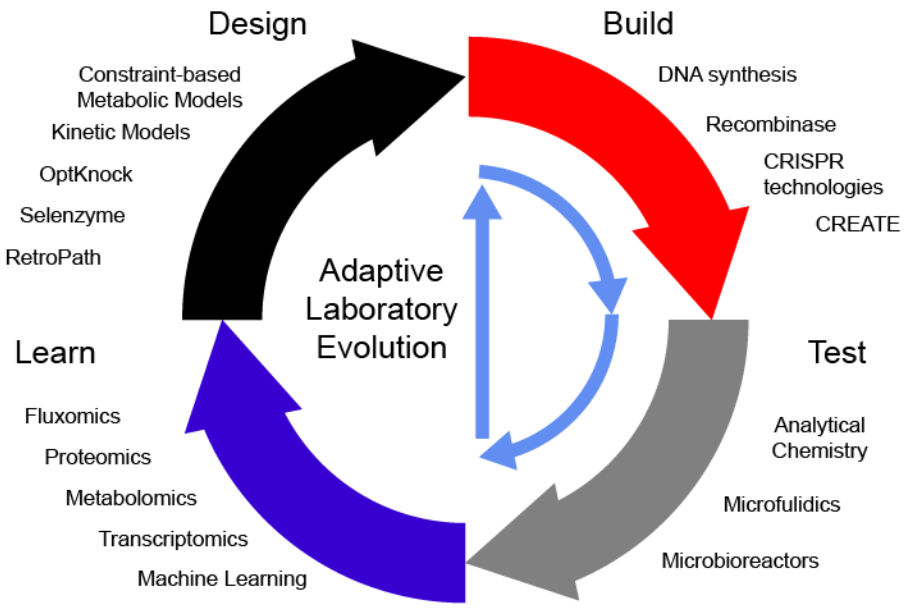


Figure 1.2. *The Design-Build-Test-Learn cycle. Adaptive Laboratory Evolution is comprised of many unsupervised iterations of the Build and Test cycles.*

enzymes have been discovered with higher efficiencies of recombination allowing for even easier genetic manipulation across diverse Gammaproteobacteria [61]. The discovery of CRISPR related proteins has made a huge impact on the field giving the capacity for multiplexed genome engineering [62,63]. CRISPR-enabled trackable genome engineering (CREATE) provides a method to create large genomic mutant libraries and has been used for protein engineering and adaptive laboratory evolution [64].

After constructing desired DNA sequences or creating genomic manipulations, the designed whole-cell catalysts then must be tested for performance. Advances in microbioreactor systems have drastically improved throughput [65]. Such systems can then be integrated with liquid handling robots to facilitate sampling for thorough phenotyping. However, scale-down culture conditions must be carefully considered so that selected genotypes perform well in large scale fermentation [66]. Throughput can be improved even further through the use of microfluidics at the pL volume scale. Microfluidic devices allow for the study of whole-cell catalysts at the single-cell level to assess system heterogeneity [67].

Data collected from testing strain performance can then be compared to expected outcomes to learn from the experiment and guide future designs. Beyond basic metrics such as product titer, growth and yield, multiple types of omics data can be collected to inform metabolic engineers about the intricacies of their metabolic host. Metabolomics is the quantification of the small molecules cellular metabolism is comprised of [68]. Fluxomics is used to quantify the relative rates chemical reactions within cellular metabolism [69]. Both metabolomics and fluxomics can help identify pathway bottlenecks and unique phenotypes [70]. Proteomics, or the measurement of protein intracellular protein concentration and transcriptomics provide snapshots of the cellular gene expression state and prove to be valuable tools for metabolic engineers [71,72]. Transcriptomics has been used for identification of gene targets by examining the transcriptional

response associated with a desirable phenotype [73]. Combined data sets from metabolomics and proteomics allowed Brunk et al. to analyze and learn from eight strains engineered for production of biofuels [74]. They were able to identify the roles of candidate genes and pathways and use this knowledge to construct a mutant strain with improved productivity. It is predicted that machine learning will have large impacts on analysis of these complex data sets and help metabolic engineers learn about whole-cell catalysts [75].

Adaptive laboratory evolution (ALE) serves as a process that condenses many iterations of the build and test portions of the cycle. During ALE, a microorganism is cultured under clearly defined conditions for prolonged periods of time, within the range of weeks to years. Through prolonged growth, the microorganism accumulates mutations and improved phenotypes are selected for through growth competition. Next generation DNA sequencing and transcriptional profiling allow for phenotype-genotype correlations to be obtained through whole genome sequencing [76]. ALE has been applied to obtain many different growth phenotypes, including growth on minimal media, utilization of non-preferred substrates, and increased tolerance to environmental stress. With careful experimental design, ALE can also be used to optimize metabolite production and increase titer [77–82]. Evolved strains can then be examined using omics approaches to learn from adaptive strategies employed by the cell.

1.4. Thesis organization

Within this work, we explored the optimization of microbial whole-cell catalysts for production of two C_4 commodity chemicals, *n*-butanol and (*R*)-1,3-butanediol. *n*-Butanol has similar properties to gasoline allowing it to serve as an alternative fuel source [83]. (*R*)-1,3-butanediol is an important chiral alcohol used for the synthesis of pheromones, fragrances and insecticides [84]. Even more, dehydration of *n*-butanol and (*R*)-1,3-butanediol yields the C_4 monomers 1-butene and butadiene respectively [85,86]. Biosynthetic pathways were constructed for production of these target molecules in *E. coli*. Through genomic manipulation, anaerobic growth was coupled to product titers, allowing us to employ ALE to improve these catalysts. Chapters 2 and 3 of this thesis seek to understand how the mutations acquired through ALE confer the improved phenotype. In Chapter 2, we thoroughly characterized anaerobic growth of the evolved strains of *E. coli* optimized for *n*-butanol production. Growth characterization revealed to us that the parent strain for ALE is not capable of sustained anaerobic *n*-butanol fermentation and suggested that metabolic burden may have a significant effect on growth. In Chapter 3, we utilized a systems level analysis in the form of transcriptomics. From this data, we developed a hypothesis that the evolved strains have reduced metabolic burden associated with pathway gene expression and that this has been achieved by maintaining the plasmids carrying our metabolic pathway at lower copy numbers. We then verified this hypothesis through further investigations into the level of pathway expression. The lessons learned from our adaptive evolution experiment add to the body of literature outlining the significant impact that metabolic burden can have on productivity of whole-cell catalysts and what strategies are employed by the metabolic host to overcome these challenges.

In Chapter 4, we turned our focus to the microbial production of polyesters belonging to the class of polyhydroxyalkanoates (PHAs). We devised a metabolic pathway for the incorporation of unsaturated monomers derived from noncanonical amino acids into PHAs within *E. coli*. Through literature searches, we identified enzyme candidates for our pathway and verified activity with our substrates of interest. Experiments outlined here lay the foundation for construction of a microbial

catalyst capable of generating PHAs decorated with unsaturated extender units that can serve as chemical handles for downstream modification.

1.5. References

1. Gogineni, V. K. Probiotics: History and evolution. in *J. Anc. Dis. Prev. Remedies* **01,** (2013).
2. Samuel, D. Investigation of ancient Egyptian baking and brewing bethods by correlative nicroscopy. *Science* **273,** 488–490 (1996).
3. Gijzen, H. J. Anaerobic digestion for sustainable development: a natural approach. *Water Sci. Technol.* **45,** 321–328 (2002).
4. Sauer, M. Industrial production of acetone and butanol by fermentation—100 years later. *FEMS Microbiol. Lett.* **363,** fnw134 (2016).
5. Kubicek, C. P., Röhr, M. & Rehm, H. J. Citric acid fermentation. *Crit. Rev. Biotechnol.* **3,** 331–373 (1985).
6. Penicillin production through deep-tank fermentation - National Historic Chemical Landmark. *Am. Chem. Soc.* at <https://www.acs.org/education/whatischemistry/landmarks/penicillin.html>
7. Elander, R. P. Industrial production of β-lactam antibiotics. *Appl. Microbiol. Biotechnol.* **61,** 385–392 (2003).
8. Boer, C. D. & Peterson, D. H. Geldanamycin and process for producing same. (1971). at <https://patents.google.com/patent/US3595955A/en?q=(antibiotic)&q=(production)>
9. Wu, J. *et al.* Toward Improvement of erythromycin A production in an industrial *Saccharopolyspora erythraea* strain via facilitation of genetic manipulation with an artificial *attB* site for specific recombination. *Appl. Environ. Microbiol.* **77,** 7508–7516 (2011).
10. França, L. T. C., Carrilho, E. & Kist, T. B. L. A review of DNA sequencing techniques. *Q. Rev. Biophys.* **35,** 169–200 (2002).
11. Berg, P. & Mertz, J. E. Personal reflections on the origins and emergence of recombinant DNA technology. *Genetics* **184,** 9–17 (2010).
12. Ensley, B. D. *et al.* Expression of naphthalene oxidation genes in Escherichia coli results in the biosynthesis of indigo. *Science* **222,** 167–169 (1983).
13. Akuzawa, R., Miura, T. & Surono, I. S. Asian fermented milks. *Encycl. Dairy Sci.* 507–511 (2011). doi:10.1016/B978-0-12-374407-4.00186-2
14. Lahue, C., Madden, A. A., Dunn, R. R. & Smukowski Heil, C. History and domestication of *Saccharomyces cerevisiae* in bread baking. *Front. Genet.* **11,** 584718 (2020).
15. Rise of the anaerobic digester. *Renew. Energy Focus* **9,** 28–34 (2008).
16. Revuelta, J. L. *et al.* Bioproduction of riboflavin: a bright yellow history. *J. Ind. Microbiol. Biotechnol.* **44,** 659–665 (2017).
17. Huang, S. *et al.* A Review of the recent developments in the bioproduction of polylactic acid and its precursors optically pure lactic acids. *Molecules* **26,** 6446 (2021).

18. RSS-Feed, bio-based eu/news-das F. für bio-basierte Ö., Biowerkstoffe und Industrielle Biotechnologie-. U.S.A.: World's first propanediol production from corn sugar opened. *Renew. Carbon News* (2007). at <https://renewable-carbon.eu/news/u-s-a-worlds-first-propanediol-production-from-corn-sugar-opened/>

19. Paddon, C. J. & Keasling, J. D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* **12,** 355–367 (2014).

20. Fraser, R. Z., Shitut, M., Agrawal, P., Mendes, O. & Klapholz, S. Safety evaluation of soy leghemoglobin protein preparation derived from *Pichia pastoris*, intended for use as a flavor catalyst in plant-based meat. *Int. J. Toxicol.* **37,** 241–262 (2018).

21. Handler, R. M., Shonnard, D. R., Griffing, E. M., Lai, A. & Palou-Rivera, I. Life cycle assessments of ethanol production via gas fermentation: anticipated greenhouse gas emissions for cellulosic and waste gas feedstocks. *Ind. Eng. Chem. Res.* **55,** 3253–3261 (2016).

22. Walsh, V. & Goodman, J. From taxol to Taxol: the changing identities and ownership of an anti-cancer drug. *Med. Anthropol.* **21,** 307–336 (2002).

23. Bailey, J. E. Toward a science of metabolic engineering. *Science* **252,** 1668–1675 (1991).

24. Anderson, S. *et al.* Production of 2-keto-L-gulonate, an intermediate in L-ascorbate synthesis, by a genetically modified *Erwinia herbicola*. *Science* **230,** 144–149 (1985).

25. Csete, M. & Doyle, J. Bow ties, metabolism and disease. *Trends Biotechnol.* **22,** 446–450 (2004).

26. Hermsen, R., Okano, H., You, C., Werner, N. & Hwa, T. A growth-rate composition formula for the growth of *E. coli* on co-utilized carbon substrates. *Mol. Syst. Biol.* **11,** 801 (2015).

27. Gottschalk, G. in *Bact. Metab.* (ed. Gottschalk, G.) 81–87 (Springer US, 1979). doi:10.1007/978-1-4684-0465-4_4

28. Kim, S. & Dale, B. E. Global potential bioethanol production from wasted crops and crop residues. *Biomass Bioenergy* **26,** 361–375 (2004).

29. Halder, P., Azad, K., Shah, S. & Sarker, E. in *Adv. Eco-Fuels Sustain. Environ.* (ed. Azad, K.) 211–236 (Woodhead Publishing, 2019). doi:10.1016/B978-0-08-102728-8.00008-5

30. Maser, A., Peebo, K., Vilu, R. & Nahku, R. Amino acids are key substrates to *Escherichia coli* BW25113 for achieving high specific growth rate. *Res. Microbiol.* **171,** 185–193 (2020).

31. Jiang, W. *et al.* Metabolic engineering strategies to enable microbial utilization of C_1 feedstocks. *Nat. Chem. Biol.* **17,** 845–855 (2021).

32. Uchida, M. & Miyoshi, T. Algal fermentation—the seed for a new fermentation industry of foods and related products. (2013).

33. Yang, Y. *et al.* Biodegradation and mineralization of polystyrene by plastic-eating mealworms: part 1. chemical and physical characterization and isotopic tests. *Environ. Sci. Technol.* **49,** 12080–12086 (2015).

34. Rawoof, S. A. A. *et al.* Production of optically pure lactic acid by microbial fermentation: a review. *Environ. Chem. Lett.* **19,** 539–556 (2021).

35. Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P. & McPhail, A. T. Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* **93,** 2325–2327 (1971).

36. Walsh, V. & Goodman, J. The billion dollar molecule: Taxol in historical and theoretical perspective. *Clio Medica Amst. Neth.* **66,** 245–267 (2002).

37. Holton, R. A. *et al.* First total synthesis of taxol. 1. Functionalization of the B ring. *J. Am. Chem. Soc.* **116,** 1597–1598 (1994).

38. Nicolaou, K. C. *et al.* Total synthesis of taxol. *Nature* **367,** 630–634 (1994).

39. US EPA, O. Presidential green chemistry challenge: 2004 Greener Synthetic Pathways Award. (2013). at <https://www.epa.gov/greenchemistry/presidential-green-chemistry-challenge-2004-greener-synthetic-pathways-award>

40. Schmid, G. & Hofheinz, W. Total synthesis of qinghaosu. *J. Am. Chem. Soc.* **105,** 624–625 (1983).

41. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496,** 528–532 (2013).

42. Zhao, L. *et al.* From plant to yeast—advances in biosynthesis of artemisinin. *Molecules* **27,** 6888 (2022).

43. Burgard, A., Burk, M. J., Osterhout, R., Van Dien, S. & Yim, H. Development of a commercial scale process for production of 1,4-butanediol from sugar. *Curr. Opin. Biotechnol.* **42,** 118–125 (2016).

44. Eichhorn, S., Hearle, J. W. S., Jaffe, M. & Kikutani, T. *Handbook of Textile Fibre Structure: Volume 1: Fundamentals and Manufactured Polymer Fibres*. (Elsevier, 2009).

45. Jarboe, L. R. Improving the success and impact of the metabolic engineering design, build, test, learn cycle by addressing proteins of unknown function. *Curr. Opin. Biotechnol.* **53,** 93–98 (2018).

46. Whitford, C. M., Cruz-Morales, P., Keasling, J. D. & Weber, T. The Design-Build-Test-Learn cycle for metabolic engineering of *Streptomycetes*. *Essays Biochem.* **65,** 261–275 (2021).

47. IV, G., Nanjannavar, P., Miller, E. & Fong, S. Integrative metabolic engineering. *AIMS Bioeng.* **2,** 93–103 (2015).

48. Ghatak, S., King, Z. A., Sastry, A. & Palsson, B. O. The y-ome defines the 35% of *Escherichia coli* genes that lack experimental evidence of function. *Nucleic Acids Res.* **47,** 2446–2454 (2019).

49. Yang, J. E. *et al.* One-step fermentative production of aromatic polyesters from glucose by metabolically engineered *Escherichia coli* strains. *Nat. Commun.* **9,** 79 (2018).

50. Mishra, P. *et al.* Genome-scale model-driven strain design for dicarboxylic acid production in *Yarrowia lipolytica*. *BMC Syst. Biol.* **12,** 12 (2018).

51. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol.* **20,** 121 (2019).

52. Harder, B.-J., Bettenbrock, K. & Klamt, S. Model-based metabolic engineering enables high yield itaconic acid production by *Escherichia coli*. *Metab. Eng.* **38,** 29–37 (2016).

53. Islam, M. M., Schroeder, W. L. & Saha, R. Kinetic modeling of metabolism: Present and future. *Curr. Opin. Syst. Biol.* **26,** 72–78 (2021).

54. Yim, H. *et al.* Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **7,** 445–452 (2011).

55. Wang, L., Dash, S., Ng, C. Y. & Maranas, C. D. A review of computational tools for design and reconstruction of metabolic pathways. *Synth. Syst. Biotechnol.* **2,** 243–252 (2017).

56. Carbonell, P. *et al.* Selenzyme: enzyme selection tool for pathway design. *Bioinformatics* **34,** 2153–2154 (2018).

57. Delépine, B., Duigou, T., Carbonell, P. & Faulon, J.-L. RetroPath2.0: A retrosynthesis workflow for metabolic engineers. *Metab. Eng.* **45,** 158–170 (2018).

58. Eisenstein, M. Enzymatic DNA synthesis enters new phase. *Nat. Biotechnol.* **38,** 1113–1115 (2020).

59. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97,** 6640–6645 (2000).

60. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008 (2006).

61. Wannier, T. M. *et al.* Improved bacterial recombineering by parallelized protein discovery. *Proc. Natl. Acad. Sci.* **117,** 13689–13698 (2020).

62. Campa, C. C., Weisbach, N. R., Santinha, A. J., Incarnato, D. & Platt, R. J. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat. Methods* **16,** 887–893 (2019).

63. Jiang, Y. *et al.* Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81,** 2506–2514 (2015).

64. Garst, A. D. *et al.* Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.* **35,** 48–55 (2017).

65. Rienzo, M. *et al.* High-throughput screening for high-efficiency small-molecule biosynthesis. *Metab. Eng.* **63,** 102–125 (2021).

66. Crater, J. S. & Lievense, J. C. Scale-up of industrial microbial processes. *FEMS Microbiol. Lett.* **365,** fny138 (2018).

67. Hemmerich, J., Noack, S., Wiechert, W. & Oldiges, M. Microbioreactor systems for accelerated bioprocess development. *Biotechnol. J.* **13,** 1700141 (2018).

68. Baidoo, E. E. K. & Teixeira Benites, V. in *Microb. Metabolomics Methods Protoc.* (ed. Baidoo, E. E. K.) 11–69 (Springer, 2019). doi:10.1007/978-1-4939-8757-3_2

69. Long, C. P. & Antoniewicz, M. R. High-resolution ^{13}C metabolic flux analysis. *Nat. Protoc.* **14,** 2856–2877 (2019).

70. Cheah, Y. E. *et al.* Systematic identification and elimination of flux bottlenecks in the aldehyde production pathway of *Synechococcus elongatus* PCC 7942. *Metab. Eng.* **60,** 56–65 (2020).

71. Kim, M., Park, B. G., Kim, J., Kim, J. Y. & Kim, B.-G. Exploiting transcriptomic data for metabolic engineering: toward a systematic strain design. *Curr. Opin. Biotechnol.* **54,** 26–32 (2018).

72. Yunus, I. S. & Lee, T. S. Applications of targeted proteomics in metabolic engineering: advances and opportunities. *Curr. Opin. Biotechnol.* **75,** 102709 (2022).

73. Kim, M., Park, B. G., Kim, J., Kim, J. Y. & Kim, B.-G. Exploiting transcriptomic data for metabolic engineering: toward a systematic strain design. *Curr. Opin. Biotechnol.* **54,** 26–32 (2018).

74. Brunk, E. *et al.* Characterizing strain variation in engineered *E. coli* using a multi-omics-based workflow. *Cell Syst.* **2,** 335–346 (2016).

75. Liao, X., Ma, H. & Tang, Y. J. Artificial intelligence: a solution to involution of design–build–test–learn cycle. *Curr. Opin. Biotechnol.* **75,** 102712 (2022).

76. Dragosits, M. & Mattanovich, D. Adaptive laboratory evolution – principles and applications for biotechnology. *Microb. Cell Factories* **12,** 64 (2013).

77. Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O. & Feist, A. M. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab. Eng.* **56,** 1–16 (2019).

78. Yomano, L. P., York, S. W. & Ingram, L. O. Isolation and characterization of ethanol-tolerant mutants of *Escherichia coli* KO11 for fuel ethanol production. *J. Ind. Microbiol. Biotechnol.* **20,** 132–138 (1998).

79. Zhang, X. *et al.* Metabolic evolution of energy-conserving pathways for succinate production in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 20180–20185 (2009).

80. Zhang, X., Jantama, K., Moore, J. C., Shanmugam, K. T. & Ingram, L. O. Production of l-alanine by metabolically engineered *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **77,** 355–366 (2007).

81. Pontrelli, S. *et al.* Directed strain evolution restructures metabolism for 1-butanol production in minimal media. *Metab. Eng.* **49,** 153–163 (2018).

82. Davis, M. A. *et al.* A Cellular platform for production of C4 monomers. 2023.01.09.523327 Preprint at https://doi.org/10.1101/2023.01.09.523327 (2023)

83. Kolesinska, B. *et al.* Butanol synthesis routes for biofuel production: trends and perspectives. *Materials* **12,** 350 (2019).

84. Zu, H. *et al.* Highly enantioselective synthesis of (*R*)-1,3-butanediol via deracemization of the corresponding racemate by a whole-cell stereoinverting cascade system. *Microb. Cell Factories* **19,** 125 (2020).

85. Morrow, N. L. The industrial production and use of 1,3-butadiene. *Environ. Health Perspect.* **86,** 7–8 (1990).

86. *Industrial biorenewables: a practical viewpoint*. (Wiley, 2016).

Chapter 2: *Physiological characterization of mutants identified through adaptive evolution*

Portions of this work were performed in collaboration with the following persons:

Vivian Y. Yu and Hongjun Dong performed genomic manipulations to create the single mutant strains. Albert Qiang assisted with fermentation and analysis.

2.1. Introduction

Whole-cell catalysis has emerged as a developing technology with the potential to replace existing chemical synthesis that is largely dependent on fossil fuels with fermentation-based approaches using renewable carbon sources derived from biomass. One example of whole-cell catalysis is the fermentation of glucose to generate butanol, a drop-in biofuel [1,2]. By combining enzymes from various organisms into a synthetic metabolic pathway, Bond-Watts et al. showed that *n*-butanol can be produced within the well-studied microbial chassis, *Escherichia coli*.

The *n*-butanol biosynthetic pathway begins with the central metabolite, acetyl-CoA (*Figure 2.1A*). Acetyl-CoA is primarily made through decarboxylation of pyruvate, the terminal metabolite of the glycolysis pathway [3]. The first enzyme in the pathway, acetoacetyl-CoA synthase (*phaA*), catalyzes the formation of acetoacetyl-CoA from two molecules of acetyl-CoA. Acetoacetyl-CoA is then reduced to (*S*)-3-hydroxybutyrate by *S*-specific NADH-dependent acetoacetyl-CoA dehydrogenase (*hbd*). (*S*)-3-hydroxybutyrate is dehydrated by crotonase (*crt*) to generate *trans*-enoyl-CoA. An exergonic reduction is then performed by the NADH dependent *trans*-enoyl-CoA reductase (*ter*), contributing to the overall driving force of the pathway. The next two enzymes, aldehyde dehydrogenase (*aldh46*) and alcohol dehydrogenase (*adh2*), reduce butyrl-CoA to butyraldehyde and then butyraldehyde to *n*-butanol respectively.

While screening aldehyde dehydrogenase candidates for this pathway, the by-product 1,3-butanediol (BDO) was discovered [4]. BDO likely arises from the reduction of an earlier pathway intermediate, 3-hydroxy-butyryl-CoA. BDO is another industrially relevant commodity chemical due to the fact that it can readily be dehydrated to butadiene, a chemical used in the production process of synthetic rubber [5]. Removing crotonase (*crt*) yields a pathway that no longer produces *n*-butanol. Exchanging the existing aldehyde dehyrdrogenase *aldh46* for *aldh7* that shows greater activity towards 3-hydroxy-butyryl-CoA creates a pathway that yields BDO production in high titers (*Figure 2.1B*).

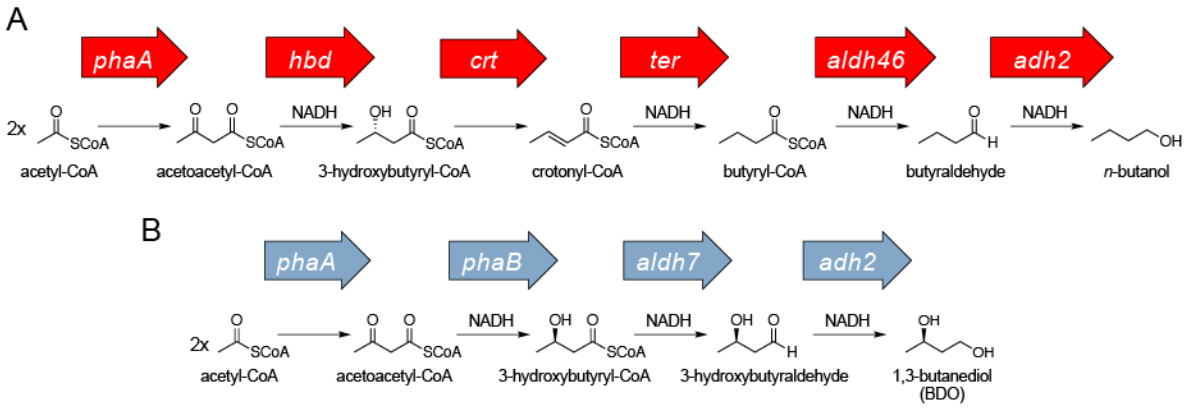


Figure 2.1 *Biosynthetic pathways for n-butanol and 1,3-butanediol. A) The butanol biosynthetic pathway is comprised of phaA (acetoacetyl-CoA synthase), hbd (S-specific NADH-dependent acetoacetyl-CoA dehydrogenase), crt (crotonase), ter (trans-enoyl-CoA reductase), aldh46 (aldehyde dehydrogenase), and adh2 (alcohol dehydrogenase). B) The 1,3-butanediol (BDO) pathway is comprised of phaA (acetoacetyl-CoA synthase), phaB (R-specific NAD(P)H-dependent acetoacetyl-CoA dehydrogenase), aldh7 (aldehyde dehydrogenase), and adh2 (alcohol dehydrogenase).*

Beyond pathway engineering, we looked to laboratory adaptive evolution as an approach to improve production of both *n*-butanol and butanediol. Perhaps the most challenging aspect of

laboratory evolution is establishing a coupling between growth and the desired phenotype. This is more straight forward in some cases than others. For example, seeking to improve tolerance to growth inhibitors [6–8] or to improve catabolism of a non-canonical substrates [9,10], part of the desired phenotype is improved growth and so the system is already set up for adaptive evolution [11]. The situation becomes more difficult when the phenotype that you want to select for is the production of a metabolite. One example of this type of adaptive evolution is the production of succinate [12]. Succinate is a natural fermentation product of *E. coli* when grown anaerobically on certain substrates, a fact that can be leveraged to create a growth selection.

E. coli is capable of growth in both aerobic and anaerobic conditions. During growth on carbohydrates, reducing equivalents in the form of NADH, FADH, and quinols are generated by the metabolic pathways of glycolysis and the TCA cycle [3]. These reduced cofactors shuttle electrons to the electron transport chain where they are eventually transferred to a terminal electron receptor, generating chemical energy through the process. In aerobic conditions, oxygen is available and serves as the preferred terminal electron acceptor due to its large reduction potential. In anaerobic conditions when oxygen is not available, alternative terminal electron acceptors can be used including nitrate, fumarate, or TMAO [13]. Even in the absence of any of these terminal electron acceptors, *E. coli* is still capable of growth. In such conditions, the metabolism of *E. coli* is altered to greatly diminish the production of reduced cofactors that no longer can be consumed by the electron transport chain [14]. The TCA cycle becomes inoperative and the activity of the pyruvate dehydrogenase complex (PDHc), which converts pyruvate to acetyl-CoA with concomitant generation of NADH, is greatly reduced. Chemical energy in the form of ATP is instead primarily produced through substrate-level phosphorylation which occurs during glycolysis at a rate of 2 moles ATP per moles of glucose. However, glycolysis also generates 2 NADH per glucose consumed. In order to provide glycolysis with sufficient levels of the NAD^+ cofactor for glycolysis to remain operational, NADH is recycled by reducing pyruvate, the terminal metabolite of glycolysis, to different fermentation products (*Figure 2.2*).

The major fermentation products of *E. coli* are ethanol, acetate, lactate, formate, and succinate to a lesser extent [14]. Lactate results from the direct reduction of pyruvate, where the other products are made through acetyl-CoA as an intermediate. In aerobic conditions, the PDHc is responsible for the conversion of pyruvate to acetyl-CoA. However, as mentioned earlier, this enzyme is no longer active in anaerobic conditions and this roll is instead filled by the enzyme pyruvate formate lyase (PFL). Instead of generating CO_2 and NADH like the PDHc, PFL instead produces formate as a byproduct, thereby not contributing to the reduced cofactor imbalance (*Figure 2.2*).

To create a growth selection, Ingram and coworkers took the approach of disrupting the other native fermentation pathways found in *E. coli*. Upon introducing genomic knockouts, they found that growth of this strain was coupled to production of the desired metabolite, succinate, in anaerobic conditions [12]. This approach has since been applied to various other fermentation products in *E. coli* [15–17].

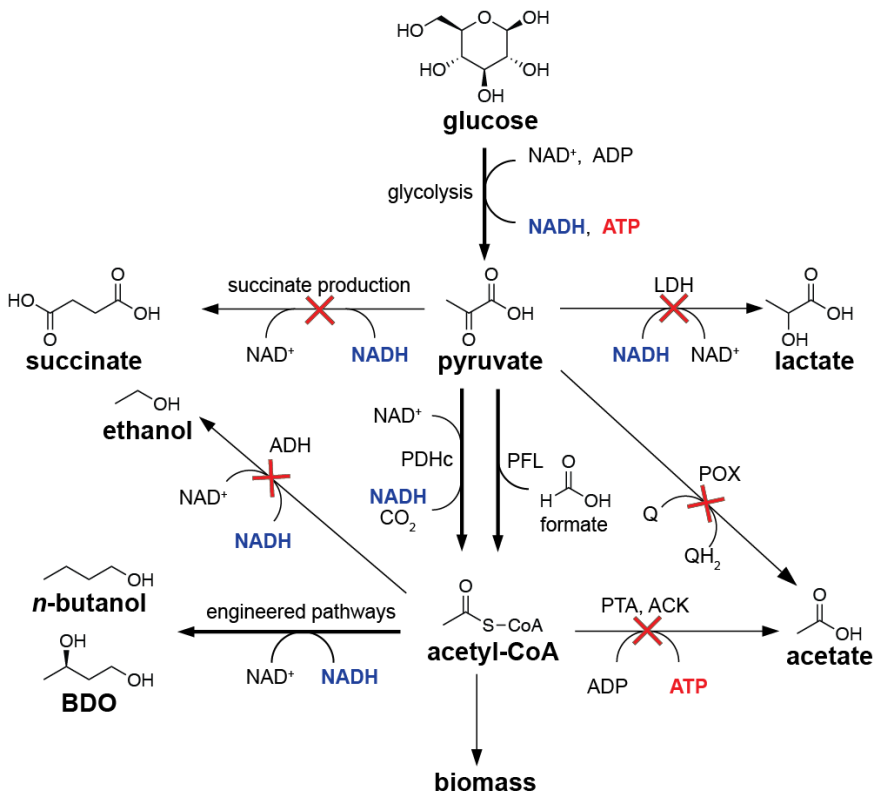


Figure 2.2. *The main metabolic pathways active in E. coli during anaerobic fermentation. Pathways and reactions that utilize NADH or ATP as a cofactor are highlighted. A series of genomic knockouts were introduced to make a strain that is unable to grow in the absence of oxygen. Introduction of an artificial biosynthetic pathway fills the role of native fermentation by oxidizing NADH and rescues anaerobic growth.*

Examining our biosynthetic pathways, we noted that NADH is oxidized through the reduction of a pyruvate derived metabolite, acetyl-CoA, which is strikingly similar to what occurs in native fermentation pathways. In the case of the *n*-butanol pathway, NADH is consumed at a ratio of 4 NADH per 2 acetyl-CoA. For the BDO pathway, NADH is consumed at a ratio of 3 NADH per 2 acetyl-CoA. The net result of glycolysis and PFL is the generation of 2 NADH per 2 acetyl-CoA [14]. Comparing this ratio to our biosynthetic pathways, we notice an imbalance where the biosynthetic pathways require more NADH per acetyl-CoA than what is natively provided. The 3 NADH per 2 acetyl-CoA ratio of the BDO pathway is close to the 2 NADH per 2 ATP generated through glycolysis and PFL, so it is feasible that this pathway can serve as a mechanism for redox balance during fermentation. However, the 4 NADH per 2 acetyl-CoA required for butanol production, will not be sufficient. This pathway has the same NADH to acetyl-CoA ratio as ethanol fermentation. Natively, 1 mol acetate is produced for every mol ethanol resulting in a net reaction of 2 acetyl-CoA with 2 NADH yielding 1 EtOH and 1 acetate, thereby matching the redox ratio resulting from glycolysis and PFL reactions. The redox ratio imbalance between the *n*-butanol pathway and native acetyl-CoA generation can be remedied if the PDHc is used instead of PFL for acetyl-CoA production. Glycolysis in conjunction with the PDHc yields a redox ratio of 4 NADH per 2 acetyl-CoA, perfectly matching the *n*-butanol pathway.

Seeing the success that Ingram and coworkers had with laboratory adaptive evolution using their fermentation-based growth selection, we decided to take an analogous approach. We

proceeded to disrupt genes in the major fermentation pathways [18]. *adhE* was deleted to inhibit ethanol fermentation, *ackA*, *pta* and *poxB* to interrupt acetate production, *ldhA* for lactate fermentation, and *frdBC,* to reduce succinate production. The knockouts were separately introduced into two strains of *E. coli*, DH1 and BW25113 yielding strains that we named DH1Δ5 and BWΔ5 respectively. These strains are dependent on the introduced biosynthetic pathways for functional fermentation. As outlined above, the genes encoding the PDHc were overexpressed in addition to the butanol pathway genes in order to satisfy redox balance during fermentation. We observed in the BWΔ5 strain expressing the PDHc and variants of the butanol pathway that titers were successfully coupled to anaerobic growth (*Figure 2.3A*). A strain containing a more active *n*-butanol pathway was enriched within a mixed culture also containing a strain with a less active pathway.

After establishing this growth selection, the adaptive evolution campaign was initiated. The strains were passaged for about two months in anaerobic conditions, allowing mutations to occur through the natural mutation rate. At the end of the evolution, evolved strains produced the desired products, butanol and BDO, at significantly higher titers than the parent strain (*Figure 2.3B*). The genomes of the evolved strains were sequenced to reveal the mutations that afforded an improved phenotype. Mutations in three genes, *pcnB*, *rne*, and *rpoC* were found to be frequently mutated among the evolved strains [19]. *pcnB* encodes for Poly(A)-polymerase which polyadenylates RNA species, an activity that has been shown to facilitate the breakdown of structured RNAs [20]. *rne* encodes for RNase E, an RNA endonuclease that initiates the decay of most mRNAs and is involved in the maturation of non-coding RNAs [21]. The gene *rpoC* codes for the β' subunit of RNA-polymerase, the essential enzyme complex responsible for transcribing DNA to RNA [22]. Curiously, all of the genes that arose as mutational hotspots in the evolved strains are related to RNA metabolism, either through the production or degradation of RNA. This result was surprising to us as a direct connection of these genes to fermentation was not immediately apparent.

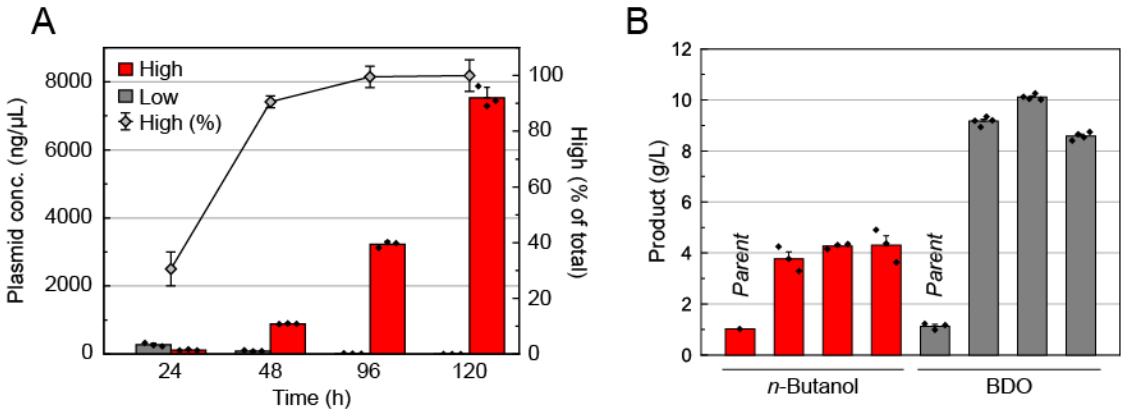


Figure 2.3. *Development of a genetic selection for evolving C_4 monomer synthesis in E. coli. **(**A) Enrichment of a high-producing n-butanol strain, DH1Δ5 pCDF3-ter.aldh46 pBT-0.3crt, was observed when mixed (0.1%) with a low-producing strain DH1Δ5 pCDF3-ter.aldh46 pBT-0.03HBD (99.9%). The mixed culture was grown anaerobically in TB with 2.5% (w/v) glucose media. Changes in the genetic population were monitored using qPCR. (B) Characterization of n-butanol and BDO production in evolved strains. The parent strain for butanol is BWΔ5 containing pPDHc, pBu1 and pBu2. Butanol evolution was carried out in M9 containing 10% (v/v) LB and 25 g/L glucose under anaerobic conditions. For BDO production, the parent strain was DH1Δ5 containing pPDHc, pBDO1 and pBDO2. Evolution was carried out in TB with 25 g/L glucose. Data are mean ± standard error of biological replicates. Figure adapted from Davis et al. [4].*

Metabolic engineering often follows a paradigm known as design-build-test-learn [23,24]. At the outset of the project, biosynthetic pathways were designed for the production of butanol and BDO. Plasmid constructs were then built containing genes coding for the enzymes in the pathway. Performance of these constructs were tested through production experiments and the results were used to inform further modifications [2,4]. The development of these biosynthetic pathways represents a cycle of design-build-test-learn. The cycle was then applied again, but this time at the organism level. A growth-based selection was designed by examining similarities to fermentation. A series of knockouts from the design were then introduced into the microbial host. The growth-based selection was then implemented, yielding a series of evolved strains with an improved phenotype [4]. In order to complete the cycle, we must learn from our adaptive evolution. Understanding of how microbial mechanism has shifted to promote butanol and BDO production may provide insight applicable to other metabolic engineering efforts or reveal new biology. Thus, we set out to understand how cellular metabolism had shifted within the mutant strains to promote butanol and BDO production and how mutations to *pcnB*, *rne* and *rpoC* yield such an effect.

2.2. Materials and methods

Commercial materials. Ammonium heptamolybdate tetrahydrate, boric acid, cobalt(II) chloride, hexanol, iron(II) sulfate heptahydrate, sodium lactate, sodium pyruvate, thiamine hydrochloride, valeric acid, 2-mercaptoethanol, 2,4-pentanediol, 1,3-butanediol, were purchased from Sigma-Aldrich (St. Louis, MO). Ammonium bicarbonate (LC-MS grade), ammonium chloride, calcium chloride dihydrate, cupric sulfate, carbenicillin disodium salt (Cb), deoxynucleotides (dNTPs), formic acid (LC-MS grade), kanamycin sulfate (Km), chloramphenicol (Cm), manganese(II) chloride, O'GeneRuler 1 kb Plus DNA Ladder, PageRuler Plus Prestained Protein Ladder, polyethylene glycol 3350, potassium chloride, potassium phosphate monobasic, sodium acetate, sodium chloride, sodium hydroxide, sodium phosphate dibasic heptahydrate, were purchased from Thermo Fisher Scientific (Waltham, MA). Glycerol, LB Miller agar, LB Miller broth, Terrific broth, anhydrous magnesium sulfate, and butanol were purchased from EMD-Millipore (Burlington, MA). Glucose was purchased from MP Biomedicals (Solon, OH). Zinc sulfate was purchased from Mallinckrodt (Paris, OH). Isopropyl ß-D-1-thiogalactopyranoside (IPTG) was purchased from Santa Cruz Biotechnology (Dallas, TX). Ethidium bromide was purchased from Bio-Rad Laboratories (Hercules, CA). Phusion polymerase, Phusion HF buffer, all restriction enzymes, restriction enzyme buffer (CutSmart), T4-PNK, T4 DNA ligase, and Taq ligase were from New England Biolabs (Ipswitch, MA). DNA purification kits were purchased from Qiagen (Redwood City, CA). Oligonucleotides and gBlocks gene fragments were synthesized by Integrated DNA Technologies (Coralville, IA).

Bacterial Strains. *E. coli* DH10B was used for DNA construction. *E. coli* DH1 (ATCC 39936), DH1Δ5, DH1Δ5 *pcnB*(R149H), DH1Δ5 *pcn*(BR149P), DH1Δ5 *pcnB*(R373S), DH1Δ5 *rpoC*(M466L), DH1Δ5 *rpoC*(K1192E), BW25113, BWΔ5, BWΔ5 *pcnB*(E108A), BWΔ5 *pcnB*(N138H), *pcnB*(L392fs), BWΔ5 *rne*(R488H V489L), and BWΔ5 Δ*pcnB* were used for fermentation experiments.

Gene and plasmid construction. Plasmid construction was carried out using standard molecular biology techniques including the Gibson protocol [25], Golden Gate Assembly [26], or quick-change PCR [27]. PCR amplifications were carried out with Phusion DNA polymerase or

GoTaq DNA polymerase, following manufacturer instructions. Primer sequences are listed in *Appendix 2.1A*. Constructs were verified by sequencing (Genewiz; South Plainfield, NJ).

pCRISPR-Gibson1 was constructed to clone constructs with specific guide sequence to target the *E. coli* chromosome for introduction of point mutants. The parent plasmid, pCRISPR-Gibson1 (#2786), was generated from pCRISPR (Addgene 42875) to introduce cut sites between sgRNA promoter and the sgRNA to facilitate the use of Gibson assembly to introduce guide sequences for the target DNA. PCR amplicons of pCRISPR using primer sets VY01/VY02 and VY03/VY04 were inserted into pCRISPR linearized with AvrII and XbaI via Gibson assembly.

pCRISPR-PcnB2409 was constructed by insertion of the annealed oligonucleotides, VY05 and VY06 into the XbaI-HindIII site of pCRISPR-Gibson1 using the Gibson protocol.

pBut. The first half of the *n*-butanol pathway, comprising genes phaA, hbd, and crt, was amplified from pBu1 using primers EK28 and EK29. The amplicon was inserted into pBu2 digested with BstZ17I via Gibson assembly.

pBAC-gg. About 1kb of genomic context 5' to *pcnB* was amplified from *E. coli* BW25113 genome with primers EK16 and EK17. About 1kb of genomic context 3' to *pcnB* was amplified from *E. coli* BW25113 genome with primers EK18 and EK19. The two amplicons were inserted into pBAC-lacZ digested with BstEII and EcoRI via Gibson assembly.

pBAC-pcnB. The pcnB gene amplified from *E. coli* BW25113 genome with primers EK20 and EK21 was inserted into pBAC-gg digested with BsaI via golden gate assembly.

pBAC-pcnB E108A. The pcnB gene amplified from *E. coli* BWΔ5 pcnB E108A genome with primers EK20 and EK21 was inserted into pBAC-gg digested with BsaI via golden gate assembly.

pBAC-pcnB N138H. The pcnB gene amplified from *E. coli* BWΔ5 pcnB N138H genome with primers EK20 and EK21 was inserted into pBAC-gg digested with BsaI via golden gate assembly.

pGFP. sfGFP was amplified from pY71.sfGFP using primers EK58 and EK59. The backbone was amplified from pBu2 with primers EK56 and EK57. The two pieces were combined via Gibson assembly.

Batch production of *n*-butanol and 1,3-butanediol. Overnight cultures of freshly transformed E. coli strains were grown overnight in LB at 37 °C and used to inoculate LB (25 ml) with 25 g/L glucose for *n*-butanol, or TB (25 ml) with 25 g/L glucose for 1,3-butanediol, with appropriate antibiotics to an optical density at 600 nm OD600 of 0.05 in a 250 mL-baffled anaerobic flask with GL45 threaded top (Chemglass). In some cases where noted, cultures were grown in LB (10 ml) with 25 g/L glucose in a 25ml Balch tube (Chemglass). The cultures were grown at 30 °C in a rotary shaker (200 rpm) for 3h before induction with IPTG (1.0 mM). Anaerobic cultures were sealed and the headspace was sparged with argon for 3 min immediately following induction. *n*-Butanol samples were collected after 3 d of cell culture. 1,3-Butanediol samples were collected after 5 days of cell culture. Time point samples were collected with a nitrogen sparges syringe.

Quantification of *n*-butanol titers. Samples (1 mL) were removed from cell culture and cleared of biomass by centrifugation at 14,000 × *g* for 5 min using an Eppendorf 5417R centrifuge. The supernatant or cleared medium sample was then (80 μL) extracted with toluene (80 μL)

containing the *n*-hexanol internal standard (1000 mg mL^{-1}). The sample was mixed using a STD VORTEX MIXER (Fisher) for 30s set at 8. The organic layer (25 µL) was diluted four-fold with toluene (75 uL). Diluted samples were ran on an Agilent 7890A GC using an HP-5MS column (0.25 mm x 30 m, 0.25 µM film thickness, J & W Scientific). The oven program was as follows: 95 °C for 4 min, ramp to 300 at 50 °C min^{-1}, 300 °C for 2 min. The injector temperature was set to 250 °C. Split injections wer performed with a split ratio of 20:1. The GC was equipped with an Agilent 5975C MSD and was ran in scan mode (*m/z* 35 – 150). *n*-Butanol and n-hexanol were quantified using the extracted ion chromatograms (*m/z* 56.1). Samples were quantified relative to a standard curve of 10, 5, 2.5, 1.25, 0.625 g L^{-1} *n*-butanol prepared freshly for each run. All standards and samples were normalized for injection volume using the *n*-hexanol internal standard.

Quantification of 1,3-butanediol (BDO). Samples (1 mL) were removed from cell culture and cleared of biomass by centrifugation at 20,817 × *g* for 2 min using an Eppendorf 5417R centrifuge. The cleared medium samples, or standards prepared in TB medium, diluted 1:1000 into water, and filtered through a 0.22 µm filter (EMD Millipore MSGVN2210). Supernatants were diluted 1- to 1,000-fold with water containing 2,4-pentanediol (10 µM) added as internal standard and analyzed on an Agilent 1290 HPLC using a Rezex ROA-Organic Acid H^+ (8%) column (150 × 4.6 mm, Phenomenex) with isocratic elution (0.5% *v/v* formic acid, 0.6 mL min^{-1}, 55 °C). Samples were detected with an Agilent 6460C triple quadrupole MS with Jet Stream ESI source, operating in positive MRM mode (*m/z* 91→73 transition; fragmentor, 50 V; collision energy, 0 V; cell accelerator voltage, 7 V; delta EMV, +400). Samples were quantified relative to a standard curve of 0.3125, 0.625, 1.25, 2.5, 5, 10 g L^{-1} 1,3-butanediol.

Generation of chromosomal point mutations via CRISPR approach I. This method was used to create BWΔ5 *pcnB*(E108A), BWΔ5 *pcnB*(N138H), BWΔ5 *pcnB*(L392fs), BWΔ5 *rne*(R488H V489L), DH1Δ5 *pcnB*(R149P), DH1Δ5 *pcnB*(R373S), DH1Δ5 *rpoC*(K1192E). The targeting vectors were constructed from pTargetF vector by reverse PCR using primer 459-pTargetF-F2 and different -target primers, and subsequent self-ligation. The repair fragments were generated by primer pairs -1 & -2 and -3 & -4 using *E. coli* DH1Δ5 genomic DNA as template, and subsequent SOE-PCR for fusion of above two fragments. Primers are listed in *Appendix 2.1B*. CRISPR/Cas9 genome editing was carried out as previously reported method by Jiang et al [28].

Generation of chromosomal point mutations via CRISPR approach II. Point mutations were made using the CRISPR Cas9 system [29,30]. Briefly, cells were transformed with the pKD46-Cas9-RecA-Cure, which allows for the expression of the Cas9 protein for double stranded DNA breaks and the RecA protein to assist homologous recombination. A single transformant was picked and inoculated in liquid culture to make electro-competent cells. Then cells that carried the pKD46-Cas9-RecA-Cure plasmid were transformed with both the pCRISPR plasmid containing the specific guide RNA and the double stranded DNA repair fragment that carried the desired sequence. The repair fragment also carries a silent mutation to remove the PAM site and a phosphorothioate modification at both the 5′- and 3′-ends. Transformations were recovered and cultured on plates with appropriate selection markers. Colonies were validated by Sanger sequencing.

DH1Δ5 *pcnB* R149H – C<u>G</u>C → C<u>A</u>C mutation at position 446 in the *pcnB* gene. DH1Δ*5* was transformed with pKD46-Cas9-RecA-Cure and plated on appropriate antibiotic resistant LB agar plates and incubated at 30 °C overnight. A single colony was picked and inoculated in 10 mL LB liquid media with appropriate antibiotics overnight at 30 °C. The overnight culture was then diluted in fresh LB media with 0.2% *w/v* of arabinose to an OD_{600} of 0.01 to

induce RecA expression. Once cultures reached an OD_{600} of 0.4, cells were harvested to make electro-competent cells. DH1 pKD46-Cas9-RecA-Cure electro-competent cells were then transformed with pCRISPR-PcnB2409 (#2784) plasmid and repair fragments VY07 and VY08 (*Appendix 2.1A*). Cells were recovered at 30 °C for 1.5 h and plated on appropriate antibiotic selection LB agar plate. Plate was incubated at 30 °C overnight. Presence of the desired genomic mutation was verified with sanger sequencing.

***pcnB* genomic knockout**. BW25113Δ5-T1R Δ*pcnB* was constructed using the lambda-red recombinase system [31]. The template for recombination was amplified from pKD4 using primers EK14 and EK15. The Km^R marker was subsequently removed using plasmid pCP20. Primers pcnB F and pcnB R were used to verify gene disruption.

***pcnB* mutant library generation.** Error-prone PCR was carried out using the GeneMorph II Random Mutagenesis Kit (Agilent) according to the supplier's instructions with primers EK20 and EK21 (*Appendix 2.1A*). Specifically, the reaction was performed in a Bio-Rad PCR machine with 500 ng of template DNA and 20 cycles of amplification. The thermal cycle was programmed for 2 min at 95 °C for initial denaturation, followed by 20 cycles of 30 s at 95 °C for denaturation, 30 s at 55 °C for annealing, 90 s at 72 °C for extension, and 10 min at 72 °C for the final extension. The PCR reaction was purified with the QIAquick PCR Purification Kit (Qiagen). The library was ligated into pBAC-gg using Golden Gate assembly.

Quantification of glucose and organic acids via HPLC. Glucose, lactate, acetate, and pyruvate were quantified on HPLC (Agilent, 1200) equipped with refractive index detector. The supernatant from fermentation culture was mixed in a 9:1 ratio with aqueous solution containing the valeric acid internal standard (100 g l^{-1}). These samples were analyzed on an Aminex HPX-87H Column (BioRad). The mobile phase was H_2SO_4 (5 mM) at a flow rate of 0.6 ml min^{-1}, oven temperature was at 60 °C. Glucose was detected with a refractive index detector. Organic acids were detected using a photodiode array detector at 210 nm. Concentrations were determined by extrapolation from standard curves.

Approximating glucose diverted to biomass. The following equation was used to approximate glucose (mol) diverted to biomass (g dry cell weight): $\frac{0.48\ g\ C\ g\ Biomass^{-1}}{0.4\ g\ C\ g\ glucose^{-1} * 180.156\ g\ glucose\ mol\ glucose^{-1}}$. The amount of carbon in dried biomass was obtained from Han et al. [32]. The mass ratio of carbon in glucose was obtained by examining the molecular formula. We experimentally determined the conversion factor between OD_{600} and dried biomass to be 0.391 g dry cell weight OD_{600}^{-1}.

2.3. Results and discussion

In order to study mutations found in the evolved strains, we reintroduced the mutations into the respective parent strain. In this way, we could examine their isolated effects on fermentation without the presence of extraneous mutations that arose from evolution. From the strains evolved for *n*-butanol production, we recreated four genomic mutations in the BWΔ5 parent strain. Three of the selected mutations were of the *pcnB* gene encoding the Poly(A) polymerase: *pcnB* E108A, *pcnB* N138H and *pcnB* L392fs. In *pcnB* E108A the conserved catalytic glutamate residue involved in coordination of the active site magnesium ion has been mutated to an alanine resulting decreased activity relative to the wild type (*Figure 2.4A*) [33]. N138 is located on the edge of a disordered region near the active site. The *pcnB* L392fs mutant contains a frameshift mutation disrupting translation of the remaining 62 residues in wild-type PcnB, a region predicted to be involved with

RNA binding. In addition to these *pcnB* mutants, the R488H and V489L mutations were introduced into the *rne* gene of BWΔ5, encoding Rnase E. We sought to obtain only the R488H mutation, but V489L incidentally also occurred. Given that V and L are both hydrophobic residues, we reasoned that this additional mutation would not have a significant impact on the phenotype. Rnase E is known to natively function as a homo-tetramer and R488 and V489 are located in the N-terminal catalytic domain near one of the dimer interfaces (*Figure 2.4B and C*) [34]. Five mutations arising from evolution of DH1Δ5 with the BDO pathway were selected for further study. Two mutations of the same residue in Poly(A) polymerase, R149H and R149L, were created in DH1Δ5. R149 is located near the active site and is involved in a structural hydrogen bonded network.[33] The R373S mutation was introduced into RNAse E, which is positioned close to where RNA is bound.[34] Two mutations, M466L and K1192E, were also introduced into rpoC, the gene encoding the β′ subunit of RNA Polymerase. M466 is located near the catalytic magnesium of RNAP and K1192 is located just downstream of the hypervariable region (*Figure 2.4D and E*) [35,36].

We compared *n*-butanol production of two of the single mutant strains, BWΔ5 *pcnB*(E108A) and BWΔ5 *pcnB*(N138H), to the evolved strains where those mutations were originally identified. We found that the single mutants actually performed better than the evolved strains (*Figure 2.5*). All of the BWΔ5 mutants showed increased final OD_{600} and *n*-butanol titer relative to BWΔ5 (*Figure 2.6A*). Similarly, the DH1Δ5 single mutants showed improved BDO production and growth relative to the parent strain, DH1Δ5 (*Figure 2.6B*). These observations verify that each of these mutations identified through adaptive evolution are in fact drivers of the improved phenotype. It is surprising that a single point mutation in one of the roughly 4,400 genes of *E. coli* can have such a large impact on pathway productivity [37].

Evolution of the *n*-butanol producing trains was performed with strains carrying the *n*-butanol pathway split between two plasmids. pBu1 carried genes *phaA*, *hbd*, *crt* while pBu2 contained the genes *ter*, *aldh46* and *adh2*. In an attempt to streamline the production system and reduce the number of plasmids required for *n*-butanol fermentation, the genes from pBu1 were transferred to pBu2 resulting in one plasmid, pBut, that contains all genes for the entire pathway. Fermentation performance of the newly created pBut plasmid was compared to the existing pBu1and pBu2 plasmid system (*Figure 2.7*). Despite the two systems carrying identical sets of genes, differences in the final OD_{600} and production titers were observed for the parent strain, BWΔ5, and two mutants, *pcnB*(N138H) and BWΔ5 *pcnB*(E108A). Production dropped slightly for BWΔ5 and even more so for BWΔ5 pcnB N138H. Strikingly, production and growth are vastly improved for BWΔ5 *pcnB*(E108A) with pBut compared to pBu1 and pBu2. Although there are differences in fermentation performance, the trends between the parent and mutant strains were preserved. We concluded that pBut could be used to further investigate *n*-butanol fermentation in the parent and mutant strains.

Noticing that a variety of *pcnB* mutations were found in the evolved strains, we wanted to identify additional mutations in this gene that could improve fermentation performance. To achieve this goal, we devised a strategy to create a mutagenesis library of pcnB on a plasmid and subject it to our growth selection. Due to challenges creating libraries of genomic mutations, we explored the feasibility of a plasmid-based system for examining the effects of pcnB mutants. For a plasmid-based strategy to work, several points needed to be addressed. For one, if the chromosomal version of *pcnB* is present, it may mask the effects of the *pcnB* mutant introduced

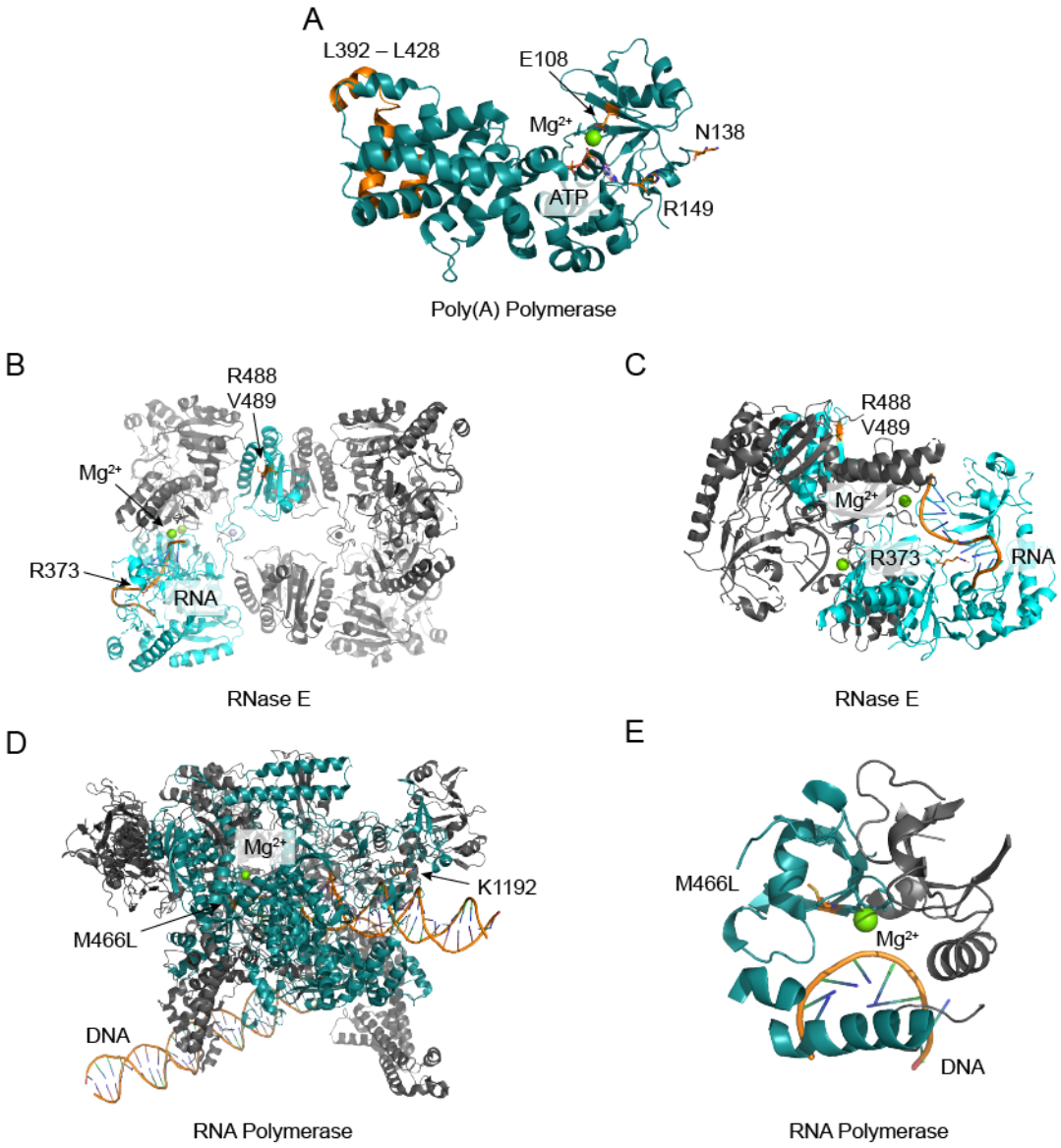


Figure 2.4. *Structures of relevant enzymes highlighting mutated residues. (A) Crystal structure or Poly(A) Polymerase encoded by pcnB. Mutated residues are indicated in orange. The orange region at the C-terminus indicates the residues affected by the L392fs mutation. ATP and magnesium are shown in the active site. (PDB: 3AQN) [33]. (B) Crystal structure of the catalytic domain of RNAse E shown in its active form as a homotetramer. One protomer is colored teal showing bound RNA and the catalytic magnesium is shown in chartreuse. Mutated residues are indicated. C) Another perspective highlighting the active site (PDB: 2BX2) [34]. (C) A cryo-EM structure of RNA polymerase bound to DNA. The mutated residues and active site magnesium are labeled. (D) Another perspective showing a close up of the active site (PDB: 7MKE) [36].*

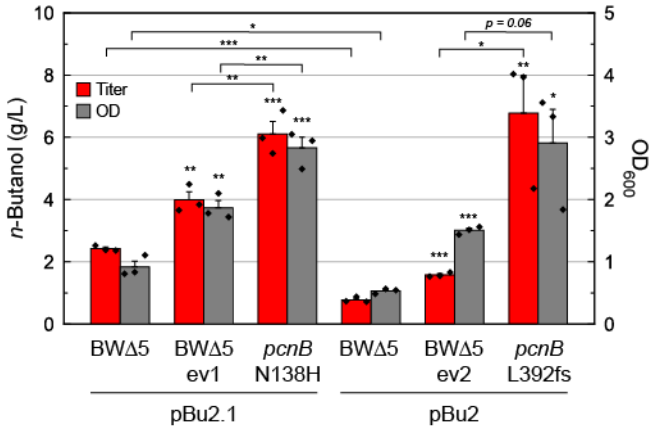


Figure 2.5. *n-Butanol production comparing evolved strains to single mutants. Introducing single mutations found through adaptive evolution into the parent strain improves fermentation saturation OD and titer. Surprisingly, the single mutants outperform the evolved strains (BWΔ5 ev1 and BWΔ5 ev2) where the mutations were initially discovered. pBu2.1 contains a different aldehyde dehydrogenase than pBu2. All strains contained pPDHc, pBu1 and either pBu2.1 or pBu2. n-Butanol fermentation was carried out for 3 d anaerobically at 30 °C in LB with 25 g/L glucose. Error bars represent standard error. Symbols above bars indicate statistical significance when compared to the parent strain for each set with the same plasmid. Statistical analysis by a two-tailed, two-sample equal variance t-test, * p < 0.05, ** p < 0.01, *** p < 0.001.*

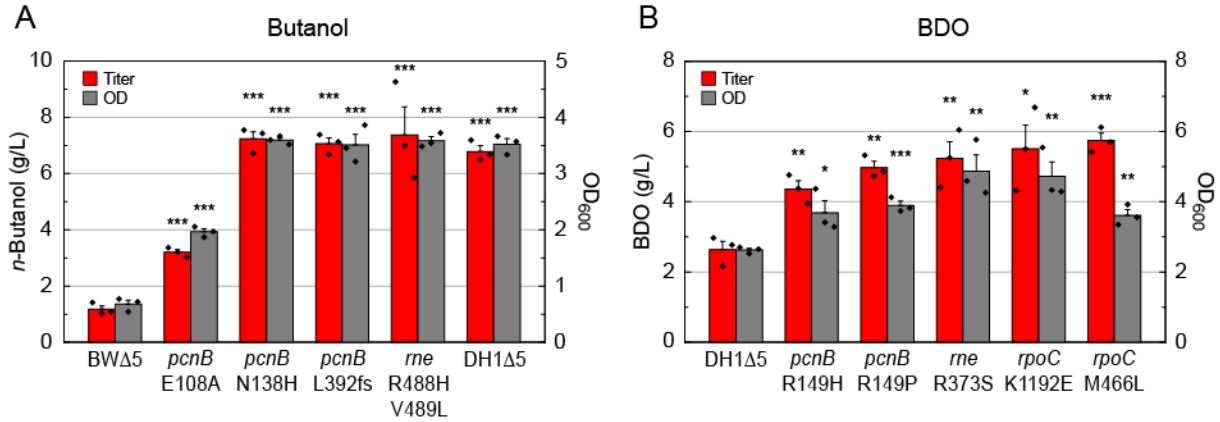


Figure 2.6. *n-Butanol and BDO production with all mutant strains. Mutations identified through adaptive evolution were introduced into the parent strain. These single mutants are sufficient to significantly improve stationary phase OD and product titer for both (A) n-butanol and (B) BDO during fermentation. n-Butanol fermentation was carried out for 3 d anaerobically at 30 °C in LB with 25 g/L glucose. BDO fermentation was conducted for 6 d anaerobically at 30 °C in TB with 25 g/L glucose. Error bars represent standard error. Symbols above bars indicate significance compared to the parent strain (BWΔ5 or DH1Δ5). Statistical analysis by two-tailed, two-sample equal variance t-test, * p < 0.05, ** p < 0.01, *** p < 0.001.*

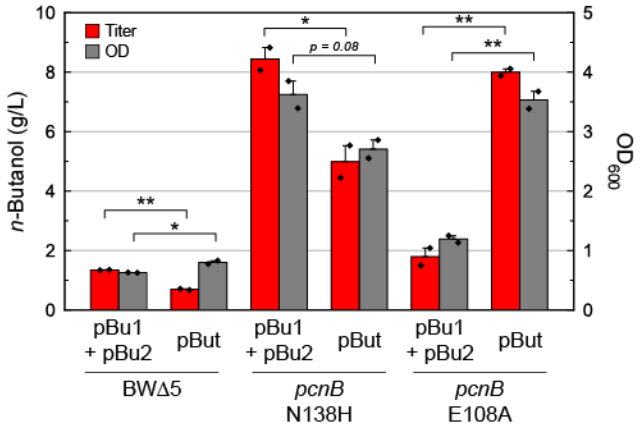


Figure 2.7. *Comparing n-butanol production with pBu1 and pBu2 to pBut. The two plasmids pBu1 and pBu2 were combined by taking the genes on pBu1 and inserting them onto pBu2 to create pBut. pBut shows a similar trend in that the mutants still have improved stationary phase OD and n-butanol titers, so further experiments could be conducted with pBut. Fermentation was performed anaerobically for 3 d at 30 °C in LB with 25 g/L glucose. Strains contained pPDHc and either pBu1 and pBu2 or pBut. Error bars represent the standard error of two replicates. Statistical analysis by a two-tailed, two-sample equal variance t-test, * p < 0.05, ** p < 0.01, *** p < 0.001.*

on a plasmid. Fortunately, *pcnB* is not an essential gene, so we inactivated *pcnB* in the genome of BWΔ5 [38]. Second, we wanted to closely simulate the native expression level of *pcnB*, which is likely much lower than expression from a high-copy plasmid with a strong promoter that is typically used for protein overexpression. Instead, we chose to use the pBAC plasmid which is known to be maintained at low copy numbers within the cell [39]. To further replicate native expression levels, we incorporated one kilobase of upstream and downstream genomic context around the pcnB insertion site to capture the native genomic elements controlling expression. Several versions of *pcnB*, wild type, *pcnB* E108A, and *pcnB* N138H, were inserted into our pBAC vector and transformed into BWΔ5 Δ*pcnB* along with pPDHc and pBut. We then compared the performance of plasmid-borne alleles to the equivalent genomic version (*Figure 2.8*). We were encouraged to see that this approach effectively recaptured the phenotype of the genomic wild type *pcnB* and two of the mutants, BWΔ5 *pcnB* E108A and BWΔ5 *pcnB* N138H.

After verifying that a plasmid-borne copy of *pcnB* recreated phenotypes similar to the equivalent chromosome mutants, we proceeded to create a library of *pcnB* mutants via error-prone PCR [40]. This library was ligated into the pBAC plasmid and growth selection was carried out in the BWΔ5 ΔpcnB strain. We then sequenced a number of colonies resulting from the growth selection (*Figure 2.9*). Surprisingly, most of the isolates contained nonsense mutations or a frameshift. The few colonies that did not have such mutations contained *pcnB* variants that were heavily mutated. This outcome is consistent the observation that both the genomic knockout of *pcnB* alone and *pcnB* E108A, where the glutamate coordinating the active site magnesium has been mutated, both improve *n*-butanol production. Taken together, these results indicate that a non-functional form of *pcnB* is beneficial for *n*-butanol fermentation. Given this conclusion, we decided to discontinue further attempts to mutagenize *pcnB*.

We next sought to examine fermentation of these strains more extensively than quantifying only *n*-butanol concentrations and the final OD reached by the cultures. BWΔ5 and a representative mutant, *pcnB* E108A, carrying either pBut, or a control plasmid coding for GFP were selected for

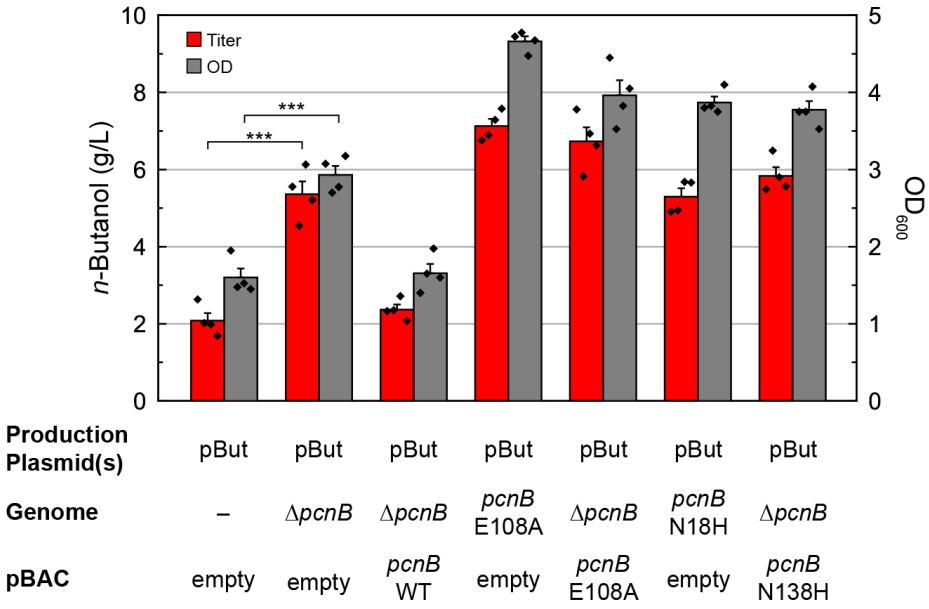


Figure 2.8. *Establishing a system to investigate the effect of pcnB mutants on a plasmid. Stationary phase OD and n-butanol titers are shown for two pcnB alleles present in either the chromosome or on a single copy plasmid. Markedly, the ΔpcnB strain had improved growth and increased n-butanol production during fermentation. Strains with a plasmid-borne copy of pcnB alleles have a similar phenotype to the equivalent chromosomal mutation. Cultures were fermented anaerobically for 3 d at 30 °C in LB with 25 g/L glucose. Strains contained pPDHc and the indicated production plasmid. Error bars represent the standard error of biological replicates. Statistical analysis by a two-tailed, two-sample equal variance t-test, *** $p < 0.001$.*

Colony #	*pcnB* mutations
1	V5A G190S P198T D303G L327fs
2	R214L L406*
3	Y53*
4	V131D Q141fs
5	Q22R Q297fs
6	R201C Q255*
7	E61*
8	P89T F272L R407*
9	V65L R281H A405S
10	K185*
11	I1T (no start)
12	T85I G259S T284P I362F
13	R105H T304A P395L G440E R453H
14	V12L A20T Y170N Q427*
15	L339P Y401*

Figure 2.9. *Mutations found in pcnB after growth selection of an error-prone PCR mutagenesis library. Twelve of the fifteen colonies sequenced contained either a nonsense mutation, or a frameshift deletion. The remaining three colonies were heavily mutated, containing three or more mutated residues. Given that our growth selection yielded several transformants with truncated or frameshifted versions, it seems that mutants of pcnB lacking enzyme activity were selected for.*

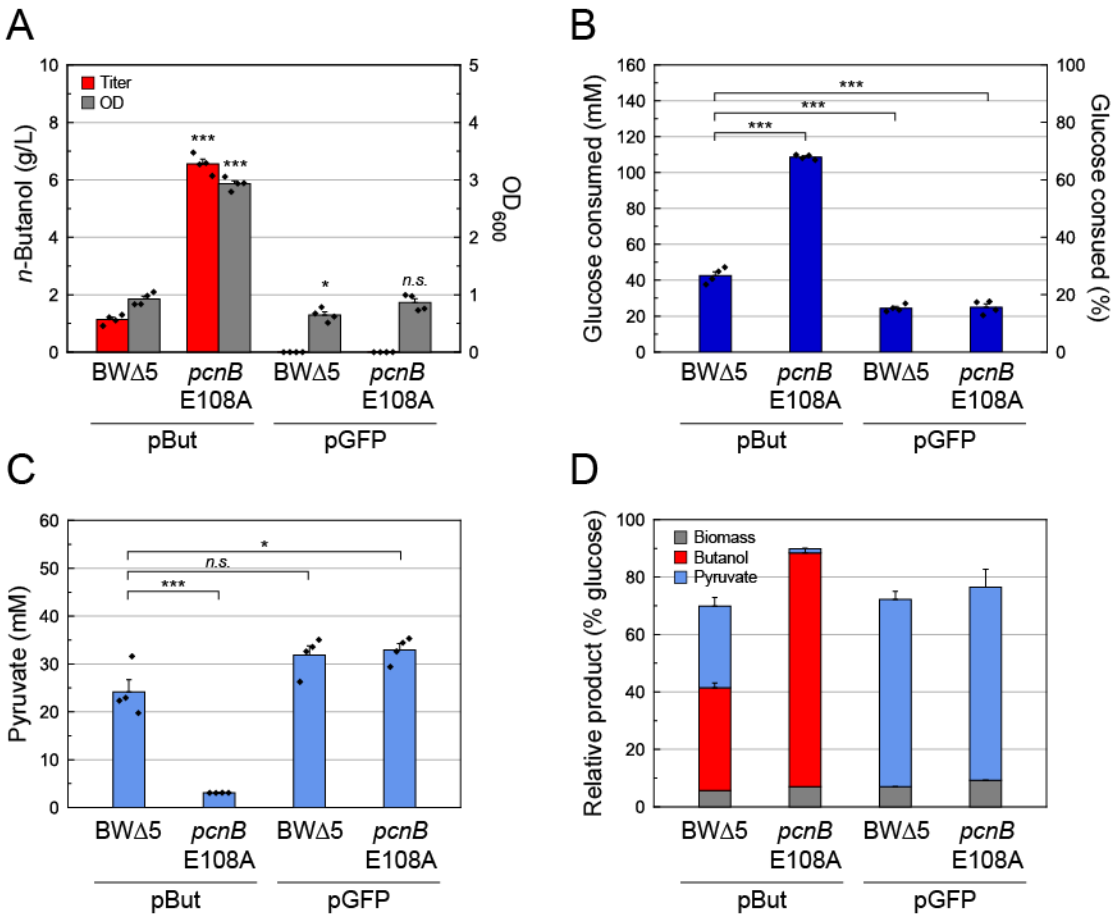


Figure 2.10. *Analysis of other fermentation products during n-butanol production.* ***(****A) The mutant strain produced more n-butanol and reached a higher final OD during fermentation than either the parent strain or pGFP controls. Symbols above the bars indicate statistical significance when compared to the parent strain, BWΔ5 with pBut. (B) A similar trend was seen when quantifying the amount of glucose consumed. The right axis shows the percent of starting glucose consumed. (C) Significantly less pyruvate was present in the media after fermentation with BWΔ5 pcnB E108A than the parent or pGFP controls. (D) Based on the amount of glucose consumed, the yield of biomass and each fermentation product was estimated. Comparing BWΔ5 pcnB E108A to BWΔ5, n-butanol yield was significantly higher ($p < 0.001$) and pyruvate yield was significantly lower ($p < 0.001$). The majority of glucose consumed was converted to pyruvate in strains carrying the pGFP control plasmid. The parent strain had a slightly lower biomass yield than pcnB E108A ($p < 0.001$). Fermentation was performed anaerobically for 3 d at 30 °C in LB with 25 g/L glucose. Strains contained pPDHc and either pBut or pGFP. Error bars represent standard error. Statistical analysis by a two-tailed, two-sample equal variance t-test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

a detailed analysis of fermentation products. The first observation was that some growth is still possible in BWΔ5 and BWΔ5 *pcnB* E108A strains carrying pGFP and lacking a functional *n*-butanol pathway (*Figure 2.10A*). This may arise from a combination of factors. For one, the strains are grown aerobically for 3 h prior to thoroughly sparging the headspace with argon. Even after sparging, a considerable amount of O_2 may still be dissolved in the liquid media supporting a short period of aerobic growth. BWΔ5 reaches an OD that is only slightly higher than the strains lacking a *n*-butanol pathway. This fact indicates that the parent strain is not capable of substantial growth via *n*-butanol fermentation. Growth differences are also evident from quantification of glucose

consumed (*Figure 2.10B*). The mutant strain consumes significantly more glucose than either the parent strain or pGFP controls. As expected from the genomic pathway knockouts, no detectable amount of the canonical fermentation products, lactate, ethanol, acetate, or succinate were detected. Instead, pyruvate was found accumulated in the fermentation media (*Figure 2.10C*). Lack of native fermentation pathways for consumption likely leads to an accumulation of intracellular pyruvate which is ultimately excreted from the cell. We see that the mutant strain, which is shown to be capable of fermentation using the *n*-butanol biosynthetic pathway, accumulates much less pyruvate in the media consistent with the above explanation. Assuming minimal catabolism of amino acids in the rich media, the yield of each fermentation product and biomass from glucose can be estimated (*Figure 2.10D*). The majority of glucose consumed by the control strains carrying pGFP was directed to pyruvate. About an equal amount of glucose was diverted towards *n*-butanol and pyruvate in the parent strain. The mutant strain, in contrast, converted nearly all of the glucose consumed to *n*-butanol, further indicating that the mutant is capable of growth via *n*-butanol fermentation. These observations suggest that BWΔ5 is unable to support anaerobic fermentation and that the small amount of growth and *n*-butanol production observed likely occurs during the aerobic growth phase and in the microaerobic conditions following.

Next, we proceeded to characterize *n*-butanol production beyond end point analysis. We began by collecting time course data for *n*-butanol production of the parent and mutant strains (*Figure 2.11*). The data revealed that the mutant strains continued growing long after BWΔ5 had reached stationary phase. Additionally, *n*-butanol production continues in stationary phase, suggesting that fermentation is still active to support cellular maintenance energy requirements [41]. Stationary phase production also occurs in the BWΔ5 strain, but to a much lesser extent.

To gain insight about the metabolic shifts that have occurred in the mutant strains, we looked into the utility of ^{13}C metabolic flux analysis (^{13}C-MFA). ^{13}C-MFA has proven to be a powerful tool to probe microbial metabolism [42,43]. The basis of this technique is that the isotopic labeling pattern found in the amino acids of a microbial culture grown on an isotopically labeled carbon source can be used to inform the activity of various central metabolic activities within the cell. In order to constrain the mathematical models used in this technique, the strains must be grown on defined minimal media so that all of the biomass is derived from a carbon source with a known labeling pattern.

We explored the feasibility of using ^{13}C-MFA to study our system. We grew relevant strains anaerobically in minimal media and quantified several different rates for comparison (*Figure 2.12*). Wild-type BW25113 was included as a control strain capable of robust anaerobic growth to allow for comparison to previously reported values. The *pcnB* E108A mutant was found to have a higher growth rate than the parent strain, consistent with the improved fermentation performance that we have observed (*Figure 2.12A*). Glucose uptake rate was found to be consistent among BW25113, BWΔ5 and *pcnB* E108A in induced conditions (*Figure 2.12B*). As expected, acetate production is seen in BW25113 and is abolished in BWΔ5 (*Figure 2.12C*). Measurable pyruvate production is only observed in the parent strain, consistent with the previous end-point analysis of fermentation in rich media (*Figure 2.12D*). Lag phase, describing the duration of time between culture inoculation and log phase growth, was much longer for BWΔ5 than the other strains (*Figure 2.12E*). The lag phase for BWΔ5 was measured to be greater than 30 h. This extremely long lag phase is likely an indicator of a lack of essential catalytic capacity resulting from cellular

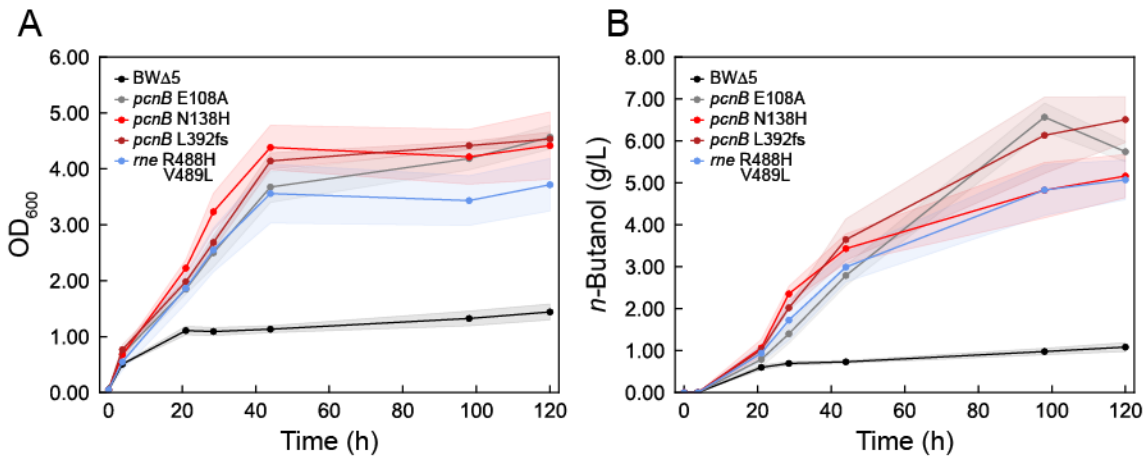


Figure 2.11. *Time course data for n-butanol production. (A) OD and (B) n-butanol titers. n-Butanol production continues for the mutant strains after they reach stationary phase. Strains carried pPDHc and pBut. Fermentation was carried out anaerobically for 3 d at 30 °C in LB with 25 g/L glucose. The shaded area represents standard deviation.*

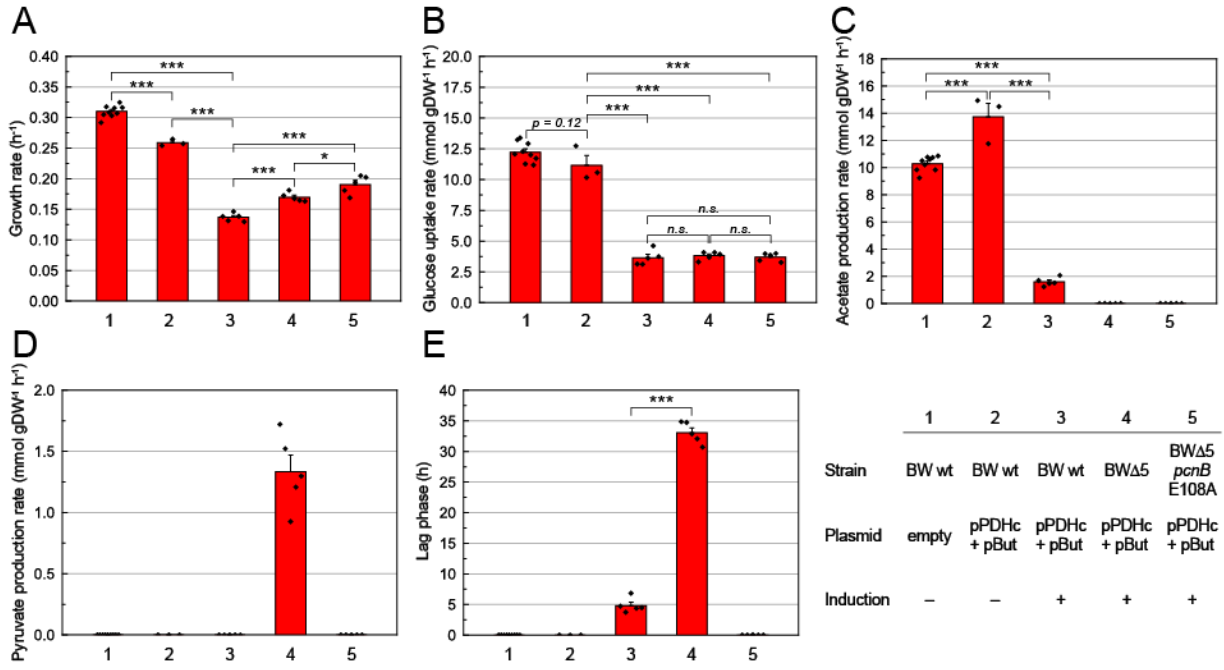


Figure 2.12. *Quantitation of growth parameters in minimal media. (A) Growth rates are reduced upon introducing the production plasmids into BW25113 (BW wt) and the effect is further accentuated upon pathway induction. (B) BW wt had a considerably longer lag phase when the n-butanol pathway was induced. BWΔ5 had a lag phase of ~30h with induction of the n-butanol pathway, whereas no lag phase was observed for BWΔ5 pcnB E108A. (C) Glucose uptake rate dropped dramatically in the wild type strain upon induction of the n-butanol biosynthetic pathway. (D) Acetate secretion is abolished in the knockout strains. (E) Only BWΔ5 shows significant pyruvate production. Strains were grown in M9 minimal media with 3 g/L glucose in anaerobic conditions. Error bars represent standard error. Statistical analysis by a two-tailed, two-sample equal variance t-test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

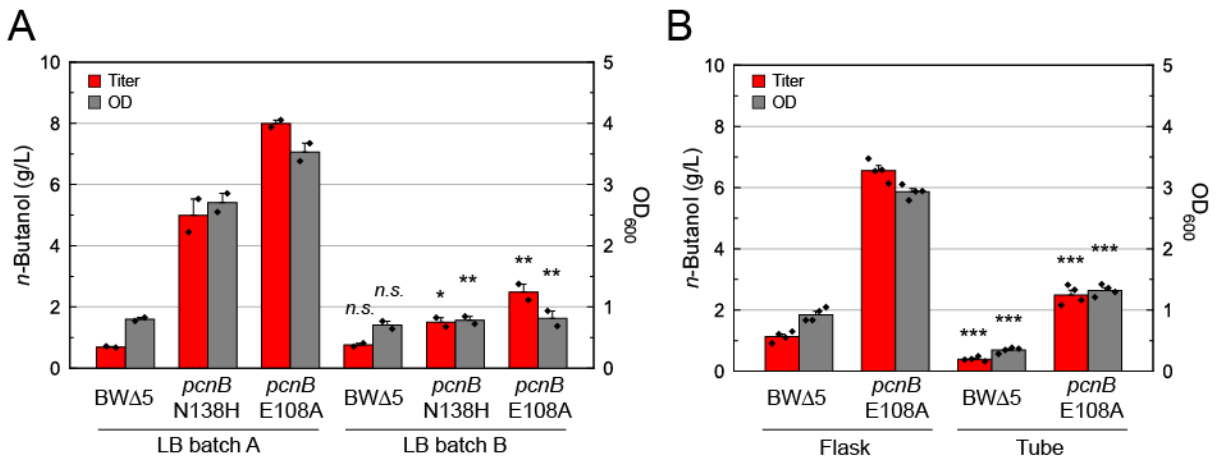


Figure 2.12. *Other factors affecting n-butanol production. (A) Different batches of LB media can have large impacts on fermentation outcomes. Symbols above bars represent significance comparing the same strain to LB batch A. (B) Two types of flasks where used for fermentation: 25ml cultures in 250ml flasks sealed with a screw lid containing a rubber septum and 10ml cultures in 25ml tubes with thick septa sealed with a crimp cap. Saturation OD and product titers were reduced when fermentation was carried out with the tubes instead of flasks. Symbols above bars represent significance comparing the same strain to the flask condition. Fermentation was performed anaerobically for 3 d at 30 °C in LB with 25 g/L glucose. Strains contained pPDHc and pBut plasmids. Error bars represent the standard error. Statistical analysis by a two-tailed, two-sample equal variance t-test, * p < 0.05, ** p < 0.01, *** p < 0.001.*

stress [44,45]. Growth following these long lag phases could be due to a spontaneous mutation, which would result in an unknown genotype for study. Additionally, large variations in lag phase between experiments proved to make sample collection technically challenging. Due to these factors, our attempts for further characterization and metabolic flux analysis were not further pursued.

Despite our inability to robustly measure relevant rates in BWΔ5, quantification of BW25113 growth with and without pBut provided useful insight. The growth rate of BW25113 dropped significantly when carrying pPDHc and pBut without induction (*Figure 2.12*). Upon inducing expression of the pathway genes with 1 mM IPTG, the growth rate dropped even further. Matching the trends in growth rate, glucose uptake rate decreased with pathway induction. A notable increase in lag phase was also observed in the wild type strain when the *n*-butanol pathway was induced. These growth defects may result from increased activity of the overexpressed PDHc disrupting redox balance [14]. Indeed, we see a higher growth rate for BWΔ5 and the *pcnB* E108A mutant than BW25113, perhaps since these strains only contain the *n*-butanol fermentation pathway which matches the redox balance resulting from PDHc activity.

Alternatively, growth defects in BW25113 with induction of the pathway genes could be attributed to the phenomenon known as metabolic burden. Metabolic burden is defined as the negative effect that overexpression of a heterologous protein can have on microbial physiology [46–49]. It is well documented that metabolic burden can cause a decrease of growth rate, as is seen for wild type BW25113 upon pathway induction [50]. If the wild-type strain with fully functioning native fermentation experiences metabolic burden, it seems reasonable to assume that the parent strain experiences a similar effect. Further, we know that the parent strain is not capable of fermentation to support anaerobic growth suggesting that the cellular energy budget is slim. The

resource expenditure required to transcribe and translate six biosynthetic enzymes may be too much for cellular metabolism to bare in an anaerobic environment.

During the course of experimentation, several extraneous factors were found to affect *n*-butanol production (*Figure 2.12*). To our surprise, batch-to-batch variation in LB (EMD Milipore) had a profound effect on *n*-butanol titers and growth for both the parent mutant strains. Additionally, the flask used for anaerobic fermentation also impacted fermentation outcomes. Production titers and stationary phase OD are significantly reduced upon switching from a 25 ml culture in a 250ml baffled flask with a screw on cap with a septum to 10 ml cultures in a 25ml Balch tube sealed with a rubber septum. These differences may result from less nutrient mixing in tubes compared to flasks, leading to nutrient gradients that cannot support growth to higher ODs.

2.4. Conclusion

As an approach to improve the performance of our *n*-butanol and BDO biosynthetic pathways, we turned to adaptive evolution. A growth-based selection was established by removing the native fermentation pathways from our *E. coli* metabolic hosts strains, BW25113 and DH, yielding BWΔ5 and DH1Δ5 respectively. Using this growth-based selection, we carried out adaptive evolution and isolated evolved strains that reach higher ODs and produce more of our desired products than the parent strains through fermentation. Sequencing these strains revealed three genes to be frequently mutated, *pcnB*, *rne*, and *rpoC*. Introducing mutations found through evolution back into the parent strain recaptures the improved phenotype, verifying that these mutations impact production (*Figure 2.6*).

This set of single mutant strains in both BWΔ5 and DH1Δ5 provide a system to study the effect that mutating *pcnB*, *rne*, and *rpoC* has on production of either *n*-butanol or BDO. In order to simplify our production system, we placed all of the *n*-butanol pathway genes onto one plasmid, pBut, and verified that it performed similarly to the original two-plasmid system consisting of pBu1 and pBu2 (*Figure 2.7*). We next sought to identify additional *pcnB* mutants that improve *n*-butanol production. To do so, we devised a system that allowed us to probe the effects of *pcnB* mutations carried on a single-copy plasmid in a *pcnB* knockout background (*Figure 2.8*). With this system, we subjected an error-prone library of *pcnB* mutants to our growth selection. The outcome of this selection indicated that loss-of-function mutations in *pcnB* are beneficial for *n*-butanol fermentation (*Figure 2.9*).

A more detailed analysis of *n*-butanol fermentation for the parent BWΔ5 strain compared to strains carrying mutations in *pcnB* and *rne* yielded several observations. A pilot study examining the introduction of a genomic *pcnB* E108A mutation into the BWΔ5 background shows that this mutation allows for higher *n*-butanol production and biomass accumulation than the parent strain, which appears to be related to increased glucose consumption and an improved capacity for conversion of pyruvate to the *n*-butanol product (*Figure 2.10*). A larger study including *pcnB* N138H, *pcnB* L392fs, and *rne* R488H V489L shows that these mutant strains continue growing long after the parent strain reaches stationary phase (*Figure 2.11*). Further analysis shows that the BWΔ5 parent strain appears to be unable to perform fermentation to support substantial anaerobic growth even in LB media based on its accumulation of pyruvate in the culture media coupled with a lack of growth and glucose consumption (*Figure 2.10*). We also quantified growth in minimal media for the wild-type BW25513, the BWΔ5 parent knockout strain for evolution, and the BWΔ5 *pcnB* E108A mutant as a representative evolved daughter strain. From these studies, we noticed

extremely long lag phases for the parent strain (*Figure 2.12*). This suggests that growth observed for the BWΔ5 culture is possibly the result of a spontaneous mutant, reducing our confidence in continuing a comparative ^{13}C-flux study in minimal media.

However, growth analysis of the wild-type BW25113 strain in minimal media provided the seed for a hypothesis that reduced metabolic burden could be important for the improvement in growth and production (*Figure 2.12*). From our data, it seems likely that BW25113 experiences metabolic burden associated with overexpression of the *n*-butanol pathway. If this phenomenon was observed in the wild-type strain capable of native fermentation, it is likely to also affect our parent BWΔ5 strain used for evolution. These observations sow the seeds for the following hypothesis: Did the mutations arise from adaptive evolution to reduce the metabolic burden of pathway overexpression? Perhaps *n*-butanol fermentation is able to support anerobic growth without this additional energy demand in the mutant strains.

2.5. References

1. Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451,** 86–89 (2008).
2. Bond-Watts, B. B., Bellerose, R. J. & Chang, M. C. Y. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nat. Chem. Biol.* **7,** 222–227 (2011).
3. Nelson, D. L., Lehninger, A. L. & Cox, M. M. *Lehninger Principles of Biochemistry*. (Macmillan, 2008).
4. Davis, M. A., Yu, V. Y., Fu, B., Wen, M., Koleski, E. J., Silverman, J., Berdan, C. A., Nomura, D. K. & Chang, M. C. Y. A Cellular Platform for Production of C_4 Monomers. 2023.01.09.523327 Preprint at https://doi.org/10.1101/2023.01.09.523327 (2023)
5. Morrow, N. L. The industrial production and use of 1,3-butadiene. *Environ. Health Perspect.* **86,** 7–8 (1990).
6. Atsumi, S., Wu, T.-Y., Machado, I. M. P., Huang, W.-C., Chen, P.-Y., Pellegrini, M. & Liao, J. C. Evolution, genomic analysis, and reconstruction of isobutanol tolerance in *Escherichia coli*. *Mol. Syst. Biol.* **6,** 449 (2010).
7. Horinouchi, T., Tamaoka, K., Furusawa, C., Ono, N., Suzuki, S., Hirasawa, T., Yomo, T. & Shimizu, H. Transcriptome analysis of parallel-evolved *Escherichia coli* strains under ethanol stress. *BMC Genomics* **11,** 579 (2010).
8. Reyes, L. H., Almario, M. P., Winkler, J., Orozco, M. M. & Kao, K. C. Visualizing evolution in real time to determine the molecular mechanisms of n-butanol tolerance in *Escherichia coli*. *Metab. Eng.* **14,** 579–590 (2012).
9. Lee, D.-H. & Palsson, B. Ø. Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a nonnative carbon source, L-1,2-propanediol. *Appl. Environ. Microbiol.* **76,** 4158–4168 (2010).
10. Conrad, T. M., Joyce, A. R., Applebee, M. K., Barrett, C. L., Xie, B., Gao, Y. & Palsson, B. Ø. Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biol.* **10,** R118 (2009).

11. Dragosits, M. & Mattanovich, D. Adaptive laboratory evolution – principles and applications for biotechnology. *Microb. Cell Factories* **12,** 64 (2013).
12. Zhang, X., Jantama, K., Moore, J. C., Jarboe, L. R., Shanmugam, K. T. & Ingram, L. O. Metabolic evolution of energy-conserving pathways for succinate production in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 20180–20185 (2009).
13. Unden, G. & Bongaerts, J. Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. *Biochim. Biophys. Acta BBA - Bioenerg.* **1320,** 217–234 (1997).
14. Clark, D. P. The fermentation pathways of *Escherichia coli*. *FEMS Microbiol. Rev.* **5,** 223–234 (1989).
15. Zhang, X., Jantama, K., Moore, J. C., Shanmugam, K. T. & Ingram, L. O. Production of L-alanine by metabolically engineered *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **77,** 355–366 (2007).
16. Yomano, L. P., York, S. W. & Ingram, L. O. Isolation and characterization of ethanol-tolerant mutants of *Escherichia coli* KO11 for fuel ethanol production. *J. Ind. Microbiol. Biotechnol.* **20,** 132–138 (1998).
17. Zhou, S., Shanmugam, K. T., Yomano, L. P., Grabar, T. B. & Ingram, L. O. Fermentation of 12% (w/v) glucose to 1.2 M lactate by *Escherichia coli* strain SZ194 using mineral salts medium. *Biotechnol. Lett.* **28,** 663–670 (2006).
18. Davis, M. A. Exploring in vivo biochemistry with C_4 fuel and commodity chemical pathways. at <https://www.proquest.com/dissertations/docview/2043901617/abstract/F815975033FA44CAPQ/1>
19. Yu, V. Y. harnessing evolution to study cellular regulation of metabolism using synthetic pathways for production of C_4 monomers. at <https://www.proquest.com/dissertations/docview/2436896585/abstract/83327E8D617E4415PQ/1>
20. He, L., Söderbom, F., Wagner, E. G., Binnie, U., Binns, N. & Masters, M. PcnB is required for the rapid degradation of RNAI, the antisense RNA that controls the copy number of ColE1-related plasmids. *Mol. Microbiol.* **9,** 1131–1142 (1993).
21. Mackie, G. A. RNase E: at the interface of bacterial RNA processing and decay. *Nat. Rev. Microbiol.* **11,** 45–57 (2013).
22. Sutherland, C. & Murakami, K. S. An introduction to the structure and function of the catalytic core enzyme of *Escherichia coli* RNA polymerase. *EcoSal Plus* **8,** (2018).
23. Woolston, B. M., Edgar, S. & Stephanopoulos, G. Metabolic engineering: past and future. *Annu. Rev. Chem. Biomol. Eng.* **4,** 259–288 (2013).
24. Volk, M. J., Tran, V. G., Tan, S.-I., Mishra, S., Fatma, Z., Boob, A., Li, H., Xue, P., Martin, T. A. & Zhao, H. Metabolic engineering: methodologies and applications. *Chem. Rev.* (2022). doi:10.1021/acs.chemrev.2c00403
25. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
26. Engler, C. & Marillonnet, S. Golden Gate cloning. *Methods Mol. Biol. Clifton NJ* **1116,** 119–131 (2014).

27. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8,** 91 (2008).
28. Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J. & Yang, S. Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81,** 2506–2514 (2015).
29. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233–239 (2013).
30. Chen, W., Zhang, Y., Yeo, W.-S., Bae, T. & Ji, Q. Rapid and efficient genome editing in *Staphylococcus aureus* by using an engineered CRISPR/Cas9 system. *J. Am. Chem. Soc.* **139,** 3790–3795 (2017).
31. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6640–6645 (2000).
32. Han, L., Enfors, S.-O. & Häggström, L. *Escherichia coli* high-cell-density culture: carbon mass balances and release of outer membrane components. *Bioprocess Biosyst. Eng.* **25,** 205–212 (2003).
33. Toh, Y., Takeshita, D., Nagaike, T., Numata, T. & Tomita, K. Mechanism for the alteration of the substrate specificities of template-independent RNA polymerases. *Structure* **19,** 232–243 (2011).
34. Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G. & Luisi, B. F. Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature* **437,** 1187–1191 (2005).
35. Ederth, J., Isaksson, L. & Abdulkarim, F. Origin-specific reduction of ColE1 plasmid copy number due to mutations in a distinct region of the *Escherichia coli* RNA polymerase. *Mol. Genet. Genomics* **267,** 587–592 (2002).
36. Saecker, R. M., Chen, J., Chiu, C. E., Malone, B., Sotiris, J., Ebrahim, M., Yen, L. Y., Eng, E. T. & Darst, S. A. Structural origins of *Escherichia coli* RNA polymerase open promoter complex stability. *Proc. Natl. Acad. Sci. U. S. A.* **118,** e2112877118 (2021).
37. Serres, M. H., Gopal, S., Nahum, L. A., Liang, P., Gaasterland, T. & Riley, M. A functional update of the *Escherichia coli* K-12 genome. *Genome Biol.* **2,** research0035.1-research0035.7 (2001).
38. Goodall, E. C. A., Robinson, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., Lund, P. A., Cole, J. A. & Henderson, I. R. The essential genome of *Escherichia coli* K-12. *mBio* **9,** e02096-17 (2018).
39. Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y. & Simon, M. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci.* **89,** 8794–8797 (1992).
40. Wilson, D. S. & Keefe, A. D. Random mutagenesis by PCR. *Curr. Protoc. Mol. Biol.* **Chapter 8,** Unit8.3 (2001).
41. McGrew, S. B. & Mallette, M. F. Energy of maintenance in *Escherichia coli*. *J. Bacteriol.* **83,** 844–850 (1962).
42. Long, C. P. & Antoniewicz, M. R. High-resolution 13C metabolic flux analysis. *Nat. Protoc.* **14,** 2856–2877 (2019).

43. de Falco, B., Giannino, F., Carteni, F., Mazzoleni, S. & Kim, D.-H. Metabolic flux analysis: a comprehensive review on sample preparation, analytical techniques, data analysis, computational modelling, and main application areas. *RSC Adv.* **12,** 25528–25548 (2022).
44. Bertrand, R. L. Lag phase is a dynamic, organized, adaptive, and evolvable period that prepares bacteria for cell division. *J. Bacteriol.* **201,** e00697-18 (2019).
45. Adkar, B. V., Manhart, M., Bhattacharyya, S., Tian, J., Musharbash, M. & Shakhnovich, E. I. Optimization of lag phase shapes the evolution of a bacterial enzyme. *Nat. Ecol. Evol.* **1,** 1–6 (2017).
46. Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S. & Koffas, M. A. G. Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications. *Trends Biotechnol.* **34,** 652–664 (2016).
47. Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* **12,** 415–418 (2015).
48. Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W. & Milo, R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci.* **110,** 10039–10044 (2013).
49. Noor, E., Eden, E., Milo, R. & Alon, U. Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. *Mol. Cell* **39,** 809–820 (2010).
50. Glick, B. R. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **13,** 247–261 (1995).

Chapter 3: *Exploring metabolic burden associated with pathway expression*

Portions of this work were performed in collaboration with the following persons:

Albert Qiang assisted with molecular cloning, fermentation, and analysis.

3.1. Introduction

In Chapter 2, mutations in *Escherichia coli* BWΔ5 were identified through adaptive evolution that result in higher butanol titers. These mutations in the *pcnB* and *rne* genes were generated in a clean BWΔ5 background in order to isolate their effects on microbial host metabolism for study. Anaerobic growth of BWΔ5 and these point mutant strains was characterized in detail in both LB and M9 media. Through this analysis, we found that BWΔ5 is only capable of a limited amount of fermentative growth, where the mutant strains are able to support anaerobic growth through fermentation of glucose to butanol. Characterization of the wild-type *E. coli* BW25113 containing the *n*-butanol pathway in minimal media revealed large growth defects upon induction of pathway expression, a hallmark sign of the phenomenon known as metabolic burden.

Metabolic burden is defined as the effect that overexpression of a heterologous protein can have on microbial physiology [1,2]. We reasoned that if the wild type strain capable of native fermentation is impacted by the energy demand of pathway expression, perhaps this metabolic load could have an equivalent if not greater impact on the ability of BWΔ5to grow anaerobically, given the native fermentation pathways have been disrupted. As such, it is possible that the initial mutations enabling growth under selective pressure arose to first reduce metabolic burden and improve growth. In order to verify this hypothesis, several outstanding questions need to be answered. Does the parent strain experience metabolic burden due to pathway expression? Is pathway expression reduced in the mutant strains? These questions will be explored in this chapter. Up to this point, most of the analysis has been focused on the series of mutations in BWΔ5 affecting butanol production. In this chapter we extend the breadth of our studies to include the series of DH1Δ5 single mutants shown to improve BDO production.

3.2. Methods

Commercial materials. Hexanol was purchased from Sigma-Aldrich (St. Louis, MO). Anhydrotetracycline, calcium chloride dihydrate, carbenicillin disodium salt (Cb), deoxynucleotides (dNTPs), kanamycin sulfate (Km), chloramphenicol (Cm), O'GeneRuler 1 kb Plus DNA Ladder, PageRuler Plus Prestained Protein Ladder, polyethylene glycol 3350, potassium chloride, sodium chloride, sodium phosphate monobasic monohydrate, were purchased from Thermo Fisher Scientific (Waltham, MA). *n*-Butanol, LB Miller agar, LB Miller broth, anhydrous magnesium sulfate, and Terrific broth were purchased from EMD-Millipore (Burlington, MA). Glucose was purchased from MP Biomedicals (Solon, OH). Isopropyl ß-D-1-thiogalactopyranoside (IPTG) was purchased from Santa Cruz Biotechnology (Dallas, TX). Ethidium bromide was purchased from Bio-Rad Laboratories (Hercules, CA). Phusion polymerase, Phusion HF buffer, all restriction enzymes and restriction enzyme buffer (CutSmart) were from New England Biolabs (Ipswitch, MA). DNA purification and RNeasy kits were purchased from Qiagen (Redwood City, CA). Oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA).

Bacterial Strains. *E. coli* DH10B was used for DNA construction. *E. coli* DH1 (ATCC 39936), DH1Δ5, DH1Δ5 *pcnB*(R149H), DH1Δ5 *pcnB*(R149P), DH1Δ5 *pcnB*(R373S), DH1Δ5 *rpoC*(M466L), DH1Δ5 *rpoC*(K1192E), BW25113, BWΔ5, BWΔ5 *pcnB*(E108A), BWΔ5 *pcnB*(N138H), BWΔ5 *pcnB*(L392fs), and BWΔ5 *rne*(R488H V489L) were used for fermentation experiments.

Gene and plasmid construction. Plasmid construction was carried out using standard molecular biology techniques including the Gibson protocol [3], Golden Gate Assembly [4], or quick-change PCR [5]. PCR amplifications were carried out with Phusion DNA polymerase or GoTaq DNA polymerase, following manufacturer instructions. Primer sequences are listed in *Appendix 2.1A*. Constructs were verified by sequencing (Genewiz; South Plainfield, NJ).

pGFP. sfGFP was amplified from pY71.sfGFP using primers EK58 and EK59. The backbone was amplified from pBu2 with primers EK56 and EK57. The two pieces were combined via Gibson assembly.

pAQ1. The pSC101 origin of replication was amplified from pBMOS1:V2 with primers AQ3 and AQ4. The *araC* cassette was amplified from pBad33 with primers AQ5 and AQ6. The Cm^r resistance cassette was amplified from pBad33 with primers AQ7 and AQ8. The sequence for RNAI was amplified from pBu2 with primers AQ14 and AQ15. The pieces were combined via Gibson assembly.

pAQ2. The R43W mutation was introduced into the *repA* gene of the pSC101 origin of pAQ1 via quick-change PCR using primers AQ18 and AQ19.

pAQ5. The E115K mutation was introduced into the *repA* gene of the pSC101 origin of pAQ1 via quick-change PCR using primers AQ24 and AQ25.

pSC101-WT-Bu2. The wild type pSC101 origin was amplified from pAQ1 with primers EK158 and EK159. AmpR was amplified from pBu2 with primers EK160 and EK161. The two pieces were inserted into pBu2 digested with SpeI and SapI via Gibson assembly.

pSC101-R43W-Bu2. The pSC101-R43W origin was amplified from pAQ2 with primers EK158 and EK159. AmpR was amplified from pBu2 with primers EK160 and EK161. The two pieces were inserted into pBu2 digested with SpeI and SapI via Gibson assembly.

pSC101-E115K-Bu2. The pSC101-E115K origin was amplified from pAQ5 with primers EK158 and EK159. AmpR was amplified from pBu2 with primers EK160 and EK161. The two pieces were inserted into pBu2 digested with SpeI and SapI via Gibson assembly.

pSC101-R43W-But. The first half of the *n*-butanol pathway, comprising genes *phaA*, *hbd*, and *crt*, was amplified from pBu1 using primers EK162 and EK163. The amplicon was inserted into pSC101-R43W-Bu2 digested with SapI via Gibson assembly.

pBu2-wRBS. The RBS in front of the *aldh46* gene was changed by amplifying the regions flanking the RBS site with overlapping primers containing the new RBS and inserting the pieces into pBu2 digested with XhoI and EcoRI via Gibson assembly. Insert I was amplified from pBu2 with primers AQ60 and AQ52. Insert II was amplified from pBu2 with primers AQ51 and AQ53.

pRNAI. The P_{Tet} cassette was amplified from pPOL3 with EK117 and EK118. RNAI was amplified from pBu2 with EK119 and EK120. The backbone containing the origin and Cm^r resistance marker was amplified from pBAD33 with EK115 and EK116. The pieces were combined via Gibson assembly.

Fermentation. Overnight cultures of freshly transformed *E. coli* strains were grown overnight in LB at 37 °C and used to inoculate LB (25 ml) with 25 g/L glucose for *n*-butanol, or TB (25 ml) with 25 g/L glucose for 1,3-butanediol, with appropriate antibiotics to an optical density at 600 nm (OD_{600}) of 0.05 in a 250 mL-baffled anaerobic flask with GL45 threaded top (Chemglass). Cultures were grown at 30 °C in a rotary shaker (200 rpm) for 3h before induction with IPTG (1.0 mM).

Anaerobic cultures were sealed and the headspace was sparged with argon for 3 min immediately following induction. *n*-Butanol and BDO samples were collected after 3 d of cell culture.

RNA sequencing and analysis. Cultures were grown according to the protocol outlined above for fermentation. Samples were collected 24 h after induction. RNA was isolated using the RNeasy RNA isolation kit (Qiagen). Total RNA samples were submitted to the DNA Technologies & Expression Analysis Cores at the University of California, Davis for library preparation and sequencing. Single-read 50 bp data was obtained on an Illumina HiSeq 4000. Reads were mapped using Kallisto [6] and statistical analysis was performed with DeSeq2 [7]. PantherDB [8] was used for statistical enrichment analysis of GO terms.

GFP fluorescence. Overnight cultures were used to inoculate 5 mL of LB or TB with 25 g/L glucose with appropriate antibiotics to an OD_{600} of 0.05 in glass culture tubes (Pyrex 9820-25 25 × 150mm 55mL). Cultures were incubated at 30°C with shaking at 200 rpm for 3 h before induction with 1mM IPTG. Samples (1 mL) were collected after 18 h and centrifuged at 8,000 × *g* for 1 min (Eppendorf MiniSpin 5452). Samples were resuspended to OD_{600} = 1 with phosphate buffered saline pH 7.4 (PBS) then diluted four-fold into PBS and transferred to a flat-bottom 96-well plate. OD_{600} and GFP fluorescence (excitation at 465 nm and emission at 510 nm) were measured in a spectrophotometer (Molecular Devices M2).

qPCR. Overnight cultures were used to inoculate 5 mL of LB or TB with 25 g/L glucose with appropriate antibiotics to OD_{600} = 0.05 in glass culture tubes (Pyrex 9820-25 25 × 150 mm 55 mL). Cultures were incubated at 30°C with shaking at 200 rpm. Samples (1 mL) were collected after 18h and centrifuged at 8,000 × *g* for 1 min (Eppendorf MiniSpin 5452). Samples were resuspended to an OD600 of 1 with phosphate buffered saline pH 7.4 (PBS) and diluted 100x into 0.2 mL PCR tubes. Samples were then incubated at 95 °C to lyse the cells and transferred immediately to the freezer at -20 °C. To prepare the RT-qPCR reaction, 3 µL cell lysate, 1 µL forward primer (10 µM), 1 µL reverse primer (10 µM), 5 µL MQ-water, and 10 µL iQ SYBR Green Supermix (Bio-Rad) were mixed in a 96-well PCR plate (Bio-Rad). RT-qPCR was carried out in a CFX3000 PCR machine (Bio-Rad) following the recommended PCR protocol for iQ SYBR Green Supermix (Bio-Rad) with an annealing temperature of 55 °C. The primers for *idnT*, *phaA*, and *ter* genes are listed in *Appendix 2.1C*. Data analysis was performed with LinReg PCR (https://www.gear-genomics.com) [9].

Growth curves. Overnight cultures were used to inoculate 200 µL of LB or TB with 25 g/L glucose with appropriate antibiotics to an OD_{600} = 0.01 in a 96-well flat bottom plate (Greiner). The plate was covered with a clear plastic lid and placed in the spectrophotometer (Molecular Devices M2). The temperature was maintained at 30 °C. Every 15 min, the plate was shaken for 5 min and then the OD_{600} was measured. OD_{600} values were adjusted to 1 cm pathlength using the scale factor 1.67. Samples were induced with 1 mM IPTG after 2 h where indicated.

Quantification of *n*-butanol titers. Samples (1 mL) were removed from cell culture and cleared of biomass by centrifugation at 14,000 × *g* for 5 min using an Eppendorf 5417R centrifuge. The supernatant or cleared medium sample was then (80 µL) extracted with toluene (80 µL) containing the *n*-hexanol internal standard (1000 mg mL^{-1}). The sample was mixed using a STD VORTEX MIXER (Fisher) for 30s set at 8. The organic layer (25 µL) was diluted four-fold with toluene (75 uL). Diluted samples were ran on an Agilent 7890A GC using an HP-5MS column (0.25 mm x 30 m, 0.25 µM film thickness, J & W Scientific). The oven program was as follows: 95 °C for 4 min, ramp to 300 at 50 °C min^{-1}, 300 °C for 2 min. The injector temperature was set to 250 °C. Split injections wer performed with a split ratio of 20:1. The GC was equipped with an

Agilent 5975C MSD and was ran in scan mode (*m/z* 35 – 150). *n*-Butanol and *n*-hexanol were quantified using the extracted ion chromatograms (*m/z* 56.1). Samples were quantified relative to a standard curve of 10, 5, 2.5, 1.25, 0.625 g/L *n*-butanol prepared freshly for each run. All standards and samples were normalized for injection volume using the *n*-hexanol internal standard.

3.3. Results and discussion

Knowing that *pcnB*, *rne*, and *rpoC* all encode for enzymes that are responsible for RNA metabolism, we wanted to examine how the total RNA pool, or transcriptome, of the mutant strains may have shifted relative to the parent strain. RNA-sequencing (RNA-seq) is a technique leveraging Illumina high-throughput sequencing that allows the researcher to obtain a quantitative snapshot on all of the RNA species present in the cell at the time of sampling [10]. Through normalization and statistical analysis, a list of genes that are differentially expressed between experimental conditions can be extracted from the data.

RNA-seq was performed with the parent, BWΔ5, and the four point mutant strains expressing the *n*-butanol pathway from the two plasmid system. From the data, a large table of genes that were differentially expressed in the mutant strain relative to the parent was created and then filtered using a statistical cut off of $p_{adj} < 0.05$. Relative to BWΔ5, 476, 323, 364, and 302 genes were found to be differentially expressed in BWΔ5 *pcnB* E108A, BWΔ5 *pcnB* N138H, BWΔ5 *pcnB* L392fs, and BWΔ5 *rne* R488H V489L, respectively. Using PantherDB, a web-based tool, the list of differentially expressed genes (DEGs) for each mutant was submitted for statistical enrichment of annotated biological process gene ontology (GO) terms [8]. GO terms represent groups of genes that are validated to contribute to the same biological process. Statistical enrichment analysis provides an approach to determine if genes contributing to a certain biological process occur more frequently in a list than a subset of genes chosen at random. Applying this type of analysis to our set of DEGs provided clues for which biological processes are impacted by transcriptome shifts in the mutants. Between 80 and 100 GOs were found to be statistically enriched in the data set for each mutant. GO terms are constructed in a hierarchical manner and often one gene belongs to many GO terms, complicating analysis. From our list of enriched GO terms, a subset was manually curated to represent the lowest-level GO terms within the hierarchy of the enriched terms. Data for genes belonging to select GO terms are shown as heatmaps (*Figure 3.1*). The color represents $\log_2$ fold change for that gene relative to BWΔ5 in the case of each respective mutant.

Something to consider when interpreting this data set is that the data differences in gene expression are relative to the parent strain cultured in anaerobic conditions, which we have shown does not grow well in these conditions. Thus, the transcriptome of the parent strain, BWΔ5, may not serve as an accurate comparison as the cells may be extremely stressed or perhaps growing after an adaptive mutation or oxygen leak has occurred. As such, differences in expression could represent changes in the mutant strain that are required to return the cell to a state capable of anerobic fermentation.

At a glance, all of the mutants showed similar trends in expression for most genes, indicating that the mutants have made similar metabolic adaptations to support anaerobic growth. The aerobic respiration biological process (GO: 0019646) was statistically enriched in all of the mutants. This GO term contains a subset of genes that comprises the glycolytic biological process (GO: 0006096). Many of these genes were found to be more highly expressed in the mutant strains relative to the parent. This is consistent with the observation that the mutant strains are able to

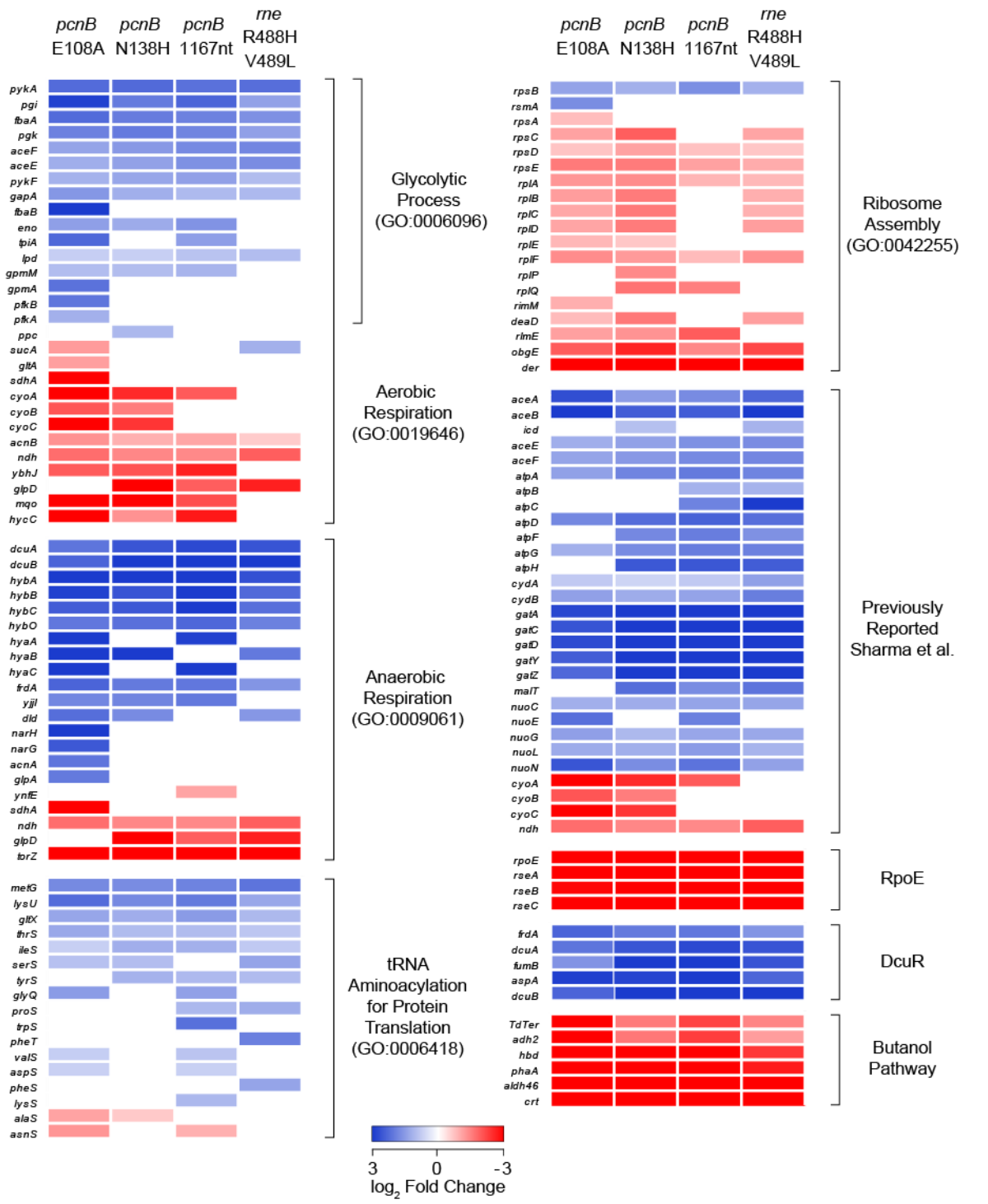


Figure 3.1 *Heatmaps of select genes in the BWΔ5 series RNA-seq data set. Several Gene Ontology (GO) terms, representing groups of genes that participate in a shared biological process, were statistically enriched within the RNA-sequencing data set [8]. Other relevant groups of genes are included. Color represents the log_2 fold change in expression relative to the parent strain according to the color bar. Strains carrying pPDHc, pBu1 and pBu2 were grown anaerobically at 30°C in LB with 25 g/L glucose.*

grow via butanol fermentation, and thus upregulate glycolysis to support energy generation for growth. The remainder of genes in the aerobic respiration GO term were found to be down regulated in the mutant strain, perhaps indicating that the parent strain has not successfully switched over from the brief aerobic growth phase prior to deoxygenation.

At a glance, all of the mutants showed similar trends in expression for most genes, indicating that the mutants have made similar metabolic adaptations to support anaerobic growth. The aerobic respiration biological process (GO: 0019646) was statistically enriched in all of the mutants. This GO term contains a subset of genes that comprises the glycolytic biological process (GO: 0006096). Many of these genes were found to be more highly expressed in the mutant strains relative to the parent. This is consistent with the observation that the mutant strains are able to grow via butanol fermentation, and thus upregulate glycolysis to support energy generation for growth. The remainder of genes in the aerobic respiration GO term were found to be down regulated in the mutant strain, perhaps indicating that the parent strain has not successfully switched over from the brief aerobic growth phase prior to deoxygenation.

Anaerobic respiration (GO:0009061) was also found to be statistically enriched. Members of the *hyb* operon, coding for hydrogenase II, were observed to be overexpressed. It has been proposed that this hydrogenase can operate bidirectionally and may function to reduce protons using a quinol as a reducing agent to evolve H_2, potentially this is serving as a mechanism to dissipate excess NADH equivalents [11,12]. We also see increased expression of the *fumB-dcuB* operon, *aspA-dcuA* operon, and *frdA*, which are all activated by the DcuR transcription factor and participate in anerobic respiration on fumarate [13]. However, anaerobic fermentation on fumarate is non-functional because it relies on *frdBC* to form the active enzyme complex, which have been knocked out in our selection strain [14]. DcuR is activated by four carbon dicarboxylate molecules including fumarate, succinate, malate, and aspartate [13]. It is possible that fumarate accumulated in the mutant strains during anaerobic growth, especially since fumarate reductase is disrupted, which is the main enzyme consuming fumarate that is operative in anaerobic conditions [14]. The gene *ndh* encoding NADH:quinone oxidoreductase II was downregulated in the mutant strains. *ndh* is known to be repressed during anaerobic growth [15], supporting the idea that expression changes in the mutant strains represent a shift to a state more functional for anaerobic growth.

There appears to be a trend of upregulation in tRNA aminoacylation for Protein Translation (GO:0006418). Haddadin et al. has previously reported that aminoacyl tRNA synthetases were down regulated in a strain of *E. coli* over expressing a recombinant gene [16]. In our data set, we observed increased expression for most genes belonging to this GO term. If we assume that expression of aminoacyl tRNA synthetase genes are reduced in BWΔ5 as a result of overexpressing heterologous proteins at high levels, increased expression may represent a shift to a more normal transcriptional state. Also related to translation, we observed a decrease in expression of genes belonging to Ribosome Assembly (GO:0042255). This observation runs counter to literature evidence supporting the idea that ribosome content is often correlated with growth rate [17], given that the mutant strains were shown to have a higher growth rate.

We also looked for evidence of metabolic burden in our RNA-seq data. Previous studies have collected transcriptomic data to examine cellular response to the metabolic burden of overexpressing heterologous genes [16,18–24]. Comparing our list of DEGs to data reported in the literature for conditions experiencing metabolic burden provides an opportunity to strengthen our hypothesis. If the mutant strains experience less metabolic burden than BWΔ5, we expect that the genes reported to be upregulated as a result of metabolic burden would be expressed at a lower

level in the mutant strains than in BWΔ5. Conversely, genes downregulated during metabolic burden would be expressed more in the mutants than in BWΔ5, indicating a return to a transcriptional state less affected by metabolic burden.

Previous studies reported increased expression of stress response proteins, including those correlated with heat-shock [23–25]. RpoE is a sigma factor known to be involved in the heat-shock response and has also been shown to be upregulated in response to other cellular stresses [26–28]. Within our data set, the sigma factor RpoE was expressed at a much higher level in BWΔ5 than in the mutants. In fact, the whole operon *rpoE-rseABC* was expressed at higher levels. RpoE is known to activate transcription of its own operon and this positive feedback likely explains the magnitude of expression differences within our data [29]. Decreased expression of the stress associated RpoE transcription factor is consistent with our hypothesis that the mutant strains experience less metabolic burden than BWΔ5.

A number of DEGs within our dataset matched the list of DEGs resulting from heterologous protein overexpression previously reported [16,18]. Genes belonging to the glyoxylate cycle (*aceA*, *aceB*, *icd*) and the *gat* operon encoding the galactose phosphotransferase system and subsequent metabolism (*gatYZABCD*) were down regulated with gratuitous protein expression. Reduced transcription of energy production genes such as two main terminal oxidases cytochrome bd (*cydAB*) and cytochrome bo (*cyoABC*), the *atp* operon encoding for ATP synthase subunits (*atpABCDFGH*), the *nuo* operon encoding for components of NADH:ubiquinone oxidoreductase I (*nuoCEGLN*), and NADH:quinone oxidoreductase II (*ndh*) was also reported. Within our dataset, we observed an increase in expression of all of these genes in the mutant strains, with the exception of *cyoABC* and *ndh*. Expression patterns of *cyoAB* and *ndh* are further complicated by transcriptional control subject to oxygen concentration, which may explain the exception to the trend [15,30]. Based on the patterns described above, the transcriptional state of the mutants appears to reflect less of an impact from metabolic burden than BWΔ5.

The result that stood out the most was that expression of all the genes in the *n*-butanol biosynthetic pathway was greatly reduced in the mutant strains. This observation further supports our metabolic burden hypothesis and is consistent with a model that pathway expression is too high in BWΔ5 and the associated metabolic burden makes it difficult for anaerobic growth to be supported. In comparison, the mutant strains with lower pathway expression are capable of increased growth and production during fermentation. The magnitude of pathway expression is illuminated by examining the proportion of transcripts originating from the production plasmids to those originating from the genome (*Figure 3.2*). In the parent strain, transcripts originating from the genome represent only about five percent of the total transcripts found in our samples, where the overexpressed genes of the butanol pathway represent about 95% of the transcriptome. This proportion of transcripts resulting from protein overexpression is even larger than previous reports [20]. Decreased pathway expression is accompanied by a significant increase in the proportion of genome derived transcripts for all of the mutant strains. Additionally, we see a greater proportion of transcripts related to the pPDHc plasmid in the mutant strains, which could allow for higher NADH regeneration rates if acetyl-CoA is made through this node rather than PFL. Increased expression of the PDHc may be important for establishing a balanced redox environment during butanol fermentation.

RNA-sequencing was also performed with DH1Δ5 and the mutant strains producing BDO (*Figure 3.3*). Between 50 and 100 genes were found to be differentially expressed in the mutants compared to the parent strain, DH1Δ5. These numbers are much lower than what was observed

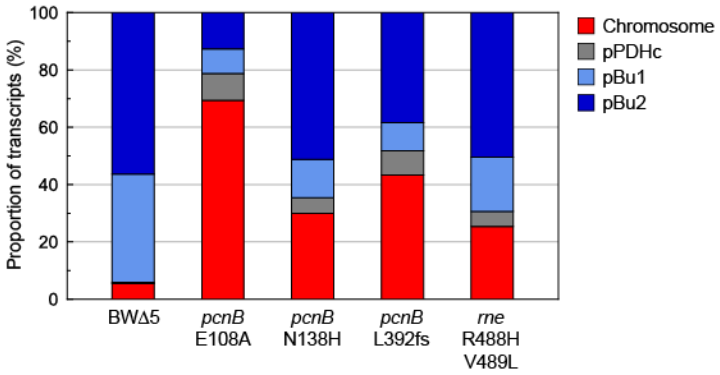


Figure 3.2. *Origin of transcripts in the BWΔ5 series RNA-seq data set. Values shown are transcripts per hundred which represents the proportion of transcripts in each sample derived from either the chromosome or one of the plasmids carrying genes for butanol production (pPDHc, pBu1, or pBu2). The proportion of chromosome-coded transcripts increased for all of the mutant strains relative to the parent strain ($p < 0.01$).*

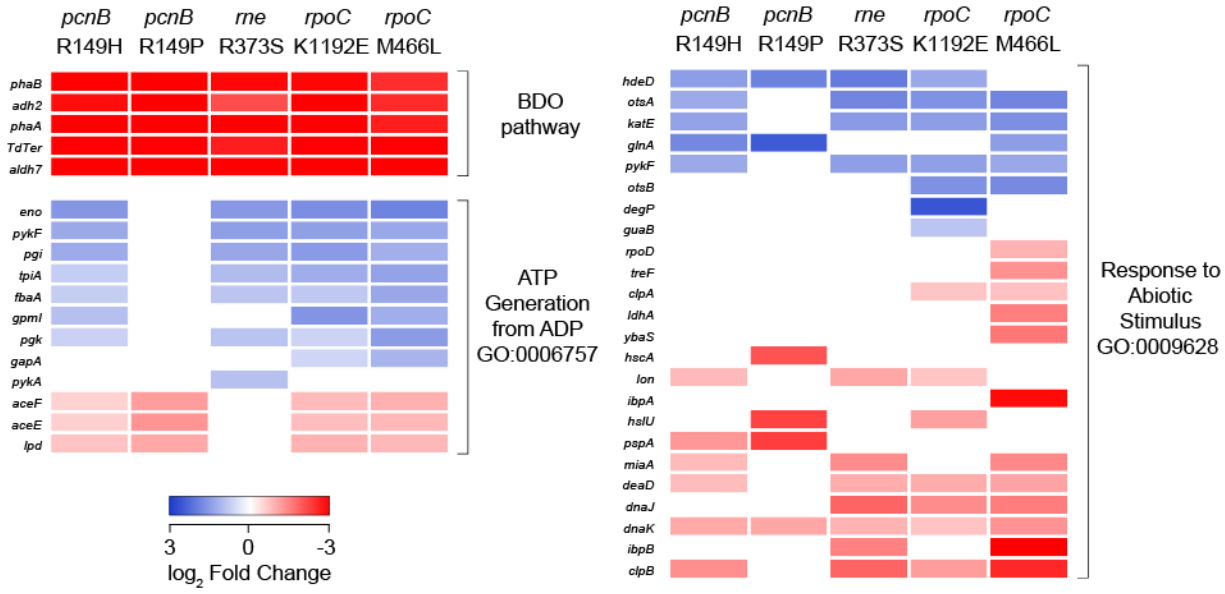


Figure 3.3. *Heatmaps of select genes in the DH1Δ5 series RNA-seq data set. Several Gene Ontology (GO) terms, representing groups of genes that participate in a shared biological process, were statistically enriched within the RNA-sequencing data set [8]. Color represents log2 fold change in expression relative to the parent strain. Strains carrying pBDO1 and pBDO2 were grown anaerobically at 30°C in TB with 25 g/L glucose.*

for the BW mutants, where roughly 400 genes were found to be differentially expressed. Only 28 GO terms were statistically enriched among the genes that were differentially expressed. From this set, two representative GO terms were selected, ATP Generation from ADP (GO:0006757) and Response to Abiotic Stimulus (GO:0009628). Similar to what was observed for the BWΔ5 series, expression of glycolytic genes was increased in the mutant strains. Increased glycolytic gene expression likely results in greater glycolytic flux, consistent with the observation that the mutant strains show improved growth in anaerobic conditions. Genes for the pyruvate dehydrogenase complex, *aceEF* and *lpd* were also found to be downregulated. Downregulation could be due to pyruvate-dependent repression by PdhR [31]. It is possible that this downregulation could lead to increased pyruvate secretion resulting from increased glycolytic activity.

Within the response to abiotic stimulus (GO:0009628) term, we found an upregulation of genes involved in acid resistance (*hdeD*) and osmotic stress response (*otsA*, *katE*) [32–34]. Increased expression of environmental resistant proteins may be in response to higher concentrations of BDO in the fermentation media. We also observe a down regulation of genes involved in cold shock (*deaD*) and heat shock (*lon*, *clpB*, *dnaJ*, *dnaK*, *pspA*) [35–37]. This observation suggests that the parent strain is under protein overexpression stress. In fact, upregulation of *lon*, *clpB*, and *dnaK* have all been previously associated with protein overexpression [25]. Downregulation of these stress response genes within the mutant strains suggests that metabolic burden has been relieved to some extent.

Similar to the BWΔ5 series data, we also found all of the BDO pathway genes to be down regulated to a large extent in each of the mutant strains relative to the parent. Examining the proportion of transcripts originating from the chromosome compared to the BDO pathway plasmids, we see that ~25% of transcripts originated from the chromosomal genes with the remainder corresponding to the heterologous pathway genes (~75%, *Figure 3.4*). In the case of the mutant strains, the proportion of chromosome coded transcripts is significantly increased to ~70%-90% depending on the mutant. Reduced expression of the pathway genes almost certainly lessens the metabolic burden placed on the cell.

To summarize, within the mutant strains we observed trends in the RNA-seq data that indicate a transcriptional state experiencing reduced metabolic burden. Further, we saw that pathway expression was significantly decreased. These observations held true for both the series of BWΔ5 and DH1Δ5 mutants. Together, these data support the hypothesis that the mutant strains experienced less metabolic burden than the parent strains, DH1Δ5 and BWΔ5, and that this has been achieved through reduced pathway gene expression.

We observed reduced transcription of plasmid-borne pathway genes in the mutant strains, but it has been shown that RNA levels do not always correlate with protein expression. We used GFP as a reporter gene to further probe if the mutant strains have reduced translation of plasmid encoded proteins. The fluorescence of GFP provides a direct method to measure protein expression of a plasmid-borne gene [38]. A GFP expression plasmid was constructed from the backbone of pBu1 containing the ColE1 origin and the promoter used to express the *ter* gene ($2 \times P_{Tac}$) in order to closely replicate the genomic elements of the pathway plasmid. Using this construct, we probed GFP expression in both the BWΔ5 and DH1Δ5 series of mutants (*Figure 3.5*). GFP fluorescence normalized by optical density was lower in all of the BW mutant strains relative to the parent, BWΔ5, matching the transcription data obtained from RNA-seq. A similar trend was seen for all of the DH1 mutants. Reduced GFP fluorescence indicates that translation of plasmid-borne genes in the mutant strains was reduced.

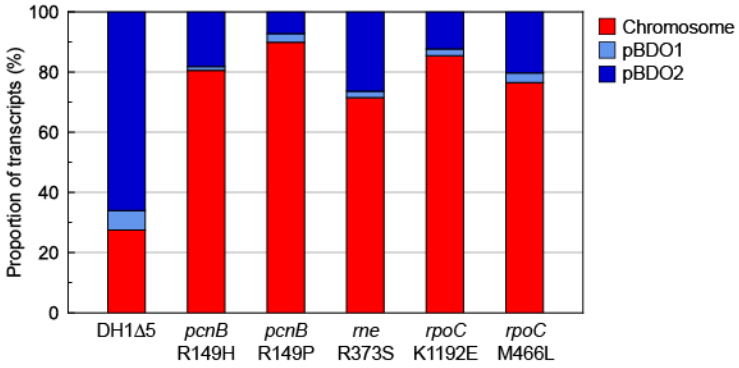


Figure 3.4. *Origin of transcripts in the DH1Δ5 series RNA-seq data set. Values shown are transcripts per hundred which represents the proportion of transcripts in each sample derived from either the chromosome or one of the plasmids carrying genes for butanol production (pBDO1, or pBDO2). The proportion of chromosome-coded transcripts increased for all of the mutant strains relative to the parent strain ($p < 0.01$).*

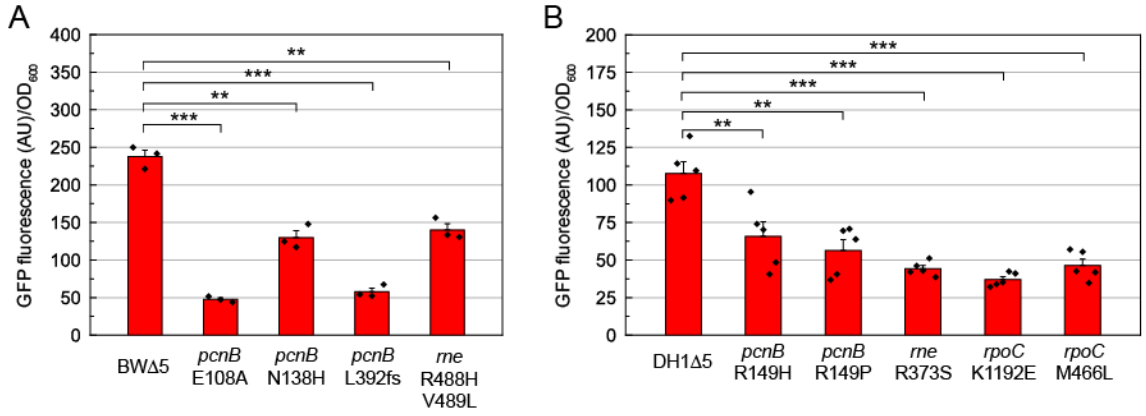


Figure 3.5. *GFP reporter gene to probe protein expression. (A) All of the mutant strains produced significantly less GFP than the parent strain BWΔ5. Strains containing pPDHc and pGFP were grown aerobically for 18 h at 30 °C in LB with 25 g/L glucose. (B) A similar trend was observed in the DH1 series, where the mutant strains showed a decrease in GFP expression. Strains containing pBDO1 and pBDO2 were grown aerobically for 18 h at 30 °C in TB with 25 g/L glucose. Error bars represent the standard error. Statistical analysis by a two-tailed, two-sample equal variance t-test, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

Previous studies have reported that mutations in *pcnB* affect the copy number of ColE1 plasmids [39,40]. The gene *pcnB* was originally identified because mutations in that genetic locus resulted in decreased copy number of plasmids with the ColE1 and related origins of replication [40]. In fact, the name *pcnB* was designated for plasmid copy number [40]. Mutations in *rne* and *rpoC* have also been shown to reduce copy number of ColE1-type plasmids [41,42]. Replication of ColE1 plasmids is initiated via an RNA pre-primer known as RNAII encoded within the origin of replication. After processing by RNAseH, a host factor, this RNA serves as the primer for plasmid replication via DNA polymerase I [43]. Replication is regulated by a short-lived, antisense RNA also encoded within the origin known as RNAI. RNAI can hybridize with RNAII and prevent RNAII from serving as a primer for plasmid replication [44,45]. *pcnB* and *rne* have both been implicated in the degradation of the regulatory RNA, RNAI [42,46]. Mutations in these genes have been shown to reduce the degradation rate of RNAI, increasing the likelihood of RNAI-RNAII hybridization, ultimately leading to less plasmid replication [42,46]. Mutations in *rpoC* have been shown to have altered expression from the RNAI and RNAII promoters, increasing the ratio of RNAI to RNAII and also leading to reduced copy numbers [41].

These previous reports seem particularly relevant as pBu2 and pBDO2 utilize ColE1 origins. The second plasmid carrying the other half of the genes in each pathway, either pBu1 or pBDO1, utilize p15A origins, which replicate in a similar manner to ColE1 [43]. Thus, it seems likely that the mutant strains containing mutations in *pcnB*, *rne*, and *rpoC*, maintain the production plasmids at lower copy numbers. Reduced plasmid-borne gene expression observed in the mutant strains from RNA-seq and the GFP reporter assay could therefore be explained by reduced plasmid copy numbers.

We tested this hypothesis by determining the copy number of production plasmids in the parent and mutant strain. We used quantitative PCR (qPCR) to measure the abundance of the production plasmids relative to the chromosome (*Figure 3.6*). Indeed, we found significantly reduced plasmid copy numbers for both butanol production plasmids, pBu1 and pBu2, in the mutant strains relative to the parent strain, BWΔ5. A similar trend was again seen in the DH1Δ5 series. All of the mutant

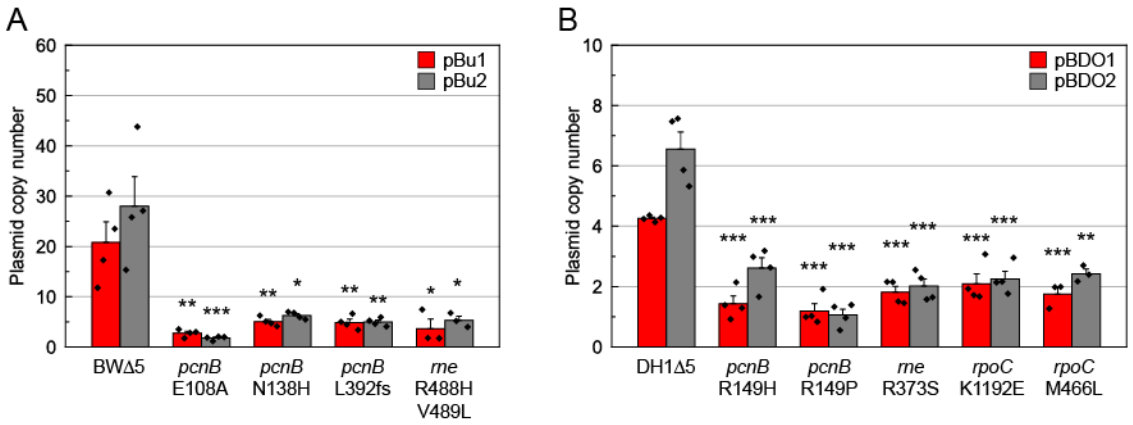


Figure 3.6. *Plasmid copy number quantified with qPCR. (A) All BW mutant strains have a decreased plasmid copy number for both pBu1 and pBu2. BW strains containing pPDHc, pBu1, and pBu2 were grown aerobically at 30 °C in LB with 25 g/L glucose (B) A similar trend was seen for the DH1Δ5 mutants, which all maintain pBDO1 and pBDO2 at a lower copy number than the parent strain. DH1 strains containing pBDO1 and pBDO2 were grown aerobically at 30ºC in TB with 25 g/L glucose. Bars represent mean ± SE. Symbols above bars indicate significance compared to the parent strain (BWΔ5 or DH1Δ5). Statistical analysis by two-tailed, two-sample equal variance t-test, * p < 0.05, ** p < 0.01, *** p < 0.001.*

strains maintained the BDO production plasmids, pBDO1 and pBDO2, at a lower copy number than the parent strain, DH1Δ5. These data support the hypothesis that all of the mutant strains in this study reduce pathway gene expression through a reduction in plasmid copy number.

After establishing reduced pathway expression in the mutant strains, we sought to investigate the effect on growth. In order to probe effects of pathway expression, we compared growth of the parent and mutant strains with or without induction. Performing this analysis in anerobic conditions is complicated by the fact that the *n*-butanol pathway is essential for anaerobic growth. Instead, we examined growth of the butanol strains in aerobic conditions. Metabolic shifts in the mutant strains that improve fermentative growth should not have an effect in aerobic conditions, allowing us to isolate the effect of relieving metabolic burden.

Growth curves of BW25113 wild type (BW) and the BWΔ5 series of strains with either pGFP or pBut were obtained in both uninduced and induced conditions (*Figure 3.7*). First, differences can be seen comparing growth curves for strains containing pGFP to those with pBut even in uninduced conditions. BW and BWΔ5 both show longer lag phases and lower growth rates when carrying pBut rather than pGFP (*Figure 3.8*). We attribute the differences in growth without

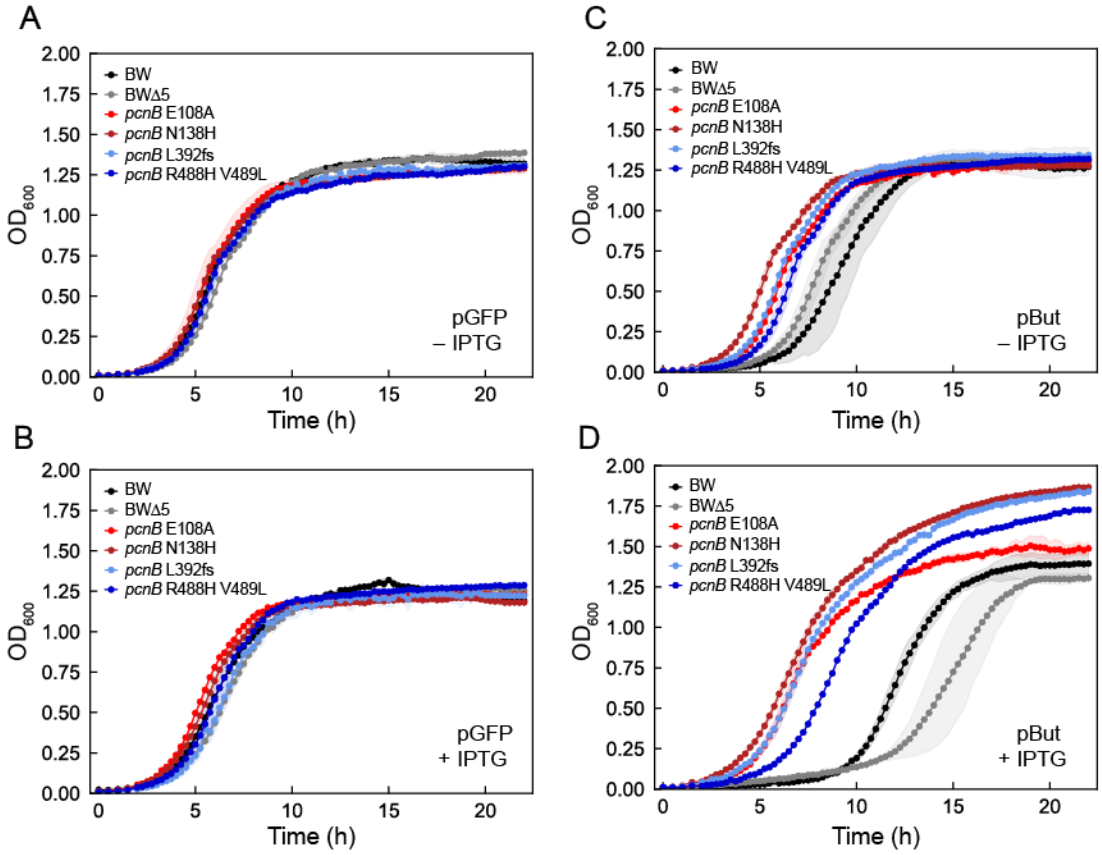


Figure 3.7. *Aerobic growth curves of the BWΔ5 series. (A) Strains grown with pPDHc and pGFP without induction. (B) Strains growth with pPDHc and pGFP with 1 mM IPTG induction. (C) Strains grown with pPDHc and pBut without induction. (D) Strains growth with pPDHc and pBut with 1 mM IPTG induction. Growth curves of the strains with pGFP without induction are indistinguishable from the induced condition. However, large differences in the growth curves of the strains containing pBut are seen upon induction. BW indicates BW25113. Cultures were grown aerobically at 30 ºC in LB with 25 g/L glucose. The error band represents the standard deviation of two replicates.*

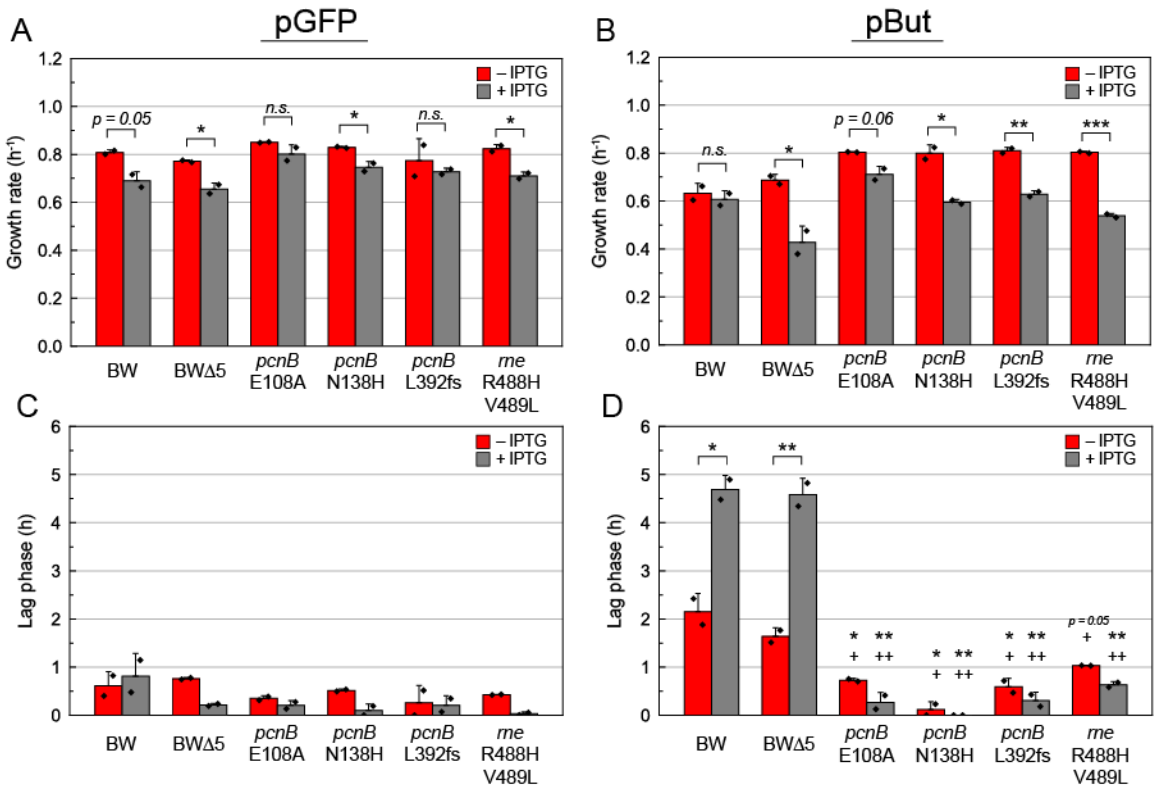


Figure 3.8. *Quantification of growth rate and lag phase for the BWΔ5 series in aerobic conditions. (A) Growth rates were calculated from growth curves of the strains carrying pPDHc and pGFP with and without 1 mM IPTG induction. (B) Growth rates were calculated from growth curves of the strains carrying pPDHc and pBut with and without 1 mM IPTG induction. (C) The lag phases were calculated from growth curves of the strains carrying pPDHc and pGFP with and without 1 mM IPTG induction. (D) The lag phases were calculated from growth curves of the strains carrying pPDHc and pBut with and without 1 mM IPTG induction. Growth rate significantly drops with induction for most strains containing either pBu2 or pGFP. For both BW and BWΔ5 strains, the lag phase was longer for cultures containing pBu2 than the control, pGFP, even in the uninduced condition. Induction lengthened the lag phase of both BW and BWΔ5 containing pBu2, an effect not seen with the pGFP control plasmid. The lag phase for mutant strains was shorter than both the wild type BW and parent BWΔ5. Cultures were grown at 30 ºC in LB with 25 g/L glucose. Error bars represent standard error. A two-tailed, two-sample t-test was used to determine significance, * p < 0.05, ** p < 0.01. In panel D, the top symbol represents the p-value for a t-test against BW wild type and the bottom for BWΔ5 parent strain (* or +: p < 0.05, ** or ++: p < 0.01).*

induction to leaky expression of the butanol pathway. Differences in growth are further accentuated in induced conditions (*Figure 3.7B and D*). We see that growth rate drops for each strain with induction. The effect of induction on growth rate is greater in strains carrying the butanol pathway than those carrying pGFP. We also noticed that the growth rate was lower and the lag phase was longer for both BWΔ5 and BW containing pBut compared to pGFP (*Figure 3.8AB*). From the observed deleterious effects of pathway induction, we conclude that expression of the *n*-butanol pathway genes from pBut places a metabolic burden on the microbial host and that this burden is greater than when only GFP is expressed. Even the wild-type BW strain demonstrates this effect although it does not have any of its fermentation pathway genes knocked out.

Comparing the growth curves of BWΔ5 and BW to the mutant strains we see that the mutant strains reach OD_{600} = 1 much faster than BWΔ5 or BW. Growth rates show a trend to be higher for the mutant strains than BWΔ5 carrying pBut in induced conditions. Additionally, lag phase for

all of the mutant strains is substantially shorter than the lag phase observed with pathway expression in BW or BWΔ5. These observations are consistent with the fact that the mutant strains are enriched through our growth selection. Improved growth rates and shorter lag phases seen in the mutant strains, indicate that the mutants do indeed experience less metabolic burden associated with pathway expression.

We also obtained growth curves for DH1 wild type and the DH1Δ5 series of strains containing pBDO1 and either pGFP or pBDO2 (*Figure 3.9*). At a glance, we notice that the wild-type strain, DH1, grows much better than the knockout strains, likely due to disruption of acetate overflow metabolism in the knockout strains [47]. This observation is in contrast to the BW wild-type strain, which showed similar growth behavior to the knockout strain, BWΔ5, representing metabolic differences between BW25113 and DH1. Growth curves for the strains containing pGFP are very similar between uninduced and induced conditions (*Figure 3.9AB*). More differences are seen between induction conditions for the strains carrying pBDO2 (*Figure 3.9CD*). Upon induction, growth rate drops for most strains regardless if they are carrying pGFP or pBDO2 (*Figure 3.10AB*). We observed a larger decrease in growth rate for strains carrying pBDO2 compared to those

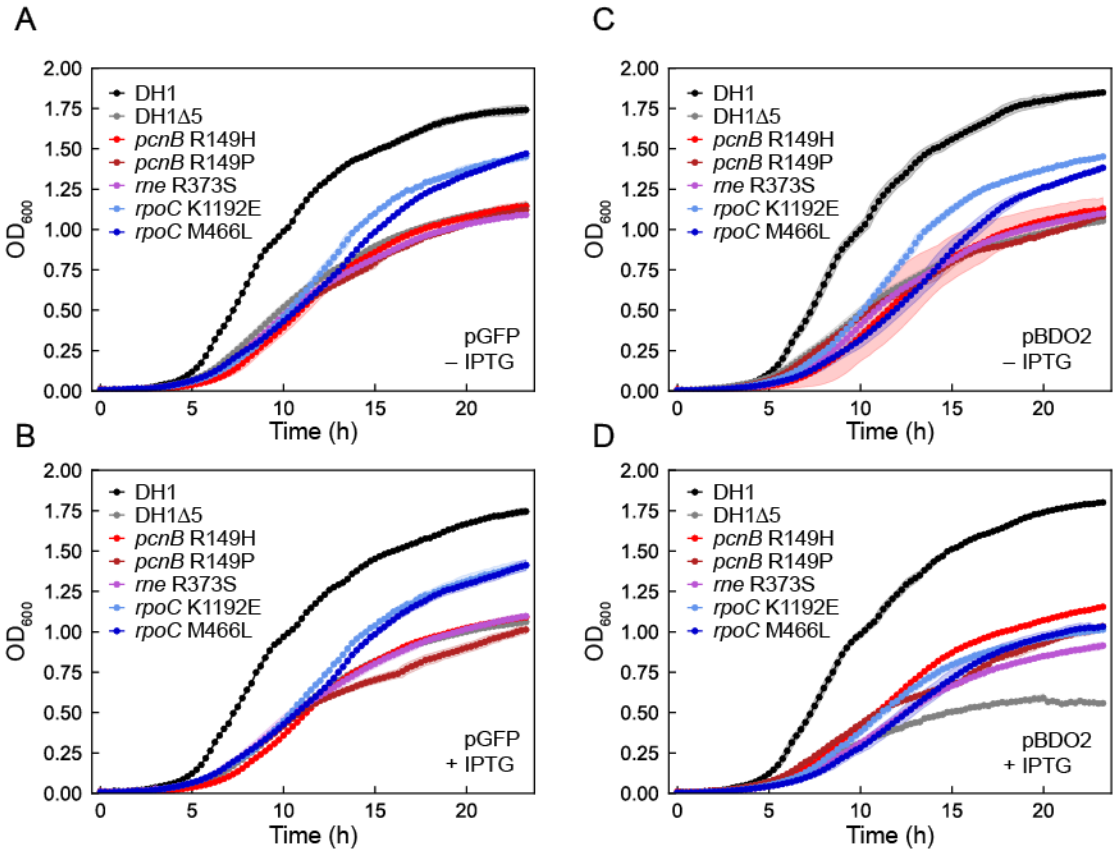


Figure 3.9. *Aerobic growth curves of the DH1Δ5 series in addition to wild-type DH1, which contained no genomic manipulation. (A) Strains grown with pBDO1 and pGFP without induction. (B) Strains growth with pBDO1 and pGFP with 1 mM IPTG induction. (C) Strains grown with pBDO1 and pBDO2 without induction. (D) Strains growth with pBDO1 and pBDO2 with 1 mM IPTG induction. Little difference can be seen between the uninduced and induced conditions when the strains are harboring pGFP. However, a larger effect from induction was observed in the strains containing pBDO2. Cultures were grown aerobically at 30 ºC in LB with 25 g/L glucose. The error band represents the standard deviation of two replicates.*

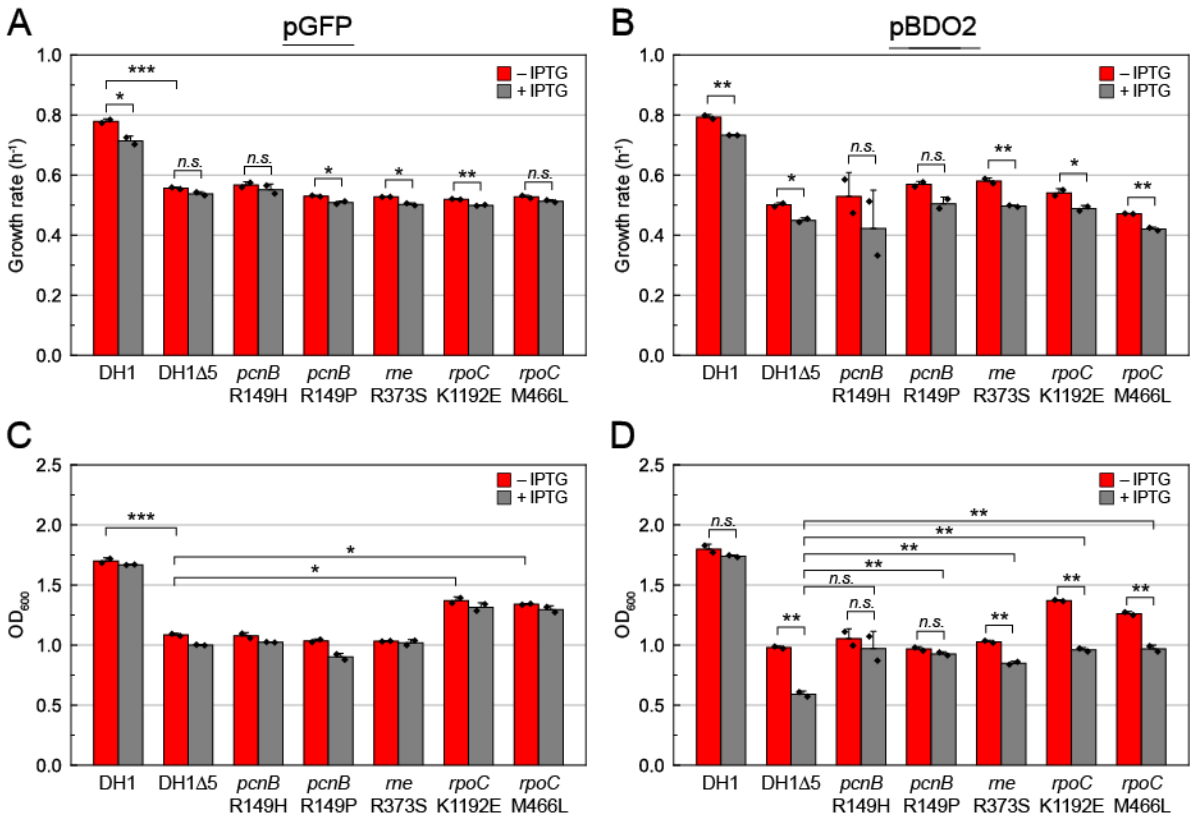


Figure 3.10. *Quantification of growth rate and stationary phase OD_{600} for the DH1Δ5 series in aerobic conditions. (A) Growth rates were calculated from the growth curves of strains carrying pBDO1 and pGFP. (B) Growth rates were calculated from the growth curves of strains carrying pBDO1 and pBDO2. (C) The OD_{600} of cultures carrying pBDO1 and pGFP 20 h post-inoculation. (C) The OD_{600} of cultures carrying pBDO1 and pBDO2 20 h post-inoculation. Growth rate drops upon induction for nearly all of the strains with either pGFP or pBDO2, with the effect more pronounced for strains carrying pBDO2. Mutants have an increased OD_{600} 20 h post-inoculation relative to the parent in induced conditions. Strains were grown aerobically at 30 ºC in LB with 25 g/L glucose containing pBDO1 and either pGFP or pBDO2. Error bars represent standard error of two replicates. Statistical analysis by Student's t-test, *p < 0.05, ** p < 0.01, *** p < 0.001.*

carrying pGFP. When induced, the growth rate for DH1Δ5 was lower and the OD_{600} measured at 20 h decreased when harboring pBDO2 instead of pGFP (*Figure 3.10CD*). These observations suggest that induction of pBDO2 places a metabolic burden on the host and that the magnitude of burden is larger than the burden associated with GFP expression.

The final OD_{600} reached in stationary phase also differed between strains (*Figure 3.10C and D*). Both *rpoC* mutants reach a higher OD_{600} after 20 h than DH1Δ5 and the other mutants when uninduced, regardless of what plasmid they contain. This difference could be caused by an altered transcriptional state resulting from the mutation in RNA polymerase [48]. The most notable difference between DH1Δ5 and the mutant strains, is that the mutant strains reach a significantly higher OD_{600} after 20 h than the parent strain when the pathway genes from pBDO2 are induced. This growth difference with induction in aerobic conditions, where fermentation performance is not involved, indicates that DH1Δ5 experiences greater metabolic burden than the mutants due to pathway expression.

We have shown that pathway gene expression is reduced in the mutant strains via a reduction in plasmid copy number. Growth curves provide evidence that the mutant strains relieve metabolic

burden resulting from pathway gene expression in aerobic conditions. In anaerobic conditions used for fermentation, metabolic burden likely has an even more deleterious effect on growth since there is less cellular energy production [49]. We have provided evidence to support the hypothesis that the main effect of the mutations identified through adaptive evolution in our study is to reduce the metabolic burden placed on the microbial host through decreased expression of the biosynthetic pathway during fermentation.

As another way to support our hypothesis, we sought to design alternative pathway constructs with reduced metabolic burden. We examined the elements affecting expression for each gene in the butanol biosynthetic pathway to identify modifications that could be made to reduce expression (*Figure 3.11A*). Promoter and ribosome binding site (RBS) strength were estimated using the Salis lab calculator (*Appendix 2.3*) [50,51]. We noticed that predicted promoter strengths matched the trends observed in transcript abundance within our RNA-seq data set. We noticed that *aldh46* was by far the most abundant transcript in our sample. Consistent with this observation, the RBS driving *aldh46* translation was predicted to have the highest strength among the RBSs present in our pathway. One of the theories for the origin of growth defects resulting from metabolic burden is the idea of ribosome sequestration. Previous work has claimed that the "free ribosome pool" is a critical factor for optimal cell growth [52–54]. We hypothesized that a transcript such as *aldh46* expressed at high levels could sequester ribosomes from the 'free ribosome pool', especially since this mRNA contains a strong RBS.

We reasoned that changing the RBS to one that is weaker would lower protein expression, mitigate ribosome sequestration, and reduce metabolic burden [55]. As such, we made a variant of pBu2 with an RBS predicted to be weaker in front of *aldh46* and tested its performance in an *n*-butanol fermentation (*Figure 3.11B*). We found that this modified plasmid increased the stationary-phase OD_{600} and *n*-butanol titer, consistent with our hypothesis. We also observed a lack of *n*-butanol production in BWΔ5 *pcnB* L392fs, likely because the combination of lowered gene transcription from a reduced copy number combined with the lower translation rate of a weaker RBS resulted in *aldh46* expression at too low of a level to support a functioning butanol fermentation pathway in the mutant strain.

We explored another strategy to reduce pathway expression by changing the origin of pBu2 from ColE1 to the lower copy pSC101 origin [56,57]. In addition to the wild-type pSC101 origin, we also cloned two mutants of the origin, R43W and E115K, that have previously been shown to increase the copy number [58]. These three plasmids have a variety of different copy numbers allowing us to probe how the copy number of pBu2 affects production (*Figure 3.11 C*). We found that production with BWΔ5 containing the pSC101-Bu2 series improved the stationary phase OD_{600} and *n*-butanol production. The R43W and E115K variants showed greater improvements than the pSC101-WT plasmid, suggesting that a copy number that is somewhere between that of the wild type pSC101 origin and ColE1 is optimal. We believe that the titers of pSC101-R43W and pSC101-E115K origins were unable to reach the level of the mutant because the copy number has not been optimized. However, the observation that a plasmid with lower copy number improves production reinforces the hypothesis that alterations in the copy number have a large impact on *n*-butanol production.

We took yet another approach to reduce the copy number in the parent strain by overexpressing RNAI. As mentioned above, RNAI is a small RNA species that regulates plasmid regulation through its interaction with RNAII, which serves as a primer for plasmid replication [42]. It has previously been shown that overexpression of RNAI can reduce the copy number of ColE1

A

Gene	Origin	Promoter	Predicted Promoter Strength	Measured Transcript Abundance	Predicted RBS Strength
phaA	p15A	T5	9,500	18	170
hbd	p15A	(T5)	9,500	10	140
crt	p15A	Trc	18,000	60	200
ter	ColE1	Tac	12,000	10	24,000
aldh46	ColE1	Trc	18,000	100	170,000
adh2	ColE1	(Trc)	18,000	21	10,000
aldh46	ColE1	Trc	18,000	*n.d.*	90,000

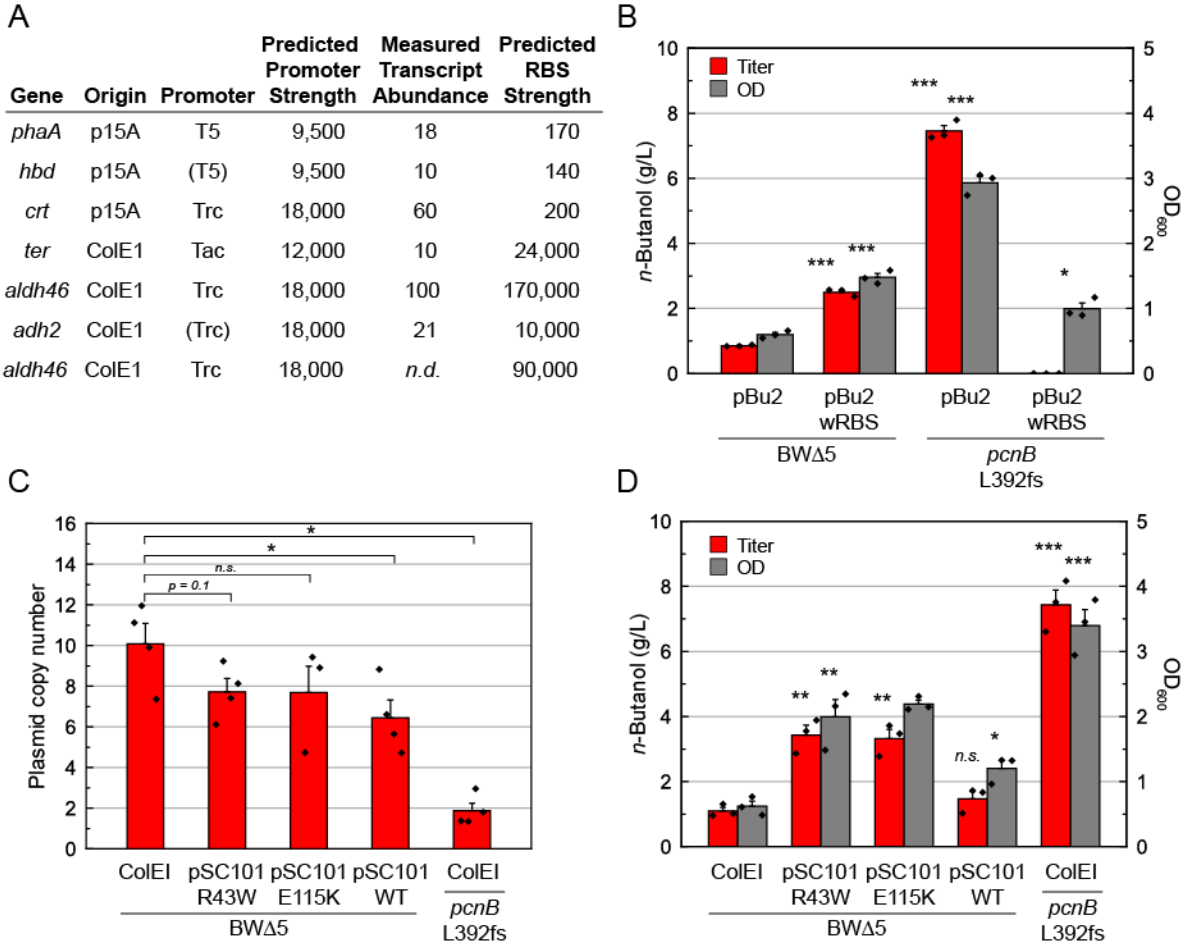


Figure 3.11. *Strategies to reduce metabolic burden. (A) The table shows the genomic elements affecting gene expression along with predicted promoter and RBS strengths. Promoter strength estimated from RNA-seq data is also included. (B) The strong RBS in front of aldh46 was exchanged for a weaker RBS. This change improved stationary phase OD and titers. Strains containing pPDHc, pBu1, and either pBu2 or pBu2-wRBS were grown anaerobically for 3 d at 30 °C in LB with 25 g/L glucose. (C) The origin of pBu2 was swapped from the medium-copy ColE1 to variants of the lower copy SC101 origin. Several point mutations were introduced to the origin of pSC101 to increase copy number. (D) These alternate constructs increased the stationary phase OD and titers. Strains containing pPDHc, pBu1, and either pBu2, pSC101R43W-Bu2, pSC101E115K-Bu2, or pSC101WT-Bu2 were grown anaerobically for 3 d at 30 °C in LB with 25 g/L glucose. Error bars represent standard error of two replicates. Statistical analysis by Student's t-test, *p < 0.05, ** p < 0.01, *** p < 0.001.*

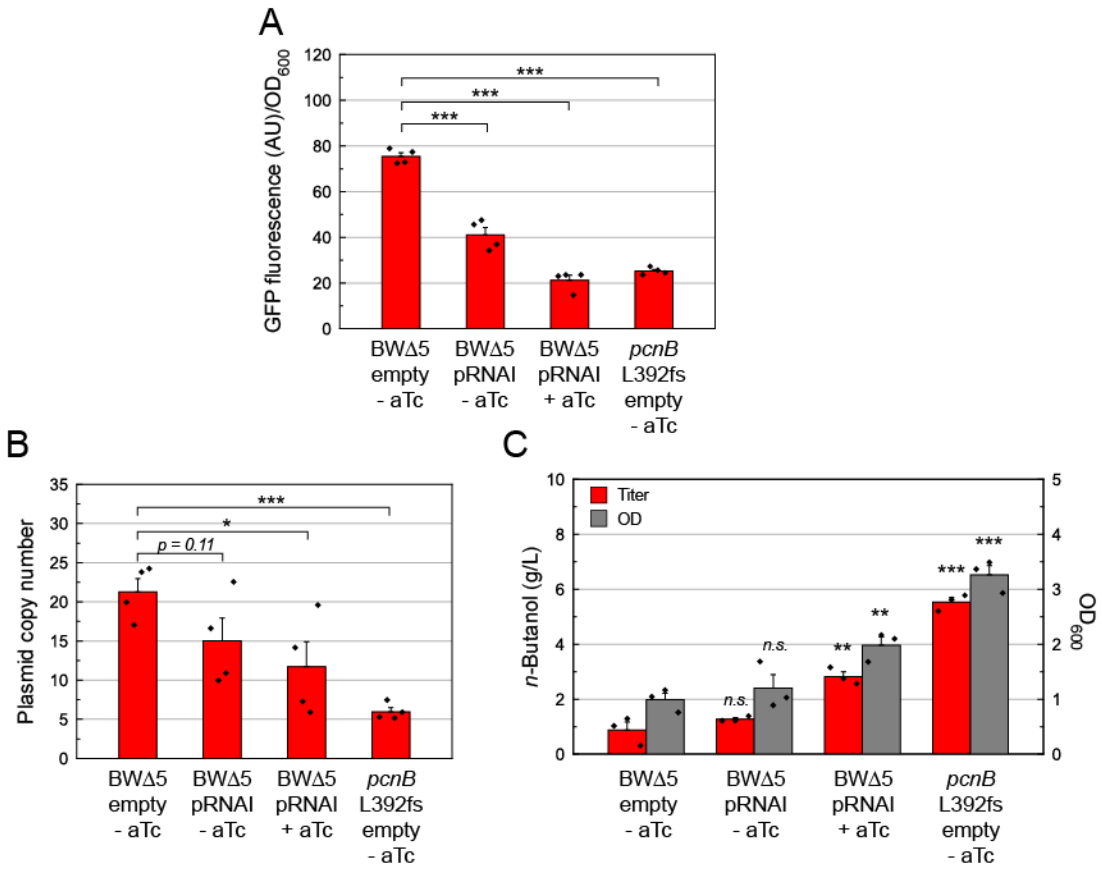


Figure 3.12 *Overexpression of RNAI to reduce ColE1 copy number. pRNAI contains RNAI gene under an aTc controlled promoter. (A) pGFP was used as a reporter for protein expression. pRNAI reduced GFP expression uninduced, and the effect was even larger when induced. (B) A similar trend is reflected in the copy number of pBut. (C) Growth and titer are significantly increased when RNAI expression from pRNAI is induced. Strains contained pPDHc with either pBut or pGFP. Fermentation was carried out anaerobically for 3 d at 30 ºC in LB with 25 g/L glucose. Error bars represent standard error of two replicates. Statistical analysis by Student's t-test, *p < 0.05, ** p < 0.01, *** p < 0.001.*

plasmids [59]. We constructed a plasmid containing RNAI under control of the inducible P_{Tet} promoter. We first investigated the effect of pRNAI on GFP expression from pGFP (*Figure 3.12A*). We see that BWΔ5 containing both pGFP and pRNAI shows reduced GFP expression, even when expression of RNAI is not induced. It is known that the presence of other plasmids can affect the copy number of ColE1 type plasmids like pGFP, and in turn, gene expression [43]. Upon RNAI induction, we see a reduction in GFP expression to levels seen in the *pcnB* L392fs mutant strain. We also determined the copy number of pBut in BWΔ5 containing pRNAI and we observed a similar trend to that seen with GFP expression (*Figure 3.12B*). The presence of pRNAI without induction lowered pBut copy number and this effect was accentuated upon expression of RNAI. Even with overexpression of RNAI, pBut is still maintained at a higher copy number than what is measured in the BWΔ5 *pcnB* L392fs mutant. GFP expression and copy number measurements suggest that pRNAI has the intended effect on gene expression from pBut. Indeed, when we perform a production with pRNAI, we see slight improvements to production without induction,

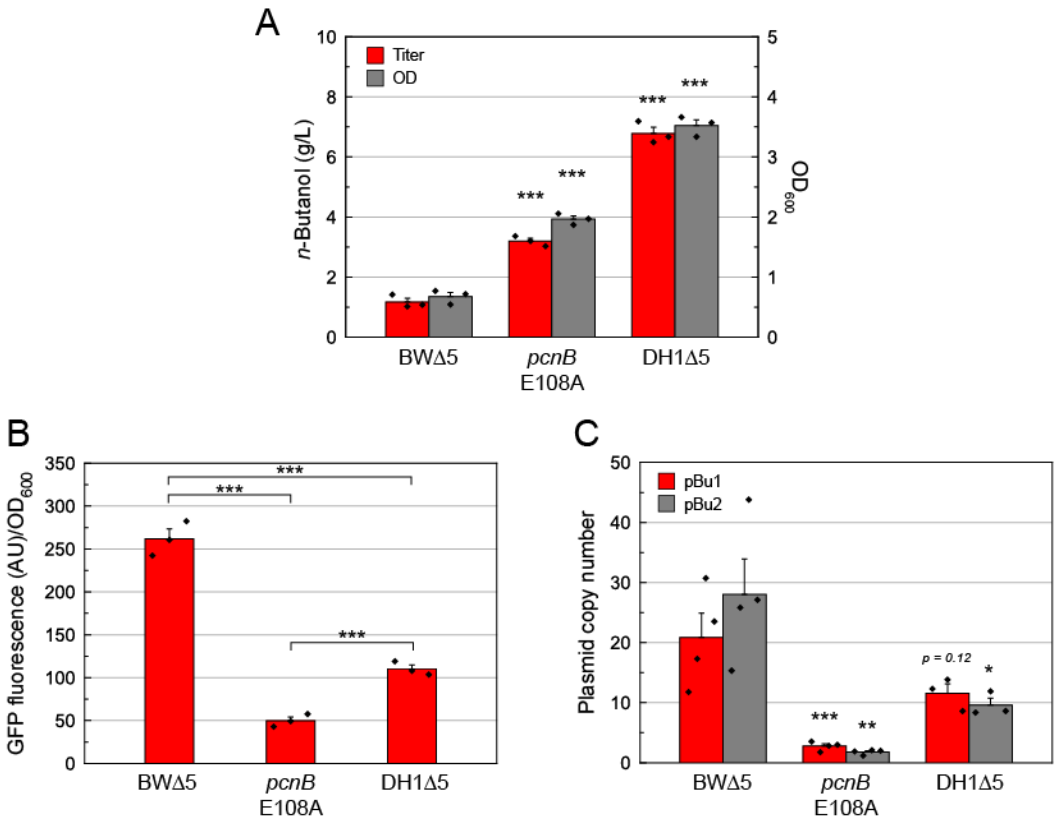


Figure 3.13. *Comparing butanol production in BWΔ5 and DH1Δ5. (A) Relative to BWΔ5, DH1Δ5 produces more butanol and reaches a higher stationary phase OD during fermentation. Significance is indicated above bars for comparison to BWΔ5. Strains containing pPDHc, pBu1 and pBu2 were grown anaerobically for 3 d at 30 °C in LB with 25 g/L glucose. (B) Using a plasmid borne copy of GFP as a reporter for gene expression, DH1Δ5 shows reduced expression compared to BWΔ5. pcnB E108A mutant has even lower GFP expression. Strains contained pPDHc and pGFP. (C) DH1Δ5 was found to maintain the plasmids pBu1 and pBu2 at a lower copy number than BWΔ5, but at a higher copy number than pcnB E108A. Significance is indicated above bars for comparison to BWΔ5. Error bars represent standard error of two replicates. Statistical analysis by Student's t-test, *p < 0.05, ** p < 0.01, *** p < 0.001.*

and even greater improvements with RNAI induction (*Figure 3.12C*). However, growth and titer does not reach the level seen for BWΔ5 *pcnB* L392fs, likely because the copy number of pBut has not been sufficiently reduced. Again, this approach shows that reducing the copy number of the plasmids used for pathway expression improves production.

We also noticed that DH1Δ5 reaches a higher OD_{600} and produces more butanol during fermentation compared to BWΔ5 (*Figure 3.13*). In order to probe these differences further, we examined GFP expression from pGFP and measured the copy number of pBu1 and pBu2 in DH1Δ5. Similar to the BWΔ5 mutant strains, DH1Δ5 expressed less GFP and maintained pBu1 and pBu2 at a lower copy number than BWΔ5. This provides some evidence that plasmid copy number is host dependent. Based on these experiments, we conclude that DH1Δ5 expresses the butanol pathway genes to a lesser extent and experiences less metabolic burden, analogous to the BWΔ5 mutant strains.

3.4. Conclusion

Building on the conclusions from Chapter 2, we formulated the hypothesis that expression of the biosynthetic pathways placed a metabolic burden on the parent strains, BWΔ5 and DH1Δ5. This hypothesis was further explored in this chapter. From a global analysis of the transcriptional state of the parent and mutant strains, we observed changes in gene expression that have previously been attributed to the cellular response to metabolic burden, supporting this hypothesis (*Figure 3.1*). Additionally, transcript levels of the genes in both the butanol and BDO biosynthetic pathways were greatly reduced in the mutant strains (*Figure 3.1* and *Figure 3.3*). This observation suggests that reduced expression of pathway genes in the mutant strains relieves metabolic burden, while still yielding a functional fermentation pathway. We next verified that the mutant strains reduce expression of GFP from a reporter plasmid (*Figure 3.5*). This reduction in heterologous protein expression appears to be derived from a reduction in plasmid copy number. Indeed, all of the mutant strains in this study maintain the production plasmids at lower levels than the respective parent strain (*Figure 3.6*). Through analysis of growth curves, we showed that the parent strains experience metabolic burden and that the mutant strains relieve metabolic burden associated with induction of the pathway genes (*Figure 3.7* and *Figure 3.9)*. We then applied several strategies to reduce the metabolic burden of pathway expression via approaches that do not rely on genomic mutations. All of these strategies improved butanol production (*Figure 3.11* and *Figure 3.12*).

Taken together, the data shown here support the hypothesis that the mutations found through adaptive evolution likely arose as a cellular response to reduce metabolic burden resulting from pathway expression. We can learn a lesson from this cellular adaptation. Coercing microbes to over express non-native proteins and pathways can have a deleterious effect on cellular performance, especially when grown in energy restricted conditions. Counterintuitively, reducing gene expression levels can actually lead to greater yields of desirable metabolites [60].

3.5. References

1. Glick, B. R. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **13,** 247–261 (1995).
2. Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S. & Koffas, M. A. G. Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications. *Trends Biotechnol.* **34,** 652–664 (2016).
3. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
4. Engler, C. & Marillonnet, S. Golden Gate cloning. *Methods Mol. Biol. Clifton NJ* **1116,** 119–131 (2014).
5. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8,** 91 (2008).
6. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34,** 525–527 (2016).
7. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

8. Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X. & Thomas, P. D. Protocol update for large-scale genome and gene function analysis with panther classification system (v.14.0). *Nat. Protoc.* **14,** 703–721 (2019).

9. Untergasser, A., Ruijter, J. M., Benes, V. & van den Hoff, M. J. B. Web-based LinRegPCR: application for the visualization and analysis of (RT)-qPCR amplification and melting data. *BMC Bioinformatics* **22,** 398 (2021).

10. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20,** 631–656 (2019).

11. Pinske, C., Jaroschinsky, M., Linek, S., Kelly, C. L., Sargent, F. & Sawers, R. G. Physiology and bioenergetics of [NiFe]-hydrogenase 2-catalyzed H_2-consuming and H_2-producing reactions in *Escherichia coli*. *J. Bacteriol.* **197,** 296–306 (2015).

12. Beaton, S. E., Evans, R. M., Finney, A. J., Lamont, C. M., Armstrong, F. A., Sargent, F. & Carr, S. B. The structure of hydrogenase-2 from *Escherichia coli*: implications for H_2-driven proton pumping. *Biochem. J.* **475,** 1353–1370 (2018).

13. Zientz, E., Bongaerts, J. & Unden, G. Fumarate regulation of gene expression in *Escherichia coli* by the DcuSR (*dcuSR* genes) two-component regulatory system. *J. Bacteriol.* **180,** 5421–5425 (1998).

14. Maklashina, E., Berthold, D. A. & Cecchini, G. Anaerobic expression of *Escherichia coli* succinate dehydrogenase: functional replacement of fumarate reductase in the respiratory chain during anaerobic growth. *J. Bacteriol.* **180,** 5989–5996 (1998).

15. Meng, W., Green, J. & Guest, J. R. FNR-dependent repression of *ndh* gene expression requires two upstream FNR-binding sites. *Microbiol. Read. Engl.* **143 (Pt 5),** 1521–1532 (1997).

16. Haddadin, F. T. & Harcum, S. W. Transcriptome profiles for high-cell-density recombinant and wild-type *Escherichia coli*. *Biotechnol. Bioeng.* **90,** 127–153 (2005).

17. Bosdriesz, E., Molenaar, D., Teusink, B. & Bruggeman, F. J. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *Febs J.* **282,** 2029–2044 (2015).

18. Sharma, A. K., Mahalik, S., Ghosh, C., Singh, A. B. & Mukherjee, K. J. Comparative transcriptomic profile analysis of fed-batch cultures expressing different recombinant proteins in *Escherichia coli*. *AMB Express* **1,** 33 (2011).

19. Ceroni, F., Boo, A., Furini, S., Gorochowski, T. E., Borkowski, O., Ladak, Y. N., Awan, A. R., Gilbert, C., Stan, G.-B. & Ellis, T. Burden-driven feedback control of gene expression. *Nat. Methods* **15,** 387–393 (2018).

20. Tan, J., Sastry, A. V., Fremming, K. S., Bjørn, S. P., Hoffmeyer, A., Seo, S., Voldborg, B. G. & Palsson, B. O. Independent component analysis of *E. coli*'s transcriptome reveals the cellular processes that respond to heterologous gene expression. *Metab. Eng.* **61,** 360–368 (2020).

21. Guleria, R., Jain, P., Verma, M. & Mukherjee, K. J. Designing next generation recombinant protein expression platforms by modulating the cellular stress response in *Escherichia coli*. *Microb. Cell Factories* **19,** 227 (2020).

22. B. Singh, A., K. Sharma, A. & J. Mukherjee, K. Analyzing the metabolic stress response of recombinant *Escherichia coli* cultures expressing human interferon-beta in high cell density fed batch cultures using time course transcriptomic data. *Mol. Biosyst.* **8,** 615–628 (2012).

23. Gill, R. T., Valdes, J. J. & Bentley, W. E. A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. *Metab. Eng.* **2,** 178–189 (2000).

24. Oh, M.-K. & Liao, J. C. DNA Microarray Detection of metabolic responses to protein overproduction in *Escherichia coli*. *Metab. Eng.* **2,** 201–209 (2000).

25. Lesley, S. A., Graziano, J., Cho, C. Y., Knuth, M. W. & Klock, H. E. Gene expression response to misfolded protein as a screen for soluble recombinant protein. *Protein Eng. Des. Sel.* **15,** 153–160 (2002).

26. Ades, S. E., Grigorova, I. L. & Gross, C. A. Regulation of the alternative sigma factor σ^E during initiation, adaptation, and shutoff of the extracytoplasmic heat shock response in *Escherichia coli*. *J. Bacteriol.* **185,** 2512–2519 (2003).

27. Bianchi, A. A. & Baneyx, F. Hyperosmotic shock induces the σ^{32} and σ^E stress regulons of *Escherichia coli*. *Mol. Microbiol.* **34,** 1029–1038 (1999).

28. Rouvière, P. E., De Las Peñas, A., Mecsas, J., Lu, C. Z., Rudd, K. E. & Gross, C. A. *rpoE*, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*. *EMBO J.* **14,** 1032–1042 (1995).

29. Klein, G., Stupak, A., Biernacka, D., Wojtkiewicz, P., Lindner, B. & Raina, S. Multiple transcriptional factors regulate transcription of the *rpoE* gene in *Escherichia coli* under different growth conditions and when the lipopolysaccharide biosynthesis is defective. *J. Biol. Chem.* **291,** 22999–23019 (2016).

30. Minagawa, J., Nakamura, H., Yamato, I., Mogi, T. & Anraku, Y. Transcriptional regulation of the cytochrome b562-o complex in *Escherichia coli*. Gene expression and molecular characterization of the promoter. *J. Biol. Chem.* **265,** 11198–11203 (1990).

31. Ogasawara, H., Ishida, Y., Yamada, K., Yamamoto, K. & Ishihama, A. PdhR (pyruvate dehydrogenase complex regulator) controls the respiratory electron transport system in *Escherichia coli*. *J. Bacteriol.* **189,** 5534–5541 (2007).

32. Woodruff, L. B. A., Boyle, N. R. & Gill, R. T. Engineering improved ethanol production in Escherichia coli with a genome-wide approach. *Metab. Eng.* **17,** 1–11 (2013).

33. Masuda, N. & Church, G. M. Regulatory network of acid resistance genes in *Escherichia coli*. *Mol. Microbiol.* **48,** 699–712 (2003).

34. Weber, A., Kögl, S. A. & Jung, K. Time-dependent proteome alterations under osmotic stress during aerobic and anaerobic growth in *Escherichia coli*. *J. Bacteriol.* **188,** 7165–7175 (2006).

35. Phadtare, S. *Escherichia coli* cold-shock gene profiles in response to over-expression/deletion of CsdA, RNase R and PNPase and relevance to low-temperature RNA metabolism. *Genes Cells Devoted Mol. Cell. Mech.* **17,** 850–874 (2012).

36. Thomas, J. G. & Baneyx, F. Roles of the *Escherichia coli* small heat shock proteins IbpA and IbpB in thermal stress management: comparison with ClpA, ClpB, and HtpG *in vivo*. *J. Bacteriol.* **180,** 5165–5172 (1998).

37. Brissette, J. L., Weiner, L., Ripmaster, T. L. & Model, P. Characterization and sequence of the *Escherichia coli* stress-induced *psp* operon. *J. Mol. Biol.* **220,** 35–48 (1991).

38. Siegele, D. A. & Hu, J. C. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci.* **94,** 8168–8172 (1997).

39. Masters, M., Colloms, M. D., Oliver, I. R., He, L., Macnaughton, E. J. & Charters, Y. The *pcnB* gene of *Escherichia coli*, which is required for ColE1 copy number maintenance, is dispensable. *J. Bacteriol.* **175,** 4405–4413 (1993).

40. Lopilato, J., Bortner, S. & Beckwith, J. Mutations in a new chromosomal gene of *Escherichia coli* K-12, *pcnB*, reduce plasmid copy number of pBR322 and its derivatives. *Mol. Gen. Genet. MGG* **205,** 285–290 (1986).

41. Ederth, J., Isaksson, L. & Abdulkarim, F. Origin-specific reduction of ColE1 plasmid copy number due to mutations in a distinct region of the *Escherichia coli* RNA polymerase. *Mol. Genet. Genomics* **267,** 587–592 (2002).

42. Lin-Chao, S. & Cohen, S. N. The rate of processing and degradation of antisense RNAI regulates the replication of ColE1-type plasmids in vivo. *Cell* **65,** 1233–1242 (1991).

43. Camps, M. Modulation of ColE1-like plasmid replication for recombinant gene expression. *Recent Pat. DNA Gene Seq.* **4,** 58–73 (2010).

44. Lacatena, R. M. & Cesareni, G. Interaction between RNA1 and the primer precursor in the regulation of Co1E1 replication. *J. Mol. Biol.* **170,** 635–650 (1983).

45. Tomizawa, J. Control of ColE1 plasmid replication: binding of RNA I to RNA II and inhibition of primer formation. *Cell* **47,** 89–97 (1986).

46. He, L., Söderbom, F., Wagner, E. G., Binnie, U., Binns, N. & Masters, M. PcnB is required for the rapid degradation of RNAI, the antisense RNA that controls the copy number of ColE1-related plasmids. *Mol. Microbiol.* **9,** 1131–1142 (1993).

47. Wolfe, A. J. The Acetate Switch. *Microbiol. Mol. Biol. Rev.* **69,** 12–50 (2005).

48. Conrad, T. M., Frazier, M., Joyce, A. R., Cho, B.-K., Knight, E. M., Lewis, N. E., Landick, R. & Palsson, B. Ø. RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 20500–20505 (2010).

49. Nelson, D. L., Lehninger, A. L. & Cox, M. M. *Lehninger Principles of Biochemistry*. (Macmillan, 2008).

50. Salis, H. M. in *Methods Enzymol.* (ed. Voigt, C.) **498,** 19–42 (Academic Press, 2011).

51. LaFleur, T. L., Hossain, A. & Salis, H. M. Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. *Nat. Commun.* **13,** 5159 (2022).

52. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330,** 1099–1102 (2010).

53. Carrera, J., Rodrigo, G., Singh, V., Kirov, B. & Jaramillo, A. Empirical model and *in vivo* characterization of the bacterial response to synthetic gene expression show that ribosome allocation limits growth rate. *Biotechnol. J.* **6,** 773–783 (2011).

54. Vind, J., Sørensen, M. A., Rasmussen, M. D. & Pedersen, S. Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes. Expression from reporter genes does not always reflect functional mRNA levels. *J. Mol. Biol.* **231,** 678–688 (1993).

55. Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* **12,** 415–418 (2015).

56. Manen, D. & Caro, L. The replication of plasmid pSC101. *Mol. Microbiol.* **5,** 233–237 (1991).

57. Shao, B., Rammohan, J., Anderson, D. A., Alperovich, N., Ross, D. & Voigt, C. A. Single-cell measurement of plasmid copy number and promoter activity. *Nat. Commun.* **12,** 1475 (2021).

58. Thompson, M. G., Sedaghatian, N., Barajas, J. F., Wehrs, M., Bailey, C. B., Kaplan, N., Hillson, N. J., Mukhopadhyay, A. & Keasling, J. D. Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. *Sci. Rep.* **8,** 1590 (2018).

59. Rouches, M. V., Xu, Y., Cortes, L. B. G. & Lambert, G. A plasmid system with tunable copy number. *Nat. Commun.* **13,** 3908 (2022).

60. Jones, K. L., Kim, S. W. & Keasling, J. D. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab. Eng.* **2,** 328–338 (2000).

Chapter 4: *Towards in vivo production of PHAs with amino-acid derived monomers*

Portions of this work were performed in collaboration with the following persons:

Sunnie Kong performed cloning of BesB PROSS variants. Albert Qiang assisted with Pra and Alg production experiments. Jason Fang synthesized 2H4Py and assisted with synthesis of 2H4Pe. Veronica Stafford cloned, purified and performed enzymatic reactions with L-HicDH, 4HBCS, and PhaC STQK.

4.1. Introduction

Due to increased fossil fuel scarcity, there is a need to transition to alternative sources for commodity chemicals. In Chapters 2 and 3, we focused on the production of two C_4-chemicals, BDO and *n*-butanol. In this chapter, we explore the utility of microbes for the production of plastics, which are traditionally made from petrochemicals. In addition to scarcity, there are also environmental impacts associated with plastics, particularly at the end of use. Globally, only about 18% of plastics waste are recycled, 24% are incinerated and the remaining 58% accumulates in landfills or the environment [1]. This is directly apparent from the large high-concentration accumulations of plastics in the South Pacific subtropical gyre and the Eastern Pacific Ocean gyre [2]. Rising concentrations of macroplastics and microplastics, their degradation products, are having significant ecological impacts [3].

With these considerations in mind, bioplastics are a promising alternative to the petrochemically-derived polymers. Bioplastics are made from biomass, a renewable carbon source, and in some cases can even be degraded by organisms in the environment preventing wide-scale accumulation [4]. The most attractive bioplastic to date is polylactic acid (PLA) due to its biodegradability, biocompatibility, clarity, and superior barrier properties for general performance plastics [5]. PLA has found use in a variety of applications including consumer goods, 3D printing, agriculture, and medicine. The main producer of commercial PLA, NatureWorks LLC, uses lactate produced by *Escherichia coli* fermentation on plant starch from corn and sugarcane [5]. PLA is then synthesized by chemical polymerization of the bio-sourced lactate [6]. In addition to PLA, other bio-based polymers have been explored. Polylactic acid belongs to a class of polyesters known as polyhydroxy alkanoates (PHAs). Recently, Danimer Scientific has commercialized a PHA for use in food packaging [7]. Another well studied PHA is poly 3-hydroxybutyrate (P3HB), which has found commercial use in medical applications [8].

Material properties of PHAs can be tuned through several approaches. Blending PHA with other biobased polymers such as cellulose and lignin can yield favorable results [9,10]. Alternatively, material properties can be modified by monomer composition [11]. P3HB and poly 3-hydroxyvalerate (P3HV) have mechanical properties that are comparable to those of PP and PE although they are more brittle and have a much lower elongation to break [12]. Copolymers of poly(3HB-co3HV) have good strength properties that can be varied depending on monomer proportions [13]. In contrast, medium-chain-length PHAs possess low melting and glass transition temperatures, often existing as amorphous liquids or semi-crystalline elastomers [14]. As an additional approach, chemical modification of PHAs can be used to alter material properties. Block copolymers have been made by chemically grafting other polymers to the end groups of PHA [15]. Additionally, the side chains of PHAs can be functionalized. It has been shown that PHAs containing unsaturated groups can be converted to carboxyl groups via epoxidation for further chemical modification [16]. Unsaturated PHAs have also been chlorinated by substitution reactions to yield polymers with higher glass transition temperatures [17]. These chlorinated PHAs can then be used for further reactions including cross-linking [18]. Additionally, olefins can also be leveraged for well-known thiol-ene click chemistry [19].

PHAs are natively made by microbes as a method of carbon and energy storage in limited growth conditions. PHB was first discovered by the French scientist Lemoigne in 1925 in *Bacillus megaterium* [20]. Since this discovery, a number of other bacteria have been shown to produce PHAs. Even more, studies have shown that a number of fed monomers can be incorporated into PHAs synthesized *in vivo* [21]. Aside from feeding monomers for incorporation, several groups

have investigated the idea of using microbial metabolism to make unique PHA monomers from glucose and incorporate them into growing PHA chains. Our lab has previously shown that this approach can be used to incorporate α-branched hydroxy acids into PHAs [22].

The enzyme responsible for *in vivo* PHA polymerization (PhaC) has been studied in detail. Mutated PhaC enzymes from various organisms have been investigated for improved enzymatic properties and altered substrate specificity. Through these engineering efforts, a mutant of PhaC from *Pseudomonas* sp. 61-3 (PhaC STQK) capable of polymerizing lactate was discovered [23]. Leveraging this new activity, metabolic pathway was created to produce PLA *in vivo* from a glucose fed culture [24]. Beyond lactate, PhaC STQK has also been shown to incorporate 2-hydroxy acids derived from leucine, phenylalanine, valine and isoleucine into PHAs *in vivo* [25].

As an analogous strategy, we looked to devise a pathway to incorporate other amino acid-derived 2-hydroxy acids. Previously, our lab characterized the biosynthetic pathway for a non-canonical amino acid, β-ethynylserine (*Figure 4.1A*)[26]. Enzymes in this pathway can be used to generate two other non-canonical amino acids from lysine, propargylglycine (Pra) and allylglycine (Alg), containing a terminal alkyne and olefin respectively. Incorporation of monomers derived from these amino acids into PHAs would provide unsaturated sites for downstream chemical modification. Alkynes can be functionalized via azide-alkyne cycloaddition in addition to other chemistries and reactive functionality of olefins was highlighted above [27].

Thus, we devised a metabolic pathway to convert lysine to one of these two non-canonical amino acids and then incorporate the 2-hydroxy form of these amino acids into a PHA (*Figure 4.1B*). Three enzymes in the β-ehtynylserine pathway, BesD, BesC, and BesB have been shown to convert lysine into Pra [26]. BesC also accepts lysine as a substrate and can perform a deamination of the epsilon amine to afford Alg. Beyond generation of the amino acids, a downstream metabolic process must be constructed for generation of the respective 2-hydroxy acids and subsequent polymerization. The α-amino group of an amino acid must be deaminated to generate a 2-oxo acid via an amino acid dehydrogenase. Then, an alcohol dehydrogenase will be needed to reduce

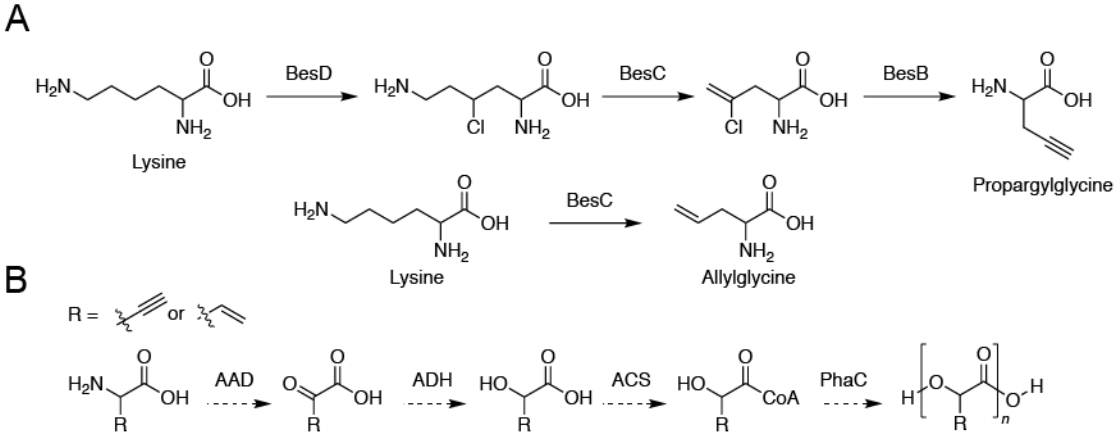


Figure 4.1. *Scheme for in vivo production of PHAs containing unique unsaturated monomers. (A) Propargyl glycine can be made using a truncated form of the β-ethynylserine pathway. The second enzyme in the pathway also has activity towards non-halogenated lysine and can be used to generate allylglycine in vivo. (B) Several enzymes are needed to create a metabolic pathway that can polymerize amino acids. (AAD: amino acid dehydrogenase, ADH: alcohol dehydrogenase, ACS: acetyl-CoA synthetase, PhaC: PHA synthase)*

the 2-oxo acid to the corresponding 2-hydroxy acid. An acetyl-CoA transferase or synthetase will be required to activate the carboxylic acid with a CoA thioester for polymerization with a polyhydroxyalkanoate synthase. Enzymes able to catalyze each of these reactions with our substrates of interest must be identified to construct a functioning pathway.

4.2. Materials and methods

Commercial Materials. Ammonium heptamolybdate tetrahydrate, boric acid, cobalt(II) chloride, Coenzyme A, DL-propargylglycine, DL-Allylglycine, DL-phenylserine, dansyl chloride, iron(II) sulfate heptahydrate, O-benzylhydroxylamine hydrochloride, ß-nicotinamide adenine dinucleotide hydrate (NAD+), ß-nicotinamide adenine dinucleotide reduced disodium salt hydrate (NADH), sodium sulfate, sodium nitrite, sulfuric acid, thiamine hydrochloride, 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB), and 3-phenylserine were purchased from Sigma-Aldrich (St. Louis, MO). Phenylmethanesulfonyl fluoride (PMSF), sodium pyruvate, ammonium bicarbonate (LC-MS grade), ammonium chloride, calcium chloride dihydrate, cupric sulfate, carbenicillin disodium salt (Cb), dithiothreitol (DTT), deoxynucleotides (dNTPs), formic acid (LC-MS grade), HEPES, imidazole, kanamycin sulfate (Km), chloramphenicol (Cm), manganese(II) chloride, O'GeneRuler 1 kb Plus DNA Ladder, PageRuler Plus Prestained Protein Ladder, polyethylene glycol 3350, potassium chloride, potassium phosphate monobasic, sodium chloride, sodium hydroxide, sodium phosphate dibasic heptahydrate, sodium phosphate monobasic monohydrate, trichloroacetic acid, urea were purchased from Thermo Fisher Scientific (Waltham, MA). Glycerol, LB Miller agar, LB Miller broth, TB broth, magnesium chloride hexahydrate, anhydrous magnesium sulfate, anhydrous sodium sulfate, Terrific broth, and triethylamine were purchased from EMD-Millipore (Burlington, MA). Zinc sulfate was purchased from Mallinckrodt (Paris, OH). Instant Blue Protein Stain was from Expedeon (San Diego, CA). Isopropyl ß-D-1-thiogalactopyranoside (IPTG) was from Santa Cruz Biotechnology (Dallas, TX). Bradford assay reagent concentrate and ethidium bromide were purchased from Bio-Rad Laboratories (Hercules, CA). Deuterium oxide was from Cambridge Isotope Laboratories (Tewksbury, MA). Phusion polymerase, Phusion HF buffer, all restriction enzymes, restriction enzyme buffer (CutSmart), T4-PNK, T4 DNA ligase, and Taq ligase were from New England Biolabs (Ipswitch, MA). Ni-NTA agarose resin and DNA purification kits were purchased from Qiagen (Redwood City, CA). Oligonucleotides and gBlocks gene fragments were synthesized by Integrated DNA Technologies (Coralville, IA).

Bacterial Strains. *E. coli* DH10B was used for DNA construction. *E. coli* BL21(DE3)- $T1^R$ was used for protein production. *E. coli* DH1, *E. coli* K-12 MG1655 (DE3), *E. coli* DH5a (DE3), *E. coli* BL21 (DE3) Tuner, *E. coli* C41 (DE3), *E. coli* BL21 (DE3) Star $T1^R$ were used for production experiments.

Gene and plasmid construction. Plasmid construction was carried out using standard molecular biology techniques including the Gibson protocol [28], Golden Gate Assembly [29], or quick-change PCR [30]. PCR amplifications were carried out with Phusion DNA polymerase or GoTaq DNA polymerase, following manufacturer instructions. Primer sequences are listed in *Appendix 2.1A* and gBlocks are listed in *Appendix 2.2*. Constructs were verified by sequencing (Genewiz; South Plainfield, NJ).

pPra2-BesB.1. MBP was amplified from pPra2 with SK1 and SK9. BesB PROSS variant 1 was amplified from a gBlock with SK5 and SK6. Both pieces were inserted into pPra2 digested with XhoI and NdeI via Gibson assembly.

pPra2-BesB.3. MBP was amplified from pPra2 with SK1 and SK9. BesB PROSS variant 1 was amplified from a gBlock with SK7 and SK8. Both pieces were inserted into pPra2 digested with XhoI and NdeI via Gibson assembly.

pET16hp-BcLeuDH. BcLeuDH was amplified from the genome of *Bacillus cereus* with primers JF1 and JF2 and then inserted into pET16hp-IMDH digested with NdeI and BamHI via Gibson assembly.

pET16hp-BsAlaDH. BcLeuDH was amplified from the genome of *Bacillus subtilis* with primers JF3 and JF4 and then inserted into pET16hp-IMDH digested with NdeI and BamHI via Gibson assembly.

pET28a-LHicDH. pET28a-(6His)LHicDH(6His) cloning intermediate was constructed by inserting LHicDH amplified from a gBlock with primers VS01 and VS02 into pET-28a-matB digested with NdeI and XhoI via Gibson assembly. pET28a-LHicDH(His) was constructed by linearizing pET28a-(6His)LHicDH(6His) with primers VS03 and VS04 and ligating the ends together using T4-PNK (New England Biolabs) and T4 DNA ligase (New England Biolabs) following the recommended protocols.

pET16hp-4HBCS. A gBlock containing the 4HBCS coding sequence and homology overhangs was inserted into pET16hp digested with NdeI and BamHI via Gibson assembly.

pET16hp-PhaC STQK. PhaC STQK was amplified from a gBlock with primers VS05 and VS06 and then inserted into pET16hp digested with NdeI and BamHI via Gibson assembly.

Propargylglycine production. Overnight cultures of freshly transformed *E. coli* strains were grown overnight in LB at 37 °C and used to inoculate TB (15 ml) with appropriate antibiotics adjusted to a pH of 6.5 at a ratio of 1:100 (*v*/*v*) in a 250 mL-baffled anaerobic flask with GL45 threaded top (Chemglass). The cultures were grown at 30 °C in a rotary shaker (200 rpm) for 2h before induction with IPTG (1.0 mM) and continued growth at 30°C in a rotary shaker (200 rpm). Samples were collected after 3 d of cell culture.

Allylglycine production. Overnight cultures of freshly transformed *E. coli* strains were grown overnight in LB at 37 °C and used to inoculate LB (15 ml) or M9 (15 ml) when indicated, with appropriate antibiotics at a ratio of 1:100 (*v*/*v*) in a 250 mL-baffled anaerobic flask with GL45 threaded top (Chemglass). The cultures were grown at 30 °C in a rotary shaker (200 rpm) for 2h before induction with IPTG (1.0 mM) and continued growth at 30°C in a rotary shaker (200 rpm). Samples were collected after 2 d of cell culture.

Minimal media. M9 minimal medium contained per liter: 3 g glucose, 12.8 g sodium phosphate dibasic heptahydrate, 3 g potassium phosphate monobasic, 0.5 g sodium chloride, 1 g ammonium chloride, supplemented with 2 mL of 1M magnesium chloride, 1 mL 0.1mM calcium chloride, 20 µL micronutrient solution containing 0.18 g ammonium heptamolybdate tetrahydrate, 1.24g boric acid, 0.36 g cobalt(II) chloride, 0.12 g cupric sulfate, 0.8 g manganese(II) chloride, 0.14 g zinc sulfate per liter, 0.1 mL of 1 mg thiamine hydrochloride per mL, 0.1 mL of 0.1 M iron(II) sulfate and appropriate antibiotics.

Quantification of propargylglycine. Samples (1 mL) were removed from cell culture and cleared of biomass by centrifugation at 14,000 × g for 5 min using an Eppendorf 5417R centrifuge. The cleared media sample (200 µL) was mixed 1:1 with a 1% formic acid methanol solution (100 µL). Samples were analyzed by LC-QQQ using an Agilent 1290 UPLC and Agilent 6460 Triple Quadrupole mass spectrometer using multiple reaction monitoring (MRM) in positive ionization

mode equipped with a Poroshell 120 Hilic column (2.7 μm, 2.1 × 100mm, room temperature, Agilent). Chromatography conditions were as follows. Solvent A: 10% acetonitrile, 10 mM ammonium formate, 0.1% formic acid; solvent B: 90% acetonitrile, 10 mM ammonium formate, 0.1% formic acid; flow rate 0.3 mL min^{-1}; timetable (%A): 10% at 0 min, 10% at 2 min, 25% at 6 min, 70% at 6.1 min, 70% at 12 min, 10% at 12.5 min. Transition used to monitor the analyte was as follows (transition, collision energy, fragmentation voltage): propargylglycine (114.1 → 68, 0, 60). Concentrations were determined by extrapolation from standard curves of commercial standards.

Dansyl chloride derivatization and quantification of allylglycine. Samples (1 mL) were removed from cell culture and cleared of biomass by centrifugation at 14,000 × *g* for 5 min using an Eppendorf 5417R centrifuge. The cleared media sample (10 μL) was diluted 1:10 with 200 μM 3-phenylserine (90 μL). The diluted samples (100 μL) was mixed 1:1 with dansyl chloride (DNC) reagent (100 μL) and incubated for 42 °C for 1 h. DNC reagent was prepared by mixing 10 mM carbonate buffer (pH 9) 1:1 with 1 mM DNC in acetone and stored at 4 °C. The derivatized sample was centrifuged for 10 min at 14,000 × *g* using an Eppendorf 5417R centrifuge. Samples were analyzed by LC-QQQ using an Agilent 1290 UPLC and Agilent 6460 Triple Quadrupole mass spectrometer using multiple reaction monitoring (MRM) in positive ionization mode equipped with a Poroshell 120 EC-C18 column (2.7 μm, 2.1 × 100mm, room temperature, Agilent). Chromatography conditions were as follows. Solvent A: 0.1% formic acid; solvent B: 100% acetonitrile; flow rate 0.6 mL min^{-1}; timetable (%A): 100% at 0 min, 100% at 1 min, 0% at 6 min, 0% at 6.5 min, 100% at 7.5 min, 100% at 8.5 min. Transitions used to monitor analytes were as follows (transition, collision energy, fragmentation voltage): DN-phenylserine (415.1 → 119, 12, 135), DN-allylglycine (349.1 → 170, 12, 135). Injection volumes were normalized using DN-phenylserine as an internal standard. Concentrations were determined by extrapolation from standard curves of commercial standards.

Expression and purification of His-tagged proteins. Expression plasmids (10 ng) were transformed into E. coli BL21(DE3)-$T1^R$ (KCM method) and a single colony was used to inoculate 25 mL of TB media with 50 μg/mL of the appropriate antibiotic. The seed culture was grown overnight with shaking at 37 ºC (200 RPM). Then, 1 L of TB in an Ultra Yield baffled flask (Thomson Instrument Company; Oceanside, CA) was inoculated with the seed culture and growth was continued to OD600 of 0.8-1.2. The culture was chilled in an ice bath for 15 min and protein expression was induced with IPTG (1 mM). Expression proceeded at 16 ºC overnight, after which the cells were harvested (7,000 × g, 5 min, 4 ºC) and either stored at -80 ºC or subjected to protein purification immediately.

Harvested cells were resuspended in 5 mL/(g wet cell wt.) of lysis buffer (20 mM Tris pH 7.5, 500 mM NaCl, 10 mM imidazole, 10% v/v glycerol). To the cells was added PMSF (1 mM) and lysozyme (1 mg/mL). After incubation at r.t. for 30 min, the sample was sonicated (QSonica Q700) with the following program: 10 s on, 20 s off, 1 min total process time, amplitude 50. Cell debris was removed by centrifugation (15,000 × g, 20 min, 4 ºC). Ni-NTA agarose resin (50% suspension in 20% EtOH, 0.2 mL/(g wet cell wt.)) was added to the samples, which were shaken gently on an orbital shaker at r.t. for 30 min and then poured into a glass column. The resin was washed with >20 column volumes (CV) of wash buffer (same as lysis buffer with 20 mM imidazole) until the effluent tested negative by the Bradford dye binding assay. Protein was eluted with >5 CV of elution buffer (same as lysis buffer with 250 mM imidazole) until the effluent tested negative

again. The elution fraction was supplemented with βME (0.1% v/v). Fractions and purified proteins were analyzed by SDS-PAGE

The proteins were concentrated in Amicon Ultra-15 centrifugal filter units with 30 kDa molecular weight cutoff (EMD Millipore; Burlington, MA). Buffer exchange was conducted with PD-10 columns (GE Healthcare; Chicago, IL) into HEPES storage buffer (20 mM HEPES-NaOH pH 7.5, 150 mM NaCl, 10% v/v glycerol, 1 mM TCEP), phosphate storage buffer (50 mM NaH2PO4, 100 nM NaCl, 10 mM imidazole, 10% glycerol, pH 8), or MOPS buffer (100 mM MOPS, 100 mM potassium chloride, pH 7.9, 20% glycerol) as indicated in *Appendix 3.2*. Protein concentrations were determined by the absorbance at 280 nm, using extinction coefficients listed in *Appendix 3.2* predicted with ExPASy ProtParam (Swiss Institute of Bioinformatics).

BcLeuDH and AlaDH kinetic assays. Assays were performed in a 96 well plate at room temperature with a total volume of 100 µL using a spectrophotometer (Molecular Devices M2). For preliminary screening, the reaction mixture contained 1 mM NAD^+, 5 mM substrate (leucine or propargylglycine), 1 µM enzyme (BcLeuDH or BsAlaDH), and 20 mM HEPES pH 7.5. For steady state kinetics of BcLeuDH with leucine, the reaction mixture contained 1 mM NAD^+, 0.05, 0.125, 0.25, 0.5, 1.25, 2.5, 5, 10, 20 mM leucine, 100 nM BcLeuDH, and 20 mM HEPES pH 7.5. For steady state kinetics of BcLeuDH with propargylglycine, the reaction mixture contained 1 mM NAD^+, 0.25, 0.5, 1.25, 2.5, 5, 10, 20 mM propargylglycine, 5 µM BcLeuDH, and 20 mM HEPES pH 7.5. Activity was monitored by the reduction of NAD^+ at 340 nm.

L-HicDH and LeuDH coupled assay and MS analysis. Reaction mixtures contained 1 mM NAD^+, 5 µM BcLeuDH, 5 µM L-HicDH, 5 mM substrate (propargylglycine or allylglycine), and 20 mM HEPES pH 7.5. Reactions were incubated at room temperature for 24 h. Samples (10 µL) were diluted with acetonitrile (90 µL) and then centrifuged at 14,000 × g for 10 min using an Eppendorf 5417R centrifuge. Samples were analyzed by LC-QQQ using an Agilent 1290 UPLC and Agilent 6460 Triple Quadrupole mass spectrometer using multiple reaction monitoring (MRM) in positive ionization mode equipped with a Poroshell 120 EC-C18 column (2.7 µm, 2.1 × 100mm, room temperature, Agilent). Chromatography conditions were as follows. Solvent A: 0.1% formic acid; solvent B: 100% acetonitrile; flow rate 0.6 mL min^{-1}; timetable (%A): 100% at 0 min, 100% at 1 min, 0% at 6 min, 0% at 7 min, 100% at 8.5 min, 100% at 9.5 min. Transitions used to monitor analytes were as follows (transition, collision energy, fragmentation voltage): 2-hydroxy-4-pentenoate (115.1 → 97, 8, 50), 2-hydroxy-4-pentynoate (113.1 → 95, 8, 50).

(*RS*)-2-Hydroxy-4-pentynoic acid Synthesis was carried out as previously described [31]. Briefly, DL-Propargylglycine (200 mg, 1.77 mmol) was dissolved in 0.5 M H_2SO_4 (7 mL, 3.5 mmol). A solution of sodium nitrite (700 mg, 10.1 mmol) in water (5 mL) was added dropwise over 15 min with stirring at 0 °C, then the reaction was stirred overnight warming to room temperature. The product was extracted with diethyl ether (5 x 10 mL), then the combined organic layers were washed with brine (10 mL), dried over Na_2SO_4, and concentrated to provide the product as a brown oil (112 mg, 55%). ^1H NMR (400 MHz, D_2O) δ 4.35 (t, J = 5.2 Hz, 1H), 2.63 (m, 2H), 2.31 (m, 1H).

(*RS*)-2-Hydroxy-4-pentenoic acid Synthesis was carried out as previously described [31]. Briefly, DL-Allylglycine (200 mg, 1.74 mmol) was dissolved in 0.5 M H_2SO_4 (7 mL, 3.5 mmol). A solution of sodium nitrite (700 mg, 10.1 mmol) in water (5 mL) was added dropwise over 15 min with stirring at 0 °C, then the reaction was stirred overnight warming to room temperature. The product was extracted with diethyl ether (5 x 10 mL), then the combined organic layers were washed with brine (10 mL), dried over Na_2SO_4, and concentrated to provide the product as a yellow

oil (112 mg, 55%). ^{1}H NMR (400 MHz, D_2O) δ 5.73 (m, 1H), 5.13-5.03 (m, 2H), 4.28 (dd, J = 6.9, 4.8 Hz, 1H), 2.52-2.33 (m, 2H).

4HBCS activity assay. Assays were performed in a 96 well plate at room temperature with a total volume of 100 µL using a spectrophotometer (Molecular Devices M2). The reaction mixture contained 1 mM substrate (2-hydroxy-4-pentenoate, 2-hydroxy-4-pentynoate, or 4-hydroxybutyrate), 150 µM coenzyme A (CoA), 2.5 mM ATP, 10 µM 4HBCS, 5 mM magnesium chloride, and 100 mM MOPS pH 7.9. Samples (80 µL) were quenched by mixing 1:1 with 400 µM 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) (80 µL). Free thiolate anions were quantified by reaction with DTNB to generate 2-nitro-5-thiobenzoate anion (TNB) which absorbs at 412 nm with an extinction coefficient of 14,150 M^{-1} cm^{-1} [32].

4HBCS activity (MS method). Reaction mixture contained 200 µM coenzyme A (CoA), 2.5 mM ATP, 10.2 µM 4HBCS, 5 mM magnesium chloride, 100 mM MOPS pH 7.9, and either 3 mM 4-hydroxybutyrate (4HB), 5.1 mM 2-hydroxy-4-pentynoate (2H4Pe), or 5.1 mM 2-hydroxy-4-pentenoic acid (2H4Py). The reaction with 4HB was incubated at room temperature for 2 h before quenching. Reactions with 2H4Pe and 2H4Py were incubated at room temperature for 24 h before quenching. Samples (10 µL) were quenched with 11% tricholoracetic acid (90 µL) to precipitate protein and then centrifuged at 14,000 × g for 10 min using an Eppendorf 5417R centrifuge. Samples were analyzed by LC-QQQ using an Agilent 1290 UPLC and Agilent 6460 Triple Quadrupole mass spectrometer using multiple reaction monitoring (MRM) in positive ionization mode equipped with a Poroshell 120 EC-C18 column (2.7 µm, 2.1 × 100mm, room temperature, Agilent). Chromatography conditions were as follows. Solvent A: 10 mM ammonium formate; solvent B: 100% acetonitrile; flow rate 0.6 mL min^{-1}; timetable (%A): 100% at 0 min, 100% at 1 min, 0% at 6 min, 0% at 6.5 min, 100% at 7.5 min, 100% at 8.5 min. Transitions used to monitor analytes were as follows (transition, collision energy, fragmentation voltage): 4-hydroxy-butyryl-CoA (854.2 → 347.1, 28, 150), 2-hydroxy-4-pentenoyl-CoA (866.2 → 359.1, 28, 150), 2-hydroxy-4-pentynoyl-CoA (864.1 → 357, 28, 150).

PhaC assay. CoA species were made using 4HBCS. The reaction mixture contained 900 µM CoA, 2.5 mM ATP, either 1 mM 2-hydroxy-4-pentenoate or 1 mM 3-hydroxybutyrate, 10 µM 4HBCS, 5 mM magnesium chloride, 50 mM sodium phosphate pH 7.0 was incubated at 75 °C for 50 min. Samples without 4HBCS or PhaC contained the respective storage buffer in place of the enzyme. Samples without the 2-oxo substrate contained water in place of the substrate. The reactions were then filtered with a 30 kDa Amicon spin concentrator (Millipore) to remove 4HBCS from the reaction mixture. Filtered samples (360 µL) were mixed with (40 µL) 35.8 µM PhaC then incubated at room temperature. Aliquots (50 µL) collected at 0, 1, 3, and 18 h were quenched by mixing 1:1 with 400 µM DTNB (50 µL). Free thiolate anions were quantified by reaction with DTNB to generate 2-nitro-5-thiobenzoate anion (TNB) which absorbs at 412 nm with an extinction coefficient of 14,150 M^{-1} cm^{-1} [32].

4.3. Results and discussion

We first set out to optimize production of propargylglycine (Pra) and allylglycine (Alg) within our metabolic host, *E. coli*. For Pra production, we overexpressed three enzymes, BesD, BesC, and BesB in. It was previously shown that BesB does not express well in *E. coli* and so we suspected that this enzyme could be the rate limiting step in production of Pra [33]. In an attempt to increase BesB solubility, we turned to the web-based Protein Repair One-Stop Shop (PROSS) tool to

recommend mutations to introduce into the BesB sequence to improve its solubility [34]. From the list of recommended mutants, we chose the entry with the least perturbations from the native sequence containing 12 mutations (PROSS 1) and the third entry in the list (PROSS 3) containing 17 mutations. We then cloned these two variants into our expression vector and tested them for Pra production (*Figure 4.2A*). Unfortunately, we found that both variants of BesB produced very small amounts of Pra, suggesting that the introduced mutations negatively affected the activity. Using the wild type BesB enzyme, we observed about 40 μM of propargylglycine present in the media at the end of the three-day production.

We tested two overexpression constructs for Alg production, one with BesC alone and the other as MBP fused to the N-terminus of BesC (*Figure 4.2B).* For the construct expressing BesC alone, lower culture temperatures increased Alg production. However, Alg production with the MBP-BesC fusion was unaffected by culture temperature. Overexpression of BesC as a MBP fusion improved titers, suggesting that the MBP fusion improved BesC solubility, leading to more activity. We chose to proceed with further production optimization using the MBP-BesC fusion construct. We decided to conduct future productions at 30 °C as it supports faster microbial growth. We next tested Alg production in several different strains of *E. coli* with two different concentrations of IPTG induction (*Figure 4.2C*). Induction level did not have much impact on Alg production. If anything, production was more variable when 1 mM IPTG was used. From this experiment, we observed that the strains K-12 and BL21 Tuner performed best, producing about 100 and 80 μM of allylglycine respectively. We also explored production in M9 minimal media supplemented with 1 mM lysine (*Figure 4.2D*). Our analytical methods for Alg quantification were more sensitive in the cleaner matrix of M9 compared to LB. However, Alg production was significantly reduced in M9 minimal media compared to LB media. Again, we observed that BL21 Tuner and K-12 strains of *E. coli* outperformed DH1. Additionally, we measured production after two and three days of production in DH1 and observed lower concentrations of Alg in the media after 3 days, perhaps due to re-uptake of Alg in stationary phase growth.

After verifying that both Pra and Alg can be produced *in vivo*, we turned to constructing our biosynthetic pathway to convert the noncanonical amino acids Alg and Pra to their respective 2-hydroxy acids and incorporate them into a PHA. The first transformation in our pathway is the conversion of an α-amino acid into a 2-oxo acid. Based on a literature search, we selected two candidate enzymes to try. Alanine dehydrogenase from *Bacillus subtilis* (BsAlaDH) and leucine dehydrogenase from *Bacillus cereus* (BcLeuDH). We reasoned that Pra and Alg are both similar to alanine, except with two additional carbons in the side chain, so BsAlaDH may have some residual activity towards these substrates [35]. BcLeuDH has previously shown to be active with a variety of nonpolar α-keto acid substrates that are chemically similar to our target substrates [36]. These two enzymes were cloned with N-terminal His_{10}-tags, heterologously expressed in *E. coli* BL21, and purified by Ni-NTA chromatography (*Figure 4.3A*). We then validated the activity of both enzymes with Ala, Leu, or Pra and we found that the enzymes were most active with their native substrate (*Figure 4.3B*). However, we observed residual activity of BsAlaDH with the non-native substrate, Pra. Promiscuous activity with Pra was even larger for BcLeuDH. We then proceeded to characterize steady-state kinetics of BcLeuDH with Leu and Pra (*Figure 4.3C*). We found a 250-fold decrease in catalytic efficiency with Pra compared to Leu resulting from about a 10-fold increase in K_M and a roughly 10-fold decrease in k_{cat}. Despite lower activity, we confirmed that BcLeuDH can achieve the desired oxidative deamination of Pra to form 2-oxopentynoate (2O4Py).

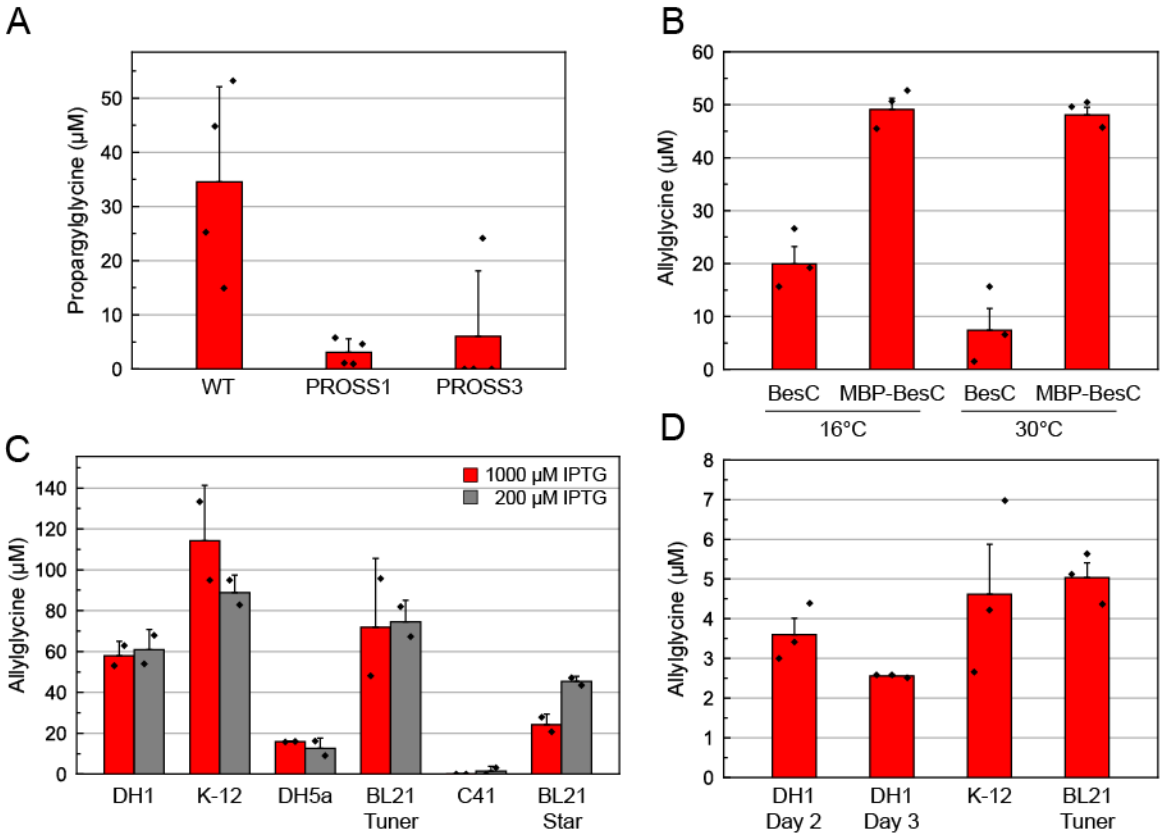
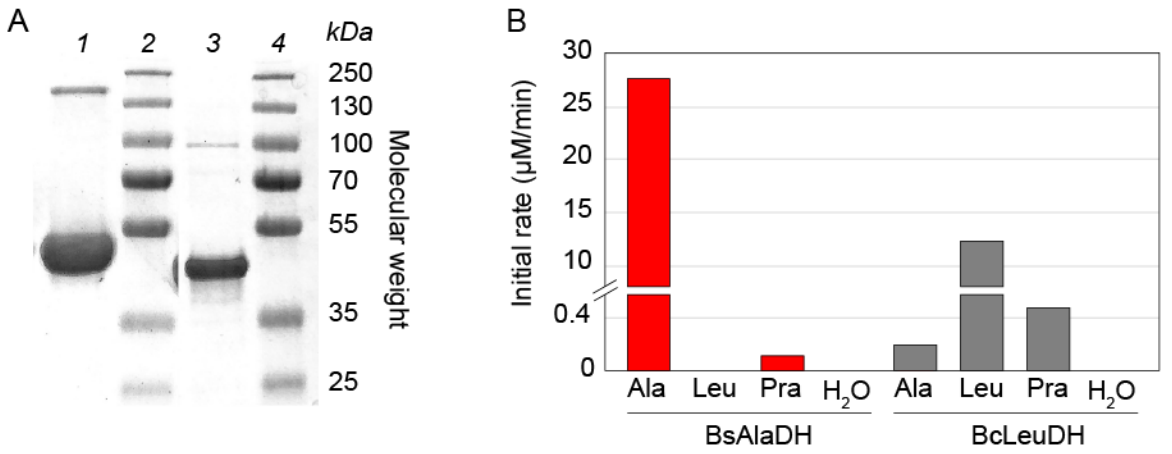
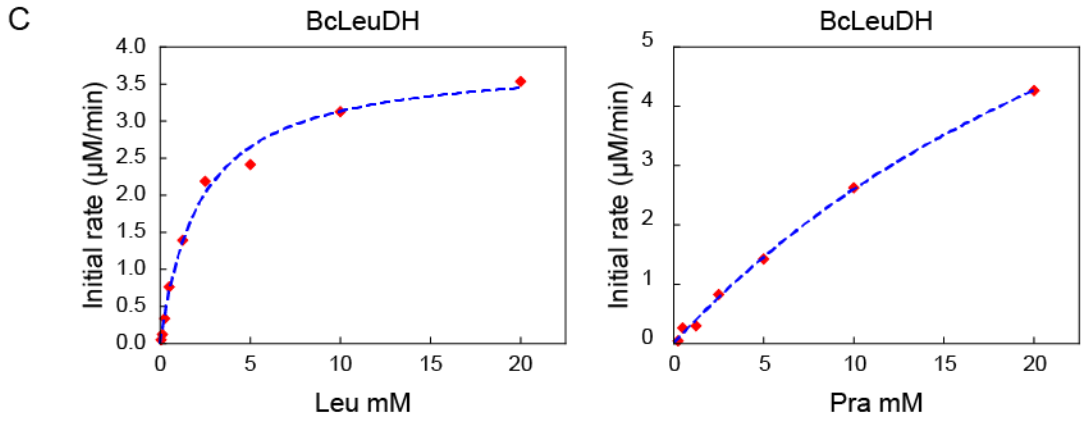


Figure 4.2*. In vivo production of Pra and Alg.* ***(****A) Overexpression of the first three enzymes in the β-ethynylserine pathway, (BesD, BesC, and BesB) leads to detectable production of propargyl glycine. Two BesB mutantswith predicted higher solubility (PROSS1 and PROSS3) were generated and expressed. E. coli BL21 with pGro7 and either pPra2, pPra2-BesB.1 or pPra2-BesB.3 was grown for 3 d at 16 °C in TB adjusted to pH 6.5. Liquid culture media was collected and Pra was quantified by LC-QQQ. (B) Allylglycine production from two different expression vectors for BesC, pET16-BesC and pSV272.1-BesC from which BesC was expressed with an N-terminal MBP fusion. BL21 carrying either pET16-BesC or pSV272.1-BesC was grown for 2 d at 16°C or 30°C in LB. (C) Several strains of E. coli were tested for Alg production. DH1, K12, and BL21 tuner strains produced the most Alg. Only marginal differences were observed between the two concentrations of inducer tested. Cultures containing pSV272.1-BesC were grown in LB at 30 °C for 2 d. (D) Alg production was also carried out in M9 media with 1 mM lysine. Cultures were grown for 3 d at 30 °C in LB unless indicated otherwise.*





Enzyme	Substrate	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M ($mM^{-1}min^{-1}$)
BcLeuDH	Leu	38 ± 1	2.2 ± 0.3	17 ± 2
BcLeuDH	Pra	2.2 ± 0.3	36 ± 6	(7 ± 1) x 10^{-2}

Figure 4.3. *Identifying an amino acid dehydrogenase. (A) Bacillus subtilis AlaDH and Bacillus cereus LeuDH were cloned from the genome of each organism with an N-terminal His_{10} tag and subsequently purified by Ni-NTA affinity chromatography. SDS-PAGE gel of the following samples: 1- BsAlaDH, 2- molecular weight marker, 3- BcLeuDH, 4- molecular weight marker. (B) Initial analysis of substrate promiscuity was assayed by monitoring the increase in NADH absorbance at 340 nm with different amino acids. (C) Michaelis-Menten kinetics of BcLeuDH reveal that BcLeuDH has about a 20-fold lower k_{cat} and 20-fold higher K_M with Pra compared to its native substrate, Leu. These differences lead to a roughly 250-fold reduction in k_{cat}/K_M.*

We next looked for an enzyme to perform the second step in our pathway, reduction of a 2-oxo acid to a 2-hydroxy acid. Searching through the literature, we came across a report on the enzyme L-HicDH from *Lactobacillus confusus* showing broad substrate specificity [37]. Additionally, the native substrate, 2-oxo-isocaproate, is derived from leucine and we have previously shown that BcLeuDH, for which leucine is the native substrate, has activity towards Pra, suggesting that Pra can fit into catalytic sites evolved to bind leucine and 2-oxo-isocaproate. L-HicDH was cloned with a C-terminal His_6-tag, heterologously expressed in *E. coli* BL21, and purified by Ni-NTA chromatography (*Figure 4.4A*). In addition to testing activity with 2O4Py, we wanted to test activity with the Alg-derived 2-oxopentenoate (2O4Pe). We were unable to obtain purified 2O4Pe or 2O4Py, so we instead chose to generate these substrates *in situ* using BcLeuDH to oxidatively deaminate Alg and Pra respectively. We analyzed the reactions by LC-QQQ for the presence of 2-hydroxypentenoate (2H4Pe) and 2-hydroxypentynoate (2H4Py) resulting from oxidative deamination and subsequent reduction of Alg and Pra respectively. We were unable to observe any enzyme-dependent formation of 2H4Py in the reaction with Pra. Given that BcLeuDH was shown to be active on this substrate, we can conclude that 2O4Py is not accepted by L-HicDH. However, in the presence of L-HicDH, we confirmed that 2H4Pe was generated, confirming that

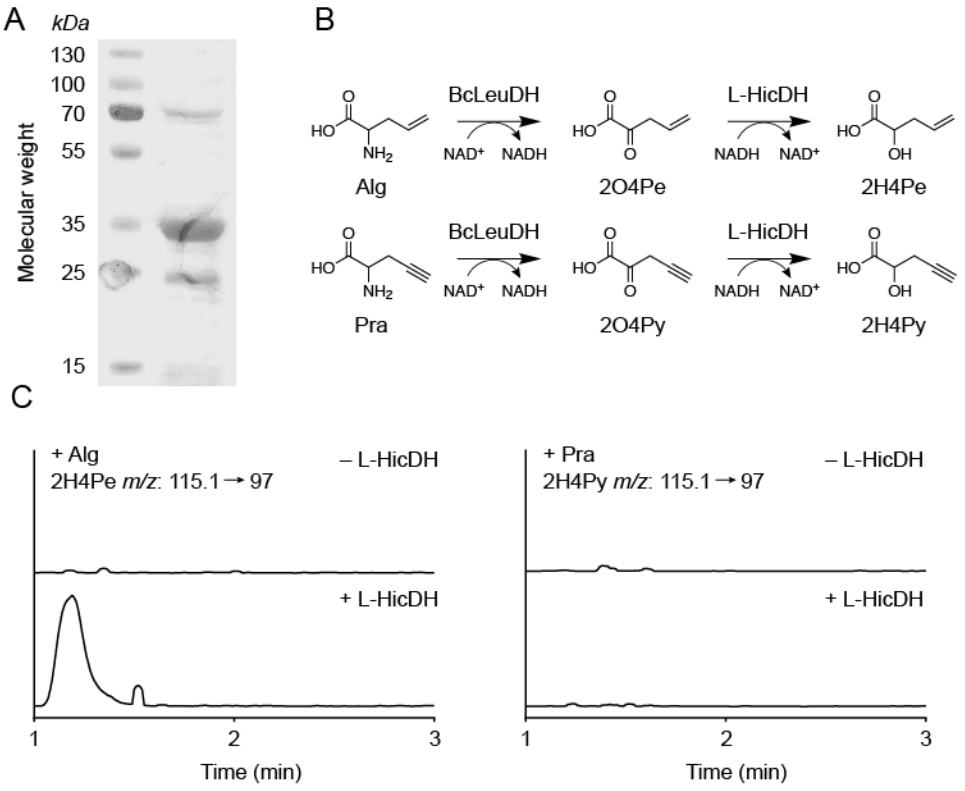


Figure 4.4. *Identifying an alcohol dehydrogenase. (A) The L-2-hydroxyisocaproate dehydrogenase (L-HicDH) was cloned with a C-terminal His_6 tag from Lactobacillus confuses and expressed in E. coli BL21 for purification by Ni-NTA chromatography. SDS-PAGE gel with samples from left to right: molecular weight marker, L-HicDH. (B) L-HicDH was assayed by coupling to BcLeuDH to generate the 2-keto substrates from Alg and propargylglycine Pra in situ. (C) Reactions were analyzed by LC-QQQ.*

L-HicDH can accept 2O4Py as a substrate. This result also verifies that BcLeuDH is active with Alg, which was not initially tested when characterizing BcLeuDH.

For PHA synthase (PhaC) to polymerize hydroxy acids, the carboxylic acids must first be activated as CoA thioesters. Therefore, we needed to find an enzyme capable of forming CoA thioesters from our 2-hydroxy acids, 2H4Pe and 2H4Py. In the metabolic pathway designed to incorporate leucine derived 2-hydroxyisocaproate into PHAs, Mizuno et al. utilized HadA 2-hydroxycaproate-CoA transferase (HadA) from *Peptoclostridium difficile* [25]. CoA. transferases catalyze formation of a thioester by hydrolysis of another CoA thioester. HadA has previously been characterized to use 2-hydroxyisocaproyl-CoA or isocaproyl-CoA as the CoA donors and does not accept other thioesters such as acetyl-CoA or butyryl-coA [38]. To avoid challenges associated with CoA transfer from rare thioesters like isocaproyl-CoA, we instead sought to find a CoA synthetase that can create CoA thioesters from free CoA in an ATP-dependent manner.

Looking for CoA synthetases, we found many enzymes that were characterized to accept fatty acids, which we felt was too dissimilar to our substrate to have residual activity [39,40]. However, we did come across a 4-hydroxybutyrate-CoA synthetase (4HBCS) discovered in the 3-hydroxypropionate/4-hydroxybutyrate cycle of a thermoacidophilic archaea, *Metallosphaera*

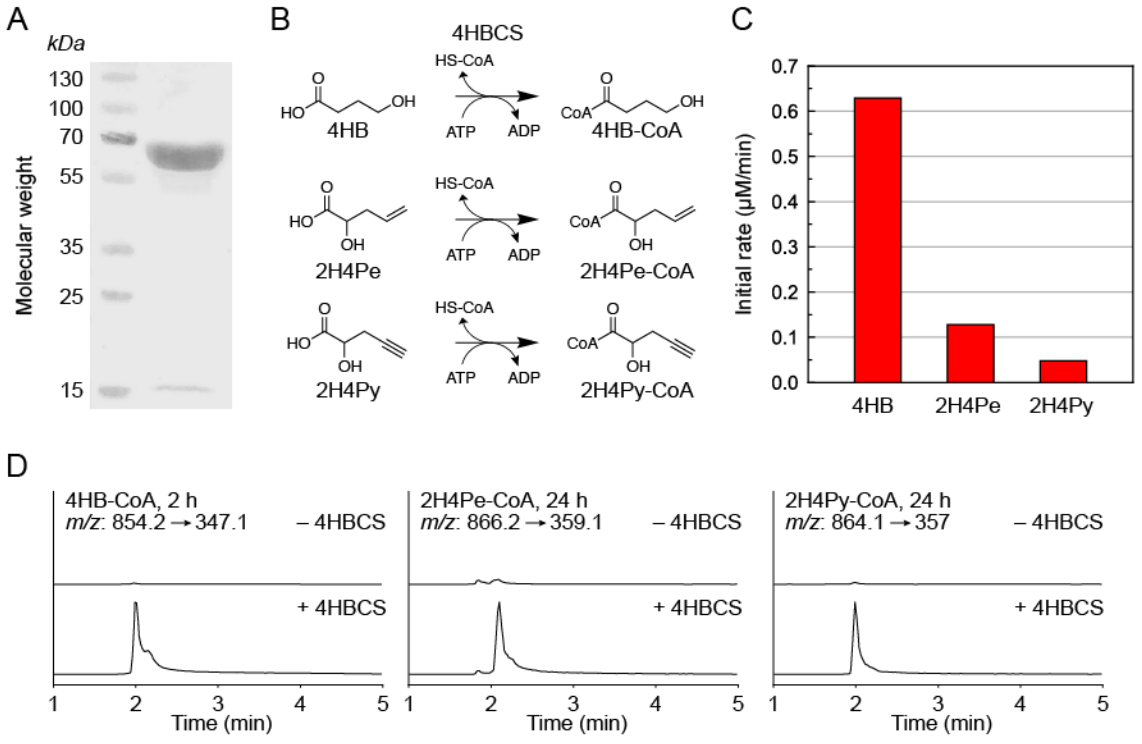


***Figure 4.5**. Identifying a coenzyme A synthetase. (A) The 4-hydroxybutyrate-CoA synthetase (4HBCS) from Metallosphaera sedula was cloned with an N-terminal His_{10}-tag and expressed in E. coli BL21 for purification by Ni-NTA chromatography. SDS-PAGE gel with samples from left to right: molecular weight marker, 4HBCS. (B) His_{10}-4HBCS was assayed with the native substrate 4-hydroxybutyrate, and 2H4e and 2H4Py. (C) CoA synthetase activity was characterized using a discontinuous DTNB assay to monitor CoA release [32]. (D) Formation of 4HB-CoA, 2H4Pe-CoA, and 2H4Py-CoA was verified with LC-QQQ.*

sedula. In addition to its native substrate, 4HBCS was also shown to have activity with another small, hydroxylated substrate, 3-hydroxypropionate [41]. We reasoned that 2H4Pe and 2H4Py are chemically similar to these substrates, so we considered this enzyme to be a good candidate. 4HBCS was cloned with a N-terminal His_{10}-tag, heterologously expressed in *E. coli* BL21, and purified by Ni-NTA chromatography (*Figure 4.5A*). Due to a lack of commercial chemical standards, we synthesized 2H4Py and 2H4Pe from Pra and Alg respectively using a previously reported procedure [31]. To monitor CoA consumption, we used 5,5′-disthiobis(2-nitrobenzoic acid) (DTNB) which is known to react with free thiols in solution to generate a product that absorbs strongly at 412 nm [32]. With this assay in hand, we obtained preliminary kinetic data for 4HBCS with 4-hydroxybutyrate, 2H4Pe, and 2H4Py (*Figure 4.5B*). We were satisfied to see that 4HBCS showed activity with 2H4Pe and 2H4Py at room temperature, with slightly more activity towards 2H4Pe. We then verified the presence of the desired 2H4Pe-CoA and 2H4Py-CoA species via LC-QQQ (*Figure 4.5C*). From these data, we conclude that we have identified a CoA synthetase capable of creating CoA thioesters from our substrates of interest.

The last remaining step in our pathway is polymerization of the CoA-thioesters. As mentioned earlier, a mutated form of PhaC from *Pseudomonas* sp. 61-3 (PhaC STQK) has been shown to have activity with lactate, another 2-hydroxy acid [23,24]. This enzyme was also shown to accommodate 2-hydroxyisocaproate incorporation *in vivo* [25]. Given that the previously characterized enzymes in our pathway also accepted leucine-derived compounds, the fact that PhaC STQK was shown to have activity with 2-hydroxyisocaproate-CoA gave us optimism that PhaC STQK could achieve the desired polymerization.

To test this hypothesis, the mutated PhaC (STQK) from *Pseudomonas* sp. 61-3 was cloned with a N-terminal His_{10}-tag, heterologously expressed in *E. coli* BL21, and purified by Ni-NTA chromatography (*Figure 4.6A*). Although the protein sample was heterogeneous, we proceeded to preliminary activity assays. Since the enzymes tested so far seemed to have higher activity with the allylglycine-derived substrates, we prioritized establishing activity with 2H4Pe-CoA. In addition to 2H4Pe-CoA, we also included 3-hydroxybutyrate-CoA as a positive control since the engineered PhaC STQK is still active with the native substrate of wild-type PhaC [24]. We used 4HBCS to generate the respective CoA thioesters from 2H4Pe and 2H4Py, then removed 4HBCS from the reaction via filtration (*Figure 4.6B*). PhaC STQK was then added to this unpurified reaction mixture and polymerization was monitored through CoA release using the DTNB assay (*Figure 4.6C*). In negative controls without the free acid or 4HBCS, no depletion of the free CoA was observed. In samples without PhaC, we observe a depletion of free CoA that suggests that the 4HBCS reaction goes to completion to form the CoA thioester. Upon addition of PhaC, we observe a time-dependent release of CoA from both 3HB-CoA and 2H4Pe-CoA, which is a promising indication of polymerization. Preliminary rate data suggests that PhaC STQK was even more active with 2H4Pe-CoA than 3HB-CoA, the native substrate of PhaC. Since CoA release can also be observed as a result of hydrolysis without polymerization, future experiments will need to be performed to verify formation of a PHA polymer.

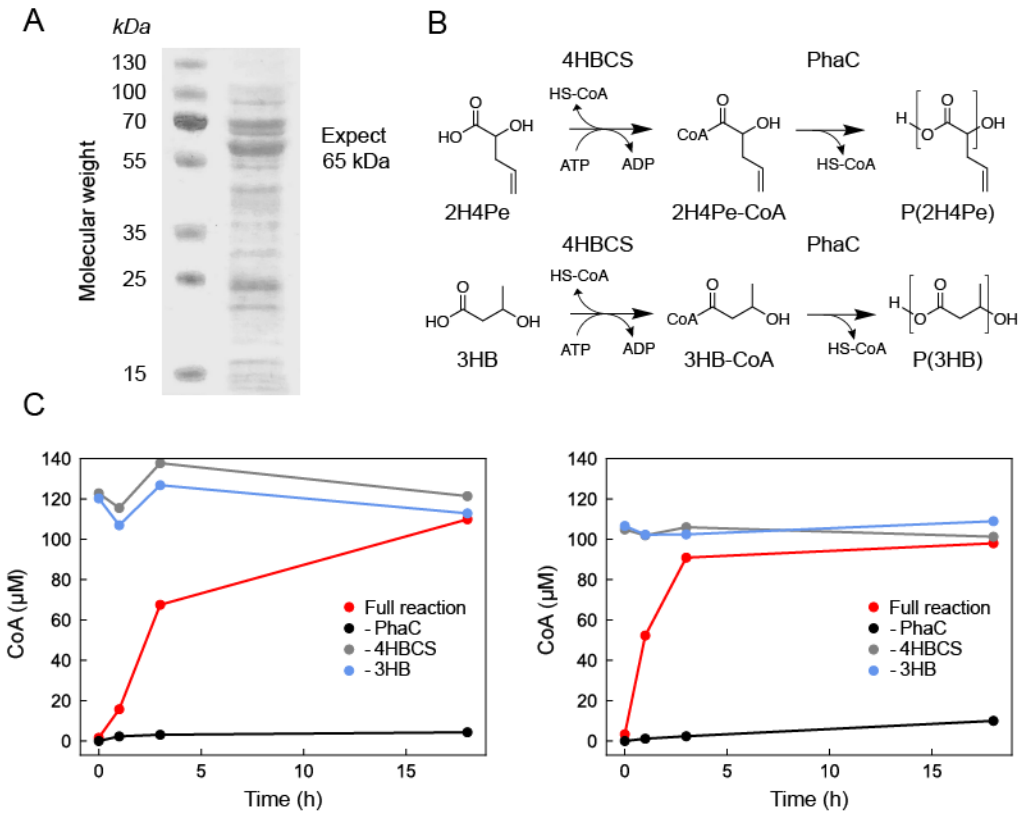


Figure 4.6. *Identifying a polyhydroxyalkanoate synthase. (A) A mutant of Pseudomonas sp. 61-3 PhaC (PhaC STQK) was cloned with an N-terminal His_{10}-tag and expressed in E. coli BL21 for purification by Ni-NTA chromatography. SDS-PAGE gel with samples from left to right: molecular weight marker, PhaC STQK. (B) His_{10}-PhaC STQK was assayed with the native substrate, 3-hydroxybutyrate-CoA (3HB-CoA), and 2-hydroxy-4-pentenoate-CoA (2H4Pe-CoA). 3HB-CoA and 2H4PE-CoA were generated in situ with 4HBCS. (C) Polyhdroxyalkanoate synthase activity was monitored using a discontinuous DTNB assay to monitor CoA release [32].*

4.4. Conclusion

In this chapter, we have identified enzymes that could be used to initiate construction of a metabolic pathway for incorporation of 2-hydroxy acids derived from the non-canonical amino acids, propargylglycine (Pra) and allylglycine (Alg). First, we showed that both of these amino acids can be produced at measurable titers by overexpressing enzymes from the β-ethynylserine biosynthetic pathway (*Figure 4.2*). Next, we devised a metabolic strategy to convert these amino acids to their respective CoA thioesters. Through literature searching, we identified candidate enzymes capable of achieving each of the reactions in our proposed pathway. These candidates were then heterologously expressed and purified to measure their activity towards our substrates of interest *in vitro*. We showed that BcLeuDH is capable of oxidative deamination of both Pra and Alg generating 2-oxo-4-pentynoate (2O4Py) and 2-oxopentenoate (2O4Pe) (*Figure 4.3*). L-HicDH is capable of reducing 2O4Pe to 2-hydroxypentenoate (2H4Pe) but is unable to achieve reduction of 2O4Py to 2-hydroxy-4-pentynoate (2H4Py) (*Figure 4.4*). We next verified that both 2H4Pe-CoA and 2H4Py-CoA formation were catalyzed by 4HBCS (*Figure 4.5*). Lastly, we showed preliminary evidence that PhaC STQK is capable of polymerizing 2H4Pe-CoA (*Figure 4.6*).

Taken together, we have successfully identified the necessary enzymes to create a functioning metabolic pathway for the generation of 2H4Pe-CoA from Alg and for incorporation of this unique monomer into PHAs. In addition to Alg, we have shown that two of the four enzymes in this pathway are compatible with Pra-derived substrates. Other alcohol dehydrogenases besides L-HicDH can be investigated for activity towards Pra-derived 2O4Py, or mutations can be introduced into the active site of L-HicDH to accommodate this unnatural substrate. Additionally, PhaC STQK activity towards 2H4Py-CoA needs to be assayed.

We next plan to construct a plasmid with the entire pathway for expression in *E. coli* to test the *in vivo* performance of our pathway. PHA pathway performance can be tested independently from Alg production by feeding commercially available Alg to the culture. After validating the PHA pathway, the Alg production module can then be added to the system. A large challenge of *in vivo* chemistry is the specificity requirement. The desired chemical reaction must be performed in the presence of the complex cellular metabolome. Given that many of the enzymes in our pathway have activity towards leucine catabolites, we expect a significant amount of 2-hydroxyisocaproate (2HIC) incorporation in our microbially induced PHA. A co-polymer of 2HIC and 2H4Pe may have interesting mechanical properties in addition to containing terminal olefins to serve as handles for downstream chemical modification of this unique PHA polymer.

4.4. References

1. Geyer, R., Jambeck, J. R. & Law, K. L. Production, use, and fate of all plastics ever made. *Sci. Adv.* **3,** e1700782 (2017).
2. Law, K. L., Morét-Ferguson, S. E., Goodwin, D. S., Zettler, E. R., DeForce, E., Kukulka, T. & Proskurowski, G. Distribution of surface plastic debris in the eastern pacific ocean from an 11-year data set. *Environ. Sci. Technol.* **48,** 4732–4738 (2014).
3. Thushari, G. G. N. & Senevirathna, J. D. M. Plastic pollution in the marine environment. *Heliyon* **6,** e04709 (2020).
4. Rosenboom, J.-G., Langer, R. & Traverso, G. Bioplastics for a circular economy. *Nat. Rev. Mater.* **7,** 117–137 (2022).
5. Choi, S. Y., Cho, I. J., Lee, Y., Park, S. & Lee, S. Y. in *Methods Enzymol.* (eds. Bruns, N. & Loos, K.) **627,** 125–162 (Academic Press, 2019).
6. Drumright, R. E., Gruber, P. R. & Henton, D. E. Polylactic acid technology. *Adv. Mater.* **12,** 1841–1846 (2000).
7. Will the biodegradable plastic PHA finally deliver? *Chem. Eng. News* at <https://cen.acs.org/business/biobased-chemicals/biodegradable-plastic-PHA-finally-deliver/99/i22>
8. Yeo, J. C. C., Muiruri, J. K., Thitsartarn, W., Li, Z. & He, C. Recent advances in the development of biodegradable PHB-based toughening materials: Approaches, advantages and applications. *Mater. Sci. Eng. C* **92,** 1092–1116 (2018).
9. Zhang, L., Deng, X. & Huang, Z. Miscibility, thermal behaviour and morphological structure of poly(3-hydroxybutyrate) and ethyl cellulose binary blends. *Polymer* **38,** 5379–5387 (1997).

10. Mousavioun, P., Doherty, W. O. S. & George, G. Thermal stability and miscibility of poly(hydroxybutyrate) and soda lignin blends. *Ind. Crops Prod.* **32,** 656–661 (2010).

11. Koller, M., Maršálek, L., de Sousa Dias, M. M. & Braunegg, G. Producing microbial polyhydroxyalkanoate (PHA) biopolyesters in a sustainable manner. *New Biotechnol.* **37,** 24–38 (2017).

12. Laycock, B., Halley, P., Pratt, S., Werker, A. & Lant, P. The chemomechanical properties of microbial polyhydroxyalkanoates. *Prog. Polym. Sci.* **38,** 536–583 (2013).

13. Madison, L. L. & Huisman, G. W. Metabolic engineering of poly(3-hydroxyalkanoates): from DNA to plastic. *Microbiol. Mol. Biol. Rev.* **63,** 21–53 (1999).

14. Rai, R., Yunos, D. M., Boccaccini, A. R., Knowles, J. C., Barker, I. A., Howdle, S. M., Tredwell, G. D., Keshavarz, T. & Roy, I. Poly-3-hydroxyoctanoate P(3HO), a medium chain length polyhydroxyalkanoate homopolymer from *Pseudomonas mendocina*. *Biomacromolecules* **12,** 2126–2136 (2011).

15. Kai, D. & Loh, X. J. Polyhydroxyalkanoates: chemical modifications toward biomedical applications. *ACS Sustain. Chem. Eng.* **2,** 106–119 (2014).

16. Hazer, B. & Steinbüchel, A. Increased diversification of polyhydroxyalkanoates by modification reactions for industrial and medical applications. *Appl. Microbiol. Biotechnol.* **74,** 1–12 (2007).

17. Arkin, A. H., Hazer, B. & Borcakli, M. Chlorination of poly(3-hydroxy alkanoates) containing unsaturated side chains. *Macromolecules* **33,** 3219–3223 (2000).

18. Arkin, A. H. & Hazer, B. Chemical modification of chlorinated microbial polyesters. *Biomacromolecules* **3,** 1327–1335 (2002).

19. Northrop, B. H. & Coffey, R. N. Thiol–ene click chemistry: computational and kinetic analysis of the influence of alkene functionality. *J. Am. Chem. Soc.* **134,** 13804–13817 (2012).

20. Anderson, A. J. & Dawes, E. A. Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiol. Rev.* **54,** 450–472 (1990).

21. Steinbüchel, A. & Valentin, H. E. Diversity of bacterial polyhydroxyalkanoic acids. *FEMS Microbiol. Lett.* **128,** 219–228 (1995).

22. Dong, H., Liffland, S., Hillmyer, M. A. & Chang, M. C. Y. Engineering *in vivo* production of α-branched polyesters. *J. Am. Chem. Soc.* **141,** 16877–16883 (2019).

23. Takase, K., Taguchi, S. & Doi, Y. Enhanced synthesis of poly(3-hydroxybutyrate) in recombinant *Escherichia coli* by means of error-prone PCR mutagenesis, saturation mutagenesis, and *in vitro* recombination of the type II polyhydroxyalkanoate synthase Gene. *J. Biochem. (Tokyo)* **133,** 139–145 (2003).

24. Taguchi, S., Yamada, M., Matsumoto, K., Tajima, K., Satoh, Y., Munekata, M., Ohno, K., Kohda, K., Shimamura, T., Kambe, H. & Obata, S. A microbial factory for lactate-based polyesters using a lactate-polymerizing enzyme. *Proc. Natl. Acad. Sci.* **105,** 17323–17327 (2008).

25. Mizuno, S., Enda, Y., Saika, A., Hiroe, A. & Tsuge, T. Biosynthesis of polyhydroxyalkanoates containing 2-hydroxy-4-methylvalerate and 2-hydroxy-3-

phenylpropionate units from a related or unrelated carbon source. *J. Biosci. Bioeng.* **125,** 295–300 (2018).

26. Marchand, J. A., Neugebauer, M. E., Ing, M. C., Lin, C.-I., Pelton, J. G. & Chang, M. C. Y. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567,** 420–424 (2019).
27. Meldal, M. & Tornøe, C. W. Cu-catalyzed azide−alkyne cycloaddition. *Chem. Rev.* **108,** 2952–3015 (2008).
28. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
29. Engler, C. & Marillonnet, S. Golden Gate cloning. *Methods Mol. Biol. Clifton NJ* **1116,** 119–131 (2014).
30. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8,** 91 (2008).
31. Feifel, S. C., Schmiederer, T., Hornbogen, T., Berg, H., Süssmuth, R. D. & Zocher, R. *In vitro* synthesis of new enniatins: probing the α-D-hydroxy carboxylic acid binding pocket of the multienzyme enniatin synthetase. *ChemBioChem* **8,** 1767–1770 (2007).
32. Collier, H. B. A note on the molar absorptivity of reduced Ellman's reagent, 3-carboxylato-4-nitrothiophenolate. *Anal. Biochem.* **56,** 310–311 (1973).
33. Benmaman, M. & A, J. Biosynthesis and translational control of terminal-alkyne amino acids. (2019). at <https://escholarship.org/uc/item/34s446zk>
34. Weinstein, J. J., Goldenzweig, A., Hoch, S. & Fleishman, S. J. PROSS 2: a new server for the design of stable and highly expressed protein variants. *Bioinformatics* **37,** 123–125 (2021).
35. Gonçalves, L. P. B., Antunes, O. A. C., Pinto, G. F. & Oestreicher, E. G. Simultaneous enzymatic synthesis of (S)-3-fluoroalanine and (R)-3-fluorolactic acid. *Tetrahedron Asymmetry* **11,** 1465–1468 (2000).
36. Zhou, J., Wang, Y., Chen, J., Xu, M., Yang, T., Zheng, J., Zhang, X. & Rao, Z. Rational engineering of *Bacillus cereus* leucine dehydrogenase towards α-keto acid reduction for improving unnatural amino acid production. *Biotechnol. J.* **14,** 1800253 (2019).
37. Feil, I. K., Hendle, J. & Schomburg, D. Modified substrate specificity of L-hydroxyisocaproate dehydrogenase derived from structure-based protein engineering. *Protein Eng.* **10,** 255–262 (1997).
38. Kim, J., Darley, D., Selmer, T. & Buckel, W. Characterization of (R)-2-hydroxyisocaproate dehydrogenase and a family III coenzyme A transferase involved in reduction of L-leucine to isocaproate by *Clostridium difficile*. *Appl. Environ. Microbiol.* **72,** 6062–6069 (2006).
39. Mashek, D. G., Li, L. O. & Coleman, R. A. Long-chain acyl-CoA synthetases and fatty acid channeling. *Future Lipidol.* **2,** 465–476 (2007).
40. Kasuya, F., Yamaoka, Y., Igarashi, K. & Fukui, M. Molecular specificity of a medium chain acyl-CoA synthetase for substrates and inhibitors: conformational analysis. *Biochem. Pharmacol.* **55,** 1769–1775 (1998).

41. Hawkins, A. S., Han, Y., Bennett, R. K., Adams, M. W. W. & Kelly, R. M. Role of 4-hydroxybutyrate-CoA synthetase in the CO_2 fixation cycle in thermoacidophilic archaea. *J. Biol. Chem.* **288,** 4012–4022 (2013).

Appendix 1: *Strains and plasmids*

Appendix 1.1. Strains

A complete list of strains used throughout this work.

Organism	Name	Description	Source
E. coli	DH10B	F- endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ-	Invitrogen
E. coli	BW25113	Δ(*araD-araB*)*567* Δ(*rhaD-rhaB*)*568* *ΔlacZ4787* (::rrnB-3) *hsdR514 rph-1*	Grenier et al. [1]
E. coli	BWΔ5	BW25113 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* Δ*fhuA*	Davis et al. [2]
E. coli	BWΔ5 *pcnB* E108A	BW25113 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* Δ*fhuA pcnB*E108A	This study
E. coli	BWΔ5 *pcnB* N138H	BW25113 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* Δ*fhuA pcnB*N138H	This study
E. coli	BWΔ5 *pcnB* L392fs	BW25113 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* Δ*fhuA pcnB*L392fs	This study
E. coli	BWΔ5 *rne* R488H V489L	BW25113 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* Δ*fhuA rne*R488H-V489L	This study
E. coli	BWΔ5 Δ*pcnB*	BW25113 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* Δ*fhuA* Δ*pcnB*	This study
E. coli	DH1	F– λ– endA1 recA1 relA1 gyrA96 thi-1 glnV44 hsdR17(rK–mK–)	Hanahan [3]
E. coli	DH1Δ5	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC*	Davis et al. [2]
E. coli	DH1Δ5 *pcnB* R149H	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* *pcnB*R149H	This study
E. coli	DH1Δ5 *pcnB* R149P	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* *pcnB*R149P	This study
E. coli	DH1Δ5 *rne* R373S	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* *rne*R373S	This study
E. coli	DH1Δ5 *rpoC* K1192E	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* *rpoC*K1192E	This study
E. coli	DH1Δ5 *rpoC* M466L	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* *rpoC*M466L	Davis et al. [2]
E. coli	K-12 MG1655 (DE3)	F- lambda- *ilvG- rfb*-50 *rph*-1	Jensen [4]
E. coli	DH5α (DE3)	F- *Δ(argF-lac)169* φ80d*lacZ58*(M15) *ΔphoA8* *glnX44*(AS) *λ- deoR481 rfbC1 gyrA96*(NalR) *recA1endA1 thiE1 hsdR17*	Thermo Fisher Scientific

E. coli	BL21 (DE3) Tuner	F- *ompT hsdS*$_B$ (r_B-, m_B-) *gal dcm lacY1* (DE3)	Thermo Fisher Scientific
E. coli	C41 (DE3)	F^- *ompT hsdS*$_B$ (r_B-, m_B-) *gal dcm* (DE3)	Millipore Sigma
E. coli	BL21 (DE3) Star $T1^R$	F^- *ompT hsdS*$_B$ (r_B-, m_B-) *gal dcm rne*131 (DE3)	Heyde et al. [5]
E. coli	BL21 (DE3) $T1^R$	F^- *ompT hsdS*$_B$ (r_B-, m_B-) *gal dcm* (DE3)	Thermo Fisher Scientific
B. cereus		ATCC 14579	ATCC 14579
B. subtilis		ATCC 6051	ATCC 6051

Appendix 1.2. Plasmids

A complete list of plasmids used throughout this work. The promoter for each gene is indicated by the name of the promoter followed by a period. A period between two genes represents an operon. Terminators are represented after a period following the gene. (pVector-promoter.gene1.gene2.terminator-promoter.gene3)

Name	Description	Number	Source
pPDHc	pBBR1-P_{Lac}.*aceE.aceF.lpd*, Km^r, pBBR1	339	Bond-Watts et al. [6]
pBu1	pT533-P_{T5}.*phaA.hbd.rrnB*-P_{Trc}.*crt*.TrrnB, Cm^r, p15A	499	Davis et al. [2]
pBu2	pCWO-2P_{Tac}.*ter*.TrrnB-P_{Trc}.*aldh46.adh2*, Cb^r, ColE1	1866	Davis et al. [2]
pBu2.1	pCWO-2P_{Tac}.*ter*. TrrnB-P_{Trc}.*aldh21.adh2*, Cb^r, ColE1	2456	Davis et al. [2]
pBut	pCWO-P_{T5}.*phaA.hbd.rrnB*-PTrc.*crt*. TrrnB-2P_{Tac}.*ter*.TrrnB-P_{Trc}.*aldh46.adh2*, Cb^r, ColE1	3283	This study
pBDO1	pT533-P_{T5}.*phaA.phaB*.TrrnB, Cm^r, p15A	1319	Davis et al. [2]
pBDO2	pCWO-2P_{Tac}.*ter*.TrrnB -P_{Trc}.*aldh7.adh2*, Cb^r, ColE1	2076	Davis et al. [2]
pTargetF	pMB1, Sm^r For expression of sgRNA	2637	Jiang et al. [7]
pCRISPR	pColE1-P_{sgRNA}. Km^r For expression of sgRNA.	2072	Addgene #42875 [8]
pCRISPR-Gibson1	pColE1-P_{sgRNA}. Km^r For expression of sgRNA. Introduced cut site to facilitate Gibson assembly.	2786	Davis et al. [2]
pKD46-Cas9-RecA-Cure	pSC101^{ts}-PTrc.tracrRNA.Cas9.*recA*-P_{BAD}.*gam.bet.exo*. Cb^r	2416	Jiang et al. [8]
pCRISPR-PcnB2409	pColE1-P_{sgRNA}. Km^r For expression of sgRNA to target *pcnB* to create the R149H mutation via CRISPR approach 1I	2784	This study
pKD4	Template plasmid to amplify Km^r cassette flanked by FRT recombinase sites with homology arms. Ap^r Km^r	152	Datsenko and Wanner [9]
pCP20	pSC101^{ts}-P_{Lambda}.*flp*, Cb^r Cm^r For expression of FLP recombinase.	154	Datsenko and Wanner [9]
pBAC-lacZ	pBAC-P_{Lac}.*lacZ*	3264	Addgene #13422
pBAC-gg	pBAC-gg, ~1kb up and down stream *pcnB* genomic context flanking a BsaI cut site for golden gate cloning, Cm^r, F-Plasmid	3348	This work

pBAC-pcnB	pBAC-*pcnB*, Cm^r, F-Plasmid	na	This work
pBAC-pcnB E108A	pBAC-*pcnB* E108A, Cm^r, F-Plasmid	3349	This work
pBAC-pcnB N138H	pBAC-*pcnB* N138H, Cm^r, F-Plasmid	3350	This work
PY71.sfGFP	pY71-PT7-RiboJ-sfGFP-MGapt, Km^r, ColE1	2925	Addgene #129119
pGFP	pCWO-2P_{Tac}.sfGFP.TrrnB, Cb^r, ColE1	3454	This work
pAQ1	pSC101-P_{BAD}.RNAI, Cm^r, pSC101	3713	This work
pAQ2	pSC101-P_{BAD}.RNAI, Cm^r, pSC101-R43W	3714	This work
pAQ5	pSC101-P_{BAD}.RNAI, Cm^r, pSC101-E115K	3717	This work
pBMOS1:V2	pSC101-P_{BmoR}.BmoR.TrrnB-P_{BMO}.GFPuv	3094	Dietrich [10]
pSC101-WT-Bu2	pSC101-2P_{Tac}.*ter*.TrrnB-P_{Trc}.*aldh46.adh2*, Cb^r, pSC101-WT	3726	This work
pSC101-R43W-Bu2	pSC101-2P_{Tac}.*ter*.TrrnB-P_{Trc}.*aldh46.adh2*, Cb^r, pSC101-R43W	3194	This work
pSC101-E115K-Bu2	pSC101-2P_{Tac}.*ter*.TrrnB-P_{Trc}.*aldh46.adh2*, Cb^r, pSC101-E115K	3725	This work
pSC101-R43W-But	pSC101-P_{T5}.*phaA.hbd*.TrrnB-P_{Trc}.*crt*.TrrnB-2P_{Tac}.*ter*.TrrnB-P_{Trc}.*aldh46.adh2*, Cb^r, pSC101-R43W	3727	This work
pBu2-wRBS	pCWO-2P_{Tac}.*ter*.TrrnB-P_{Trc}.(wRBS)*aldh46.adh2*, Cb^r, ColE1	3193	This work
pPOL3	pColE1-P_{Trc}.PhaB.PhaC.TrrnB-P_{Tet}.*nphT7*, Cb^r	1627	Thuronyi [11]
pBAD33	Cm^r, p15A. Empty vector for gene expression.	1	Guzman et al. [12]
pRNAI	pBAD33-P_{Tet}.RNAI, Cm^r, p15A	3681	This work
pPra2	pETDuet-P_{T7}.BesD-PT7.*besC*-P_{T7}.(mpb)besB, Cb^r, ColE1	3219	Marchand [13]
pPra2-BesB.1	pETDuet-P_{T7}.BesD-P_{T7}.BesC-P_{T7}.(MPB)BesB.1, Cb^r, ColE1, BesB PROSS1	3687	This work
pPra2-BesB.3	pETDuet-P_{T7}.BesD-P_{T7}.BesC-P_{T7}.(MPB)BesB.3, Cb^r, ColE1, BesB PROSS1	3663	This work
pGro7	pBAD-P_{BAD}.*groES.groEL*, Cm^r, pACYC	2070	Nishihara et al. [14]
pET16-BesC	pET16-P_{T7}.BesC, Cb^r, ColE1	3489	Marchand [13]

pSV272.1-BesC	pSV272.1-P_{T7}.(His_6)(MPB)BesC, Km^r, ColE1	3525	Marchand [13]
pET16hp-IMDH	pET16-P_{T7}. (His_{10})IMDH, Cb^r, ColE1	2378	Fang [15]
pET28a-matB	pET28a-P_{T7}. (His_6)matB(His_6), Km^r, ColE1	905	Thuronyi [11]
pET16hp-BcLeuDH	pET16-P_{T7}. (His_{10})BcLeuDH, Cb^r, ColE1	3661	This work
pET16hp-BsAlaDH	pET16- P_{T7}. (His_{10})BsAlaDH, Cb^r, ColE1	3658	This work
pet28-LHicDH	pET28a- P_{T7}.LHicDH(His_6), Km^r, ColE1	3706	This work
pET16hp-4HBCS	pET16- P_{T7}.(His_{10})4HBCS, Cb^r, ColE1	3188	This work
pET16hp-PhaC STQK	pET16- P_{T7}.(His_{10})PhaCSTQK, Cb^r, ColE1	3720	This work

References

1. Grenier, F., Matteau, D., Baby, V. & Rodrigue, S. Complete genome sequence of *Escherichia coli* BW25113. *Genome Announc.* **2,** e01038-14 (2014).
2. Davis, M. A., Yu, V. Y., Fu, B., Wen, M., Koleski, E. J., Silverman, J., Berdan, C. A., Nomura, D. K. & Chang, M. C. Y. A cellular platform for production of C_4 monomers. 2023.01.09.523327 Preprint at https://doi.org/10.1101/2023.01.09.523327 (2023)
3. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166,** 557–580 (1983).
4. Jensen, K. F. The *Escherichia coli* K-12 'wild types' W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. *J. Bacteriol.* **175,** 3401–3407 (1993).
5. Heyde, S. A. H. & Nørholm, M. H. H. Tailoring the evolution of BL21(DE3) uncovers a key role for RNA stability in gene expression toxicity. *Commun. Biol.* **4,** 1–9 (2021).
6. Bond-Watts, B. B., Bellerose, R. J. & Chang, M. C. Y. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nat. Chem. Biol.* **7,** 222–227 (2011).
7. Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J. & Yang, S. Multigene editing in the *Escherichia coli genome* via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81,** 2506–2514 (2015).
8. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233–239 (2013).
9. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6640–6645 (2000).
10. Dietrich, J. A. Transcription factor-based small-molecule screens and selections. (2011). at <https://escholarship.org/uc/item/7j6805gf>

11. Thuronyi, B. W. Engineered biosynthesis of fluorinated polyketides. at <https://www.proquest.com/dissertations/docview/1917761242/abstract/AFEF041A9AF847EBPQ/2>

12. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.* **177,** 4121–4130 (1995).

13. Benmaman, M. & A, J. Biosynthesis and translational control of terminal-alkyne amino acids. (2019). at <https://escholarship.org/uc/item/34s446zk>

14. Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. & Yura, T. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli. *Appl. Environ. Microbiol.* **64,** 1694–1699 (1998).

15. Fang, J. Aldolase-Catalyzed Synthesis of Chiral Organofluorines. (2022). at <https://www.proquest.com/dissertations/docview/2718833349/abstract/CEEBB104454F4667PQ/7>

16. Taguchi, S., Yamada, M., Matsumoto, K., Tajima, K., Satoh, Y., Munekata, M., Ohno, K., Kohda, K., Shimamura, T., Kambe, H. & Obata, S. A microbial factory for lactate-based polyesters using a lactate-polymerizing enzyme. *Proc. Natl. Acad. Sci.* **105,** 17323–17327 (2008).

Appendix 2: *DNA sequences*

Appendix 2.1. Oligonucleotides

A. Cloning Primers

A complete list of oligonucleotides used for cloning.

Primer Name	Sequence
VY01	ATACCGCTCGCCGCAGCCGAACGCCCTAGGTCTAGGGCGGCGGATTTGTC
VY02	ATTCAAAACAGCATAGCTCTAAAACAAGCTTCCGCGGGTTTTGGGACCATTCAAAACAGCAT AG
VY03	GCTGTTTTGAATGGTCCCAAAACCCGCGGAAGCTTGTTTTAGAGCTATGCTGTTTTGAATGG TC
VY04	ATTCAAAACAGCATAGCTCTAAAACTCTAGAGTTTTGGGACCATTCAAAACAGC
VY05	GTTTTGAATGGTCCCAAAACCAGCCTGTATTACAGCGTAGGTTTTAGAGCTATGCTGTTT
VY06	AAACAGCATAGCTCTAAAACCTACGCTGTAATACAGGCTGGTTTTGGGACCATTCAAAAC
VY07	GCTGCGCGACAACATTTTCGCCTCCATCGAAGAAGACGCCCAGCACCGCGATTTCACTATCA ACAGCCTGTATTACAGCGTAGCAGATTTTACCGTCCGTGATTACGTTG
VY08	CAACGTAATCACGGACGGTAAAATCTGCTACGCTGTAATACAGGCTGTTGATAGTGAAATCG CGGTGCTGGGCGTCTTCTTCGATGGAGGCGAAAATGTTGTCGCGCAGC
VY09	GTTTTGAATGGTCCCAAAACGATTGGTTCGCCGTTCGCCGGTTTTAGAGCTATGCTGTTT
VY10	AAACAGCATAGCTCTAAAACCGGCGAACGGCGAACCAATCGTTTTGGGACCATTCAAAAC
VY11	CTTCGATGGTGACCAGCTGGCTGTTCACGTACCGCTGACGCTGGAAGCCCAGCTGGAAGCGC GTGCGCTGATGATGTCTACCAACAACATCCTGTCACCGGCGAACGGCGAACCAATCATCGTT CCGTCTCAGGA
VY12	TCCTGAGACGGAACGATGATTGGTTCGCCGTTCGCCGGTGACAGGATGTTGTTGGTAGACAT CATCAGCGCACGCGCTTCCAGCTGGGCTTCCAGCGTCAGCGGTACGTGAACAGCCAGCTGGT CACCATCGAAG
EK11	CGGCGTGCGCGACCTGTTACTTGGCAAAAAGCCGAAAGATTGTAGGCTGGAGCTGCTTC
EK12	CGCAAGGCCAACAGGTCATAAGCCGCACGGAACTTAGGATTGGGAATTAGCCATGGTCC
pcnB F	TTCAATTTCTGGGGCAGCAGG
pcnB R	TTACCATTTGTTTAATTTGTCAAATGCTCTTGTATGTAAGATTT
EK14	GCCCTGAAGGTAATGTACAGGCTCAATAAAGCGGGATACGAAGCCTGGCTTGTAGGCTGGAG CTGCTTC
EK15	CGGAAACCTGGAACTCACCCCACCATTTCACCAGACGCTGCAGTTCAGCGTGGGAATTAGCC ATGGTCC
EK28	GCCGCATAGTTAAGCCAGTAGGCGTATCACGAGGCCCTTTC
EK29	CGTAGCGATAGCGGAGTGTACCAGCCAGCCCTTCCGGCAA
EK16	CCCTTTCGTCTTCAAGAATTACAGACACACAGTATATTGGCCGCT

EK17	TGAGACCACAAGTTATCAGATCTGGAAAGCGGTCTCGAGTACACCTCGGTAGTGTCAAACAT CATTCAGGACAA
EK18	CGAGACCGCTTTCCAGATCTGATAACTTGTGGTCTCACAGTGGCGTATATTGCCATAGGCAG CAATCTGGCCTC
EK19	TGGACAAAATACCTGGTTACCCATCCTTGTTACCGCCGTCAACGC
EK20	ACTGCTGGTCTCGTACTATTTTTACCCGAGTCGCTAATTTTTGCCG
EK21	ACTGCTGGTCTCGACTGTCATGCGGTACCCTCACGACG
EK56	CCACTTCTGCGCTCGGCCCTTCCGGCCACTAGTATCGATGATAAGCTGTCAAACATGAGC
EK57	CGGTAAACAGTTCTTCACCTTTGCTCATATGACCTCCTAAGCATCGATGGATCC
EK58	GGATCCATCGATGCTTAGGAGGTCATATGAGCAAAGGTGAAGAACTGTTTACCG
EK59	CCGCCAAAACAGCCCTCGAGCTTATTTTTCGAACTGCGGATGGCTCC
AQ3	AGGTATTTATTCGGCGCAAACTAGGGTACGGGTTTTGCTG
AQ4	AGTAGCATAGGGTTTGCAGAACTAGGGACAGTAAGACGGG
AQ5	CCCGTCTTACTGTCCCTAGTTCTGCAAACCCTATGCTACTCC
AQ6	TAAGGCAGTTATTGGTGCCCACAGCCAAGCTTGCATGCCT
AQ7	AGGCATGCAAGCTTGGCTGTGGGCACCAATAACTGCCTTA
AQ8	CAGCAAAACCCGTACCCTAGTTTGCGCCGAATAAATACCT
AQ14	TTATTGGTGCCCACAGCCAAGCTCTGCTGCTTGCAAACAAAAAAACCACC
AQ15	TCGGTACCCGGGGATCCTCTAGACAGTATTTGGTATCTGCGCTCTGCT
AQ18	GTTCTTTCCATGTAGGGTTTTCAATCGTGGGGTTGAGTAGT
AQ19	ACCCTACATGGAAAGAACGGACGGTATCGTTCACTTATAACCA
AQ24	GGCAATATCTTTTCACTAAAAACTAATTCTAATTTTTCGCTTGAGAACTT
AQ25	TTAGTGAAAAGATATTGCCTTATCTTTTCCAGTTAAAAAAATTCATAAAATATAATCTGG
AQ51	ATTACCCGCATCACAAGGAAAAGCCCAATATGAATAAAGACACCCTGATTCCGACCA
AQ52	ATTGGGCTTTTCCTTGTGATGCGGGTAATCTGTTTCCTGTGTGAAATTGTTATCCGC
AQ53	TTCTTTACCTTTGCCGAATTCGATACGGGTTGGATTGCAG
AQ60	AAGTTGAACGTTTCGATCGTATTTAAGCTCGAGGGCTGTTTTGGCGGATGAG
EK158	CTGTCAGACCAAGTTTACTCCTAGGGTACGGGTTTTGCTGCCC
EK159	GCATCAGGCGCTCTTCCGCTACTAGGGACAGTAAGACGGGTAAGCC
EK162	GTCTTACTGTCCCTAGTAGCGGCGTATCACGAGGCCCTTTCG

EK163	GCATCAGGCGCTCTTCCGCTCCAGCCAGCCCTTCCGGCAA
EK115	TTTTTGTTTGCAAGCAGCAGCCGCTCATGAGACAATAACCCTGATAAATGC
EK116	CCCTTTAGATTCATAAAGCGATTCACCACCCTGAATTGACTCTCTTCC
EK117	GTCAATTCAGGGTGGTGAATCGCTTTATGAATCTAAAGGGTGGTTAACTCGA
EK118	GCGCAGATACCAAATACTGTGGATCCTGAAGACGAAAGGGCCT
EK119	CCCTTTCGTCTTCAGGATCCACAGTATTTGGTATCTGCGCTCTGCT
EK120	GGTTATTGTCTCATGAGCGGCTGCTGCTTGCAAACAAAAAAACCACC
SK01	ATAAGAAGGAGATATACATATGGGTTCTTCTCACCATC
SK09	CATGCCCTGAAAATACAGGTTTTCCTCG
SK05	AAAACCTGTATTTTCAGGGCATGTCGCAGGCGGTCTCTG
SK06	TCGAGTTAGTCTGGGCGCAAGGCG
SK07	AAAACCTGTATTTTCAGGGCATGTCACAGGCCGTATCTGGCA
SK08	TCGAGTTAATCCGGGCGTAAAGCACGC
JF01	CATCTAGAAGTGCTTTTTCAGGGCCCGCATATGACATTAGAAATCTTCGAATACTTAG
JF02	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAGCGACGGCTAATAATATCG
JF03	CATCTAGAAGTGCTTTTTCAGGGCCCGCATATGATCATAGGGGTTCCTAAAGAG
JF04	TCCTTTCGGGCTTTGTTAGCAGCCGGATCCTTAAGCACCCGCCACAG
VS01	CTGGTGCCGCGCGGCAGCCATATGGCACGTAAGATCGGGATCATC
VS02	CAGTGGTGGTGGTGGTGGTGCTCGAGCGTGTCAACGATTTCGTC
VS03	AGCAGCGGCCTGGTGC
VS04	GCTGCTGCCCATGGTATATCTCCT
VS05	AAGTGCTTTTTCAGGGCCCGCATATGAGTAACAAGAATAGCGATGACTTGAATCGT
VS06	TTCGGGCTTTGTTAGCAGCCGGATCCTCAACGTTCATGCACATACG

B. Primers for CRISPR approach I

Oligonucleotides used for generation of chromosomal point mutations via CRISPR approach I.

Primer	Target mutation	Sequence
459-pTargetF-F2	*NA*	ACTAGTATTATACCTAGGACTGAGCTAGCTGTCAAG
338-V5-target	*rpoC* K1192E	TCCTAGGTATAATACTAGTCGGATCGCTACCGTCTACCGGTTTTAGAGCTAGAAATAGC
343-V10-target	*pcnB* R149P	TCCTAGGTATAATACTAGTGGCTGTTGATAGTGAAATCGGTTTTAGAGCTAGAAATAGC
344-V11-target	*pcnB* E108A	TCCTAGGTATAATACTAGTGGCTCATGTAATGTTTGGCCGTTTTAGAGCTAGAAATAGC
345-V12-target	*pcnB* N138H	TCCTAGGTATAATACTAGTGCGCTGGGCGTCTTCTTCGAGTTTTAGAGCTAGAAATAGC
346-V13-target	*pcnB* L392fs	TCCTAGGTATAATACTAGTGGAGCATCCTAAGTTCCGTGGTTTTAGAGCTAGAAATAGC
347-V14-target	*rne* R373S	TCCTAGGTATAATACTAGTTCTGCGTGAAGCGGTGCGTCGTTTTAGAGCTAGAAATAGC
350-V17-target	*rne* R488H V489L	TCCTAGGTATAATACTAGTCGTGCTGCGCGTGCGTAAAGGTTTTAGAGCTAGAAATAGC
375-V5-1	*rpoC* K1192E	GGATCAGACGACCCACGATAACGTTCTCTT
376-V5-2	*rpoC* K1192E	GGTTATCACGCCGGTAGACGGTAGCGATCCGTACGAAGAGATGATTCCGGAATGGCGTC
377-V5-3	*rpoC* K1192E	GACGCCATTCCGGAATCATCTCTTCGTACGGATCGCTACCGTCTACCGGCGTGATAACC
378-V5-4	*rpoC* K1192E	CGCACACCATGCCGGTTATCACCGAAGTAA
395-V10-1	*pcnB* R149P	GAGGTGTACTATTTTTACCCGAGTCGCTAA
396-V10-2	*pcnB* R149P	AATACAGGCTGTTGATAGTGAAATCGCGTGGCTGGGCGTCTTCTTCGATGGAGCCGAAA
397-V10-3	*pcnB* R149P	TTTCGGCTCCATCGAAGAAGACGCCCAGCCACGCGATTTCACTATCAACAGCCTGTATT
398-V10-4	*pcnB* R149P	ATCTTCTGTGCCGTCTCCAGCAGTGGGTAC
399-V11-1	*pcnB* E108A	CACACTGGCAGGATTTCAGCGTCGAGCAAA
400-V11-2	*pcnB* E108A	CGTGGTGTCCACGGAAGGTCGCAACTGCGATAATCTCTGGGCCAAACATTACATGA
401-V11-3	*pcnB* E108A	TCATGTAATGTTTGGCCCAGAGATTATCGCAGTTGCGACCTTCCGTGGACACCACG
402-V11-4	*pcnB* E108A	CAATGATCCGCTCCATCGGGCTGTCGCCAT
403-V12-1	*pcnB* N138H	GTCCTGAATGATGTTTGACACTACCGAGGTG
404-V12-2	*pcnB* N138H	CGTCTTCTTCGATAGAGCCGAAAATGTGGTCGCGCAGCAACATGCCGTTTTGCCCG

405-V12-3	*pcnB* N138H	CGGGCAAAACGGCATGTTGCTGCGCGACCACATTTTCGGCTCTATCGAAGAAGACG
406-V12-4	*pcnB* N138H	CGGGTTCACGCGCATATCGTTATGGATACG
407-V13-1	*pcnB* L392fs	ATATCCCACCGGCACGCCTGTTTGAAGAATC
408-V13-2	*pcnB* L392fs	TTACGCTCAACTTCAGCTCGCAAGGCCAAAGGTCATAAGCGGCACGGAACTTAGGATG
409-V13-3	*pcnB* L392fs	CATCCTAAGTTCCGTGCCGCTTATGACCTTTGGCCTTGCGAGCTGAAGTTGAGCGTAA
410-V13-4	*pcnB* L392fs	CGTTCAGTATTTATCACTTCATTACCAAACAG
477-V14-1new	*rne* R373S	CACCCGCATTGAACCGAGTCTGGAAGCTGCTTTTG
478-V14-2new	*rne* R373S	GAGAAATATGGCTGATTTGAATACTCGCACGGTCTTGGCGAACGGCTTCACGCAGACG
479-V14-3new	*rne* R373S	CGTCTGCGTGAAGCCGTTCGCCAAGACCGTGCGAGTATTCAAATCAGCCATATTTCTC
480-V14-4new	*rne* R373S	CTTGTTGCGCCTGACGTTTATCATCATTACGGCGGC
485-V17-1new	*rne* R488H V489L	GGCATTTCTCGCCGTATCGAAGGCGACGACC
486-V17-2new	*rne* R488H V489L	CTTAAGGTTGGGGTTTCTTCACCTTTGCGAAGGTGCAGCACGTGGTAGTGCGGGGTTTC
487-V17-3new	*rne* R488H V489L	GAAACCCCGCACTACCACGTGCTGCACCTTCGCAAAGGTGAAGAAACCCCAACCTTAAG
488-V17-4new	*rne* R488H V489L	ACTTTGCCAGAGGCCAGTTCCGGAGACGC

C. qPCR Primers

Primers used for RT-qPCR.

Name	**Sequence**
idnT F	CTGTTTAGCGAAGAGGAGATGC
idnT R	ACAAACGGCGGCGATAGC
phaA F	GATGAATTTGCAGTCGGCTC
phaA R	CTGTGGAATCAGAACCGGAA
ter F	AACAAGGAAAACCCGAGCAT
ter R	GCGGAATGACCGGAATTACT

Appendix 2.2. gBlocks

Nucleotide sequences of ordered gBlocks.

Name	Sequence
L-HicDH	caTATGGCACGTAAGATCGGGATCATCGGTCTCGGGAATGTTGGTGCTGCCGTAGCGCACG GCCTGATCGCCCAAGGTGTGGCGGACGACTACGTGTTTATCGACGCGAACGAAGCCAAGGT CAAGGCAGACCAGATTGATTTTCAAGACGCCATGGCGAACCTCGAAGCGCATGGGAATATC GTCATTAACGACTGGGCTGCCTTAGCCGATGCGGATGTTGTTATTTCTACACTTGGCAATA TCAAACTTCAACAAGACAATCCGACTGGGGATCGTTTTGCCGAACTTAAATTTACATCTTC AATGGTCCAGTCGGTCGGGACTAACCTCAAAGAATCTGGCTTCCATGGCGTACTCGTCGTC ATTTCCAACCCTGTCGATGTGATTACGGCACTCTTTCAACATGTAACTGGTTTTCCGGCGC ACAAGGTAATCGGCACAGGGACTTTATTAGATACGGCGCGGATGCAGCGGGCCGTCGGTGA GGCCTTTGATTTAGATCCTCGGTCAGTATCCGGTTACAACCTGGGGGAGCACGGGAACAGT CAATTCGTAGCCTGGAGCACAGTCCGGGTCATGGGGCAACCAATTGTTACACTGGCGGACG CTGGTGACATCGACCTTGCAGCAATCGAAGAAGAGGCTCGCAAAGGGGGTTTCACCGTGCT GAATGGCAAAGGGTACACATCGTATGGTGTGGCAACTTCCGCCATCCGGATTGCTAAGGCC GTGATGGCAGACGCCCACGCAGAATTGGTAGTGAGCAATCGGCGTGATGACATGGGGATGT ATTTATCGTATCCAGCGATTATCGGTCGGGATGGGGTTCTCGCGGAAACTACTTTAGACCT GACCACAGACGAACAAGAGAAGCTTTTACAGTCGCGCGACTATATCCAACAGCGCTTTGAC GAAATCGTTGACACGCtcgag
4HBCS	AGCGGCCATATCGAAGGTCGTCATATGGTGACCGTGCAGGATTTCTTTCGGAAATTTATCG AATTTCAGAACTCGCCTAACGAGAAATCTCTTCAAGAGATAGTAAAGTTAGTCGGGCAATT AGACCTTAGACGGTTCAATTGGGTGCGGGACGTATTTGAGGATATACATGTTAAGGAAAGA GGCTCCAAAACGGCCTTGATATGGCGCGACATAAATACCGGTGAGGAGGCCAAATTATCTT ATCACGAACTGAGTTTAATGTCCAATCGTGTGTTGTCAACCTTGAGAAAACACGGTTTGAA AAAAGGGGACGTTGTCTACTTGATGACCAAGGTACACCCGATGCACTGGGCAGTTTTTCTG GCTGTTATTAAGGGTGGGTTTGTTATGGTCCCAAGCGCGACGAACCTTACTGTTGCTGAAA TGAAATACCGTTTTTCTGATTTAAAACCTAGCGCGATTATCTCAGATTCCCTGAGAGCTTC GGTCATGGAAGAAGCGTTGGGCAGCCTGAAGGTAGAAAAGTTCTTGATTGACGGCAAGCGG GAGACGTGGAACTCCCTGGAAGATGAATCCAGTAATGCTGAGCCCGAAGATACGCGGGGCG AGGATGTTATAATTAATTATTTTACTTCTGGGACGACCGGCATGCCCAAGAGAGTCATCCA CACGGCTGTGAGCTACCCTGTAGGATCAATAACCACTGCGAGTATCGTAGGCGTGCGCGAG AGCGACTTGCACTTGAATCTTTCAGCAACGGGATGGGCAAAATTTGCCTGGAGTTCATTTT TTTCACCGCTTCTTGTTGGCGCGACTGTTGTTGGCATTAACTACGAAGGGAAATTAGATAC TCGGCGCTACCTGGGGGAGGTTGAGAATCTGGGGGTAACCAGCTTCTGCGCCCCGCCAACG GCTTGGCGTCAATTTATTACCCTTGATTTAGATCAGTTTCGGTTCGAGCGGTTAAGAAGTG TGGTTTCAGCAGGCGAGCCGTTGAATCCAGAAGTTATCAAGATATGGAAAGACAAATTCAA TCTTACTATACGGGATTTCTACGGTCAAACGGAAACAACCGCGATGGTTGGTAATTTCCCC TTTCTGAAGGTAAAACCCGGCTCCATGGGAAAGCCTCACCCGCTTTACGACATTAGACTTT TAGACGACGAGGGCAAGGAAATTACCAAGCCTTACGAGGTAGGTCATATTACGGTCAAGCT TAACCCGCGTCCTATAGGCCTGTTCCTGGGATACTCAGACGAAAAGAAAAACATGGAATCA TTTAGAGAAGGTTACTACTACACGGGTGACAAGGCGTATTTTGACGAAGAGGGCTACTTTT ATTTTGTTGGTCGGGGCGACGACGTCATTAAAACATCAGATTATAGAGTCGGGCCTTTTGA AGTGGAGTCTGCGCTGCTGGAGCACCCTGCGGTCGCTGAGGCCGCCGTAGTGGGAGTCCCG GATACAGTTCGGTGGCAGCTGGTAAAGGCGTACATTGTCCTGAAGAAAGGATATATGCCTT CCAAGGAATTAGCGGAGGAGATAAGAGAAAAAATGAAGACCTTGCTGTCCCCATATAAGGT ACCCCGTATAATAGAGTTTGTCGACGAGCTGCCCAAGACAATTTCCGGTAAGATTCGTCGC GTGGAATTACGCAAGCGCGAAGAAGAGAAGAGAAAAAAGGGCGAGGTCGGACAAAATGAAT ATGTTTTCTGAgatccggctgctaacaaagccc
PhaC STQK	catATGAGTAACAAGAATAGCGATGACTTGAATCGTCAAGCCTCGGAAAACACCTTGGGGC TTAACCCTGTCATCGGCCTGCGTGGAAAAGATCTGCTGACTTCTGCCCGAATGGTTTTAAC CCAAGCCATCAAACAACCCATTCACAGCGTCAAGCACGTCGCGCATTTTGGCATCGAGCTG AAGAACGTGATGTTTGGCAAATCGAAGCTGCAACCGGAAAGCGATGACCGTCGTTTCAACG ACCCCGCCTGGAGTCAGAACCCACTCTACAAACGTTATCTACAAACCTACCTGGCGTGGCG CAAGGAACTCCACGACTGGATCGGCAACAGCAAACTGTCCGAACAGGACATCAATCGCGCT

CACTTCGTGATCACCCTGATGACCGAAGCCATGGCCCCGACCAACAGTGCGGCCAATCCGG
CGGCGGTCAAACGCTTCTTCGAAACCGGCGGTAAAAGCCTGCTCGACGGCCTCACACATCT
GGCCAAGGACCTGGTAAACAACGGCGGCATGCCGAGCCAGGTGGACATGGGCGCTTTCGAA
GTCGGCAAGAGTCTGGGGACGACTGAAGGTGCAGTGGTTTTCCGCAACGACGTCCTCGAAT
TGATCCAGTACCGGCCGACCACCGAACAGGTGCATGAGCGACCGCTGCTGGTGGTCCCACC
GCAGATCAACAAGTTTTATGTGTTTGACCTGAGCCCGGATAAAAGCCTGGCGCGCTTCTGC
CTGAGCAACAACCAGCAAACCTTTATCGTCAGCTGGCGCAACCCGACCAAGGCCCAGCGTG
AGTGGGGTCTGTCGACTTACATCGATGCGCTCAAAGAAGCCGTCGACGTAGTTTCCGCCAT
CACCGGCAGCAAAGACATCAACATGCTCGGCGCCTGCTCCGGTGGCATTACCTGCACCGCG
CTGCTGGGTCACTACGCCGCTCTCGGCGAGAAGAAGGTCAATGCCCTGACCCTTTTGGTCA
CCGTGCTCGACACCACCCTCGACTCCCAGGTTGCACTGTTCGTCGATGAGAAAACCCTGGA
AGCTGCCAAGCGTCACTCGTATCAGGCCGGCGTGCTGGAAGGCCGCGACATGGCCAAAGTC
TTCGCCTGGATGCGCCCTAACGACCTGATCTGGAACTACTGGGTCAACAACTACCTGCTGG
GTAACGAGCCACCGGTCTTCGACATTCTTTTCTGGAACAACGACACCACCCGGTTGCCTGC
TGCGTTCCACGGCGATCTGATCGAAATGTTCAAAAATAACCCACTGGTGCGCGCCAATGCA
CTCGAAGTGAGCGGCACGCCGATCGACCTCAAACAGGTCACTGCCGACATCTACTCCCTGG
CCGGCACCAACGATCACATCACGCCCTGGAAGTCTTGCTACAAGTCGGCGCAACTGTTCGG
TGGCAAGGTCGAATTCGTGCTGTCCAGCAGTGGGCATATCAAAAGCATTCTGAACCCGCCG
GGCAATCCGAAATCACGTTACATGACCAGCACCGACATGCCAGCCACCGCCAACGAGTGGC
AAGAAAACTCAACCAAGCACACCGACTCCTGGTGGCTGCACTGGCAGGCCTGGCAGGCCGA
GCGCTCGGGCAAACTGAAAAAGTCCCCGACCAGCCTGGGCAACAAGGCCTATCCGTCAGGA
GAAGCCGCGCCGGGCACGTATGTGCATGAACGTTGAggatcc

Appendix 2.3. Pathway RBS sequences

Ribosome binding site sequences driving translation of enzymes in the n-butanol and BDO pathways. The predicted strength for each RBS.

Gene	RBS strength	RBS
PhaA	170	GAGCTCAAGGAGATATACAT
Hbd	140	GAATTCAAGGAGATATATA
Crt	200	TTTCACACAGGAAACAGACC
Ter	24,000	ATCGATGCTTAGGAGGTCAT
Aldh46	170,000	TTCAAAAAAGGAGGTAAAAA
Adh2	10,000	AAACTCTAAGCGAGGAATAC
Aldh46 (wRBS)	9,000	ATCACAAGGAAAAGCCCAAT

Appendix 3: *Proteins*

Appendix 3.1. Amino acid sequences

Amino acid sequences and accession numbers for relevant proteins.

Protein	Accession	Sequence
PhaA	WP_010810132.1	MTDVVIVSAARTAVGKFGGSLAKIPAPELGAVVIKAALERAGVKPEQ VSEVIMGQVLTAGSGQNPARQAAIKAGLPAMVPAMTINKVCGSGLKA VMLAANAIMAGDAEIVVAGGQENMSAAPHVLPGSRDGFRMGDAKLVD TMIVDGLWDVYNQYHMGITAENVAKEYGITREAQDEFAVGSQNKAEA AQKAGKFDEEIVPVLIPQRKGDPVAFKTDEFVRQGATLDSMSGLKPA FDKAGTVTAANASGLNDGAAAVVVMSAAKAKELGLTPLATIKSYANA GVDPKVMGMGPVPASKRALSRAEWTPQDLDLMEINEAFAAQALAVHQ QMGWDTSKVNVNGGAIAIGHPIGASGCRILVTLLHEMKRRDAKKGLA SLCIGGGMGVALAVERK*
PhaB	WP_010810131.1	MTQRIAYVTGGMGGIGTAICQRLAKDGFRVVAGCGPNSPRREKWLEQ QKALGFDFIASEGNVADWDSTKTAFDKVKSEVGEVDVLINNAGITRD VVFRKMTRADWDAVIDTNLTSLFNVTKQVIDGMADRGWGRIVNISSV NGQKGQFGQTNYSTAKAGLHGFTMALAQEVATKGVTVNTVSPGYIAT DMVKAIRQDVLDKIVATIPVKRLGLPEEIASICAWLSSEESGFSTGA DFSLNGGLHMG*
Hbd	WP_010965995.1	MKKVCVIGAGTMGSGIAQAFAAKGFEVVLRDIKDEFVDRGLDFINKN LSKLVKKGKIEEATKVEILTRISGTVDLNMAADCDLVIEAAVERMDI KKQIFADLDNICKPETILASNTSSLSITEVASATKRPDKVIGMHFFN PAPVMKLVEVIRGIATSQETFDAVKETSIAIGKDPVEVAEAPGFVVN RILIPMINEAVGILAEGIASVEDIDKAMKLGANHPMGPLELGDFIGL DICLAIMDVLYSETGDSKYRPHTLLKKYVRAGWLGRKSGKGFYDYSK *
Crt	WP_010965999.1	MELNNVILEKEGKVAVVTINRPKALNALNSDTLKEMDYVIGEIENDS EVLAVILTGAGEKSFVAGADISEMKEMNTIEGRKFGILGNKVFRRLE LLEKPVIAAVNGFALGGGCEIAMSCDIRIASSNARFGQPEVGLGITP GFGGTQRLSRLVGMGMAKQLIFTAQNIKADEALRIGLVNKVVEPSEL MNTAKEIANKIVSNAPVAVKLSKQAINRGMQCDIDTALAFESEAFGE CFSTEDQKDAMTAFIEKRKIEGFKNR*
Ter	WP_002685379.1	MIVKPMVRNNICLNAHPQGCKKGVEDQIEYTKKRITAEVKAGAKAPK NVLVLGCSNGYGLASRITAAFGYGAATIGVSFEKAGSETKYGTPGWY NNLAFDEAAKREGLYSVTIDGDAFSDEIKAQVIEEAKKKGIKFDLIV YSLASPVRTDPDTGIMHKSVLKPFGKTFTGKTVDPFTGELKEISAEP ANDEEAAATVKVMGGEDWERWIKQLSKEGLLEEGCITLAYSYIGPEA TQALYRKGTIGKAKEHLEATAHRLNKENPSIRAFVSVNKGLVTRASA VIPVIPLYLASLFKVMKEKGNHEGCIEQITRLYAERLYRKDGTIPVD EENRIRIDDWELEEDVQKAVSALMEKVTGENAESLTDLAGYRHDFLA SNGFDVEGINYEAEVERFDRI*
Aldh46	WP_012059995.1	MNKDTLIPTTKDLKVKTNGENINLKNYKDNSSCFGVFENVENAISSA VHAQKILSLHYTKEQREKIITEIRKAALQNKEVLATMILEETHMGRY EDKILKHELVAKYTPGTEDLTTTAWSGDNGLTVVEMSPYGVIGAITP STNPTETVICNSIGMIAAGNAVVFNGHPCAKKCVAFAVEMINKAIIS CGGPENLVTTIKNPTMESLDAIIKHPSIKLLCGTGGPGMVKTLLNSG KKAIGAGAGNPPVIVDDTADIEKAGRSIIEGCSFDNNLPCIAEKEVF VFENVADDLISNMLKNNAVIINEDQVSKLIDLVLQKNNETQEYFINK KWVGKDAKLFLDEIDVESPSNVKCIICEVNANHPFVMTELMMPILPI VRVKDIDEAIKYAKIAEQNRKHSAYIYSKNIDNLNRFEREIDTTIFV KNAKSFAGVGYEAEGFTTFTIAGSTGEGITSARNFTRQRRCVLAG*
Aldh7	WP_012425099.1	MERNLSVLSQTNDLKITKRTEGDKSNNKESYLGVFKKVENAITKAIY AQKKLSLYYTKEDRERIIKSIRKATLENKEILAKMIVDETHMGRYED KILKHELVAKYTPGTEDLITTAWSGDQGLTLVEMSPYGVIGAITPST NPTETVICNSIGMIAAGDSVVFNGHPGAKKCVAFAVDMINKAVIREG GPENLVTTVENPTMESLNVIMKHPYIKLLCGTGGPGLIKTLLNSGKK

		AIGAGAGNPPVIVDDSADIDKAAKNIIEGCSFDNNLPCIAEKEVFVF ENVANDLIQNMIKNNAVLINENQVSKLLDLVLLERKDETLEYAINKK WVGKDAKLFLDKIGIKASDNVRCIICEVDANHPFVMTELMMPILPIV RVKDVDEAIECAKTAEQRKRHSAYMYSKNIDNLNRFEKEIDTTIFVK NAKSFAGVGFGAEGFTTFTIAGPTGEGITSARNFTRQRRCVLAG*
Aldh21	WP_004143947.1	MNTAELETLIRTILSEKLAPTPPAPQQEQGIFCDVGSAIDAAHQAFL RYQQCPLKTRSAIISALRETLAPELATLAEESATETGMGNKEDKYLK NKAALENTPGIEDLTTSALTGDGGMVLFEYSPFGVIGAVAPSTNPTE TIINNSISMLAAGNSVYFSPHPGAKKVSLKLIARIEEIAYRCSGIRN LVVTVAEPTFEATQQMMSHPLIAVLAITGGPGIVAMGMKSGKKVIGA GAGNPPCIVDETADLVKAAEDIISGAAFDYNLPCIAEKSLIVVASVA DRLIQQMQDFDALLLSRQEADTLRAVCLPDGAANKKLVGKSPAALLA AAGLAVPPRPPRLLIAEVEANDPWVTCEQLMPVLPIVRVADFDSALA LALRVEEGLHHTAIMHSQNVSRLNLAARTLQTSIFVKNGPSYAGIGV GGEGFTTFTIATPTGEGTTSARTFARLRRCVLTNGFSIR*
Adh2	WP_011732148.1	MVNFSYCNPTRIEFGKGKENSIGEYLNEYGAKNVLILFGSDRVKKDG LFDKATASLTKFGIKFSELGDIVSNPVLSKVYEAINLARKNGVDSVL AIGGGSVLDTAKSVAAGAKYDGDVWDLFLAKAPIKDALMVFDIMTLA ATGSEMNSFAVVTNEDTKEKISITSSLVNPKVSVINPELMKSISKNY LVYSAADIIAHSIEGYLTATHHPEIISKLVEANISTIIKTTEILLAD PDNYDARAEFAWAATCALNGTTYVGVGGYSYPNHMIEHSISALYGVP HGAGLSVVMPAWMKWYKDKNEAQFSRFAKVIFGKNSADEGIEALKTW FKKIGTPTKLRDFGLDMSVSDITTAALHHAKAFGIADIYTKDVLEEI LNLAY*
sfGFP	QLI61463.1	MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFI CTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQ ERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKL EYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPI GDGPVLLPDNHYLSTQTVLSKDPNEKGTRDHMVLHEYVNAAGITWSH PQFEK*
BsAlaDH	WP_003243280.1	MGHHHHHHHHHHSSGHLEVLFQGPHMIIGVPKEIKNNENRVALTPGG VSQLISNGHRVLVETGAGLGSGFENEAYESAGAEIIADPKQVWDAEM VMKVKEPLPEEYVYFRKGLVLFTYLHLAAEPELAQALKDKGVTAIAY ETVSEGRTLPLLTPMSEVAGRMAAQIGAQFLEKPKGGKGILLAGVPG VSRGKVTIIGGGVVGTNAAKMAVGLGADVTIIDLNADRLRQLDDIFG HQIKTLISNPVNIADAVAEADLLICAVLIPGAKAPTLVTEEMVKQMK PGSVIVDVAIDQGGIVETVDHITTHDQPTYEKHGVVHYAVANMPGAV PRTSTIALTNVTVPYALQIANKGAVKALADNTALRAGLNTANGHVTY EAVARDLGYEYVPAEKALQDESSVAGA*
BcLeuDH	WP_000171355.1	MGHHHHHHHHHHSSGHLEVLFQGPHMTLEIFEYLEKYDYEQVVFCQD KESGLKAIIAIHDTTLGPALGGTRMWTYDSEEAAIEDALRLAKGMTY KNAAAGLNLGGAKTVIIGDPRKDKSEAMFRALGRYIQGLNGRYITAE DVGTTVDDMDIIHEETDFVTGISPSFGSSGNPSPVTAYGVYRGMKAA AKEAFGTDNLEGKVIAVQGVGNVAYHLCKHLHAEGAKLIVTDINKEA VQRAVEEFGASAVEPNEIYGVECDIYAPCALGATVNDETIPQLKAKV IAGSANNQLKEDRHGDIIHEMGIVYAPDYVINAGGVINVADELYGYN RERALKRVESIYDTIAKVIEISKRDGIATYVAADRLAEERIASLKNS RSTYLRNGHDIISRR*
L-HicDH	WP_003607654.1	MARKIGIIGLGNVGAAVAHGLIAQGVADDYVFIDANEAKVKADQIDF QDAMANLEAHGNIVINDWAALADADVVISTLGNIKLQQDNPTGDRFA ELKFTSSMVQSVGTNLKESGFHGVLVVISNPVDVITALFQHVTGFPA HKVIGTGTLLDTARMQRAVGEAFDLDPRSVSGYNLGEHGNSQFVAWS TVRVMGQPIVTLADAGDIDLAAIEEEARKGGFTVLNGKGYTSYGVAT SAIRIAKAVMADAHAELVVSNRRDDMGMYLSYPAIIGRDGVLAETTL DLTTDEQEKLLQSRDYIQQRFDEIVDTLE*
4HBCS	WP_012020371.1	MVTVQDFFRKFIEFQNSPNEKSLQEIVKLVGQLDLRRFNWVRDVFED IHVKERGSKTALIWRDINTGEEAKLSYHELSLMSNRVLSTLRKHGLK KGDVVYLMTKVHPMHWAVFLAVIKGGFVMVPSATNLTVAEMKYRFSD

		LKPSAIISDSLRASVMEEALGSLKVEKFLIDGKRETWNSLEDESSNA EPEDTRGEDVIINYFTSGTTGMPKRVIHTAVSYPVGSITTASIVGVR ESDLHLNLSATGWAKFAWSSFFSPLLVGATVVGINYEGKLDTRRYLG EVENLGVTSFCAPPTAWRQFITLDLDQFRFERLRSVVSAGEPLNPEV IKIWKDKFNLTIRDFYGQTETTAMVGNFPFLKVKPGSMGKPHPLYDI RLLDDEGKEITKPYEVGHITVKLNPRPIGLFLGYSDEKKNMESFREG YYYTGDKAYFDEEGYFYFVGRGDDVIKTSDYRVGPFEVESALLEHPA VAEAAVVGVPDTVRWQLVKAYIVLKKGYMPSKELAEEIREKMKTLLS PYKVPRIIEFVDELPKTISGKIRRVELRKREEEKRKKGEVGQNEYVF *
PhaC STQK	QJR97783.1	MSNKNSDDLNRQASENTLGLNPVIGLRGKDLLTSARMVLTQAIKQPI HSVKHVAHFGIELKNVMFGKSKLQPESDDRRFNDPAWSQNPLYKRYL QTYLAWRKELHDWIGNSKLSEQDINRAHFVITLMTEAMAPTNSAANP AAVKRFFETGGKSLLDGLTHLAKDLVNNGGMPSQVDMGAFEVGKSLG TTEGAVVFRNDVLELIQYRPTTEQVHERPLLVVPPQINKFYVFDLSP DKSLARFCLSNNQQTFIVSWRNPTKAQREWGLSTYIDALKEAVDVVS AITGSKDINMLGACSGGITCTALLGHYAALGEKKVNALTLLVTVLDT TLDSQVALFVDEKTLEAAKRHSYQAGVLEGRDMAKVFAWMRPNDLIW NYWVNNYLLGNEPPVFDILFWNNDTTRLPAAFHGDLIEMFKNNPLVR ANALEVSGTPIDLKQVTADIYSLAGTNDHITPWKSCYKSAQLFGGKV EFVLSSSGHIKSILNPPGNPKSRYMTSTDMPATANEWQENSTKHTDS WWLHWQAWQAERSGKLKKSPTSLGNKAYPSGEAAPGTYVHER*
BesD	WP_177050793.1	MNYVLDEARLQEHHATNFPESSVFALRHEFARNGFIKVRNIVDDDLR EKITREVNSLIDRQLERRDLHLATTDNTPRYMSVVRSEFIAENSTLI NTLSKSKGLLETLSQIAGTQLIASVSKDEEYLITKQERKGDTHGWHW GDYSFALIWIIETPSIAKGGMLQCVPHTSWDKSNPRIHELLCSNPIA TYGFVTGDIYFLRTDTTLHRTIPLNEDATRIILNMTWAAEKDLSRNL HGNDRWWEDQHVEAAKSLT*
BesC	WP_051065928.1	MSITQETFQSETAYVVRNEGVDLELKLFMDEQIEYILKHRATEHPFL NAYAEHGLPPEQSQVLYLETLHYFKYLPFYVCGISTITRDEAVLRTI AFNARDELGETHSHSDLYRKFLHDKGISEEQIEAYKCLPSTQALNDG ICALYSKPPLQKALGGLFADEAMSASMVSKYNDGLIKEGVGERGRFF WTLHMEVEVGHSNAVFNVMEKHLQTPQERRLFAEGIEQYLHLMEVYW DGIERKLNAGGRQ*
BesB	WP_030410683.1	MTTEASGELGAACSLRHTAVGRPVPGSLHSVSVSIPDVAAVIGYESG DAATRSCISWGYPRFRTHPYVTRVAGLLAQDGIGPANGLLLTRSARA AGAAATYAGLTPEAVFERNGLHGVRLPADGPSAARARAYVQHTGSHL SSREAEDILLDAGLIENRQAEAAVEEAPAEAVRAVLAQAYGVRDAAD VSLHNSGMNAVAAALAAVSGIQRENGRRRWLQLGWIFFDTMSLLEKR VIDVEHTTVPDPFDLSEIARVADAHAGELAGIIAEVPSNPSLRTPDI PALREIATRAGCALVVDATIATPYNVEVLPYADVVCESLTKYATGSA DVLMGAAVVNPASSFGPDLRTELRRHGDEPYHRDTARVAARIRGYAE RMDRVNANALALAACLARHDNVVRGVGWAYDAASRGNYRKVERRPDA PGGLLMVDLKVPLEQVYDRLAVAKGPSFGAEFTMASPQIFIAHYDLL STPEGRAALRARGLHRDMLRISVGTEEPERIVETFEQALRPL*
BesB.1	BesB PROSS1	MSQAVSGTTGSADGLRHIAAGRPVPGSVHSVSVSIPDVASVIGYESN DAATLSRISWGYPRFRPHPYVVRVAELAAREDPAGEPGGALLLTRSA RAARAAAAYAGLPPGAARDLTLGGHVLSGVRLPDRGPAAARARAFVQ HTGGHLSSRQAEDVLWDAGLIDGRQVEETADDSPARAVAQALAGAYG VPGPRYVFLRNSGMNAVYAAVEAVTEIQRDRGRRHWLQLGWIFFDTM HLFEKKVVNVGHTTVPDPFDLAEVARVAAAHAGRLAGIIAEIPSNPG MGVPDLPALREIADRAGCALVVDATIATPHNVDVVPYADVVCESLTK YATGSADVLAGAVVVNPGSPFAADLLTVLPRYGDEPYRRDTARVAAR IRGYAERMVRVNANALALAECLRRHPDVVRDVSWALDTRSAANYRKV ARDSGGPGGLLMVDLRVPLELVYDRLAVAKGPSFGAEFTMASPQVFV AHYDLLTTPRGRAALRARGLHRDMLRVSVGTEPPELIVETFERALRP D*

BesB.3	BesB PROSS3	MSQAVSGTTGSADGLRHIAAGRPVPGSVHSVSVSIPDVASVIGYESN DAATLSRISWGYPRFRPHPYVVRVAELAAREDPAGEPGGALLLTRSA RAARAAAAYAGLPPGAARDLTLGGHVLSGVRLPDRGPAAARARAFVQ HTGGHLSSRQAEDVLWDAGLIDGRQVEETADDSPARAVRQALAGAYG VPGPRYVFLRNSGMNAVYAAIEAVTEIQRDRGRRHWLQLGWIFFDTM HLFEKKVVNVGHTTVPDPFDLAEVARVAAAHAGRLAGIIAEIPSNPL MGVPDLPALREIADRAGCALVVDATIATPHNVDVVPYADVVCESLTK YATGSADVLAGAVVVNPGSPFAADLLTVLPRYGDEPYRRDTARVAAR IRGYAERMRRVNANALALAECLRRHPDVVRDVSWALDTRSAANYRKV ARDSGGPGGLLMVDLRVPLELVYDRLAVAKGPSFGAEFTMASPQVFV AHYDLLTTPRGRAELRARGLHRDMLRVSVGTEPPELIVETFERALRP D*

Appendix 3.2. Information on purified proteins

The UniProt ID or a reference is given for each of the enzymes purified in this work. The extinction coefficient used to determine protein concentration and the storage buffer used during purification are also included.

Protein	UniProt or Reference	A280 (M^{-1} cm^{-1})	Storage Buffer
BsAlaDH	Q08352	21,890	HEPES
BcLeuDH	P0A393	33,810	HEPES
L-HicDH	P14295	21,430	Phosphate
4HBCS	A4YDT1	82,280	MOPS
PhaCSQKT	[16]	104,850	Phosphate