Production of Non-Multiple-of-Five Terpenes Using the Lepidoptera Mevalonate Pathway and ATP Positive Pathway for Five-Carbon Alcohol Production

by

Christopher Brian Eiben

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jay D Keasling, Chair
Professor Tanja Kortemme
Professor Matthew B Francis

Spring 2019

Production of Non-Multiple-of-Five Terpenes Using the Lepidoptera Mevalonate Pathway and ATP Positive Pathway for Five-Carbon Alcohol Production

By Christopher Brian Eiben

Abstract

Production of Non-Multiple-of-Five Terpenes Using the Lepidoptera Mevalonate Pathway and ATP Positive Pathway for Five-Carbon Alcohol Production

by

Christopher Brian Eiben
Doctor of Philosophy in Bioengineering
University of California, Berkeley and University of California, San Francisco
Professor Jay D Keasling, Chair

Metabolism represents at least three and a half billion years of research and development in chemical synthesis and degradation. Human have been taking advantage of this resource for thousands of years, first with the advent of agriculture and the invention of cheese, beer and other fermentation products. More recently selective breeding, metabolic engineering and synthetic biology have given humanity more control over organisms' metabolism, enabling food, chemical and fuel production on massive scales. Given the extensive previous work optimize existing metabolism to make products, the two most exciting frontiers for metabolic engineering are 1) making new chemicals and 2) making old chemicals with new and improved pathways. To explore the area of new chemical production, we have heterologously expressed the lepidopteran mevalonate pathway in *Escherichia coli* to make sixteen carbon homoterpenes. Because the pathway is modular, we demonstrate that making novel homoterpenes requires only altering one gene. As terpenes are an economically valuable class of molecules, we hope that this pathway will enable the exploration of new chemical space difficult to search with traditional organic synthesis techniques. To explore the area of new pathways we designed a pathway to make the five-carbon alcohols prenol and isopentanol. While there are already pathways to five-carbon alcohols, the novel pathway does not require ATP. In fact, if glucose is the carbon source, the pathway is ATP positive when redox balanced

with non-oxidative glycolysis. Because of this it is theoretically possible to ferment these five carbon alcohols the same way microbes ferment ethanol, potentially easing many engineering problems around titer improvements and scaling to industrial production levels. Finally, the challenges of CO_2 emissions will be considered. Many technologies will need to be developed to sustain a prosperous global society. While some of these technologies will be based in traditional engineering fields, biology and metabolism offer unique opportunities and will be critical to enabling a CO_2 neutral civilization in the future.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank my parents for their unending support, patience and understanding, even when I might have been difficult to understand. I dedicate my dissertation to you.

I would like to thank Professor Jay D Keasling for fostering an environment of intellectual freedom and encouraging me to pursue creative and difficult lines of scientific inquiry. His unending optimism and perspective were also great bulwark against the inevitable difficulties of exploratory science.

I would like to thank Professor Tanja Kortemme and Professor Matt Francis for thoughtful input and support on my projects and advice on future endeavors.

And finally, I would like to thank my coauthors and collaborators from who I have learned so much, and who have sped my research progress tremendously. Tristan de Rond, Clayton Bloszies, Jennifer Gin, Jennifer Chiniquy, Edward EK Baidoo, Christopher J Petzold, Nathan J Hillson, Oliver Fiehn and Jay D Keasling were a great help with the production of novel terpenes. Tian Tian, Nurgul Kaplan, Garima Goyal, Jennifer Chiniquy, Nathan J Hillson, Take Soon Lee, and Jay D Keasling were extremely helpful on the pathway for C5 alcohol production.

Chapter 1. Introduction

1.1. Synthetic biology for a growing world

As world population and standards of living increase, new technologies to address world resource constraints will be of increasing importance. Synthetic biology offers the potential to address many of these challenges given that humanity already uses biology on massive scales to produces food, materials, medicines, treat wastewater, and remediate environmental pollution. Even the air that we breathe is of biological origin. Displacement of the world's reliance on petrochemicals and fossil fuels would be a strong step forward in making the world more sustainable and ultimately more prosperous. Metabolic engineering of microbes to produce chemicals from renewable sources (e.g. sugars), has the right properties to scale and make significant impact in reducing society's dependence on petrochemicals.

Since the petrochemical industry has been highly optimized over the last hundred years, near term metabolic engineering efforts should focus on areas which are difficult for petrochemicals to compete such as the production of enantiomerically pure chemicals often found in bioactive molecules. As our abilities to perform metabolic engineering become more sophisticated and rational, higher volume/lower value chemicals and fuels will become more attractive targets for metabolic engineering. For products produced at the largest scales, namely commodity chemicals and fuels, the development of novel pathways that require little to no ATP, release minimal CO_2 and are viable under anaerobic conditions with high titer, rates and yields will be required to displace the current petrochemical incumbents.

1.2. Dissertation organization

In **Chapter 2** I will discuss my work on engineering the lepidopteran mevalonate pathway into *Escherichia coli* to produce "sesquiterpenes" containing sixteen carbons as opposed to the canonical fifteen carbons. Although the exact structure of these compounds remains to be elucidated, given their sesquiterpene nature they contain many carbon-carbon bonds and stereocenters which make them difficult to produce using traditional chemical synthesis based on petrochemicals. I hope that future work will show

these novel terpenes to have useful properties. The most promising area of application for homoterpenes is for generating novel pharmaceutical leads.

In **Chapter 3** I will present my work on the metabolic engineering of a carbon efficient pathway to make five-carbon branched alcohols that is ATP positive from glucose, even when redox balanced with non-oxidative glycolysis, and is thermodynamically feasible. These characteristics should enable anaerobic growth of *E. coli* with the five-carbon alcohol as a waste product much the same as ethanol and lactic acid. Such pathways are needed if biology is to compete with petrochemicals at producing the largest volume chemicals and fuels, which will be necessary to have the biggest impact on reducing CO_2 emissions and reducing resource constraints on society.

In **Chapter 4** I will detail my thoughts around CO_2 emissions and relevant technologies which might make a significant impact to reduce them. Given my synthetic biology training a focus on biological opportunities will be presented.

Chapter 2. Exploring non-canonical terpene space with a heterologous lepidopteran mevalonate pathway

2.1. Abstract

Lepidoptera (butterflies and moths) make the six-carbon compounds homoisopentenyl pyrophosphate (HIPP) and homodimethylallyl pyrophosphate (HDMAPP) that are incorporated into sixteen, seventeen and eighteen carbon farnesyl pyrophosphate (FPP) analogues. In this work we heterologously expressed the lepidopteran modified mevalonate pathway, a propionyl-CoA ligase, and terpene cyclases in *E. coli* to produce several novel terpenes containing sixteen carbons. Changing the terpene cyclase generated different novel terpene product profiles. To further validate the new compounds we confirmed ^{13}C propionate was incorporated, and that the masses and fragmentation patterns were consistent with novel sixteen carbon terpenes by GC-QTOF. Based on the available FPP analogues lepidoptera produce, this approach should greatly expand the reachable biochemical space with applications in areas where terpenes have traditionally found uses.

2.2. Significance statement

Terpenes are a rich source of drugs and pharmaceutical leads. Because terpenes are biosynthesized from five-carbon building blocks, most terpenes are multiples of five carbons which constrains their structural diversity. In this work we developed a modular pathway that leverages a six-carbon building block to produce several new sixteen carbon terpenes. This new six carbon building block is also the first step to enable production of terpenes that are multiples of six carbons or a combination of five- and six-carbon precursors. This strategy significantly expands the reachable terpene space, an area traditional chemical synthesis struggles to traverse, with potential application in generating pharmaceutical leads.

2.3. Introduction

With an estimated 55,000 chemical species, terpenoids are the largest natural products class(1). Terpenoids fulfill many important biological roles including cell wall components (sterols)(2), quinones in the electron transport chain(3), protein trafficking/localization regulators (prenylation)(4), signaling molecules (estrogen)(5),

and bioactive molecules for interspecies competition(6). In addition terpenoids have important economic uses as pharmaceuticals (taxol), (7), flavors (linalool)(8), fragrances (patchoulol)(9), commodity chemicals (isoprene, limonene)(10),(11), and biofuels (bisabolane and isopentanol) (12),(13),(14).

Despite their diversity, terpenoids canonically derive from isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), interconvertible five-carbon metabolites which can combine forming longer prenyl pyrophosphates (Fig. 2-1A). Each IPP ligation yields a molecule five carbons larger, giving the progression of geranyl (ten carbons), farnesyl (fifteen carbons), geranylgeranyl (twenty carbons) pyrophosphates, and so on. Terpene synthases can catalyze the conversion of prenyl pyrophosphates into diverse terpene scaffolds *via* carbocation rearrangements and cyclizations, producing a stunning molecular diversity further magnified by downstream tailoring enzymes. Despite the tailoring, most terpene backbones retain the characteristic five, ten, fifteen, etc. carbons.

When necessary, nature uses several strategies to produce terpene backbones outside the multiple of five carbon sequence. The most common strategies are carbon removal from larger terpene backbones, as in geosmin biosynthesis(15), or attaching another carbon containing molecule, as in bitter acids biosynthesis in *Humulus lupulus L.* (hops)(16). Although uncommon in nature, a few methyltransferases add methyl groups from S-adenosyl methionine to prenyl-phosphate molecules as in 2-methylisoborneol biosynthesis(17). Recently, a paper utilizing a SAM methyltransferase to make the eleven carbon compound HomoGPP has been coupled with monoterpene cyclases to produce novel monoterpenes(18).

Lepidoptera employ an alternate strategy to create endocrine molecules named Juvenile Hormones (JHs) containing terpene backbones of sixteen, seventeen, and eighteen carbons (Fig. 2-1D). Lepidoptera incorporate the six-carbon building blocks homo-IPP (HIPP) and homo-DMAPP (HDMAPP) into JHs(19) through a modified mevalonate pathway. The first lepidopteran mevalonate pathway step condenses propionyl-CoA and

acetyl-CoA (Fig. 2-1B), instead of the canonical two acetyl-CoAs (Fig. 2-1A). The extra methyl from propionyl-CoA is retained through the lepidopteran mevalonate pathway forming FPP analogues with sixteen, seventeen, and eighteen carbons en route to JH biosynthesis (Fig. 2-1B).

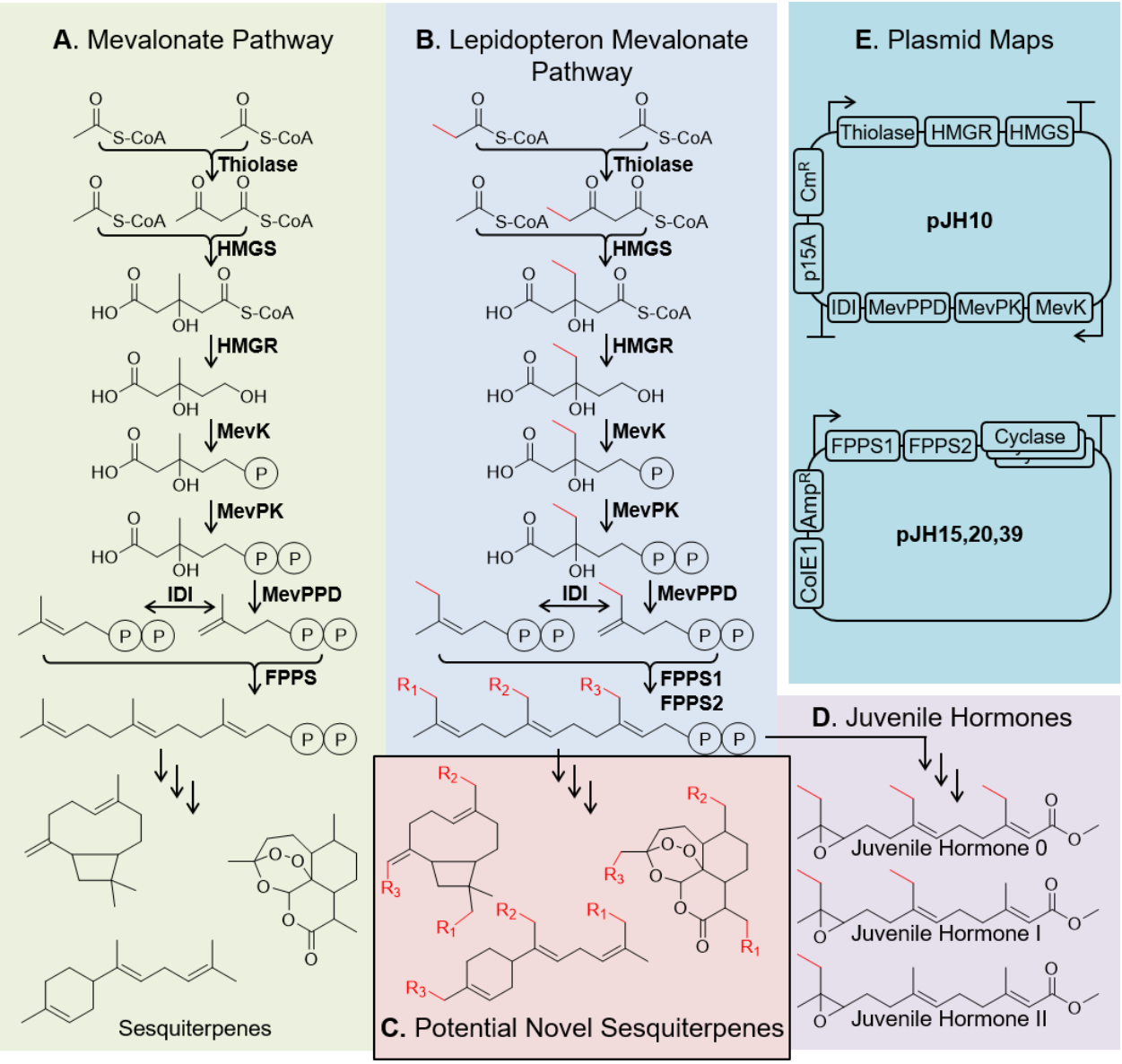

Figure 2-1 The mevalonate pathway and lepidopteran mevalonate pathway. Overview of the canonical mevalonate pathway, the lepidopteran mevalonate pathway, chemical diversity of juvenile hormones and plasmid architecture of pJH10, pJH15, pJH20, and pJH39. **A**. The canonical mevalonate pathway. The pathway starts with the condensation

of two acetyl-CoA and ultimately produces FPP, an essential metabolite. **B**. The lepidopteran mevalonate pathway. Production of FPP analogues for juvenile hormone biosynthesis begin with the condensation of a propionyl-CoA and acetyl-CoA instead of the usual two acetyl-CoA. The extra methyl group is shown in red for emphasis. The core enzymes of the lepidopteran mevalonate pathway also produce FPP via the canonical reactions. **C**. Examples of potential sesquiterpenes that could be derived from the lepidopteran FPP analogs. **D**. Examples of several lepidopteran Juvenile Hormones used as endocrine molecules showcasing some of the different locations the extra methyl can incorporate. **E**. Plasmid maps of pJH10, pJH15, pJH20, and pJH39. pJH15, pJH20 and pJH39 only differ in terpene cyclase. Enzyme abbreviations: [HMGS: 3-hydroxy-3-methylglutaryl-coenzyme A synthase], [HMGR: 3-hydroxy-3-methylglutaryl-coenzyme A reductase], [MevK: mevalonate kinase], [MevPK: mevalonate phosphate kinase], [MevPPD: mevalonate pyrophosphate decarboxylase], [IDI: isopentenyl diphosphate delta isomerase], [FPPS: farnesyl pyrophosphate synthase]. [R_1 = H, Me], [R_2 = H, Me], [R_3 = H, Me]

Several qualities make the lepidopteran mevalonate pathway attractive for exploring novel terpene space that synthetic chemistry struggles to efficiently sample (potential examples Fig. 1C). Relatively few FPP analogues can theoretically produce significant chemical diversity by varying a single enzyme, the terpene cyclase. In previous work terpene cyclases have successfully cyclized chemically synthesized prenyl-phosphate analogues similar to the lepidopteran FPP analogues(20). In addition, the lepidopteran mevalonate pathway has a limited phylogenetic distribution, meaning many potential terpene skeletons remain unsampled. To our knowledge, only three non-JH terpenes likely incorporate HIPP in nature. Though their full biosynthetic pathways remain unknown, 3-methyl-α-himachalene, a *Lutzomyia longipalpis* (sandfly) sex pheromone(21) has sixteen carbons, and homofarnesene and bishomofarnesene, Myrmica Ants trail pheromones(22), have sixteen and seventeen carbons, respectively. For these technical reasons, and the potential medicinal and industrial uses novel terpenes offer, we investigated novel terpene production in *E. coli* using the lepidopteran mevalonate pathway paired with various terpene cyclases.

2.4. Results

2.4.1. Bioprospecting lepidopteran mevalonate pathway

Lepidopteran mevalonate pathway enzyme bioprospecting and plasmid design. Lepidoptera have only one apparent enzyme for each mevalonate pathway step except the thiolase and FPP synthase steps(23). The two FPP synthases and the IPP isomerase from the lepidopteran *Choristoneura fumiferana* (eastern spruce budworm) have been previously expressed recombinantly and purified from *E. coli*, thus we selected them(24),(25). It was noted that combining purified cfFPPS1 and cfFPPS2, thought to form a heterodimer, increased FPP analogue production in previous work(24) so our designs express both enzymes concurrently. For the remaining mevalonate pathway enzymes we selected those predicted by Kinjoh et al. from the lepidopteran *Bombyx mori* (silkworm)(23). Finally we selected the well-characterized terpene cyclase epi-isozizaene synthase(26).

The predicted *Bombyx mori* thiolase did not express in *E. coli*, so we used the well-characterized thiolase PhaA from *Acinetobacter* strain RA3849. Soluble expression of the *Bombyx mori* 3-hydroxy-3-methylglutaryl-coenzyme A (bmHMGR) required truncating the membrane-associated N-terminal domain, similarly required for the *S. cerevisiae* HMGR homologue over expression in other work(27).

We cloned the genes into two plasmids named pJH10 and pJH15 based on plasmids pJBEI-2999(12) and pBbE1a-GPPS-LS(11), respectively. Annotated genbank files for the plasmid are supplied on the JBEI ICE open registry (see Supplementary Information)(28). Briefly, pJH10 expresses the first mevalonate pathway section, from thiolase to isopentenyl-pyrophosphate isomerase, on a medium-copy plasmid (Fig. 2-1E). pJH15 has the two FPP synthases and epi-isozizaene synthase driven by a strong promoter on a high copy plasmid (Fig. 2-1E).

2.4.2. Heterologous pathway expression in *E. coli* and small molecule analysis

We used the *E. coli* expression strain Bap1 because it lacks propionate catabolism and has a strong T7 inducible promoter driving the genomic propionyl-CoA synthetase

(*prpE*)(29). We grew *E. coli* Bap1 cells harboring both pJH10 and pJH15 plasmids in sodium propionate supplemented Terrific Broth medium. After induction we reduced the temperature to 18°C and added a nonane overlay to continuously extract terpenes into the hydrophobic layer. After 36 hours we harvested cell pellets for proteomic analysis and the overlay for Gas Chromatography - Mass Spectrometry (GC-MS) small molecule analysis. Proteomics confirmed protein expression, finding all pJH10 and pJH15 proteins expressed in *E. coli* Bap1 cells (Fig. 2-2) above the no plasmid control. In the nonane overlay several new peaks with m/z = 218 molecular ions were observed, consistent with novel $C_{16}H_{26}$ terpenes (Fig. 2-3A). Replacing pJH15 with the control plasmid pBbE1a-RFP(30) (synonym JBEI-2469), which encodes red fluorescent protein rather than the FPP synthases and a terpene cyclase, abolished the new peaks, as did sodium propionate omission. When feeding ^{13}C propionate, singly-labeled at the carboxylic acid position, a corresponding +1 m/z parent ion mass shift (Fig. 2-3A) was seen in the novel compounds via GC-MS.

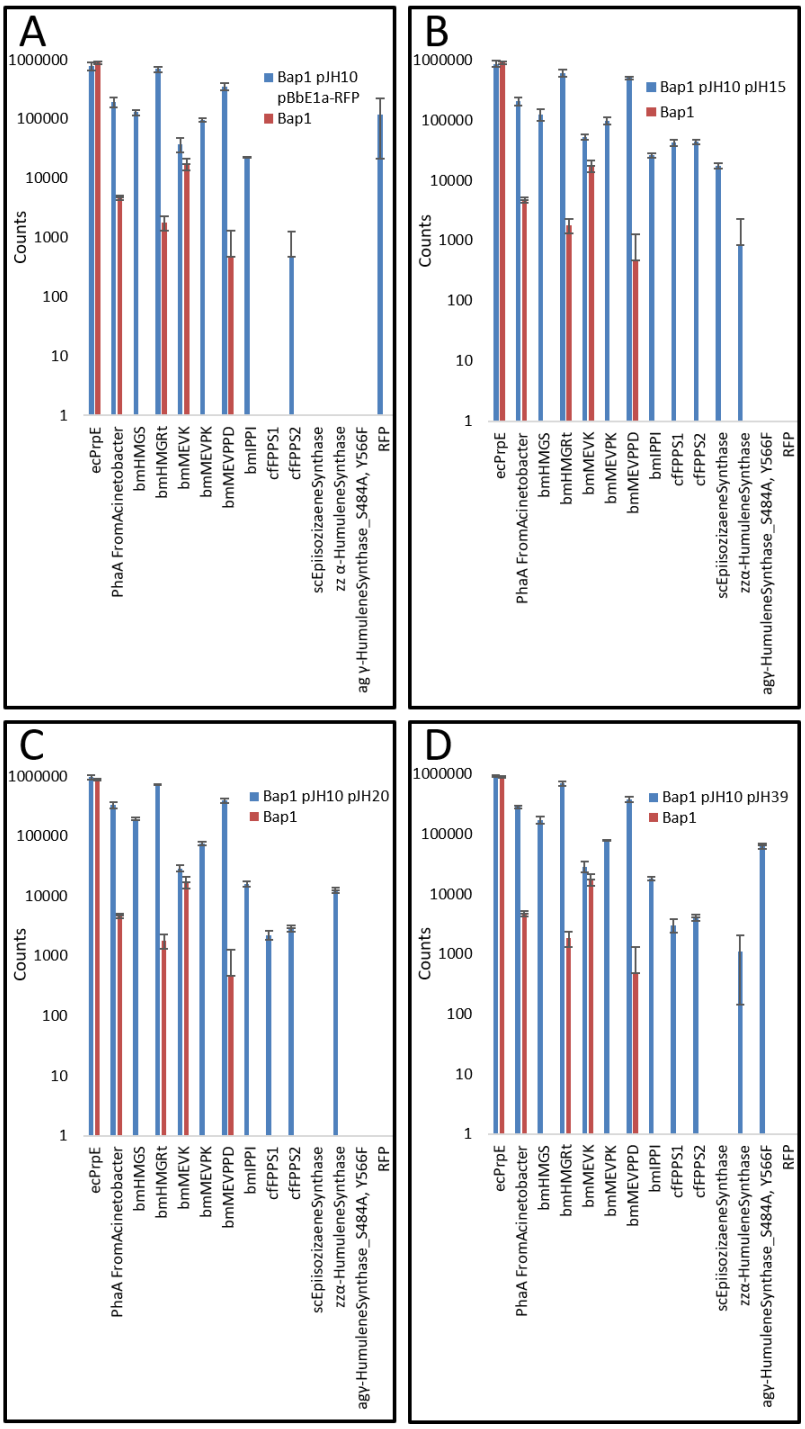

Figure 2-2 Proteomics of lepidopteran mevalonate pathway. Targeted proteomics of the lepidopteran pathway enzymes. Y axis represents counts of selected peptide ions. Colored bars represent the average of 3 biological replicates, while error bars represent

standard deviations. **A.** Targeted proteomics of condition Bap1 pJH10 pBbE1A-RFP. **B.** Targeted proteomics of condition Bap1 pJH10 pJH15. **C.** Targeted proteomics of condition Bap1 pJH10 pJH20. **D.** Targeted proteomics of condition Bap1 pJH10 pJH39. Enzyme abbreviations: [PrpE: Propionate CoA ligase], [Thiolase: PhaA thiolase], [HMGS: 3-Hydroxy-3-Methyl-Glutaryl-coenzyme A Synthase], [HMGRt: N-terminal truncated 3-Hydroxy-3-Methyl-Glutaryl-coenzyme A Reductase], [MevK: Mevalonate Kinase], [MevPK: Mevalonate Phosphate Kinase], [MevPPD: Mevalonate PyroPhosphate Decarboxylase], [IDI: Isopentenyl-diphosphate Delta Isomerase], [FPPS1: Farnesyl-Pyrophosphate Synthase 1], [FPPS2: Farnesyl-Pyrophosphate Synthase 2], [EpiIsoSyn: Epi-isozizaene synthase], [α HumSyn: α -Humulene Synthase], [MBP γ HumSyn: n-terminal Maltose Binding Protein (MBP) fusion to γ -Humulene Synthase mutant S484A, Y566F], [RFP: Red Fluorescent Protein]. Data has been uploaded to ProteomeXchange ID: PXD012697

We then tested if alternate terpene cyclases could also produce $C_{16}H_{26}$ terpenes. We built additional plasmids pJH20 and pJH39 based on pJH15, replacing epi-isozizaene synthase with, respectively, the alternate terpene cyclases α-humulene synthase from *Zingiber zerumbet* (bitter ginger)(31) and γ-humulene synthase mutant S484A, Y566F(32) from *Abies grandis* (grand fir). The γ-humulene synthase mutant was expressed as a maltose binding protein (MBP) fusion to aid solubility (Fig. 2-1E). These terpene cyclases also produced novel peaks compared to the RFP control (Fig. 2-3B, 3C), with each synthase producing a distinct product profile. All the novel peaks elute off the GC column after our sesquiterpene β-caryophyllene standard, consistent with the novel peaks having a higher molecular weight than our fifteen-carbon standard. Although theoretically possible, we could not confirm novel peaks consistent with seventeen or eighteen carbon terpenes.

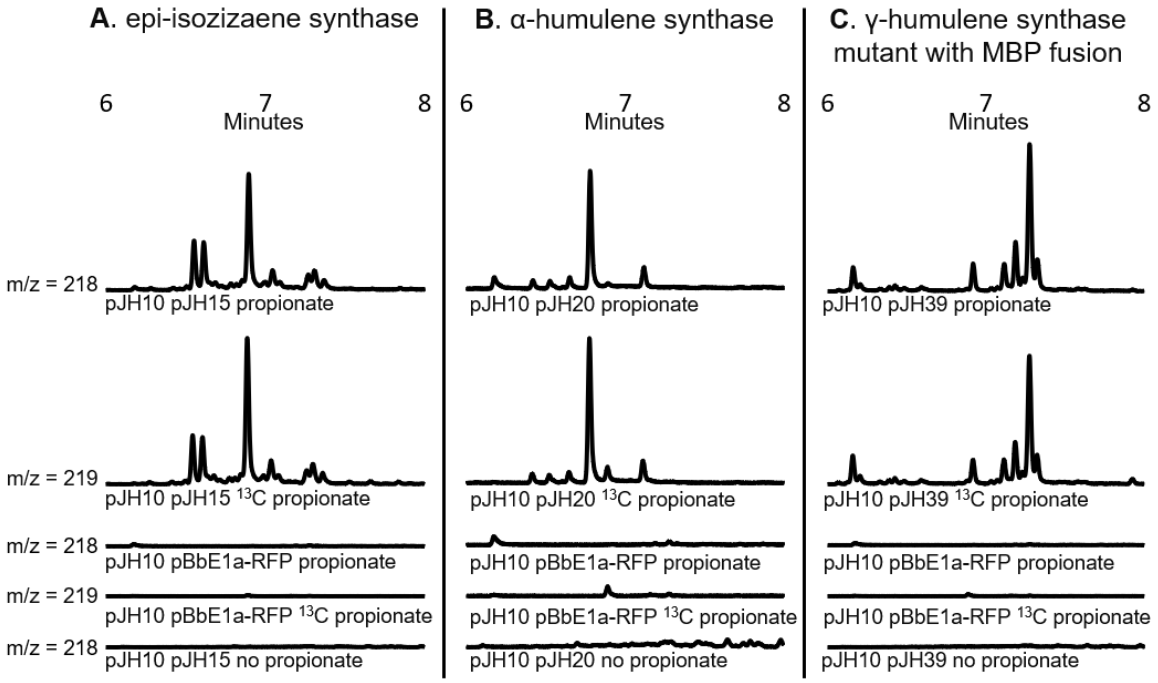

Figure 2-3 Production of sixteen carbon terpenes. GC-MS traces of nonane overlays of Bap1 *E. coli* cells expressing the lepidopteran mevalonate pathway and a terpene cyclase. Plasmid pJH10 contains the core enzymes of the pathway up to but not including the FPP synthases or a terpene cyclase. Plasmid pJH15, pJH20, and pJH39 all express the FPP synthases and a single terpene cyclase, either epi-isozizaene synthase (**A**), *α*-humulene synthase (**B**), or a N-terminal maltose binding protein (MBP) fusion to γ-humulene synthase mutant S484A, Y566F (**C**). Plasmid pBbE1a-RFP expresses only RFP as a negative control. Fermentations were conducted with either no propionate, ^{12}C propionate or singly labeled ^{13}C propionate at 1 g/L. Traces show Selected Ion Monitoring (SIM) of m/z = 218 for no propionate and propionate conditions, while m/z = 219 is used for singly labeled ^{13}C propionate conditions.

To further investigate these novel compounds we scaled up our best producing strain (Bap1 pJH10 pJH15) to 500-mL culture volume with a 50 mL nonane overlay. After expressing per the usual protocol, we rotovapped the nonane down to approximately 100 μL and analyzed with accurate mass GC-QTOF mass spectrometry after tenfold dilution. We confirmed the novel peaks have precursor and fragment ions consistent with $C_{16}H_{26}$ terpenes. The heaviest predicted ion has a m/z = 218.2029 (Fig. 3A). The 203.1794 m/z

fragments correspond to sixteen carbon terpenes with loss of a methyl group. Similarly, the 189.1638 m/z fragments corresponds to an ethyl group loss and 175.1481 m/z to a propyl group loss (Fig. 2-3B).

The mass error of the novel terpenes parent ion on the GC-QTOF were [peak A 3.2 ppm], [peak B 1.37 ppm], and [peak C -.92 ppm] with an average error of 1.83 ppm over the three. We estimated the mass error across the observed fragmentation patterns against the theoretical fragmentations from m/z = 218.2029, 203.1794, 189.1638 down to m/z 91.054 (loss of nine carbons). The average mass error was found to be [peak A 2.11 ppm], [peak B 1.28 ppm] and [peak C 1.05 ppm], with an average of 1.48 ppm between the three.

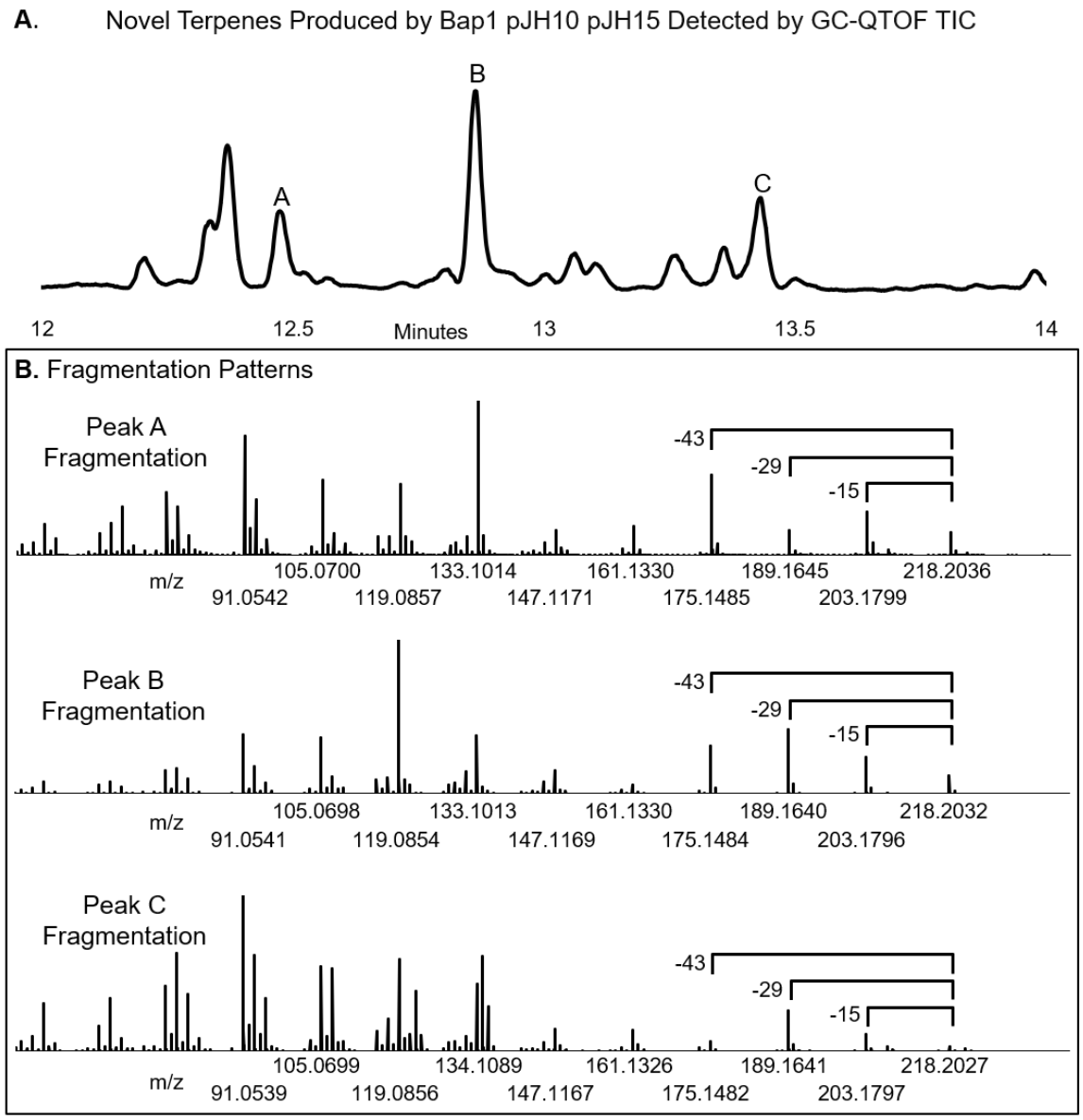

Figure 2-4 GC-QTOF confirms sixteen carbon terpenes. GC-QTOF MS assays of novel terpenes produced by Bap1 pJH10 pJH15 cells. **A**. Total ion chromatogram of concentrated nonane overlay of Bap1 E. coli cells expressing the lepidopteran mevalonate pathway and an epi-isozizaene synthase from plasmids pJH10, pJH15. **B**. Fragmentation patterns of selected chromatographic peaks. Fragmentation patterns were averaged over the whole chromatographic peak. Ions consistent with C16H26 fragmentation have predicted m/z of 218.2029, 203.1794, 189.1638, 175.1481 corresponding to the loss of an

electron, a methyl group, an ethyl group and a propyl group respectively. Fragmentation patterns for the other chromatographic peaks can be found in the Dataset S1.

2.4.3. Titer estimation

Estimating the novel compound titers is difficult because we cannot buy standards, nor have we successfully purified the novel compounds to homogeneity. A rough estimate can be made using β-caryophyllene as a standard, but we stress the error in this crude estimate is hard to know *a priori*. Given this caveat, we estimate that the combined $C_{16}H_{26}$ fraction for pJH15, pJH20, pPJH39 per L of cell culture to be .684 mg +/- .143 mg (SD), .081 mg +/- .007 mg, and .398 mg +/- .042 mg.

2.4.4. Production of homoterpenes from valerate

The most common compound we produced with plasmids pJH10 and pJH15 was still the fifteen-carbon sesquiterpene epi-isozizaene (Fig. 2-5A). This is perhaps unsurprising because *E. coli* has a relatively high acetyl-CoA concentration and chemically the thiolase is trying to distinguish only a one carbon difference between condensing two acetyl-CoA and condensing an acetyl-CoA and a propionyl-CoA. To get around this problem we devised an alternate strategy to create 3-ketovaleryl-CoA from valerate via beta oxidation. By using an *E. coli* strain that lacks appreciable endogenous thiolase activity, as JW5578 from the Keio collection (33) which lacks FadA, the main thiolase gene in *E. coli* for fatty acid degradation, we should be able to reduce the acetoacetyl-CoA to 3-ketovaleryl-CoA ratio in the cell, thus hopefully increase the ratio of our desired sixteen carbon terpenes to fifteen carbon terpenes.

For this strategy we designed two plasmids, pJH53 and pJH55. pJH53 is based on pJH10, but with the thiolase PhaA replaced by a hexanoyl-CoA ligase from *Cannabis sativa* codon optimized for *Saccharomyces cerevisiae* (the DNA template had been used in yeast for an alternate project (34)). pJH55 is based on pJH15, but with an additional operon for beta-oxidation added from the plasmid pEG1780 (previously unpublished work by Dr. Ee-Been Goh, Ph.D.). The beta-oxidation operon is driven by the IPTG inducible pLacUV5 promoter and contains FadB from *E. coli* and an acyl-CoA oxidase (ACO) from *Micrococcus luteus* codon optimized for *E. coli*.

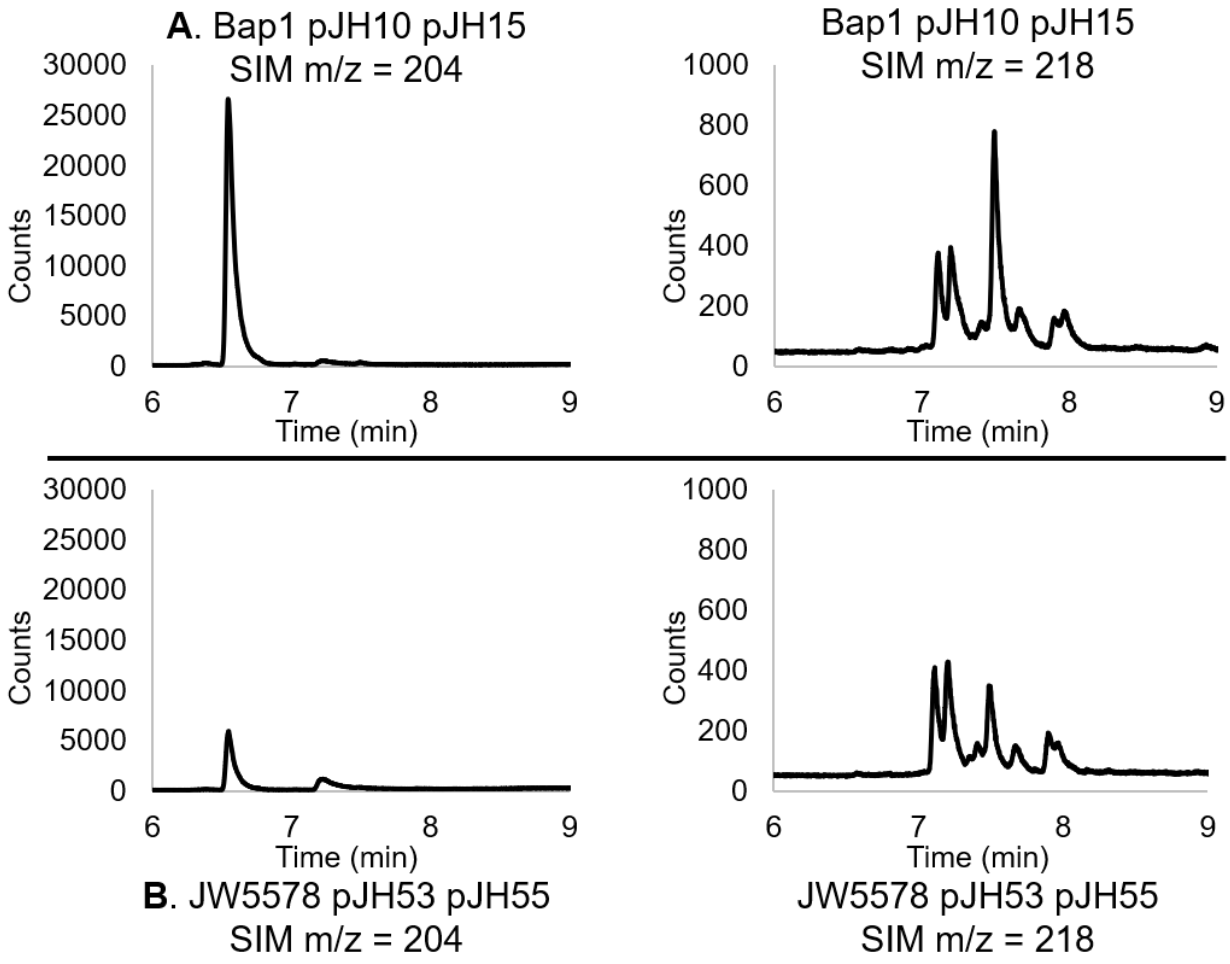

Figure 2-5 Improved ratio of sixteen carbon terpenes compared to fifteen carbon terpenes with valerate pathway. A. Plasmids pJH10 and pJH15 expressed in *E. coli* Bap1 cells with 1 g/L sodium propionate. **B**. Plasmids pJH53 and pJH55 expressed in the *E. coli* strain JW5578 fed with 1 g/L sodium valerate.

Plasmids pJH53 and pJH55 we were able to reduce the production of epi-isozizaene compared to the production of the sixteen carbon terpenes as compared to plasmids pJH10 and pJH15 (Fig. 2-5). However, with the pJH53 and pJH55 system we also saw a decrease in the production of one of the sixteen carbon terpenes compared to the other sixteen carbon terpenes. This is surprising because if the precursor FPP analogue in the pJH53/pJH55 system is at a lower or higher concentration compared to the pJH10/pJH15 system, one would expect all homoterpenes to be decreased or increased equally. It is

possible that there are two FPP analogues the FPP synthases make, each with the extra methyl at a different position. If this is the case, perhaps changing the HIPP and IPP ratio favors the production of one FPP analogue over the other, thus causing the change in abundance between the sixteen carbon terpene species. This is a rather speculative hypothesis, and further study will be required to fully understand this surprising result.

2.4.5. Production of seventeen carbon terpenes

In addition to sixteen carbon terpenes, we sporadically see chromatographic peaks which are consistent with seventeen carbon terpenes ($C_{17}H_{26}$) when doing the small scale homoterpene expression. It is likely the intermittent detection of these compounds is related to instrument sensitivity, in which the compounds are close to the limit of detection, and previous analytical protocols which involved a one to ten dilution step in ethyl acetate before sample analysis on GC-MS. We see the same peaks when feeding propionate and singly labeled ^{13}C propionate with the expected m/z shift of +2 in the case of two singly labeled ^{13}C propionate incorporation into these new compounds (Fig. 2-5). In addition, different cyclases produce different bishomoterpene product profiles as expected (Fig. 2-5A vs Fig. 2-5B). With improved methods and time it is likely we could get exact masses for these compounds using GC-QTOF instruments as we did for the homoterpenes.

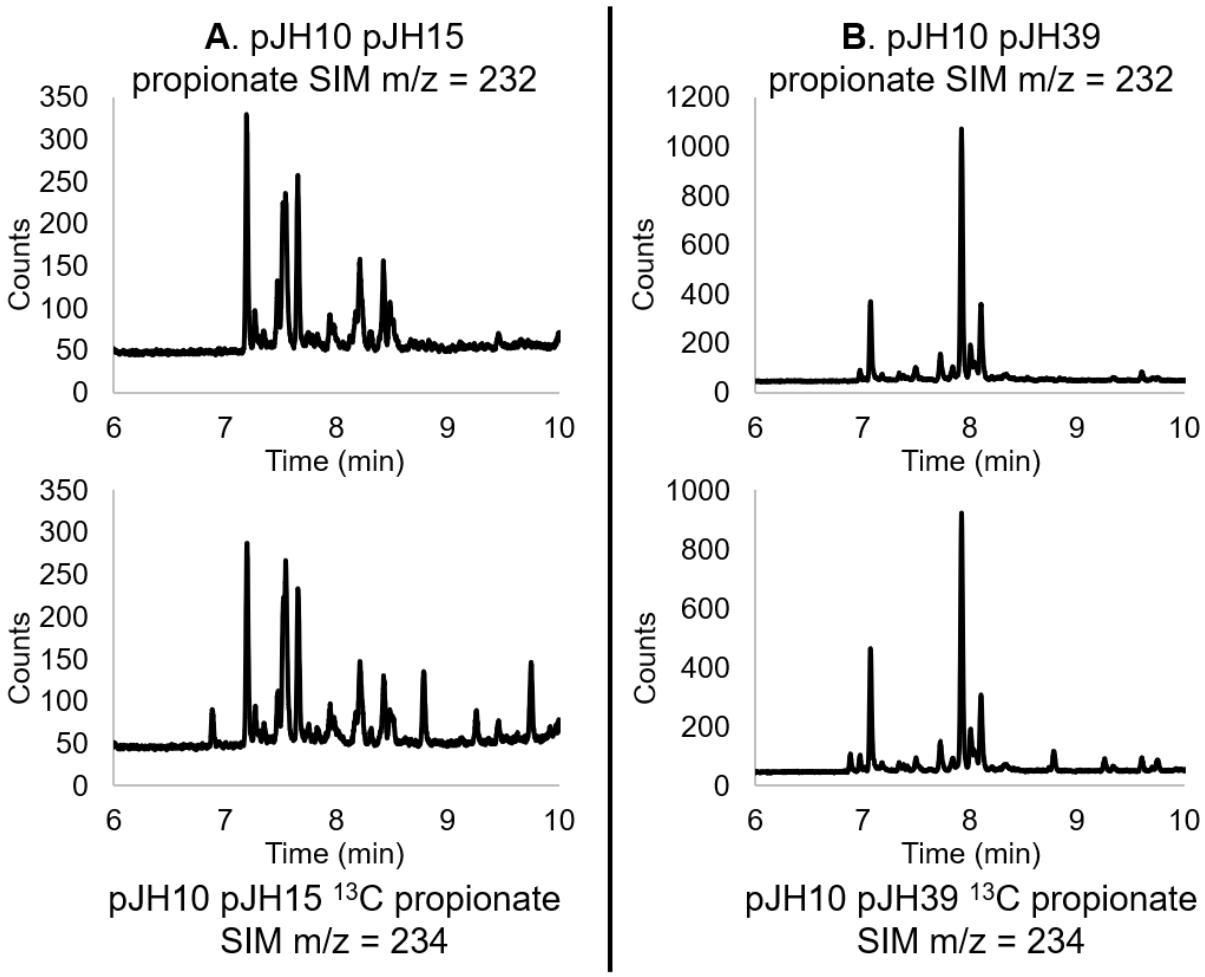

Figure 2-6 Production of seventeen carbon terpenes. GC-MS traces of nonane overlays of *E. coli* Bap1 cells expressing the lepidopteran mevalonate pathway and a terpene cyclase. Fermentations were conducted with either no propionate, ^{12}C propionate or singly labeled ^{13}C propionate at 1 g/L. **A**. Plasmids pJH10 and pJH15 (contains epi-isozizaene synthase). **B**. Plasmids pJH10 and pJH39 (contains N-terminal maltose binding protein (MBP) fusion to γ-humulene synthase mutant S484A, Y566F). (**A**), *α*-humulene synthase (**B**), or a N-terminal maltose binding protein (MBP) fusion to γ-humulene synthase mutant S484A, Y566F.

2.5. Discussion

Although terpenoid compounds containing non-multiple of five carbons are promising for pharmaceuticals, flavors, fragrances, commodity chemicals, and fuels, historically, efficiently sampling this space has been difficult. In this paper we developed a

homosesquiterpene diversity platform that can explore this novel chemical space. The most immediate application is sesquiterpene drug diversification. For instance, emerging resistance to the sesquiterpene antimalarial drug artemisinin will require new drug development to combat and underscores the importance of methods like ours which generate new analogues and eventually drug leads(35).

Several methods confirmed the novel chemicals we produced here as homosesquiterpens. The chromatographic peaks incorporated ^{13}C propionate (parent ion +1 m/z shift compared to ^{12}C propionate), revealing the extra carbon originates from propionate as predicted. Unsurprisingly, propionate omission abolished the novel chromatographic peaks, as did replacing the terpene cyclase plasmid with an RFP expressing plasmid. Replacing epi-isozizaene synthase with alternate terpene cyclases generated novel product profiles demonstrating this pathway can potentially generate large chemical diversity with predictable and minimal alterations.

We initially tried purifying the novel terpenes from the best producing strain (Bap1 pJH10 pJH15) for nuclear magnetic resonance (NMR) analysis but separating them from the nonane overlay and each other proved technically challenging. Instead we used GC-QTOF to validate the new molecules' chemical formulae, albeit without the atomic spatial configuration. We observed the parent ions (average error of 1.83 ppm) and the fragmentation ions (average error of 1.48) in great agreement with the predicted m/z. This confirmed the novel molecules exact composition as $C_{16}H_{26}$ and thus homosesquiterpenes.

We found that the pathway can accept dissimilar terpene cyclases and still produce novel compounds, a beneficial property for terpene cyclase library screening. Epi-isozizaene synthase and α-humulene synthase canonically produce very narrow product profiles(26),(36), while γ-humulene synthase mutant S484A, Y566F has a much broader product profile(32). The three terpene cyclases also catalyze various degrees of FPP cyclization, with epi-isozizaene synthase producing a tricyclic compound, γ-humulene synthase mutant S484A, Y566F primarily making a variety of tricyclic compounds, and

α-humulene synthase primarily making a monocyclic compound. Despite these differences in fifteen carbon product profile, all made many sixteen carbon products. Given Juvenile Hormone natural diversity (Fig. 2-1D)(37), extending the strategy to seventeen and eighteen carbon terpenes should be possible, although we did not confirm their production here. To our knowledge this work is the first report using the lepidopteran mevalonate to heterologously produce novel sixteen carbon terpenes, unlocking new space in the economically important terpenes compound class.

2.6. Materials and methods

2.6.1. Chemical sources

Chemicals were ordered as follows: nonane and sodium propionate from Alfa Aesar; DpnI and T4 DNA ligase (5 Weiss U/μL) from Thermo Scientific; Molecular grade BSA, type IIS restriction enzyme BsaI and Gibson Assembly Master Mix from New England BioLabs; TB dry powder from EMD Millipore; Isopropyl β-D-1-thiogalactopyranoside (IPTG) from Fisher Scientific; Glucose from VWR Analytical; all other chemical reagents, including ^{13}C sodium propionate, from Sigma Aldrich. Qiagen QIAquick Gel Extraction Kits/PCR Purification kits, Zymo Research DNA Clean and Concentrator Kits and Millipore Nitrocellulose 0.025μM VSWP dialysis membranes were used for DNA manipulations.

2.6.2. DNA synthesis and sources

DNA encoding bmHMGS, bmHMGR, bmMevK, bmMevPK, bmMevPPD, bmIPPI, and cfFPPS1 was codon optimized for *E. coli*, synthesized and cloned by Genscript into pET29b+ vector between the NdeI and XhoI cut sites harboring a TEV protease sequence before a C-terminal 6xHis tag. The amino acid sequences for "bm" constructs originate from *Bombyx mori* (silkworm), and those for "cf" from *Choristoneura fumiferana* (eastern spruce budworm). We sourced the thiolase phaA from *Acinetobacter* strain RA3849 plasmid pMMB-pha(38). Epi-Isozizaene synthase from *Streptomyces coelicolor* A3(2) was codon optimized using GeneDesign(39) and synthesized by IDT as a gBlock. cfFPPS2 native sequence was ordered through IDT as a gBlock(40).

2.6.3. PCR and plasmid assemblies

2.6.4. Primers for DNA amplification and construction were designed using DeviceEditor(41) or j5(42) and are listed in Dataset S1. Touchdown PCR was used to amplify DNA fragments for plasmid construction of pET28cfFPPS2, pET28scEpiisozizaene, pET29bmHMGRt, pJH10, pJH15, pJH20, and pJH39. iProof was used as the polymerase using the manufacturer's recommended buffer conditions with the addition of 5% dimethyl sulfoxide (DMSO). PCR reaction cycle was as follows: 98°C incubation for 30 seconds for one cycle. 10 cycles of 98°C for 10 seconds, annealing temperature (various, see Dataset S1) for 15 seconds, decreasing .5°C each cycle, then an extension time of 22.5 seconds per kilobase at 72°C. Then 25 cycles starting with a 98°C for 10 seconds, a constant annealing temperature 5°C lower than the initial annealing temperature for 15 seconds then an extension time of 22.5 seconds per kb at 72°C. One final extension step of 10 min at 72°C was included at the end of the cycle before the samples were chilled to 4°C.

PCR products were purified via a QIAquick PCR Purification kit, eluted with supplied EB buffer, DpnI digested at 37°C for 1 hour, followed by gel electrophoresis on a .5% agarose gel using GelGreen nucleotide stain. Gel imaging was done with an Invitrogen Safe Imager. DNA bands of the correct size were excised and purified via a QIAquick Gel Extraction kit, then concentrated to 10 μL using a Zymo Research DNA Clean and Concentrator kit. DNA concentration was measured using a NanoDrop at 260 nm.

Gibson Assembly was used to construct plasmids pET28cfFPPS2, pET28scEpiisozizaene and, pET29bmHMGRt(43) using Gibson Assembly Master Mix purchased from New England Biolabs. Equal molar amounts of inserts and backbone were added and incubated at 50°C for 1 hour. 1 μL of reaction mixture was then electroporated using Electromax DH10B and plated on appropriate antibiotic. Compared to the full length sequence of bmHMGR, the first 390 amino acids of bmHMGRt were truncated, yielding MKKRPMV... as the translational start of the protein based on structural intuition from the human HMGR1 crystal structure 2Q6B(44).

pJH10 and pJH15 were constructed using Golden Gate Assembly(45) using the BsaI enzyme. Assembly reaction condition consisted of .33 μL BsaI, .33 μL T4 DNA ligase, 0.05 μL BSA, .5 μL 10x T4 ligase buffer, 33 ng of backbone, and the other assembly pieces were added in equimolar concentration to the backbone. Water was then added to 5 μL total volume.

The assembly mixture used a thermocycling protocol of 45°C for 2 minutes, 16°C for 5 minutes, cycled 25 times, followed by 50°C for 5 minutes, 80°C for 5 minutes, and 4°C until samples were recovered from the PCR blocks. pJH15 was used as the template to produce pJH20 and pJH39 via Golden Gate Assembly. The epi-isozizaene synthase from pJH15 was replaced with the terpene cyclases α-humulene synthase from *Zingiber zerumbet* (bitter ginger) from plasmid pTrcHumulene (expressing ZSS1) for pJH20(31). A N-terminal Maltose Binding Protein (MBP) fusion from pET28a-MBP (previously unpublished plasmid made by Dr. Andrew Hagen, Ph.D.) was cloned in front of a γ-humulene synthase mutant S484A, Y566F from *Abies grandis* (grand fir) from plasmid pTrcHUM_Y566F_484A (AYG from table 1)(32) for pJH39. Plasmids will be available upon request from authors or addgene. Primer sequences are listed in table 2-1, and assembly notes in table 2-2.

Primer Name	Sequence
CBE123	CCACCACCACTGATGAGATCCGGCTGCTAACAA
CBE124	CATATGTATATCTCCTTCTTAAAGTTAAACA
CBE126	CAGCCGGATCTCATCAGTGGTGGTGGTGGTGGTGC
CBE128	GTTTAACTTTAAGAAGGAGATATACATATGAAAAAGCGCCCAATGGTTG
CBE205	CGGCTGCTAACAAATGAGATCCGGCTGCTAACAAAGCCCG
CBE206	CCAGTCATGCTAGCCATATGGCTGCCGCGCGGC
CBE207	CGGCAGCCATATGGCTAGCATGACTGGTGGAC
CBE208	GCAGCCGGATCTCATTTGTTAGCAGCCGGATCTC
CBE222	TGTTCTTTCCTGCGTTATCCCC
CBE223	CGTAAGGAGAAAATACCGCATCAG
CBE227	AGGTGAGAAATGACTCGAGCACCACCACCAC
CBE228	TGGCTGCTGCCCATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTT
CBE229	AAGGAGATATACCATGGGCAGCAGCCATCATCA
CBE230	TGGTGGTGCTCGAGTCATTTCTCACCTGCCGCTTCGT
CBE236	CACACCAGGTCTCAATTTTGCTTTCCTCCTTCATCAGTCACGTTCAACTGCAAGTGC
CBE237	CACACCAGGTCTCAAAATGGGCGGCAAAGTTGAAAACGTGG

CBE238	CACACCAGGTCTCAATGATGTTTTCCTCCTACTAGTTAGTGTTTTCTATCATAGGTGCGTCTACGTTGG
CBE244	CACACCAGGTCTCAGCTCATATGATCTCCCTCCTAGATCCTTACATTTCGATTGACACGCCCGGGC
CBE245	CACACCAGGTCTCAGAGCCCGAAGATTATTTTACTGTTCTCCGG
CBE246	CACACCAGGTCTCAGACATCTCATAATCCTCCTAGATCCTCACAAGTTAGAAACCAATTTCAAGATACTGTCC
CBE247	CACACCAGGTCTCATGTCAAACATTGTCACCGTTATCGCACC
CBE250	CACACCAGGTCTCACCTTATTTTTCCATAAACTTATGAATCTTTTCAGGTTC
CBE251	CACACCAGGTCTCAAAGGATCCAAACTCGAGTAAGGATCTCC
CBE255	CACACCAGGTCTCATACATGTTCTCAACTACCCGCTCCCTGG
CBE256	CACACCAGGTCTCAATCTCATAATCCTCCTAGATCCTCAGTGGTGGTGGTGGTGGTGC
CBE257	CACACCAGGTCTCAAGATGGGCAGCAGCCATCATCATCATCATCACAGC
CBE258	CACACCAGGTCTCAATTATCATTACCTCCTAGATCCTTATCAATGATTCCTCTTATAAATCATATGAATAAGCCG
CBE261	CACACCAGGTCTCAATAGGGATCCAAACTCGAGTAAGG
CBE262	CACACCAGGTCTCATGTATATCTCCTTCTTAAAAGATCTTTTGAATTCTG
CBE263	CACACCAGGTCTCATAATGGGCAGCAGCCATCATCA
CBE264	CACACCAGGTCTCACTATCATTTCTCACCTGCCGCTTCGT
CBE307	CACACCAGGTCTCAAGAAGGAGATATACATATGAAAGATGTTGTGATTGTTGCAGC
CBE309	CACACCAGGTCTCATCATGAAAAAGCGCCCAATGGTTG
CBE310	CACACCAGGTCTCATTGGCGGTTTGGACGTTCAGC
CBE311	CACACCAGGTCTCACCAATTGATGAGGATCTTTTTAAGGATCTCCAGGC
CBE312	CACACCAGGTCTCAGGTTATTTCCTCCTAGATCCTGAAATTGTTATCCGC
CBE313	CACACCAGGTCTCAAACCATGTTAACCGCAGTCGATAATTTGTTTACCG
CBE314	CACACCAGGTCTCATTACCTCCTAGATCCTTAGATTTTGTTAATTTCGCTAACAGGCAGACC
CBE315	CACACCAGGTCTCAGTAATGATAATGTTGGCCCGTCATTTGTCCC
CBE316	CACACCAGGTCTCATTCTTAAAAGATCTTTTGAATTCTGAAATTGTTATC
CBE373	CACACCAGGTCTCAATACATGGAGCGTCAATCGATGGCACTGG
CBE374	CACACCAGGTCTCATAGATCAAGAAGCTCTCGACAAAAATCAAAGAAACG
CBE375	CACACCAGGTCTCATCTAGGGATCCAAACTCGAGTAAGG
CBE382	CACACCAGGTCTCAGTATATCTCCTTCTTAAAAGATCCTTATCAATGATTCCTCTTATAAATCATATGAATAAGCCG
CBE405	CACACCAGGTCTCAGACCCACGCTCACCGGCTCC
CBE406	CACACCAGGTCTCAGGTCCCGCGGTATCATTGCAGCACTGG
CBE419	CACACCAGGTCTCACCCACGCTCACCGGCTCCAGA
CBE420	CACACCAGGTCTCACTTTATCAATGATTCCTCTTATAAATCATATGAA
CBE421	CACACCAGGTCTCAAAAGGAGGATAAAGAAATGAAAATCGAAGAAGGTAAACTGGTAATCTGG
CBE422	CACACCAGGTCTCACGCCGCTGCTAGTCTGCGC
CBE425	CACACCAGGTCTCAGAGGATCCAAACTCGAGTAAGGATCTCCAGGC
CBE426	CACACCAGGTCTCATGGGTCCCGCGGTATCATTGCAGCACTGGGG
CBE427	CACACCAGGTCTCAGGCGCTCAAATCAGCGAATCAGTGTCTCCA
CBE428	CACACCAGGTCTCACCTCATTATATAGGAACCGGAACGATTAGAACTTTCTGGACC

CBE209	GCTAGCATGACTGGTGGACAGCAAATGGGTCGCGGATCCAGGCTAAACAAGTGCTTGCGTCTTGCCT CTCAATGCAGAAATGTAAATTTAAAATCACTTTTACACACCAATGTGAATTTCAAAATAAATTACAA AACTTCAAATATCCGGTGTGTACAACAATTCGATTTTTCCAAGATGTCAACGAAAGCATCACCAACT CAGACTAAGAAAGAATCCTTTGAAGATGTGCTTCCTAGTATTCTGAACACAATCACAACCAACAGTG AGCTTACAGAAGTACCTGAAGTTGCAAATTGGCTGAAAAAGGTACTTGAATACAATTTAGCAGGTGG CAAAAAAGCGAGAGGACTGACGACATTGTTTGCGTATGAAATGCTAGAGAAGCCAGAAAATATAACC GAGGAAACTATTTATTTGGCGAAAACGCTGGGATGGTGTGTAGAAATATTACAAGGATTCCTAGTAA TGCTAGATGATATAATGGACGGGTCGACAACCCGGCGTGGGGTGCCCTGCTGGTACCAACTCCCTGA AGTGGGATTAGCCGCTGTAAACGACTCCTCACTCATGTTCTCGTCGATATTCTACGTGCTACACGCG CATTTTGCCGATAAGAAGATTTATACCAACTTAGTGGAACTCTTTAACGAGAGTCTCATGCACACTT CTATCGGGCAACACTTAGACGTTACAATGGAGCGCCGTCAGAAAAGCGACTACAGTCTGTTCACGAT AGAGCGGTACAACGCTATCGTCAAGTACAAGACTGCTTATTACACGTACCAGCTCCCCGTCTGTCTT GGCATGCTGCTTGCCAATATATCTGATCCTGTGTTGCATCAGAAGGCTGAAGACATGTGCCTCGAAA TAGGAAAATTCTTTCAAATTCAGGACGACTACATAGACTGTTACGGCGACGAGTCACTAACAGGTAA AATGGGTACAGACATACAAGAAGCGAAGTGCTCTTGGCTAGCTGTAATGGCTTTGCAGCGGTGTAGC GCCTCCCAAAAAATCGTCTTCACGACTTGCTACGGTAGTAAAGAGCCAGCGCACATCGAGCGTATTA AGGAATTATACAAGCAACTTCAGCTGCCTGAACTGTATGCTCAAGAGGAAACAAGGATGTACGAGTC GTTGATAAAGCAGGCTCACGGCTTGCCTAGCGAACTGTCACCGGCGTTGTTCGTACGGCTTATTCAT ATGATTTATAAGAGGAATCATTGAGATCCGGCTGCTAACAAA
CBE226	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGTGC ATGCTTTCCCACACGGCACCACAGCGACACCGACCGCGATCGCAGTACCGCCATCGCTCCGTCTCCC GGTGATCGAGGCGGCATTTCCCCGGCAACTGCACCCGTATTGGCCGAAGCTCCAGGAGACAACCCGG ACCTGGCTGCTCGAAAAGCGGCTCATGCCGGCGGACAAGGTCGAGGAATATGCCGACGGTCTGTGCT ACACGGACCTCATGGCGGGCTACTACCTGGGCGCCCCCGACGAGGTCCTGCAGGCGATAGCGGACTA CAGCGCGTGGTTCTTCGTCTGGGACGACCGTCACGACCGTGACATCGTCCACGGCCGGGCCGGTGCC TGGCGGCGACTGCGCGGCCTGCTGCACACGGCACTGGACTCACCCGGGGACCACCTGCACCACGAGG ACACGCTGGTTGCCGGGTTCGCCGACAGTGTGCGGCGGCTCTACGCCTTCCTGCCGGCCACCTGGAA CGCCCGGTTCGCCCGGCACTTCCACACGGTGATCGAGGCGTACGACCGTGAATTCCACAACCGCACG CGCGGAATCGTGCCCGGCGTGGAGGAGTACCTCGAACTGCGCCGGCTCACGTTCGCGCACTGGATCT GGACCGACCTGCTGGAGCCGAGTTCGGGCTGCGAACTCCCCGACGCCGTACGGAAACACCCGGCATA TCGCCGGGCGGCGTTGCTGAGTCAGGAATTCGCCGCCTGGTACAACGACCTCTGCTCACTGCCAAAG GAAATAGCGGGAGACGAGGTCCACAATCTTGGCATCAGTCTCATCACCCACCATTCACTCACCCTGG AAGAAGCCATCGGAGAAGTCCGGCGTCGTGTCGAGGAATGCATCACGGAATTCCTCGCCGTCGAGCG GGACGCCTTACGGTTCGCCGACGAACTCGCCGACGGAACCGTACGCGGAAAGGAACTGAGCGGCGCC GTGCGGGCCAACGTCGGCAATATGCGGAACTGGTTCAGTTCCGTCTACTGGTTCCACCACGAGTCCG GCCGGTACATGGTCGACAGCTGGGACGACCGGTCCACGCCCCCGTACGTCAACAACGAAGCGGCAGG TGAGAAATGA

Table 2-1 Primer list for plasmid construction. The last two sequences (CBE209, CBE226) were used as templates for plasmid construction.

For Assembly	Template	Template Part ID	Forward Primer	Reverse Primer	Anneal Temp Start	Assembly method
pET28cfFPPS2	pET28B+	JPUB_011969	CBE205	CBE223	72 °C	Gibson
pET28cfFPPS2	pET28B+	JPUB_011969	CBE222	CBE206	72 °C	Gibson
pET28cfFPPS2	CBE209	none	CBE207	CBE208	72 °C	Gibson
pET28scEpiisozizaene	pET28B+	JPUB_011969	CBE227	CBE228	69 °C	Gibson
pET28scEpiisozizaene	CBE226	none	CBE229	CBE230	69 °C	Gibson
pET29bmHMGRt	pET29B+	JPUB_011960	CBE123	CBE124	69 °C	Gibson
pET29bmHMGRt	pET29bmHMGR	JPUB_011971	CBE126	CBE128	63 °C	Gibson

pJH10	pBbA5c-MevT(CO)-T1-MBIS(CO, ispA)	JPUB_008473	CBE251	CBE316	66 °C	GG
pJH10	pMMB-PHA	JPUB_012020	CBE307	CBE236	72 °C	GG
pJH10	pET29bmHMGS	JPUB_011989	CBE237	CBE238	72 °C	GG
pJH10	pET29bmHMGRt	JPUB_011987	CBE309	CBE310	72 °C	GG
pJH10	pBbA5c-MevT(CO)-T1-MBIS(CO, ispA)	JPUB_008473	CBE311	CBE312	72 °C	GG
pJH10	pET29bmMEVK	JPUB_011985	CBE313	CBE244	72 °C	GG
pJH10	pET29bmMEVPK	JPUB_011983	CBE245	CBE246	72 °C	GG
pJH10	pET29bmMEVPPD	JPUB_011981	CBE247	CBE314	72 °C	GG
pJH10	pET29bmIPPI	JPUB_011979	CBE315	CBE250	72 °C	GG
pJH15	pBbE1a-GPPS-LS	JPUB_004929	CBE261	CBE262	66 °C	GG
pJH15	pET29cfFPPS1	JPUB_011977	CBE255	CBE256	72 °C	GG
pJH15	pET28cfFPPS2	JPUB_011975	CBE257	CBE258	72 °C	GG
pJH15	pET28scEpiisozizaene	JPUB_011973	CBE263	CBE264	72 °C	GG
pJH20	pJH15	JPUB_011964	CBE405	CBE382	72 °C	GG
pJH20	pTrcHumulene	JPUB_011959	CBE373	CBE374	72 °C	GG
pJH20	pBbE1a-GPPS-LS	JPUB_004929	CBE375	CBE406	72 °C	GG
pJH39	pJH15	JPUB_011964	CBE425	CBE426	69 °C	GG
pJH39	pJH15	JPUB_011964	CBE419	CBE420	69 °C	GG
pJH39	pET28a-MBP	JPUB_007093	CBE421	CBE422	69 °C	GG
pJH39	pTrcHUM_Y566F_484A	JPUB_011957	CBE427	CBE428	69 °C	GG

Table 2-2 Plasmid construction conditions. GG denotes Golden Gate assembly method.

2.6.5. Plasmid confirmation via MiSeq sequencing

Plasmid deep sequencing was performed as in previous work(46).

2.6.6. *E. coli* KCM competent cell prep

Competent cells were prepared starting from a single colony or glycerol stock on day one into an LB overnight. On day two the overnight was diluted 1:100 into 50 mL LB and grown at 37°C shaking at 200 RPM or higher to an OD 600 of 0.35 as measured with a Beckman Coulter DU800 spectrophotometer. Cells were then chilled on ice for 20 minutes before centrifugation at 8000 rcf for 8 minutes. The supernatant was decanted, and cells resuspended in 5 mL of ice-cold sterile-filtered TSS (LB base with 10% PEG 3350 (w/v), 5% DMSO (v/v), 20 mM $MgCl_2$). 100 µL of TSS and cell mixture were then aliquoted and used immediately or flash frozen using liquid nitrogen and then stored at -80°C till use. 1 µL of each desired plasmid was added to the TSS cell mixture and mixed with 100 µL of sterile 2x KCM (0.2 M KCl, 0.06 M $CaCl_2$, 0.1 M $MgCl_2$) and incubated on ice for 20

minutes. Cells were then heat shocked at 42°C for 90 seconds before being recovered with 200 μL TB at 37°C for one hour. Cells were then plated onto LB agar plates with appropriate antibiotics (carbenicillin at final concentration of 100 μg/mL, and chloramphenicol at a final concentration of 25 μg/mL).

2.6.7. *E. coli* pathway expressions

Freshly transformed colonies were picked into 10 mL Terrific Broth medium (yeast extract 24 g/L, tryptone 20 g/L, glycerol 4 mL/L, 17 mM KH_2PO_4, 72 mM K_2HPO_4) with 0.4% glucose and incubated shaking overnight in glass test tubes. Cells from overnights were diluted 1:100 into fresh TB with 0.4% glucose and 1 g/L sodium propionate (unlabeled or 1-^{13}C labeled). Cells were grown at 200 RPM and 37°C until OD600 of 0.8. IPTG was then added to a final concentration of 1 mM and a nonane overlay was added. Temperature was reduced to 18°C and the cells were incubated for 36 hours in small scale experiments, and 30 hours for the large scale experiment before harvesting. For small scale expressions, the culture volume was 10 mL in glass test tubes with 1 mL nonane overlay, and large scale expressions were done at 500 mL in baffled flasks with 50 mL nonane overlay.

2.6.8. Small scale homoterpene fermentation

1.4 mL of the nonane/culture emulsions were harvested and centrifuged at 18000 rcf for 1 min to separate the organic layer. 100 μL of the organic layer was collected and 1 μL analyzed directly by GC-MS. Samples were stored at -20°C for short term (one to two days), or at -80°C if longer term storage was necessary in amber glass vials.

2.6.9. Proteomics of pathway enzymes

Cells were grown as in the expressions above. 20 OD mLs of cells were collected and proteomics carried out essentially as described in previous work(47). Briefly, lysis of E. coli cell pellets combined with protein precipitation was achieved by the addition of 80 μL of methanol, 20 μL of chloroform, and 60 μL of water, with vortexing. The samples were centrifuged at 20,817 x g for 1 minute for phase separation. The methanol and water (top) layer was removed, then 100 μL of methanol was added and the sample was

vortexed briefly followed by centrifugation for 1 minute to isolate. e the protein pellet. The protein pellet was resuspended in 100 mM ammonium bicarbonate with 20% methanol and quantified by the Lowry method. A total of 100 μg of protein was reduced by adding tris 2-(carboxyethyl) phosphine (TCEP) to a final concentration of 5 mM, alkylated by adding iodoacetamide at a final concentration of 10 mM, and digested overnight at 37 ºC with trypsin at a ratio of 1:50 (w/w) trypsin:total protein.

Peptides were analyzed using an Agilent 1290 liquid chromatography system coupled to an Agilent 6460QQQ mass spectrometer (Agilent Technologies, Santa Clara, CA) operating in MRM mode. Peptide samples (10 μg) were separated on an Ascentis Express Peptide ES-C18 column (2.7 μm particle size, 160 Å pore size, 50 mm length x 2.1 mm i.d., 60 ºC; Sigma-Aldrich, St. Louis, MO) by using a chromatographic gradient (400 μL/min flow rate) with an initial condition of 95% Buffer A (99.9% water, 0.1% formic acid) and 5% Buffer B (99.9% acetonitrile, 0.1% formic acid) then increasing linearly to 65% Buffer A/35% Buffer B over 5.5 minutes. Buffer B was then increased to 80% over 0.3 minutes and held at 80% for two minutes followed by ramping back down to 5% Buffer B over 0.5 minutes where it was held for 1.5 minutes to re-equilibrate the column for the next sample. The data were acquired using Agilent MassHunter, version B.08.02, processed using Skyline version 4.1, and peak quantification was refined with mProphet in Skyline. All data and skyline files are available via the Panorama Public repository at this link: https://panoramaweb.org/Heterologous-Lepidopteran-Mevalonate-Pathway.url. Data are also available via ProteomeXchange with identifier PXD012697.

2.6.10. Method for novel terpene titer estimation

Estimating titers of the homo-terpene compounds is difficult because we cannot buy standards, nor have we been able to successfully purify them to homogeneity. To make rough estimates we used β-caryophyllene as a standard. However, we stress the error in this estimate is hard to know a priori as we must assume the novel terpenes ionize like parent ion of β-caryophyllene. 1 μL of β-caryophyllene was diluted into 99 μL of ethyl acetate. This mixture was then diluted 1.5 μL into 1500 μL ethyl acetate yielding a vial of 0.01 μL β-caryophyllene per mL of ethyl acetate. 750 μL of this mixture was serially

diluted two-fold nine times to generate the standard curve. The standard curve was run from the smallest concentration vial to the vial with highest concentration using GC-MS in single ion monitoring mode detecting m/z = 204. β-caryophyllene density was taken to be 0.9052 g/cm3 at room temperature (collected from https://en.wikipedia.org/wiki/Caryophyllene).

To estimate the purity of β-caryophyllene, a sample of 0.01 μL β-caryophyllene per mL of ethyl acetate was analyzed using scan mode on GC-MS. The sample contained primarily β-caryophyllene but also some amount of two oxidized β-caryophyllene compounds. By taking the ratio of the peak areas we concluded that standard was 94.4% β-caryophyllene, and the 5.6% other compounds. The excel function LINEST was used to generate an equation to convert peak area to mg/L, we then corrected for the purity of our β-caryophyllene standard.

To estimate the homosesquiterpene peak areas we averaged the ^{12}C propionate and ^{13}C propionate fermentation runs, each in biological triplicate thus representing six observations.

2.6.11. Large scale homoterpene fermentation and enrichment

The cell culture was added to a separation funnel and allowed to sit until the overly and a foamy layer separated from the aqueous layer. The aqueous layer was removed, leaving a foamy layer and an organic layer. These layers were washed with approximately 200 mL NaCl saturated water. After the layers separated the aqueous layer was removed. This wash step was repeated one time. The foamy layer and organic layer were put into 50 mL conical tubes and spun at 10,000 rcf for 3 min. After the centrifugation the foamy layer was greatly reduced and instead a white "goopy" layer appears between the organic layer and aqueous layer. The organic layer was then transferred to a round bottomed flask and rotovapped to approximately 100 μL. This liquid was sent through a glass wool mesh to remove any particulate. Samples were stored at -80°C or on dry ice under a nitrogen atmosphere until analysis.

2.6.12. GC-MS analysis method

Samples were run on an Agilent 6890 GC and Agilent 5973 MS using a 30-meter DB-5ms column with a helium flow of 1 mL/min. The GC was run in splitless mode with a constant inlet temp and transfer line temp of 280°C. The oven starting temperature was 100°C held for 1 minute. The temperature was increased at 20°C per minute until 160°C. Then the temperature was increased at 10°C per minute until 225°C. Then the temperature was increased at 30°C per minute until 300°C where it was held for 2 min. Single ion monitoring was carried out at 218 or 219 m/z as appropriate to detect ^{12}C or ^{13}C labeled products. A relative gain factor of 2 was used. Data was analyzed via Chemstation Enhanced Data Analysis program. For peak integration the RTE integrator was used.

2.6.13. GC-QTOF analysis method

Accurate mass GC-QTOF data was collected on an Agilent 7890A GC coupled to an Agilent 7200 QTOF mass spectrometer (Agilent Technologies, CA, USA). For chromatographic separation, a Restek RTx-5Sil MS column (30 m x 0.25 mm, 0.25um; Restek Corporation, PA, USA) was used with Helium (99.999%; Airgas, PA, USA) flow maintained at 1 mL/min. Oven temperature started at 60°C for 1 min, increased at a rate of 10°C/min to 325°C and held constant for 9.5 min, with a transfer line temperature of 270 °C. Spectra were collected over a range of 50-800 m/z at a rate of 5 spectra per second for electron ionization (EI). The source was maintained at 230 °C, with 70 eV electron energy. The transfer line temperature was 270°C. Data was analyzed using MassHunter Qualitative Analysis software (Agilent Technologies, CA, USA).

2.7. Miscellaneous

2.7.1. Acknowledgements

C.B.E. would like to thank Marlin Eiben for thoughtful conversation on statistics, as well as Nurgul Kaplan and Garima Goyal for plasmid construction related to, but ultimately not included, in this manuscript. C.B.E. was supported by National Science Foundation (NSF) Graduate Research Fellowship Program (DGE-1106400) and National Institute of Health training grant T32 GM008295. NSF grant 0540879 supported the materials for this work. This work was enabled by the DOE Joint BioEnergy Institute

(https://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through Contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

2.7.2. Author contributions

C.B.E. helped conceive and planned the project including, plasmid design, construction, experimental design, sample preparation, GC-MS assays as well as writing the primary draft of the paper. T.D.R. helped with GC-MS method development and homoterpene enrichment protocol as well as providing chemistry insights. C.B. provided GC-QTOF method development and analysis. J.G. performed the proteomics assays. J.C. aided in deep sequencing of plasmids. E.E.K.B. provided data analysis and data representation help. C.J.P. oversaw the proteomics effort. N.J.H. contributed to manuscript revisions. O.F. oversaw GC-QTOF analysis. J.D.K. helped conceive and guide the project, as well as helping with the manuscript writing.

2.7.3. Data availability

Genbank files of all plasmids are available at https://public-registry.jbei.org/folders/421. Plasmids will be available upon request from authors or addgene. Dataset S1 is also available from the Experiment Data Depot (48) https://public-edd.jbei.org/accounts/login/?next=/s/exploring-non-canonical-terpene-space-with-a-heter/overview/

The proteomics data has been uploaded to ProteomeXchange ID: PXD012697

2.7.4. Competing interests

The pathway has been patented by JDK and CBE. JDK has a financial interest in Amyris, Lygos, Demetrix, Constructive Biology, Maple Bio, and Napigen.

Chapter 3. ATP and carbon efficient route to second generation biofuel isopentanol and commodity chemical prenol

3.1. Abstract

Climate change necessitates the development of CO_2 neutral or negative routes to chemicals currently produced from fossil carbon. In this paper we demonstrate a pathway requiring zero ATP to make the next generation biofuel isopentanol and the commodity chemical prenol from sugars by pairing the isovaleryl-CoA biosynthesis pathway from *Myxococcus xanthus* and a butyryl-CoA reductase from *Clostridium acetobutylicum*. We can control which five carbon alcohol is produced by inclusion or omission of the gene AibC that reduces 3-methyl-2-butenyl-CoA to isovaleryl-CoA. Aerobically we make 73.4 mg/L ± 5.51 (SD) of prenol and 25.9 g/L ± 2.97 of isopentanol after 48 hours with our best constructs. Furthermore, we demonstrate that microaerobically we maintain prenol titers when normalized to OD cells.

3.2. Introduction

Anthropogenic climate change represents a significant challenge for modern society, with transportation activities and petrochemical production contributing about 7 gigatonne CO_2eq annually (49). Sugar derived biofuels and bioproducts offer an alternate means to these molecules, potentially lowering CO_2 emissions (50). First generation biofuels, ethanol and biodiesel, have drawbacks compared to fossil fuels in internal combustion engines, necessitating the development of newer and better molecules which can be derived from renewable sources (51) (52) (53). Five carbon alcohols have better fuel properties for internal combustion engines than ethanol as they have a higher blending limit with gasoline, are less hygroscopic and have a higher energy density (54). In addition, it was recently found that prenol can increases Research Octane Number (RON) of a fuel blend above the RON numbers of the individual components, a desirable and uncommon property in a fuel (55).

Two main pathways for five carbon alcohol production have received considerable attention in the past, namely the mevalonate pathway to isoprenol (Fig. 3-1B) (13) (56) (14), and the leucine pathway to isopentanol (Fig. 3-1D) (57). The mevalonate pathway

requires three ATP, three acetyl-CoA and 2 Nad(P)H to make isopentenyl pyrophosphate (IPP). The pyrophosphate is then removed by two successive phosphatases yielding isoprenol. Recently a more ATP efficient route to isoprenol has been developed saving one ATP, dubbed the IPP bypass pathway (Fig 3-1C)(58). The pathway uses a mutated mevalonate pyrophosphate decarboxylase to accept mevalonate phosphate instead of mevalonate pyrophosphate. Thus, one ATP is used to make mevalonate phosphate, and one ATP is used to decarboxylate that mevalonate phosphate to isopentenyl phosphate. The isopentenyl phosphate can then be dephosphorylated to isoprenol.

Another route to a branched five carbon alcohol is through leucine biosynthesis ultimately yielding isopentanol (59) (Fig. 3-1D). The pathway starts with two pyruvate making 2-acetolacetate via a decarboxylation. After several steps, including a condensation with acetyl-CoA, the pathway yields 4-methyl-2-oxopentanoate, the direct precursor of leucine. Decarboxylation of 4-methyl-2-oxopentanoate makes 3-methylbutanal, which can then be reduced to 3-methylbutanol, more commonly known as isopentanol. This 8 step pathway uses no ATP, but does require the oxygen sensitive ilvD gene (60).

At first glance the leucine pathway appears more efficient from an ATP standpoint and equal from a carbon standpoint. However, the potential production of acetyl-CoA from non-oxidative glycolysis complicates the matter. Non-oxidative glycolysis (61) converts glucose into three acetyl-phosphate which can then be converted to three acetyl-CoA without any CO_2 loss, and has been demonstrated *in vitro* and *in vivo* (62) (63). Because the mevalonate pathway uses acetyl-CoA exclusively, it is possible to use non-oxidative glycolysis to redox balance the production of isoprenol. The leucine pathway on the other hand cannot be redox balanced using non-oxidative glycolysis because it requires two pyruvate and one acetyl-CoA. The single acetyl-CoA is not oxidized enough to balance the pathway, and even with non-oxidative glycolysis the production of isopentanol yields an extra reducing equivalent.

As biofuels and commodity chemicals have low economic margins, an optimal pathway for five carbon alcohols would combine the benefits of the leucine degradation pathway and the mevalonate pathway. Mainly a pathway which is both ATP positive, but also redox balanced. It is also convenient if the pathway can be easily modulated to produce different five carbon alcohols for various uses. Here we demonstrate the 3-methyl-2-butenyl-CoA pathway to the five carbon alcohols prenol and isopentanol which fulfills these requirements.

Myxococcus xanthus has previously been noted to produce isovaleryl-CoA for branched fatty acid biosynthesis in an unusual via the intermediate 3-methlybuteneyl-CoA (64). The pathway uses the first two steps of the mevalonate pathway to make HMG-CoA, and then uses a dehydratase to make 3-methylgultaconyl-CoA (Fig. 3-1A). This is then decarboxylated to make 3-methlybuteneyl-CoA which can be reduced to isovaleryl-CoA which serves as the starting unit for branched fatty acid biosynthesis. In our work here, we couple this pathway to a butyryl-CoA reductase from *Clostridium acetobutylicum* to make prenol, or isopentanol depending on included genes. A similar pathway has been proposed in the patent literature, but actual production is not relevantly disclosed (65).

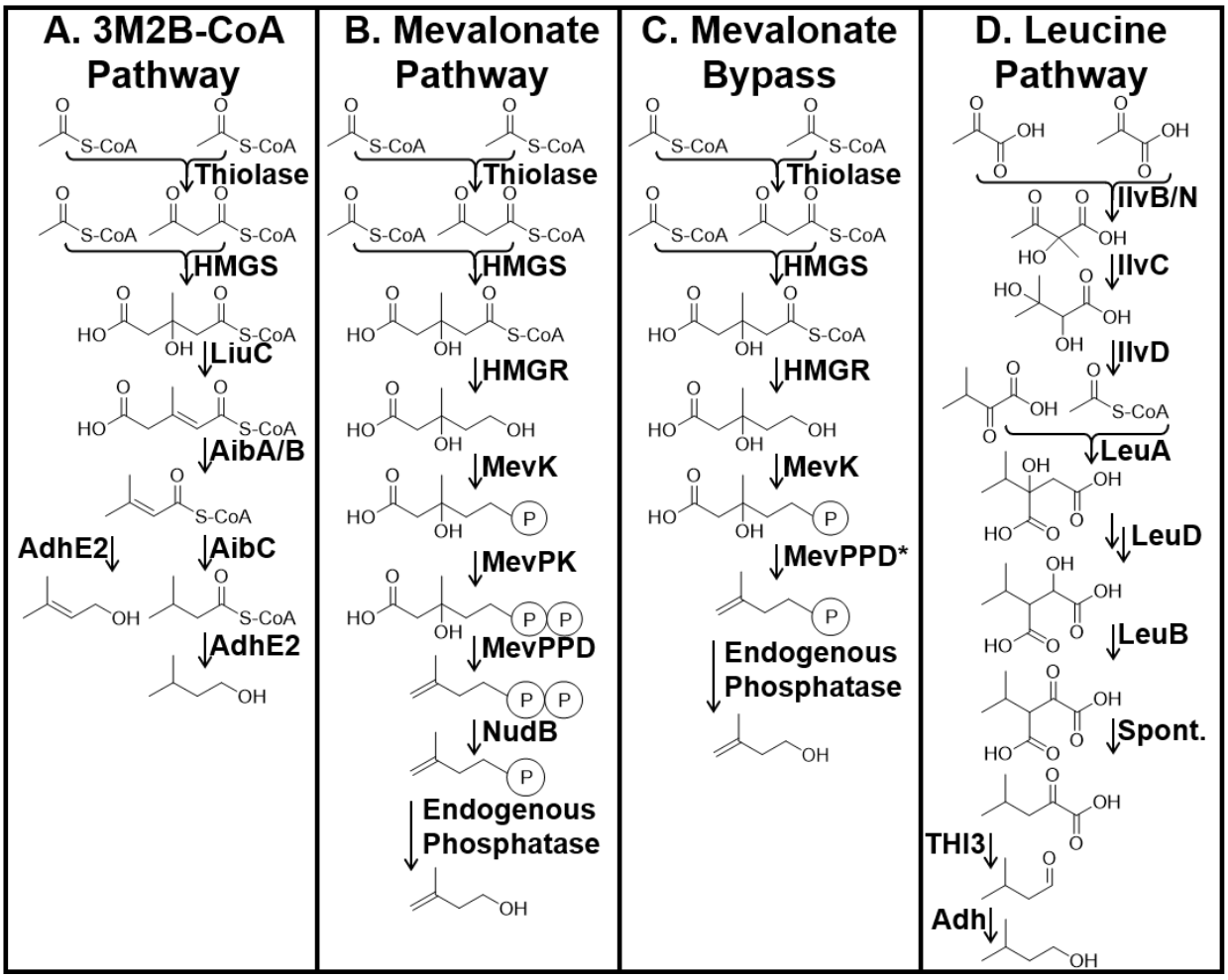

Figure 3-1 C5 Alcohol pathways. A. The 3-methyl-2-butenyl-CoA pathway to either prenol or isopentanol. **B**. The mevalonate pathway to isoprenol. **C**. The mevalonate pathway with IPP bypass to isoprenol. **D**. The leucine pathway to isopentanol. Enzyme abbreviation [HMGS: 3-hydroxy-3-methylglutaryl-coenzyme A synthase], [LiuC: 3-methylglutaconyl-CoA hydratase], [AibA/B: 3-methylglutaconyl-CoA decarboxylase complex (heterodimer)], [AdhE2: aldehyde dehydrogenase/alcohol dehydrogenase], [AibC: 3-methyl-2-butenyl-CoA reductase], [HMGR: 3-hydroxy-3-methylglutaryl-coenzyme A reductase], [MevK: mevalonate kinase], [MevPK: mevalonate phosphate kinase], [MevPPD: mevalonate pyrophosphate decarboxylase], [NudB: dihydroneopterin triphosphate diphosphatase], [MevPPD*: mevalonate pyrophosphate decarboxylase mutant], [IlvB/N: acetohydroxy acid synthase complex (heterodimer)], [IlvC: ketol-acid reductoisomerase], [IlvD: dihydroxy-acid dehydratase], [LeuA: 2-isopropylmalate

synthase], [LeuB: 3-isopropylmalate dehydrogenase], [Spont.: reaction reported to happen spontaneously], [Thi3: ketoisocaproate decarboxylase], [Adh: alcohol dehydrogenase]

3.3. Results

3.3.1. *In silico* modeling of the 3M2B-CoA pathway thermodynamics

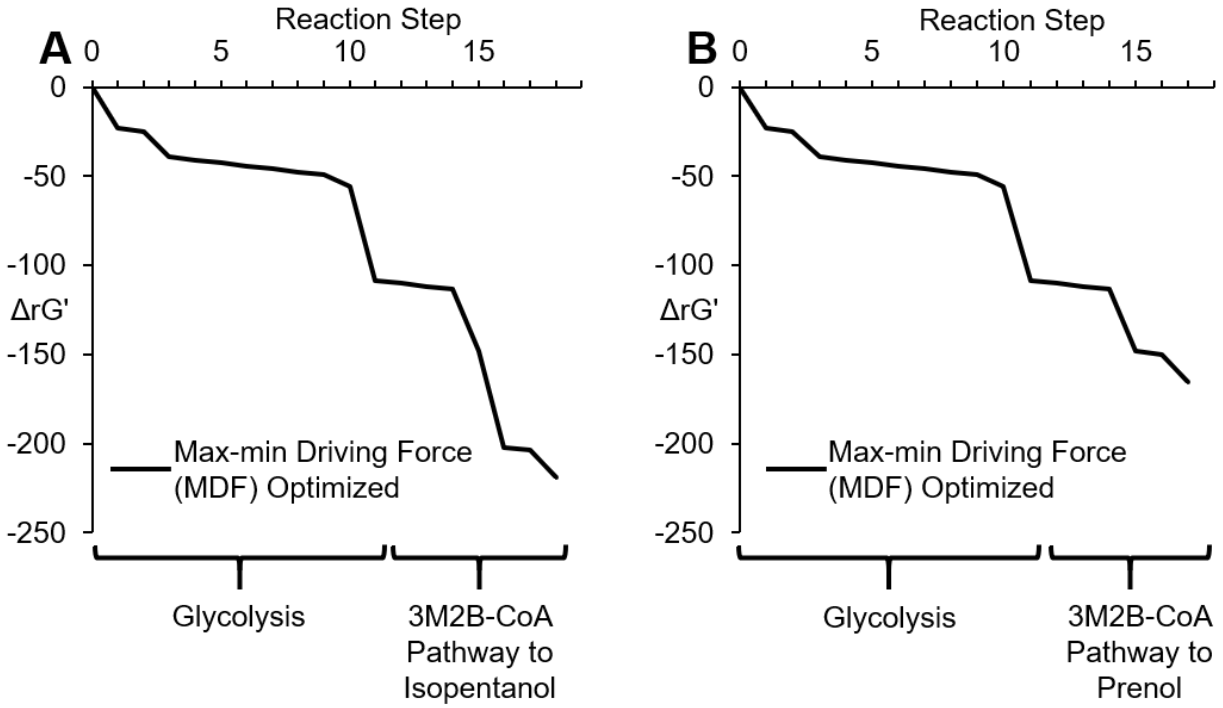

Figure 3-2 *In silico* modeling of 3M2B-CoA pathway thermodynamics. Modeling starts with glucose, proceeds through glycolysis and into the 3M2B-CoA pathway. Pathway represent the worst case scenario in which the thiolase and HMG-CoA synthase don't have substrate channeling as is possible in some configurations (66). **A**. 3M2B-CoA pathway ending with isopentanol. **B**. 3M2B-CoA pathway ending with prenol.

To confirm the thermodynamic viability of the pathway we used eQuilibrator (67). We modeled the worst case scenario where the thiolase and HMG-CoA synthase don't have substrate channeling as is possible in some cases (66). The modeling shows the pathway to be thermodynamically feasible, with glycolysis having the lowest Min-max Driving Force (MDF) section of the pathway, and therefore expected to be limiting, at least from a thermodynamic perspective. It should be noted that both pathways, either to

isopentanol or prenol produce more reducing equivalents from glucose than are consumed by C5 alcohol production uses. In the long run we propose to balance these extra reducing equivalents with non-oxidative glycolysis, already known to be thermodynamically possible and thus we did not model it. With the thermodynamics confirmed, we moved onto building a building the pathway on plasmids.

3.3.2. Bioprospecting and plasmid design

We took the p15A origin, the chloramphenicol resistance marker and the IPTG inducible pLacUV5 driven operon containing the native sequence of the *E. coli* thiolase atoB and the native sequence of *Staphylococcus aureus* HMGS from the plasmid MevTsa-T1002-Ptrc-MKco-PMKco (JBEI-6831) for the expression of the upper portion of the pathway (56). We based our designs off this plasmid because it showed high IPP flux previously, and the first two genes are common to our pathway. The pTrc promoter drives the lower portion of the pathway containing AdhE2 from *Clostridium acetobutylicum* (68) as the isovaleryl-CoA reductive thioesterase (native substrate butyryl-CoA), followed by the AibC gene encoding the 3-methyl-2-butenyl-CoA reductase from *Myxococcus xanthus*, followed by the LiuC from *Pseudomonas putida* (69) encoding the 3-methylglutaconyl-CoA hydratase, followed by the two halves of the heterodimer 3-methylglutaconyl-CoA decarboxylase from *Myxococcus xanthus*, AibA and AibB (64). All the genes in the pTrc driven operon were codon optimized *E. coli* expression with Boost (https://boost.jgi.doe.gov/) (70). We named this plasmid pII3.

3.3.3. C5 alcohol production in *E. coli* DH1

We transformed pII3 into the common *E. coli* production strain DH1 (12), and expressed the genes virtually as previously reported (58). pII3 did not make isopentanol but made prenol instead at 59.34 ± 1.49 mg/L and 74.11 ± 1.94 mg/L after 24 and 48 hours respectively (Fig. 3-3). To troubleshoot isopentanol production we cloned several different AibC homologues codon optimized to express in *E. coli* to replace the *Myxococcus xanthus* gene behind a new stronger RBS. The AibC genes in pII5 through pII7 are from *Cystobacter fuscus*, *Myxococcus fulvus*, and *Myxococcus stipitatus*, respectively. These genes were found with Blast (71) and selected for testing based on their relatively

high sequence identity with AibC from *M. xanthus* indicating they probably have the same enzymatic activity. The AibC gene in pII8 is an alternately codon optimized gene from *M. xanthus* as compared to pII3. pII5 produced the most isopentanol at 25.89 ± 1.7 mg/L after 24 hours, and 25.88 ± 2.97 mg/L after 48 hours. pII5 also produced only minimal prenol at 2.6 ± .12 mg/L and 11.23 ± .39 mg/L after 24 and 48 hours respectively (Fig. 3-3B).

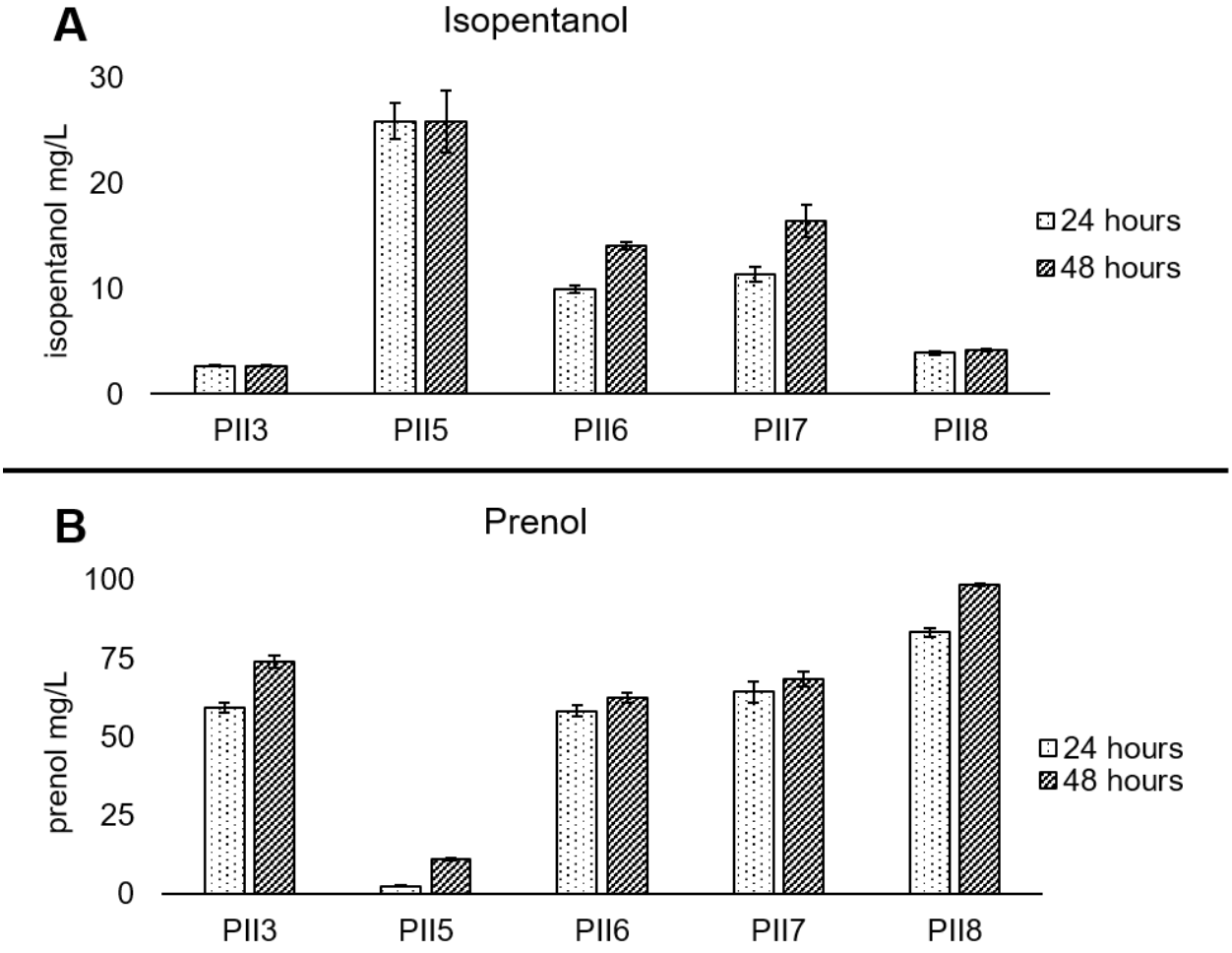

Figure 3-3 Alternate AibC genes enable isopentanol production. Bars with dots fill represent production after 24 hours. Bars with slanted line fill represents production after

48 hours. Error bars represent standard deviation of biological triplicate. **A**. Isopentanol production. **B**. Prenol production.

Based on these results we designed a plasmid exclusively for prenol production modeled after pII3. The new plasmid, named pII14 lacks the aibC gene. In addition, we replaced the origin of replication and resistance marker of pII3 with a high copy number origin of replication, ColE1, and an Amp^R resistance marker. We tested this strain in two conditions, aerobic shaking at 200 rpm, and microaerobic without shaking (Fig. 3-4A). The aerobic condition produced more prenol with 67.76 ± 3.62 mg/L and 73.43 ± 5.51 mg/L after 24 hours and 48 hours respectively. The microaerobic condition produced 31.98 ± .9 mg/L and 37.69 ± .26 mg/L after 24 and 48 hours (Fig. 3-4A). The cells grew much better aerobically than microaerobically. However, after correcting for cell growth (OD_{600}) the prenol production in both conditions is the same at ~20 mg/L/OD (Fig. 3-4B). A two tailed t-test cannot distinguish the two condition either at 24 hours or 48 hours as it gives p-values of .42 and .37 respectively.

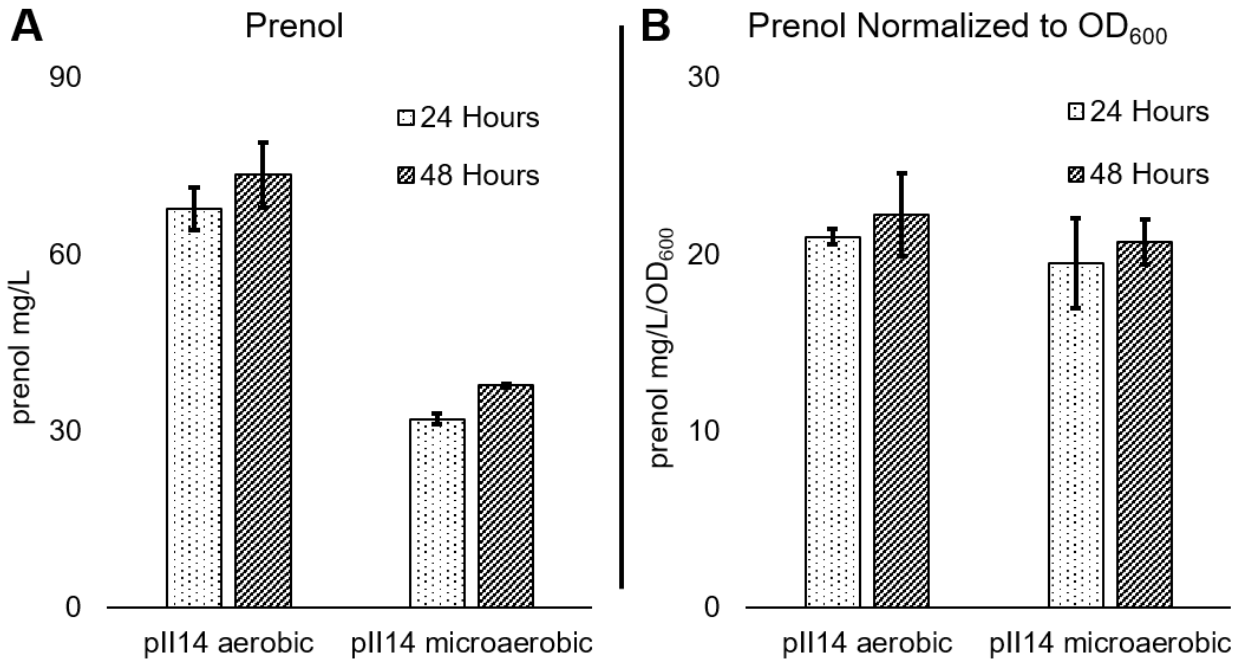

Figure 3-4 Best prenol production in aerobic and microaerobic condition. Dot fill bars represent measurements at 24 hours. Diagonal line fill bars represent measurements at 48 hours. Error bars represent standard deviation of biological triplicate. **A**. Prenol

production in mg per L of cell culture volume. **B**. Prenol production in mg per L of cell culture volume normalized to OD_{600}.

3.4. Discussion

In this work we demonstrate the utility of the 3-methyl-2-butenyl-CoA to produce prenol and isopentanol. As an early report for this pathway, much improvement could yet be made in titers and yields. The best way to quickly improve the pathway is probably to couple it to growth in anaerobic (or suitably microaerobic) conditions. To do this, a non-oxidative glycolysis strain that also lacks other anaerobic fermentation routes (ethanol, lactic acid, formate, etc.) would need to be constructed. This should enable the screening of large libraries for improved production by simply growing the cells anaerobically.

A promising directed approach to improve flux and yields would be to replace the first two enzymes of the pathway with a newly discovered enzyme complex. Recently a paper demonstrated that some thiolases and HMG-CoA synthase form a complex with a DUF35 protein that enables both enzymes to act before releasing the product (66). This overcomes the large thermodynamic barrier of two acetyl-CoA condensation to acetoacetyl-CoA, which should enable more flux for less enzyme. Given the thiolase and HMG-CoA synthase we used it is unlikely the pair take advantage of substrate channeling, as they are from two different organisms and we don't overexpress the DUF35 protein that helps hold the complex together.

Another area to explore rational improvements is the LiuC gene. There are two different classes of LiuC enzymes which catalyze the reaction (72). The *Myxococcus xanthus* LiuC can hydrate the pathway intermediate 3-methyl-2-butenyl-CoA to 3-methyl-3-hydroxybutyryl-CoA, a nonproductive side product (64). The LiuC gene from *Pseudomans putida* uses a different reaction mechanism which cannot hydrate 3-methyl-2-butenyl-CoA and thus a nonproductive side product is avoided. However, 3-methyl-2-butenyl-CoA acts as a competitive inhibitor of LiuC. We used the LiuC from *Pseudomans putida* in the current work, but perhaps the other class of LiuC will have better results. Provided 3-methyl-2-butenyl-CoA concentrations remain suitably low, then the current approach is probably the best option to iterate around in efforts to improve flux and titers.

The 3-methlybuteneyl-CoA pathway may also find uses as a precursor to higher isoprenoids. Two sequential phosphorylation steps from prenol yield dimethylallyl pyrophosphate by way of dimethylallyl phosphate. Both enzymatic steps have been reported in the literature thought they are not the canonical reactions for the enzymes. IspE (4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase) from *E. coli* performs the first reaction, and IPK (isopentenyl phosphate kinase) from *Thermoplasma acidophilum* performs the second (73) (74). It may be possible to create a selection to find mutants with higher activity because isoprenoid production is necessary for growth. At least one of the two enzymes has been reported to have been engineered for higher activity on an unnatural substrate (75).

The caveat to the selection strategy is that prenol is membrane permeable, so it might be tricky to find conditions where prenol is not secreted and thus becomes a "common good". Secreted prenol could lead to "cheater" cells, which evolve to not produce prenol, but can still use it to make isoprenoids. One potential solution is to use a plate-based selection so secreted prenol stays close to the colony which made it, or to pick colonies into individual wells so that all the cells in that well have the same genotype.

The advantage of making higher isoprenoids using the 3-methlybuteneyl-CoA pathway is that typically the mevalonate pathway requires three ATP per five carbon unit produced, while isoprenoid production via the 3-methylbuteneyl-CoA pathway would only require two ATP. As previously, the membrane permeability of prenol might be problematic for engineering very high flux pathways. Another consideration is the identification of a modified mevalonate pathway in Archaea which can achieve the production of higher isoprenoids for two ATP per five carbon unit (76). The Archaeal pathway requires two newly discovered enzymes. The first is mevalonate-phosphate dehydratase which converts mevlaonte-phosphate to tran-anhydromevalonate 5-phosphate using an iron-sulfur cluster. The second enzyme is trans-anhydromevalonate 5-phosphate decarboxylase which converts trans-anhydromevalonate 5-phosphate to isopentenyl phosphate. While this pathway keeps a charged intermediate, which is

presumably membrane impermeable, it does require an oxygen sensitive iron-sulfur cluster which may restrict or complicate it's use.

3.5. Materials and methods

3.5.1. Chemicals

Carbenicillin disodium salt, EZ-rich media, and LB antibiotic plates were ordered from Teknova. Chloramphenicol was ordered from MP Biomedicals, Inc. Restriction enzyme BsaI, DpnI, T4 DNA ligase, isopropyl β-D-1-thiogalactopyranoside (IPTG), and ElectroMAX DH10B Cells were ordered from Thermo Fisher Scientific. Molecular grade bovine serum albumin (BSA) was obtained from New England BioLabs. Dextrose (anhydrous) and ethyl acetate was purchased from VWR International. Isoprenol (3-methylbut-3-en-1-ol) was ordered from Acros Organics. Stellar chemically competent cells were purchased from Takara Bio USA. LB (Lennox) and TB powder by Merck KGaA, as well as all other chemicals, were ordered from Sigma-Aldrich.

3.5.2. *In silico* modeling

Min-max Driving Force (MDF) modeling was done with the eQuilibrator (67) web interface (http://equilibrator.weizmann.ac.il/pathway/) using the default parameters for all fields (minimum substrate concentration of .001 mM, maximum substrate concentration of 10 mM, pH 7.0, ionic strength .1 M). We tested the conversion of three glucose to two five carbon alcohols (and resulting NadH, ATP, H_2O and CO_2production). Modeling files are included at https://public-edd.jbei.org/s/eiben-et-al-c5-alcohol-production/overview/.

3.5.3. DNA transformation and cell culturing

E. coli DH1 was transformed using the KCM method. On day one cells were picked from colonies or glycerol stock into Lysogeny Broth (LB) medium and allowed to grow overnight shaking at 200 RPM at 37°C. On day two cells were back diluted 1:100 into 50 mL fresh LB and allowed to grow as previously until OD_{600} 0.4. Cells were then chilled on ice for 20 minutes before centrifuging them at 8000 RCF for 8 minutes. Supernatant was removed, and cells were resuspended in 5 mL 4°C TSS (recipe below). 100 μL of cells

were aliquoted into .6 mL Eppendorf tubes and either used immediately or flash frozen with liquid nitrogen and stored at -80°C until future use.

For DNA transformation 1 μL of DNA was added to 100 μL of cells, and 100 μL of 2x KCM (60 mM KCl, 200 mM $CaCl_2$, 100 mM $MgCl_2$) was added. Cells were incubated on ice for approximately 20 minutes, before being heat shocked for 90 seconds at 42°C in a water bath. 200 μL of TB was added and the cells were allowed to recover, without shaking, at 37°C for an hour before being plated onto LB Lennox agar plates of the correct antibiotic. When appropriate, carbenicillin was added at a final concentration of 100 μg/ml, and chloramphenicol 30 μg/ml.

To prepare 100 mL TSS, add 10 g poly ethylene glycol (molecular weight 3350), 5 mL dimethyl sulfoxide (DMSO), 2 mL of 1 M $MgCl_2$, add LB to 100 mL. Sterile filter and store at 4°C.

Optical density (OD) measurements were taken with a Beckman Coulter DU 800 spectrophotometer at 600 nm.

3.5.4. Five carbon alcohol fermentation conditions

On day one overnight seed cultures were grown in LB at 37°C shaking at 200 RPM. On day two the overnights were diluted into EZ-Rich with 1% glucose and allowed to grow as before until they reached an OD_{600} between .6 and .8. Cells were then induced with a final concentration of .5 mM IPTG and temperature reduced to 30°C. All media contained the appropriate antibiotic.

3.5.5. Five carbon alcohol extractions

Five carbon alcohol extraction was conducted essentially as elsewhere (58). Briefly, 1-butanol was added to a final concentration of 32 mg/L as internal standard in ethyl acetate. 250 μL of cell culture and 250 μL ethyl acetate with 1-butanol were vortexed in microcentrifuge tubes for 15 minutes at max speed. Tubes were then centrifuged in an Eppendorf centrifuge 5424 at max speed (approximately 21100 RCF) for two minutes. 100

μL of the upper, nonpolar, ethyl acetate layer was then collected and added to 400 μL of ethyl acetate with 1-butanol in a glass GC vial for analysis.

3.5.6. Analytical method for five carbon alcohol analysis

Analysis of five carbon alcohols was performed essentially as reported elsewhere (14) on a Thermo Scientific Focus GC-FID with a DB-WAX column (15 meter long, .32mm diameter, .25 micron film thickness). The carrier pressure was set to constant at 300 kPa and inlet temperature was 200°C. Oven temperature was initially set to 40°C for 90 seconds, and then increased at 15°C a minute till 110°C. 1 μL of sample injections were used.

3.6. Miscellaneous

3.6.1. Acknowledgements

C.B.E. was supported by National Science Foundation (NSF) Graduate Research Fellowship Program (DGE-1106400). NSF grant 0540879 supported the materials for this work. This work was enabled by the DOE Joint BioEnergy Institute (https://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through Contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

3.6.2. Data availability

Plasmid maps are available in the JBEI public registry (https://public-registry.jbei.org/folders/434) (34). Supplemental data is available at the Experiment Data

Depot (EDD) (https://public-edd.jbei.org/s/eiben-et-al-c5-alcohol-production/overview/) (35).

3.6.3. Collaborator contributions

CBE conceived the project, designed and cloned plasmids, helped design the experiments and wrote the manuscript. TT designed and cloned plasmids and designed and performed experiments. NK and GG performed PCRs for plasmid construction. JC deep sequenced plasmids. NJH oversaw plasmid construction efforts. TSL oversaw TT and experiments. JDK provided project guidance.

3.6.4. Competing interests

JDK has a financial interest in Amyris, Lygos, Demetrix, Constructive Biology, Maple Bio, and Napigen.

Chapter 4. Perspectives on global warming and technological options for reducing impact

4.1. Introduction

Anthropogenic carbon emissions are an existential threat to global civilization (77). The threat is worsened by the associated plethora of technical, economic and political problems. Because of the scale and energetic barriers of concentrating CO_2 out of ambient air, carbon capture is a difficult engineering challenge. From an economic perspective there are very few products that can be made from CO_2 profitably (other than agricultural goods). This limits the capitalist solution of businesses mining CO_2 out of the air and selling it directly or upgrading it into products. In addition, climate change is politically difficult to address because it represents a danger that isn't immediate, is slow to reverse, expensive, challenges many peoples deeply held worldviews, and is a tragedy of the commons. Furthermore, the entrenched oil and gas industry is currently inviable without CO_2 emissions. As they represent a non-trivial percentage of world GDP and are represented in every country this is a substantial constituency committed to continuing emitting CO_2.

4.2. Chemical CO_2 capture process

Given these constraints, the most elegant option for curbing global warming is to develop technical solutions which are profitable without government intervention. While a carbon tax or other governmental regulations would speed the acceptance of such a technology, most of the political problems surrounding global warming are avoided if profitable technological solutions could be developed. Three scientific areas seem particularly attractive to search for such inventions; carbon concentration technologies coupled with CO_2 conversion to useful products, improving biological carbon fixation, and new sources of extremely cheap electricity generation.

4.3. Direct CO_2 capture from ambient air

Recently a chemical processes has been demonstrated at pilot scale for removing CO_2 from the air (78). Briefly, the process is based on fans blowing air past a KOH solution which captures CO_2 as carbonic acid because of the high pH as K_2CO_3. This K_2CO_3 solution is then mixed with a $Ca(OH)_2$ which regenerates KOH and makes $CaCO_3$ that is insoluble and precipitates into pellets. These pellets are heated releasing CaO and CO_2 which is then captured. The $Ca(OH)_2$ is then regenerated by adding water to the CaO. This demonstration is significant because it provides a more accurate estimate the cost of CO_2 removal than laboratory scale experiments can.

While the cost estimates from this work are more hopeful than many, they represent only the cost of an intermediate chemical since CO_2 has only a small market and must be converted into something more useful. Proposals of using Fischer–Tropsch reaction to convert concentrated CO_2 into fuels or other petrochemicals requires large amounts of H_2. While the H_2 can be generated from splitting water using electricity, this process is energy intensive. In addition, a great deal of oxygen must be liberated from the CO_2 as water in this process, which will represent a loss of around ⅔of the weight of the CO_2 to convert it to a highly reduced finished product.

If taking this route to CO_2 emission reductions, the final products also need to have enough demand to make an impact. Most CO_2 emissions from oil are for energy production and not to produce materials. Although materials like plastics can be

relatively energy intensive to make, the volume is so much lower than fuels that they represent a relatively small proportion of CO_2 emissions overall from the petrochemical industry. This puts a constraint on what to turn CO_2 into if the goal is to reduce CO_2 emissions as much as possible. If the CO_2 can be turned into liquid carbon fuels without the conversion itself releasing much CO_2, that has the biggest impact to reduce emissions toward neutral since liquid fuels represent 14% of global CO_2 emissions (49). For reference energy production for electricity and heat is 26% from the same source. Alternatively, if CO_2 is to go into durable products, the biggest products by volume represent the biggest opportunities for CO_2 sinks that might be profitable to produce without subsidies. Finding materials that can be made from CO_2 that replace concrete, steel, aluminum and plastics are the best to focus on from a CO_2 sink standpoint.

4.4. Improving photosynthesis

Oxygenic photosynthesis is arguably the most important biological process on Earth. Significantly improving the efficiency or speed of photosynthesis is the holy grail of metabolic engineering, but the obstacles are significant. Photosynthesis is conceptually divided into two halves; splitting water to generate reducing equivalents and energy powered by photons; and fixing CO_2 into biologically useful molecules. Improving the biochemistry of water splitting seems very difficult as they are already very efficient (26% of solar energy captured in the best scenario (79), and that includes that half of solar radiation is in the wrong energy range to be used to split water biologically) and would require precise protein engineering which is extremely difficult. Perhaps the incorporation of additional pigments which can absorb underutilized wavelengths of light, especially in the shaded leaves of plants, would be a potential strategy. The carbon fixation portion of the pathway looks more hopeful to find improvements.

There are six described (80) (possibly seven (81)) natural carbon fixation pathways, which have been optimized for millennia. Of the pathways, only three are oxygen tolerant enough to be found in aerobic environments, and thus potentially interesting for CO_2 fixation from ambient air. The 3-hydroxypropionate bicycle, the 3-hydroxypropionate/4-hydroxybutyrate bicycle and the much more famous Calvin-Benson-Bassham cycle (CBB cycle). The 3-hydroxypropionate bicycle and the 3-hydroxypropionate/4-

hydroxybutyrate bicycle share the same enzymes at the start of the pathway and only vary in how they regenerate the starting substrates, the first being more ATP expensive but having higher flux and less steps, the second being less ATP expensive but more steps and lower flux. Both use the biotin containing enzymes acetyl-CoA carboxylase and propionyl-CoA carboxylase for carboxylation.

The most important carbon fixation cycle is by far the CBB cycle, estimated to fix more than 90% of CO_2 in the biological carbon cycle annually (82). In the oceans, the most common cyanobacteria genus *Prochlorococcus* is estimated to have $2.9 \pm .1 \times 10^{27}$ individuals, and the second most common genus *Synechococcus* $7 \pm .3 \times 10^{26}$ (83). Land plants also use the CBB cycle and are thoughts to contain 450 gigatonnes carbon of the 550 gigatonnes carbon in all living things (84). This represents an enormous population sampling the CBB cycle looking for improvements.

Given this backdrop, why is there any hope of finding or creating a better carbon fixation cycle? The main reason is that other carboxylases have better enzymatic properties than Rubisco, the main carboxylating enzyme of the CBB cycle. Rubisco is a relatively slow enzyme compared to other central carbon metabolism enzymes and can fix O_2 by accident (thought to be up to ⅓of total Rubisco flux in some circumstances) which produces toxic molecules that must be removed at the cost of energy. Given this fact Rubisco has evolved to optimize the tradeoff between low carboxylase activity but low O_2 side reaction, or high carboxylation activity but high O_2 side reaction activity. Photosynthetic organisms without carbon concentrating mechanisms tend to have evolved slow but selective Rubisco variants, while organisms with carbon concentrating mechanisms use the faster but sloppy Rubisco variants. Increases in temperature tend to make Rubisco catalyze the O_2 side reaction faster than the carboxylation reaction, another downside.

C3 plants, all of which lack carbon concentrating mechanism, dedicate up to 28% of the leaf total nitrogen content to Rubisco to make up for its relatively slow carboxylation rate compared to the rest of the enzymes in the CBB cycle (85). Some plants have evolved C4 metabolism to concentrate CO_2 around Rubisco, and thus lowering their Rubisco

requirement to on the high end 8% of the nitrogen in the leaf (85). However, C4 plants require additional structures and enzymes to drive the carbon concentrating mechanism, which also add some burden to the cell not accounted for in this estimate.

Algae and cyanobacteria use an alternate carbon concentrating strategy, encasing Rubisco in a proteinaceous compartment named the carboxysome next to carbonic anhydrase (86). By pumping bicarbonate into the cell using integral membrane proteins, these organisms can then increase the CO_2 concentration in the carboxysome where carbonic anhydrase converts bicarbonate into CO_2. There is some ongoing work to express the carboxysome in land plants and also to convert important C3 plants into C4 plants, such as *Oryza sativa* (rice), but neither strategy has met with much success (79).

There is one last observation which gives hope to improving carbon fixation in plants. At low intensities of light, plants are efficient with photons sufficiently energetic enough to drive photosynthesis, using up to 80% of them. However, at half of the full illumination of the sun less than 25% of these photons are used, and at full sunlight levels, less than 10% (79). I would argue this is because the carbon fixation portion of photosynthesis isn't fast enough to keep up with the reducing equivalents being generated by the light reactions. Given this constraint, increasing the speed of carbon fixation is probably more important that increasing the energy efficiency (ATP hydrolysis or burning of reducing equivalents in redox reactions) because so much light is available but cannot be used productively. Between the carbon concentrating mechanisms, or side product removal of photorespiration, the cost of fixing a CO_2 into glyceraldehyde-3-phosphate is between 12 photons (C4 plants, algae, and cyanobacteria) and 13 photons (C3 plants) on the low end.

There are many available carboxylating enzymes in nature to select from if trying to design a new CO_2 fixation pathway. Acetyl-CoA carboxylase, propionyl-CoA carboxylase, pyruvate carboxylase, crotonyl-CoA carboxylase/reductase, PEP carboxykinase and PEP carboxylase all have better K_{cat} than Rubisco, some over ten times faster (87). These enzymes also don't have the problem of accidentally incorporating molecular oxygen into their products like Rubisco. In the absence of a carbon

concentrating mechanism, the catalytic efficiency (K_{cat}/K_m) is a better indicator of an enzymes speed since atmospheric concentrations of CO_2 are relatively low compared to the K_m of the potential carboxylases available. Since all carbon concentrating mechanisms will ultimately be based on enzymes in some capacity, a concentrating mechanism is only helpful if the enzymes doing the concentrating are faster than the enzymes doing the CO_2 fixing. When comparing carboxylases, the molecular weight of the enzymes also needs to be considered since flux per mg of protein of the pathway is ultimately the relevant metric.

There are a few other constraints when considering enzymes for a novel pathway. The final pathway should produce sugars efficiently since all the major organisms to integrate a new CO_2 fixation pathway into are evolved to build themselves out of sugars. Fewer enzymatic steps in the pathway are desirable to reduce overhead, and especially few if any novel enzymes since they are so difficult to engineer. The pathway needs to be thermodynamically feasible, with few or no toxic intermediates. Finally, I would argue that making single cellular organisms more efficient isn't as attractive as making land plants more efficient.

This is because CO_2 fixing microbes require vast quantities of water, some sort of facility to grow them in, and usually a CO_2 source other than ambient air. Open ponds are cheaper to build and run but are much less efficient at producing biomass since they are prone to predation and are water inefficient. Closed bioreactors have the problem of being hard to sterilize, more expensive to build and run, and requires a concentrated CO_2 source. If the CO_2 is coming from a fossil fuel (e.g. coal burning power plant) it's hard to imagine a scenario where the net CO_2 fixed in the process is zero or net negative, especially if no subsidies/regulations are involved. If going into land plants, finding ways of reducing the number of genetic engineering steps necessary to integrate the pathway is also important because plants are so hard to engineer and have long-life cycles. Also, limitations of the host's metabolism become important since certain cofactors and vitamins may be limiting or not present, thus making some enzymes or chemical reactions unattractive to pursue.

Recently a pathway that potentially fulfills most of these requirements has been developed. The crotonyl-CoA/ethylmalonyl-CoA/hydroxybutyryl-CoA (CETCH) cycle uses one of the fastest carboxylating enzymes from a specific activity metric (K_{cat}/mg protein) (82). The most ATP efficient version of the pathway hasn't been demonstrated where the FadH cofactor regeneration is linked to ATP synthesis via the electron transport chain. But since speed is more important than efficiency in land plants, the demonstrated pathway that regenerates FadH by making hydrogen peroxide from molecular oxygen seems promising since hydrogen peroxide can be detoxified by one of the fastest enzymes known, catalase. Unfortunately, the pathway produces glyoxylate as the final product, which is difficult to turn into sugars efficiently. In addition, many of the pathway intermediates are attached to CoA and may require CoA biosynthesis upregulation, which adds to the list of changes that may be required to beat the CBB cycle. Despite these concerns, the CETCH cycle is an impressive, and potentially world changing carbon fixation pathway.

4.5. Cheap electricity generation

A technology that could produce very cheap electricity with minimal, or zero, CO_2 emissions would probably have the highest impact on anthropogenic CO_2 emissions. Electricity and heat generation represent 26% of GHG emissions directly, but other sectors, including transport representing 14% GHG emissions can be largely electrified (air transport and waterway shipping notwithstanding) (49). In addition, cheap electricity would subsidize processes to directly remove CO_2 from the air (as discussed in "Direct CO_2 Capture from Ambient Air" section). Solar is already about 15-20% efficient and wind is 30-40% efficient from a photo gathering perspective. People frequently compare these efficiencies with the chemical energy stored in sugars by photosynthesis which is about 1-2% efficient (77), but this is an unfair comparison. As mentioned earlier photosynthesis is about 26% efficient at the photon gathering level. Additional losses in efficiency occur when capturing the CO_2 and converting it to sugars, problems both wind and solar will also incur but are frequently not considered when making comparisons.

Solar and wind represent an expanding portion of electricity generation. Although the energy production per unit area is relatively poor, both technologies are scalable at their current level of advancement, and economically viable without subsidies in many regions. The main problems with these technologies are the distributed nature of their production, which the power grid was not designed or optimized for; their intermittent production which is hard to match with demand; the geographic distribution of production and consumption frequently being far apart and capital costs.

Nuclear is another mature technology that could be scaled today. Although nuclear energy has many problems, including weapons proliferation and serious waste product generation, the biggest for energy is the limited amount of uranium, thought to be between 3.4 and 17 million metric tons, representing only 6-30 years' worth of world energy demand (77). Thorium reactors, which have recently gained renewed interest, are not expected to be economic to run while plutonium is still minable given current technologies and ultimately thorium is only four times as abundant as uranium.

Finally, fusion represents the most difficult, but potentially also the most world changing technology for cheap energy production. In theory it could provide enormous amounts of energy in an extremely small area. Of course, the promise of fusion has existed since the 1940s and suffers from the old joke that "Fusion has always been 50 years away, and always will be". Only recently has a fusion reactor been capable of releasing more energy than was put into it (although not when accounting for total energy, only energy that makes it into the reactor) (88). However if fusion could be fully developed, it might make electrical energy extremely cheap which would solve a great deal of resource dependent problems humans will encounter as we move towards 9 billion people in 2050 (89).

4.6. Conclusion

The technical side to solve CO_2 emissions is monumental but not unsolvable. However large amounts of sustained investment will be necessary to develop the necessary technologies. Should the global society fail in this endeavor, either standards of living will decrease, world population, or most likely both. History is replete with examples of societies that run into resource constraints, usually in the form of famine, and these times

are not fondly remembered to say the least. Given the technological development of the modern world we are at an unusual time in history when we can see the potential disasters long before they arrive. This is a truly unique capability humans have as a species. We should use it.

Chapter 5. References

Bibliography

1. Ajikumar PK, et al. (2008) Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. Mol Pharm 5(2):167–190.
2. Horton JD (2002) Sterol regulatory element-binding proteins: transcriptional activators of lipid synthesis. Biochem Soc Trans 30(Pt 6):1091–1095.
3. Unden G, Bongaerts J (1997) Alternative respiratory pathways of Escherichia coli: energetics and transcriptional regulation in response to electron acceptors. Biochim Biophys Acta 1320(3):217–234.
4. Zhang FL, Casey PJ (1996) Protein prenylation: molecular mechanisms and functional consequences. Annu Rev Biochem 65:241–269.
5. Prossnitz ER, Maggiolini M (2009) Mechanisms of estrogen signaling and gene expression via GPR30. Mol Cell Endocrinol 308(1-2):32–38.
6. Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol 3(7):408–414.
7. Kingston DGI (2007) The shape of things to come: structural and synthetic studies of taxol and related compounds. Phytochemistry 68(14):1844–1854.
8. Denby CM, et al. (2018) Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer. Nat Commun 9(1):965.
9. Henke NA, et al. (2018) Patchoulol Production with Metabolically Engineered Corynebacterium glutamicum. Genes (Basel) 9(4).
10. Lindberg P, Park S, Melis A (2010) Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism. Metab Eng 12(1):70–79.
11. Alonso-Gutierrez J, et al. (2013) Metabolic engineering of Escherichia coli for limonene and perillyl alcohol production. Metab Eng 19:33–41.
12. Peralta-Yahya PP, et al. (2011) Identification and microbial production of a terpene-based advanced biofuel. Nat Commun 2:483.
13. George KW, et al. (2015) Metabolic engineering for the high-yield production of isoprenoid-based C_5 alcohols in E. coli. Sci Rep 5:11128.

14. Chou HH, Keasling JD (2012) Synthetic pathway for production of five-carbon alcohols from isopentenyl diphosphate. Appl Environ Microbiol 78(22):7849–7855.
15. Jiang J, He X, Cane DE (2007) Biosynthesis of the earthy odorant geosmin by a bifunctional Streptomyces coelicolor enzyme. Nat Chem Biol 3(11):711–715.
16. Tsurumaru Y, et al. (2012) HlPT-1, a membrane-bound prenyltransferase responsible for the biosynthesis of bitter acids in hops. Biochem Biophys Res Commun 417(1):393–398.
17. Wang C-M, Cane DE (2008) Biochemistry and molecular genetics of the biosynthesis of the earthy odorant methylisoborneol in Streptomyces coelicolor. J Am Chem Soc 130(28):8908–8909.
18. Ignea C, et al. (2018) Synthesis of 11-carbon terpenoids in yeast using protein and metabolic engineering. Nat Chem Biol 14(12):1090–1098.
19. Schooley DA, Judy KJ, Bergot BJ, Hall MS, Siddall JB (1973) Biosynthesis of the Juvenile Hormones of Manduca sexta: Labeling Pattern from Mevalonate, Propionate, and Acetate. Proc Natl Acad Sci USA 70(10):2921–2925.
20. Cascón O, et al. (2012) Chemoenzymatic preparation of germacrene analogues. Chem Commun 48(78):9702–9704.
21. Hamilton JGC, Hooper AM, Pickett JA, Mori K, Sano S (1999) 3-Methyl-α-himachalene is confirmed, and the relative stereochemistry defined, by synthesis as the sex pheromone of the sandfly Lutzomyia longipalpis from Jacobina, Brazil. Chem Commun (4):355–356.
22. Attygalle AB, Morgan ED (1982) Structures of homofarnesene and bishomofarnesene isomers from Myrmica ants. J Chem Soc, Perkin Trans 1:949.
23. Kinjoh T, et al. (2007) Control of juvenile hormone biosynthesis in Bombyx mori: cloning of the enzymes in the mevalonate pathway and assessment of their developmental expression in the corpora allata. Insect Biochem Mol Biol 37(8):808–818.
24. Sen SE, et al. (2007) Purification, properties and heteromeric association of type-1 and type-2 lepidopteran farnesyl diphosphate synthases. Insect Biochem Mol Biol 37(8):819–828.

25. Sen SE, et al. (2012) Cloning, expression and characterization of lepidopteran isopentenyl diphosphate isomerase. Insect Biochem Mol Biol 42(10):739–750.
26. Aaron JA, Lin X, Cane DE, Christianson DW (2010) Structure of epi-isozizaene synthase from Streptomyces coelicolor A3(2), a platform for new terpenoid cyclization templates. Biochemistry 49(8):1787–1797.
27. Polakowski T, Stahl U, Lang C (1998) Overexpression of a cytosolic hydroxymethylglutaryl-CoA reductase leads to squalene accumulation in yeast. Appl Microbiol Biotechnol 49(1):66–71.
28. Ham TS, et al. (2012) Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. Nucleic Acids Res 40(18):e141.
29. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. Science 291(5509):1790–1792.
30. Lee TS, et al. (2011) BglBrick vectors and datasheets: A synthetic biology platform for gene expression. J Biol Eng 5:12.
31. Xie X, Kirby J, Keasling JD (2012) Functional characterization of four sesquiterpene synthases from Ricinus communis (castor bean). Phytochemistry 78:20–28.
32. Yoshikuni Y, Ferrin TE, Keasling JD (2006) Designed divergent evolution of enzyme function. Nature 440(7087):1078–1082.
33. Baba T, et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006.0008.
34. Luo X, et al. (2019) Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. Nature 567(7746):123–126.
35. Haldar K, Bhattacharjee S, Safeukui I (2018) Drug resistance in Plasmodium. Nat Rev Microbiol 16(3):156–170.
36. Yu F, et al. (2008) Molecular cloning and functional characterization of alpha-humulene synthase, a possible key enzyme of zerumbone biosynthesis in shampoo ginger (Zingiber zerumbet Smith). Planta 227(6):1291–1299.
37. Kamita SG, Samra AI, Liu J-Y, Cornel AJ, Hammock BD (2011) Juvenile hormone (JH) esterase of the mosquito Culex quinquefasciatus is not a target of the JH analog insecticide methoprene. PLoS One 6(12):e28392.

38. Aldor IS, Kim S-W, Prather KLJ, Keasling JD (2002) Metabolic engineering of a novel propionate-independent pathway for the production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) in recombinant Salmonella enterica serovar typhimurium. Appl Environ Microbiol 68(8):3848–3854.
39. Richardson SM, Nunley PW, Yarrington RM, Boeke JD, Bader JS (2010) GeneDesign 3.0 is an updated synthetic biology toolkit. Nucleic Acids Res 38(8):2603–2606.
40. Cusson M, et al. (2006) Characterization and tissue-specific expression of two lepidopteran farnesyl diphosphate synthase homologs: implications for the biosynthesis of ethyl-substituted juvenile hormones. Proteins 65(3):742–758.
41. Chen J, Densmore D, Ham TS, Keasling JD, Hillson NJ (2012) DeviceEditor visual biological CAD canvas. J Biol Eng 6(1):1.
42. Hillson NJ, Rosengarten RD, Keasling JD (2012) j5 DNA assembly design automation software. ACS Synth Biol 1(1):14–21.
43. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6(5):343–345.
44. Pfefferkorn JA, et al. (2007) Design and synthesis of novel, conformationally restricted HMG-CoA reductase inhibitors. Bioorg Med Chem Lett 17(16):4531–4537.
45. Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3(11):e3647.
46. Thompson MG, et al. (2018) Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. Sci Rep 8(1):1590.
47. Batth TS, et al. (2014) A targeted proteomics toolkit for high-throughput absolute quantification of Escherichia coli proteins. Metab Eng 26:48–56.
48. Morrell WC, et al. (2017) The Experiment Data Depot: A Web-Based Software Tool for Biological Experimental Data Storage, Sharing, and Visualization. ACS Synth Biol 6(12):2248–2259.
49. Intergovernmental Panel on Climate Change Working Group III,, Edenhofer O (2014) Climate change 2014: Mitigation of climate change : Working Group III contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge University Press, New York, NY).

50. Neupane B, Konda NVSNM, Singh S, Simmons BA, Scown CD (2017) Life-Cycle Greenhouse Gas and Water Intensity of Cellulosic Biofuel Production Using Cholinium Lysinate Ionic Liquid Pretreatment. ACS Sustain Chem Eng 5(11):10176–10185.
51. Hansen AC, Zhang Q, Lyne PWL (2005) Ethanol-diesel fuel blends -- a review. Bioresour Technol 96(3):277–285.
52. Zargar A, et al. (2017) Leveraging microbial biosynthetic pathways for the generation of "drop-in" biofuels. Curr Opin Biotechnol 45:156–163.
53. Haushalter RW, et al. (2014) Production of anteiso-branched fatty acids in Escherichia coli; next generation biofuels with improved cold-flow properties. Metab Eng 26:111–118.
54. Yang Y, Dec JE, Dronniou N, Simmons B (2010) Characteristics of Isopentanol as a Fuel for HCCI Engines. SAE Int J Fuels Lubr 3(2):725–741.
55. Monroe E, et al. (2019) Discovery of novel octane hyperboosting phenomenon in prenol biofuel/gasoline blends. Fuel 239:1143–1148.
56. George KW, et al. (2014) Correlation analysis of targeted proteins and metabolites to assess and engineer microbial isopentenol production. Biotechnol Bioeng 111(8):1648–1658.
57. Atsumi S, Hanai T, Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451(7174):86–89.
58. Kang A, et al. (2016) Isopentenyl diphosphate (IPP)-bypass mevalonate pathways for isopentenol production. Metab Eng 34:25–35.
59. Gronenberg LS, Marcheschi RJ, Liao JC (2013) Next generation biofuel engineering in prokaryotes. Curr Opin Chem Biol 17(3):462–471.
60. Flint DH, Emptage MH, Finnegan MG, Fu W, Johnson MK (1993) The role and properties of the iron-sulfur cluster in Escherichia coli dihydroxy-acid dehydratase. J Biol Chem 268(20):14732–14742.
61. Bogorad IW, Lin T-S, Liao JC (2013) Synthetic non-oxidative glycolysis enables complete carbon conservation. Nature 502(7473):693–697.

62. Lin PP, et al. (2018) Construction and evolution of an Escherichia coli strain relying on nonoxidative glycolysis for sugar catabolism. Proc Natl Acad Sci USA 115(14):3538–3546.
63. Meadows AL, et al. (2016) Rewriting yeast central carbon metabolism for industrial isoprenoid production. Nature 537(7622):694–697.
64. Li Y, Luxenburger E, Müller R (2013) An alternative isovaleryl CoA biosynthetic pathway involving a previously unknown 3-methylglutaconyl CoA decarboxylase. Angew Chem Int Ed Engl 52(4):1304–1308.
65. Boisart C, Letellier G, Others New biosynthesis pathway for prenol in a recombinant microorganism.
66. Vögeli B, et al. (2018) Archaeal acetoacetyl-CoA thiolase/HMG-CoA synthase complex channels the intermediate via a fused CoA-binding site. Proc Natl Acad Sci USA 115(13):3380–3385.
67. Flamholz A, Noor E, Bar-Even A, Milo R (2012) eQuilibrator--the biochemical thermodynamics calculator. Nucleic Acids Res 40(Database issue):D770–5.
68. Bond-Watts BB, Bellerose RJ, Chang MCY (2011) Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. Nat Chem Biol 7(4):222–227.
69. Wong BJ, Gerlt JA (2004) Evolution of function in the crotonase superfamily: (3S)-methylglutaconyl-CoA hydratase from Pseudomonas putida. Biochemistry 43(16):4646–4654.
70. Oberortner E, Cheng J-F, Hillson NJ, Deutsch S (2017) Streamlining the Design-to-Build Transition with Build-Optimization Software Tools. ACS Synth Biol 6(3):485–496.
71. Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. J Comput Biol 7(1-2):203–214.
72. Eberhard ED, Gerlt JA (2004) Evolution of function in the crotonase superfamily: the stereochemical course of the reaction catalyzed by 2-ketocyclohexanecarboxyl-CoA hydrolase. J Am Chem Soc 126(23):7188–7189.

73. Lange BM, Croteau R (1999) Isopentenyl diphosphate biosynthesis via a mevalonate-independent pathway: isopentenyl monophosphate kinase catalyzes the terminal enzymatic step. Proc Natl Acad Sci USA 96(24):13714–13719.
74. Kang A, Meadows CW, Canu N, Keasling JD, Lee TS (2017) High-throughput enzyme screening platform for the IPP-bypass mevalonate pathway for isopentenol production. Metab Eng 41:125–134.
75. Liu Y, Yan Z, Lu X, Xiao D, Jiang H (2016) Improving the catalytic activity of isopentenyl phosphate kinase through protein coevolution analysis. Sci Rep 6:24117.
76. Hayakawa H, et al. (2018) Modified mevalonate pathway of the archaeon Aeropyrum pernix proceeds via trans-anhydromevalonate 5-phosphate. Proc Natl Acad Sci USA 115(40):10034–10039.
77. Hoffert MI, et al. (2002) Advanced technology paths to global climate stability: energy for a greenhouse planet. Science 298(5595):981–987.
78. Keith DW, Holmes G, St. Angelo D, Heidel K (2018) A Process for Capturing CO 2' ' from the Atmosphere. Joule 2(8):1573–1594.
79. Zhu X-G, Long SP, Ort DR (2010) Improving photosynthetic efficiency for greater yield. Annu Rev Plant Biol 61:235–261.
80. Berg IA (2011) Ecological aspects of the distribution of different autotrophic CO2 fixation pathways. Appl Environ Microbiol 77(6):1925–1936.
81. Figueroa IA, et al. (2018) Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO2 fixation pathway. Proc Natl Acad Sci USA 115(1):E92–E101.
82. Schwander T, Schada von Borzyskowski L, Burgener S, Cortina NS, Erb TJ (2016) A synthetic pathway for the fixation of carbon dioxide in vitro. Science 354(6314):900–904.
83. Flombaum P, et al. (2013) Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. Proc Natl Acad Sci USA 110(24):9824–9829.
84. Bar-On YM, Phillips R, Milo R (2018) The biomass distribution on Earth. Proc Natl Acad Sci USA 115(25):6506–6511.

85. Raven JA (2013) Rubisco: still the most abundant protein of Earth? New Phytol 198(1):1–3.
86. Kinney JN, Axen SD, Kerfeld CA (2011) Comparative analysis of carboxysome shell proteins. Photosyn Res 109(1-3):21–32.
87. Cotton CA, Edlich-Muth C, Bar-Even A (2018) Reinforcing carbon fixation: CO2 reduction replacing and supporting carboxylation. Curr Opin Biotechnol 49:49–56.
88. Hurricane OA, et al. (2014) Fuel gain exceeding unity in an inertially confined fusion implosion. Nature 506(7488):343–348.
89. Lutz W, Sanderson W, Scherbov S (2001) The end of world population growth. Nature 412(6846):543–545.