

# Reverse $\beta$ -oxidation pathways for efficient chemical production

Katia Tarasava<sup>1</sup>, Seung Hwan Lee<sup>1</sup>, Jing Chen<sup>1</sup>, Michael Köpke<sup>2</sup>, Michael C. Jewett<sup>3</sup>, and  
Ramon Gonzalez<sup>1\*</sup>

<sup>1</sup>Department of Chemical, Biological, and Materials Engineering, University of South Florida,  
Tampa, FL, USA

<sup>2</sup>LanzaTech Inc., Skokie, IL, USA

<sup>3</sup>Department of Chemical and Biological Engineering and Center for Synthetic Biology,  
Northwestern University, Evanston, IL, USA

\* To whom correspondence should be addressed:

Department of Chemical, Biological, and Materials Engineering, University of South Florida,  
ENB 118, 4202 E Fowler Ave, Tampa, FL 33620, USA.

Phone: (813)-974-1223; Email: ramongonzalez@usf.edu

## Abstract

Microbial production of fuels, chemicals and materials has the potential to reduce greenhouse gas emissions and contribute to a sustainable bioeconomy. While synthetic biology allows readjusting of native metabolic pathways for the synthesis of desired products, often these native pathways do not support maximum efficiency and are affected by complex regulatory mechanisms. A synthetic or engineered pathway that allows modular synthesis of versatile bioproducts with minimal enzyme requirement and regulation while achieving high carbon and energy efficiency could be an alternative solution to address these issues. The reverse  $\beta$ -oxidation (rBOX) pathways enable iterative non-decarboxylative elongation of carbon molecules of varying chain lengths and functional groups with only four core enzymes and no ATP requirement. Here we describe recent developments in rBOX pathway engineering to produce alcohols and carboxylic acids with diverse functional groups, along with other commercially important molecules such as polyketides. We discuss the application of rBOX beyond the pathway itself by its interfacing with various carbon-utilization pathways and deployment in different organisms, which allows feedstock diversification from sugars to glycerol, carbon dioxide, methane and other substrates.

## Introduction

Living systems can perform complex chemical reactions, a property that can be exploited for the synthesis of industrial products from renewable resources. Often the products that present commercial interest are higher chain length compounds synthesized from smaller (2-3 carbon atom) precursor metabolites through carbon chain elongation pathways. Among these are fatty acids and alcohols (Cho et al., 2020), aldehydes (Kunjapur & Prather, 2015), hydrocarbons

(Kang & Nielsen, 2017), isoprenoids (Lange et al., 2000), polyketides (Cai & Zhang, 2018) and other molecules that represent valuable fuels and commodity and fine chemicals. At its core, carbon chain elongation requires condensation reactions often carried out through the ketoacid pathway (Atsumi et al., 2008; Marcheschi et al., 2012), fatty acid (Lennen & Pfeleger, 2013; Marella et al., 2018), polyketide (Cai & Zhang, 2018; Staunton & Weissman, 2001; Zhou et al., 2020) and isoprenoid (Lange et al., 2000) biosynthesis. These natural pathways have inherent limitations to the product yield due to inefficient carbon and energy utilization, low flux, precursor competition with other essential metabolic pathways and complex intrinsic regulation.

The reverse  $\beta$ -oxidation (rBOX) pathway circumvents many of these challenges using non-decarboxylative Claisen condensation reactions catalyzed by 3-ketoacyl-CoA thiolases (hereafter referred to as thiolases) to catalyze the condensation of acyl-CoA substrates and provide a carbon and energy efficient way to manufacture important molecules of commercial interest. The iterative nature of rBOX pathways make them amenable to synthesizing an array of different chain length molecules, from C3 up to C16 (and theoretically beyond) (Cintolesi et al., 2014; Mehrer et al., 2018; Vick et al., 2015), while choice of termination pathway allows to tailor the product chain length and class. The most accessible product classes of rBOX pathways are alcohols and carboxylic acids; however, rBOX can be engineered to incorporate functionalized primer and extender units, as well as to incorporate novel biochemistries. This expanded functionality makes the rBOX capable of synthesizing a wide range of products, such as  $\alpha,\omega$ -dicarboxylic acids and diols,  $\omega$ -hydroxyacids and other functionalized fatty acid derivatives,  $\beta$ -hydroxy- $\omega$ -lactones, methyl alcohols and others (Cheong et al., 2016).

Furthermore, elements of the rBOX carbon chain elongation pathway can be combined with

other biochemistries to create new synthetic pathways for the efficient production of more complex molecules such as polyketides (Tan et al., 2020).

This review will focus on advances in expanding the functionality of rBOX-based products. We will cover the basic principles for design of rBOX pathways, strategies for diversifying the array of products within the established rBOX framework, as well as repurposing of this platform via integration with other metabolic pathways, creating novel synthetic routes for production of diverse molecule classes. We will also provide a future perspective on the rBOX architecture in the context of genome and protein engineering, dynamic and orthogonal regulation, and its transferability to other organisms.

### **Concept and Reactions of the reverse $\beta$ -Oxidation (rBOX) Pathway**

As the name implies, the rBOX function relies on reversing the direction of the  $\beta$ -oxidation cycle. The native  $\beta$ -oxidation can be functionally reversed by circumventing the regulatory system of the fatty acid degradation pathway and generating a thermodynamic pull in the direction of chain elongation. The initial rBOX system engineering accomplished constitutive expression the  $\beta$ -oxidation system through introduction of *fadR* and *atoC(c)* mutations, deletion of *arcA* gene and circumventing the carbon catabolite repression with cAMP-independent *crp*\* mutant (Dellomonaco et al., 2011). While effective, this system-level approach made it difficult to determine which of the many deregulated enzymes are responsible for product synthesis, which in turn limits efforts to fine-tune the synthesis of specific products and prevents the transfer of the engineered pathway to other organisms. Therefore, the system was subsequently redesigned using a bottom-up synthetic biology approach that allowed to isolate and elucidate

the function of each individual component and create a modular rBOX framework that can be readily engineered and controlled (Clomburg et al., 2012).

The rBOX framework offers some unique advantages over other biosynthetic pathways. Compared to fatty acid and polyketide biosynthesis pathways, it circumvents the loss of CO<sub>2</sub>, increasing carbon yield and efficiency. Moreover, the rBOX cycle reactions do not require ATP, contributing to its energy efficiency. There are no inherent limitations to flux, as acetyl-CoA is a common precursor, and allows utilization of different carbon sources. Each cycle requires two reducing equivalents in the form of NADH, which alleviates the need for supply of NADPH through the pentose phosphate pathway or other means. The biochemistry of the pathway enzymes allows for flexible precursor input and a diverse range of exit points, providing excellent flexibility and combinatorial capabilities to produce a wide range of products. This type of reaction circumvents the loss of CO<sub>2</sub> associated with the fatty acid and polyketide biosynthesis pathways, which are based on decarboxylative Claisen condensations, thus providing carbon economy. The rBOX also avoids complex regulatory mechanisms associated with the generation and/or utilization of malonyl-CoA/ACP.

### **Modular Framework for rBOX Pathways**

The general framework of the rBOX includes functional modules for priming, elongation and termination with each cycle adding two carbons to the growing chain (Figure 1). The priming module initiates the cycle by combining one primer and extender unit in a non-decarboxylative Claisen condensation reaction catalyzed by a 3-ketoacyl-CoA thiolase (Liu et al., 2020). During the reaction a new carbon-carbon bond is formed between an acyl-CoA primer and an acetyl-CoA extender unit to form a  $\beta$ -keto ester. The standard primer and extender units

of the rBOX pathway are two acetyl-CoA molecules; however, native and engineered thiolases can accept different types of functionalized acyl-CoAs as both primers and extenders (Bonk et al., 2018; Cheong et al., 2016; Kallscheuer et al., 2017; Liu et al., 2020). The thiolase-catalyzed reaction also acts as one of the control points for specifying the desired product by limiting the chain length based on its substrate specificity (Bonk et al., 2018; Dekishima et al., 2011; Kim et al., 2015; Kim & Gonzalez, 2018). The reversible thiolase condensation reaction is also a potential pathway bottleneck due to unfavorable thermodynamics (Kim, Cheong, Chou, et al., 2016). Therefore, to drive the reaction in the desired direction of condensation, the levels of intracellular acetyl-CoA should be high and 3-ketoacyl-CoA low.

The elongation module consists of the four “core” rBOX enzymes that perform the four steps of the elongation cycle. The first of these steps is also catalyzed by a 3-ketoacyl-CoA thiolase, which performs the same condensation reaction as the priming step but for subsequent cycles of elongation. The contextual distinction here lies in the difference of the priming and elongation units, which allows to incorporate different functional groups into the growing molecule. By using thiolases with unique substrate specificity, it is possible to segregate these two reactions and use primer and extender units with different functionalization (Cheong et al., 2016). The subsequent  $\beta$ -reduction steps of the rBOX cycle are catalyzed by hydroxyacyl-CoA dehydrogenase (HACD), enoyl-CoA hydratase (ECH) and enoyl-CoA reductase (ECR). These reactions generate an acyl-CoA intermediate two carbons longer than the initial primer, which can enter subsequent cycles of elongation. Different elongation enzyme variants possess different substrate length specificities, which can be exploited to customize product chain length (Machado et al., 2012). It has also been shown that the combination of type II fatty acid biosynthesis enzymes and thiolases supports a functional rBOX (Clomburg et al., 2018).

Replacing the  $\beta$ -oxidation enzymes HACD, ECH, and ECR with the equivalent enzymes from the bacterial type II fatty acid biosynthesis (FAB II) system (FabG, FabZ and FabI) allows to exploit promiscuity of FAB enzymes, as they not only accept the ACP but also CoA thioesters. The ability of some  $\beta$ -reduction enzymes to accept functionalized substrates makes it possible to create small molecules that were previously inaccessible by native pathways.

The exit from the cycle is facilitated by various termination enzymes, such as thioesterases, acyl-CoA reductases, aldehyde dehydrogenases, etc. that specify the resulting product. In terms of termination module, fatty acid formation by thioesterases may be considered the “standard” pathway for exit from the rBOX cycle. Thioesterases are ubiquitous and abundant, producing a range of different length (C4-C14) carboxylic acids (Clomburg et al., 2012; Vick et al., 2015). This presents a problem for targeting higher molecular weight products or producing other product classes, such as alcohols, as native thioesterases can compete with acyl-CoA reductases/alcohol dehydrogenases for flux. However, several strategies have been successfully employed to customize the product length and type. Kim *et al.* 2015 achieved high-yield production of medium-chain alcohols through deletion of native *E. coli* thioesterases *yciA*, *ybgC*, *ydiI*, *tesA*, *fadM* and *tesB*. This combined with overexpression of a selection of acyl-CoA reductase (ACR) variants from different organisms showed that rBOX strains are capable of producing alcohols in the C6-C10 range (Kim et al., 2015) (Table 1).

### **Overlapping Chain Length Specificity of Multiple Enzymes**

Since rBOX is an iterative pathway and can generate an array of different chain-length products, one of the biggest challenges is tailoring the product length. As mentioned above, almost every rBOX enzyme represents a point of control that can either allow or restrict flux

through the associated nodes. The ability to mix-and-match specificity range of different rBOX modules allows for tailoring of product length, as was demonstrated in the production of decanoic acid (Kim & Gonzalez, 2018). This was achieved by selecting rBOX elongation enzymes that support the synthesis of acyl-CoA molecules of up to 10 carbons (using a broad-range BktB thiolase from *Cupriavidus necator* (previously known as *Ralstonia Eutropha*) (Chen et al., 2015; Slater et al., 1998) and egTER enoyl-CoA reductase from *Euglena gracilis*) and a termination module (FadM thioesterase) that exhibited high activity toward decanoyl-CoA and longer-chain acyl-CoAs. The final engineered strain produced 2.1 g/L of decanoic acid as the primary fermentation product with a yield of 0.1 g/g glycerol (Table 1).

### Pathway Termination at Different Nodes

An alternative approach to customizing rBOX products is cycle termination from different nodes which result in formation of alternative product classes with different degrees of reduction and functionalization. For example, termination at the first intermediate in the rBOX cycle (3-ketoacyl-CoA), followed by a CoA group transfer and decarboxylation can produce methyl ketones. This was shown by Lan *et al.* 2013 using a combination of BktB (thiolase), Hbd (HACD), Crt (ECH) and Ter (ECR) and a termination module consisting of 3-oxoadipate CoA-succinyl transferase PcaIJ from *Pseudomonas putida*, which acts on 3-ketohexanoyl-CoA, and an acetoacetate decarboxylase (Ade). Two rounds of rBOX followed by transferase and decarboxylase steps yielded 0.24 g/L of 2-pentanone. Similarly, using a CoA transferase Reut\_1331 and Reut\_1332 led to production of 0.7 g/L of acetone (Lan et al., 2013).

Cycle termination at the 3-hydroxyacyl-CoA node results in a number of important products, including 3-hydroxycarboxylic acids and 1,3-diols. For example, an incomplete rBOX cycle terminating after the initial condensation (by thiolases BktB, Thl or PhaA) and



dehydrogenation (by HACDs Hbd or PhaB) steps, followed by hydrolysis of CoA thioester by a thioesterase resulted in production of enantiomerically pure (*R*)-3-hydroxybutyrate and (*S*)-3-hydroxybutyrate to concentrations of 2.92 g/L and 2.08 g/L, respectively (Tseng et al., 2009). This route was subsequently modified to produce  $22.5 \pm 5.9$  mg/L 3-hydroxyhexanoate by supplying precursor substrate butyrate and expressing a broad substrate specificity CoA activation enzyme propionyl-CoA transferase (Pct) from *Megasphaera elsdenii* (Martin et al., 2013). Clomburg *et al.* 2012 also reported 2.5 g/L of (*S*)-3-hydroxybutyric acid (using glycerol as a carbon source) by overexpression of *atoB* (thiolase) and *fadB* (HACD) in a mixed acid fermentation deficient strain, with a 0.29 g/g yield (Clomburg et al., 2012). Termination at the 3-hydroxyacyl-CoA node can result in the production of 1,3-butanediol (1,3-BDO). Gulevich reported 27 mg/L of 1,3-BDO using aldehyde dehydrogenase (*mhpF*) from *E. coli* to reduce 3-hydroxybutyryl-CoA into 3-hydroxybutyraldehyde, followed by the use of alcohol dehydrogenase (*fucO* or *yqhD*) to yield 1,3-BDO (Gulevich et al., 2016).

The third node of the rBOX pathway, trans-enoyl-CoA, can be used as an exit point from the cycle generating  $\alpha,\beta$ -unsaturated carboxylic acids or to regenerate the acyl-CoA intermediate that enters into the next round of rBOX. A native *E. coli* thioesterase YdiI was used to convert trans-enoyl-CoA to the corresponding  $\alpha,\beta$ -unsaturated carboxylic acids. Its expression in combination with a partial rBOX cycle overexpressing thiolase BktB and native FadB (performing dual HACD and ECH) resulted in crotonic acid production at titers reaching 3.2 g/L with a yield of 0.12 g crotonic acid/g glycerol. The omission of enoyl-CoA reductase (ECR) enzyme prevented competition for substrate for crotonic acid production. Alternatively, including egTER (ECR) led to production of 2-hexenoic acid, 2-octenoic acid and 2-decenoic

acid at a titer of 0.2 g/L demonstrating for the first time microbial production of unsaturated carboxylic acids (Kim, Cheong, & Gonzalez, 2016).

### Using Alternative Primer and Extender Units

The standard rBOX pathway using acetyl-CoA as both primer and extender units results in even-numbered carbon chain products. This apparent limitation can be overcome by removing a carbon as a part of the termination pathway (Lan et al., 2013) or by supplying odd-chain primer molecules. For example, Tseng et al. 2010 engineered a pathway for production of chiral 3-hydroxyvalerate (3HV) through condensation of propionyl-CoA and acetyl-CoA (Tseng et al., 2010). They showed that propionate could be supplied exogenously and activated by a Ptb-Buk complex or endogenously produced from glucose or glycerol via overexpressing threonine biosynthesis pathway. Propionyl-CoA was fed into a one-turn incomplete rBOX cycle expressing thiolase BktB (which was demonstrated to accept C3 substrates), PhaB or Hbd (HACD) and TesB to produce 0.31 g/L and 0.50 g/L of (*S*)-3HV and (*R*)-3HV from glucose.

This approach was further expanded to produce odd-chain alcohols (pentanol) (Tseng & Prather, 2012) and C3-C10 alkanes by including a complete list of rBOX cycle enzymes (thiolases BktB from *Cupriavidus necator* or ThlA from *Clostridium acetobutylicum*, reductase PhaB and dehydratase PhaJ from *Cupriavidus necator* and enoyl-CoA reductase Ter from *Treponema denticola*) and using a termination module consisting of endogenous thioesterases, a broad-specificity carboxylic acid reductase (CAR) and aldehyde decarbonylase (AD) to convert  $C_n$  fatty acids into corresponding  $C(n-1)$  alkanes via a  $C_n$  fatty aldehyde intermediate (Sheppard et al., 2016). Another example of utilizing propionate as a building block activated by *Megaphaera elsdenii* PCT (mePCT) and with AtoB (thiolase), FadBA (thiolase, HACD, ECH) and FabI (ECR) as priming and elongation modules, produced odd-chain carboxylic acids

ranging from C5 to C11 at titers from 187 to 30.9 mg/L, respectively (Vick et al., 2015). This study also showed that endogenous *E. coli* enoyl-ACP reductase FabI can act as the ECR supporting rBOX function.

Besides straight-chain aliphatic compounds, rBOX can also be utilized for production of branched or other functionalized compounds by feeding primer and/or extender units with various functional groups. While the use of propionyl-CoA as primer leads to the synthesis of straight, odd-chain products, when used as extender unit it yields  $\alpha$ -branched products such as 2-methyl-2-enoic acids and 2-methyl acids. Up to 3.9 g/L tiglic acid was produced in a controlled bioreactor through an rBOX pathway consisting of FadAx, FadB1x and FadB2x along with the mePCT transferase for activation of supplemented precursor propionate (Cheong et al., 2016). In a separate study, ketoreductase was identified to be the key driver for selectivity, forming predominantly  $\alpha$ -branched products even when paired with a thiolase that highly prefers unbranched products. Leveraging the specificity of this ketoreductase from *Ascaris suum* allowed production of chiral 2-methyl-3-hydroxy acids ( $1.1 \pm 0.2$  g/L) or branched enoic acids ( $1.12 \pm 0.06$  g/L) at 44% and 87% yield of fed propionate, respectively (Blaisse et al., 2017). Synthesis of  $\omega$ -1 branched product is also possible by using a branched acyl-CoA primer such as isobutyryl-CoA. For example, Martin et al. 2013 reported production of 3-hydroxy-4-methylvalerate through an rBOX pathway consisting of BktB and PhaB along with the mePCT for activation of isobutyrate (Martin et al., 2013). Similarly, Sheppard et al. 2014 produced 4-methylpentanol at  $192 \pm 23$  mg/L (Sheppard et al., 2014). Clomburg et al. 2018 reported that *E. coli* type II fatty acid biosynthesis enzymes (FabG, FabZ and FabI) can produce  $\omega$ -1 branched product, 4-methylvalerate, through rBOX at higher titer than rBOX enzymes (FadB, egTER and PhaJ) (Clomburg et al., 2018).

Another variant of the rBOX pathway composed of PaaJ (thiolase), PaaH1 or Hbd (HACD), Crt or Ech (ECH) and *Euglena gracilis* Ter (TER) can use  $\omega$ -carboxylated succinyl-CoA as the primer to produce adipic acid, an important bulk chemical used as the monomer of nylon-6,6. Using Ptb (phosphate butyryltransferase) and Buk1 (butyryl kinase) as a termination module allowed to carry out full biosynthesis of adipic acid from glucose in *E. coli* (Yu et al., 2014). Similarly, natively produced malonyl-CoA can be used as a primer to form an odd-chain dicarboxylate, glutarate, expressing rBOX pathway genes from *Thermobifida fusca* specific for dicarboxylate synthesis (Zhao et al., 2018). A more comprehensive study reported a modular rBOX approach for using different  $\omega$ - and  $\omega$ -1-functionalized primers and  $\alpha$ -functionalized extender units in combination with various termination pathways (Cheong et al., 2016). It demonstrated using 7 different primers and 3 extender units to synthesize 18 functionalized compounds belonging to 10 different classes: phenylalkanoic, dicarboxylic,  $\omega$ -hydroxy, 4-oxo, 4-methyl, 2-methyl, 2-methyl-2-enolic and 2,3-dihydroxy acids,  $\beta$ -hydroxy- $\omega$ -lactones and 4-methyl alcohols (Cheong et al., 2016).

### Termination Through Alpha and Omega Oxidation

Additional modifications to the rBOX termination module allow to further expand its functionality. In one example of this approach, the core rBOX modules (BktB/Thiolase, FadB/HACD-ECH and egTer/TER), along with thioesterase YdiI for cycle termination, were combined with  $\omega$ -oxidation pathways by expressing  $\omega$ -hydroxylase (AlkBGT), alcohol dehydrogenase (ChnD or YjgB) and aldehyde dehydrogenase (ChnE) to convert carboxylic acids to  $\omega$ -hydroxyacids and/or dicarboxylic acids (Clomburg et al., 2015). AlkBGT expression resulted in production of > 0.8 g/L of C<sub>6</sub>, C<sub>8</sub>, and C<sub>10</sub>  $\omega$ -hydroxyacids from glycerol (Clomburg

et al., 2015). In a second  $\omega$ -oxidation step, overexpression of alcohol dehydrogenase ChnD and aldehyde dehydrogenase ChnE further lead to production of nearly 0.5 g/L of C<sub>6</sub>-C<sub>10</sub> diacids.

The rBOX platform can also be extended by  $\alpha$ -oxidation. This pathway is characterized by the shortening of fatty acids by one carbon, resulting in odd-chain fatty aldehydes, which can be further converted to other products. Although this pathway has not been demonstrated with rBOX, it has been applied to modification of fatty acids produced through fatty acid biosynthesis pathway. For example, Kaehne *et al.* 2011 engineered *E. coli* expressing  $\alpha$ -dioxygenase to convert C<sub>10</sub>-C<sub>18</sub> fatty acids into shortened fatty aldehydes (Kaehne et al., 2011). Cao *et al.* 2015 went further by engineering *E. coli* to use  $\alpha$ -dioxygenase in combination with an aldehyde reductase to produce about 100 mg/L odd chain C<sub>11</sub>-C<sub>15</sub> fatty alcohols from glycerol (Cao et al., 2015).

### Combination Approach

Some of the examples listed above involve a combination of different approaches to synthesize specific rBOX-derived products. For example, Martin et al. 2013 used functionalized primers (like glycolyl-CoA) and a termination at the 3-hydroxyacyl-CoA node to produce diverse dihydroxyacids, such as 3,4-DHBA. The system was supplemented with glycolate, which was converted to glycolyl-CoA by a broad substrate specificity mePCT and a propionyl-CoA synthetase (PrpE) from *Salmonella typhimurium* LT2, in addition to Ptb-Buk. This approach allowed synthesis of five novel products: 3,4-DHBA, 3HBL, 2,3-dihydroxybutyric acid (2,3-DHBA), 3-hydroxyhexanoic acid (3HH) and 3-hydroxy-4-methylvaleric acid (3H4MV) (Martin et al., 2013). Another example of a combination approach is using an alternative termination pathway expressing carboxylic acid reductase (Car) from *Nocardia iowensis* to convert fatty acids to alcohols. This was demonstrated by producing 4-methyl-pentanol from endogenously

generated isobutyryl-CoA as a primer (Sheppard et al., 2014). Combination of biotic and abiotic catalysis provides novel opportunities for product diversification. Wang *et al.* 2021 utilizes 3-hydroxyacids produced through rBOX as precursor to produce olefins using heterogenous Lewis acidic catalysts. This abiotic approach enables production of olefins from fatty acids without any additional redox input (Wang et al., 2021).

### **Interfacing rBOX with Various Metabolic Pathways and its Deployment in Different Organisms**

The ubiquitous nature of  $\beta$ -oxidation and the building block acetyl-CoA allows interfacing with other metabolisms and microbial hosts. The rBOX pathway has been successfully implemented in various industrial chassis such as *Saccharomyces cerevisiae* (Lian & Zhao, 2015) and *Corynebacterium glutamicum* (Shin et al., 2021) illustrating the cross-platform capabilities of the pathway. Moreover, native rBOX capabilities of *Clostridium* species have been leveraged in the context of expanding substrate diversity. For example, lactate (B. Liu et al., 2020), acetate and lactate (Detman et al., 2019) and methanol and propionate (de Smit et al., 2019) have been utilized as feedstock to produce butyrate, caproate and isobutyrate, and n-valerate, respectively. rBOX can also be implemented in pathways and organisms that utilize one-carbon (C1) molecules such as CO<sub>2</sub> and methane to contribute to global carbon mitigation efforts. Various native and synthetic CO<sub>2</sub>-fixation pathways, such as Wood-Ljungdahl pathway (WLP) (Ragsdale, 2008), reductive TCA cycle (rTCA) (Berg, 2011) and CETCH cycle (Schwander et al., 2016) either produce acetyl-CoA as product (WLP and rTCA) or utilize various acyl-CoAs as intermediates (CETCH), which can be directed into rBOX as primer or extender units (Figure 3). As an example, production of up to 2.4 g/L hexanol from syngas

through WLP followed by rBOX has been demonstrated in *Clostridium carboxidivorans* P7 (Kottenhahn et al., 2021). Other C1 utilization pathways that do not involve CoA thioesters as direct intermediate or product can still be rewired into rBOX via central metabolism (Figure 3). For example, *Methylobacterium buryatense* 5GB1C, a methanotroph that utilizes methane via the Ribulose Monophosphate (RuMP) cycle, was engineered to produce 4-carbon carboxylic acids via rBOX by directing metabolic flux toward acetyl-CoA generation (Garg et al., 2018). In another study, production of polyhydroxyalkanoates (PHA) from methanol through rewiring the native serine cycle and ethylmalonyl-CoA pathway in conjunction with heterologously expressed rBOX pathway enzymes was demonstrated in *Methylobacterium extorquens* AM1 as host chassis (Orita et al., 2022). Other native and synthetic C1 utilization pathways can either directly generate acyl-CoAs or produce precursor metabolites that connect to central metabolism to generate acetyl-CoA. For example, the recently developed formyl-CoA elongation (FORCE) pathways (Chou et al., 2021) convert C1 substrates in the form of formyl-CoA to C2 and longer-chain 2-hydroxyacyl-CoAs, which can also be further reduced to acyl-CoAs, and serve as both primer and extender units for the rBOX pathway (Figure 3).

Diversifying substrates for primer and extender units or transferring rBOX into a different organism often requires extensive enzyme and metabolic flux optimization. Recent studies show that substrate specificity and activity of rBOX pathway enzymes can be expanded and optimized via bioprospecting (Mehrer et al., 2018; Tan et al., 2020; Wang et al., 2021) and protein engineering (Blaisse et al., 2018; Clomburg et al., 2018). Incorporating rBOX pathway in non-model organisms, such as *Clostridium*, is more challenging than in common industrial hosts, such as *E. coli*, *S. cerevisiae* and *C. glutamicum*. A platform developed by Karim *et al.*, named *in vitro* prototyping and rapid optimization of biosynthetic enzymes (iPROBE), addresses this

problem by using the cell-free system for rapid optimization of the pathway enzymes and expression levels and showing that it can be transferred to *in vivo* (*Clostridium*) with a strong correlation (Karim et al., 2020). Optimization of one-turn rBOX pathway for 3-hydroxybutyrate and n-butanol production has been demonstrated by high-throughput screening of more than 200 pathway combinations which were then engineered into *Clostridium autoethanogenum* to demonstrate 20-fold increase in 3-hydroxybutyrate titer to ~15 g/L (Karim et al., 2020). iPROBE was further used to accelerate carbon negative manufacturing of acetone and isopropanol in *C. autoethanogenum* (Liew et al., 2021).

### Products of rBOX as Building Blocks for Downstream Pathways

To further exploit the carbon and energy economy of the rBOX pathway, products derived from this platform could be utilized to feed into other metabolic routes. One example of using rBOX products as building blocks is the synthesis of bio-derived polymers such as polyhydroxyalkanoates (PHAs). Dong et al. 2019 used acetyl-CoA and propionyl-CoA to produce (3R)-hydroxyacyl-CoA which can be copolymerized by a PHA polymerase to ultimately form  $\alpha$ -branched polyesters (Dong et al., 2019). In this study, the substrate selectivity of thiolase (AsAcat3) and HACD (RePhaB) plays a key role in controlling monomer availability for incorporation into the final copolymer by the PHA polymerase. Similarly, Insomphun et al. 2016 constructed an artificial pathway for the production of PHA containing a DHBA monomer by introducing the *pct* gene from *M. elsdenii*, along with the *bktB* (thiolase) and *PhaB1* (HACD) genes from *C. necator* (Insomphun et al., 2016). By feeding glycolate, the production of PHA with 1 mol % DHBA reached 1% of the dry cell weight.

Another example is polyketide biosynthesis (Figure 4). Elements of rBOX have been successfully integrated into the polyketide synthesis framework to produce olivetolic acid



(OLA), a plant secondary metabolite sourced from Type III polyketide synthase (PKS). Here the core rBOX modules were integrated into the chromosome (thiolase BktB, HACD-ECH FadB, and ECR egTER) for production of hexanoate precursor, along with a native *E. coli* fatty acyl-CoA synthetase FadD to generate hexanoyl-CoA (Tan et al., 2018). The rBOX framework for precursor synthesis was combined with expression of PKS enzymes OLA synthase and OLA cyclase from *Cannabis sativa*, demonstrating for the first time the microbial synthesis of olivetolic acid from a single carbon source at 80 mg/L.

Utilizing elements of the rBOX pathway in a novel way was explored by Tan et al. 2020 for production of polyketide backbones (Tan et al., 2020) (Figure 4). The canonical polyketide biosynthesis pathway proceeds via polyketide synthases (PKSs), which catalyzes iterative decarboxylative Claisen condensation using malonyl-CoA as the extender unit. However, Tan et al. 2020 discovered that certain thiolases, which they named polyketoacyl-CoA thiolases (PKTs), catalyze polyketide backbone formation via iterative non-decarboxylative Claisen condensations, hence offering a synthetic and efficient alternative to PKSs. Additionally, the basic catalytic mechanism and structural features enabling this previously unknown activity were elucidated. For the first time, synthesis of polyketide triacetic acid lactone (TAL) of up to 0.86 g/L was demonstrated only with the expression of a PKT (BktB) and in the absence of PKS. It was further demonstrated that PKT can replace PKS in higher number of iterations, showing alkylresorcinolic acid (ORA) and orcinol production from a *in situ* generated 3,5,7-trioxooctanoyl-CoA backbone (Tan et al., 2020).

The PKT-based framework was also used to demonstrate *in vitro* synthesis of 6-MSA and m-cresol (decarboxylated form of 6-MSA), which is normally catalyzed by a large, multi-modular Type I 6-MSA PKS. *In vitro* production was demonstrated using acetoacetyl-CoA

primer and acetyl-CoA extender by combined expression of PKT BktB with thioester hydrolase (TH) and ketoreductase (KR) domains from *Aspergillus terreus*. Additionally, BktB mutants were shown to accept butyryl-CoA primer and acetyl-CoA as the extender unit, further expanding the range of possible polyketide backbones that can be produced by exploiting the PKT activity of thiolases. Given the promiscuous nature of some thiolase variants, it is possible that they could be engineered to incorporate functionalized molecules into the polyketide backbone beyond the limited PKS building blocks (acetyl-CoA or propionyl-CoA as primers and malonyl-CoA or methylmalonyl-CoA extender units), creating new-to-nature chemistries.

### **Optimization of Pathway Efficiency and Productivity**

Optimization of rBOX pathway involves strain engineering to improve acetyl-CoA availability and balance redox cofactor, enzyme identification and engineering for better kinetic properties, and fermentation optimization for improved titer, rate, and yield. Initial rBOX strain engineering focused on eliminating alternative acetyl-CoA utilization pathways by deleting the mixed acid fermentation (Kim, Cheong, Chou, et al., 2016). Additionally, deletion of endogenous thioesterases to prevent premature hydrolysis of CoA thioesters has been shown to improve product titers and yields (Kim et al., 2015). Alternatively, the same objective was achieved by downregulating fermentative pathway genes using CRISPRi and rewiring prematurely hydrolyzed acetate back to rBOX using acetyl-CoA synthetase (Wu et al., 2017). Moreover, incorporating a mutant pyruvate dehydrogenase (PDH) with improved activity under microaerobic conditions with high NADH/NAD<sup>+</sup> ratio improved flux toward acetyl-CoA from central metabolism (Wu et al., 2019).

Identification of optimal pathway enzymes is key for improving specificity towards the desired product in an iterative pathway like rBOX. Research shows that thiolases and termination

enzymes are two key nodes to control chain length and functionality of the desired products. Thiolases and thioesterases with desired activity have been identified by molecular phylogenetic analysis (Mehrer et al., 2018; Tan et al., 2020; Wang et al., 2021) or constructed via protein engineering (Blaisse et al., 2018). Moreover, *in vitro* prototyping allows for high throughput screening of a large number of enzyme variants like thioesterases to identify special functional group and chain length specificity (McMahon & Prather, 2014). Alternatively, substrate specificity of core rBOX enzymes like ketoreductase could influence specificity of upstream thiolase, leading to improved selectivity (Blaisse et al., 2017). Furthermore, optimization of relative expression levels is critical for debottlenecking the pathway flux, which could be facilitated by rapid *in vitro* prototyping (Karim et al., 2020) and independent gene expression control using orthogonal inducible promoters (Meyer et al., 2019) for *in vivo* implementation.

Since medium-to-long chain carboxylic acids and alcohols are more reduced than sugars, NADH availability and sufficient NADH/NAD<sup>+</sup> ratio are important for high productivity. Optimization of oxygen transfer rate is one approach to address this. Maximizing flux toward rBOX while circumventing regulatory systems under the microaerobic condition was shown to be important in improving the product titer. For example, upregulating pyruvate dehydrogenase (PDH) activity under high NADH/NAD<sup>+</sup> ratio by utilizing mutant PDH and downregulating fermentative pathways through CRISPRi were crucial for improving mid-chain fatty acid titer under microaerobic condition (Wu et al., 2019). Another key consideration is addressing the product toxicity and volatility with fatty acids and alcohols of various chain lengths having distinctive toxic effects on *E. coli* (Wilbanks & Trinh, 2017). Two-phase fermentations with an organic phase serving as extraction solvent during the fermentation was shown as an effective approach to address this issue (Kim et al., 2015; Kottenhahn et al., 2021).

## Conclusions and Future Directions

The rBOX platform offers high energy and carbon efficiency when compared to other pathways for carbon chain elongation and has the potential to be orthogonal to the host metabolism, thus minimizing unproductive crosstalk with life-sustaining reactions. In addition, the rBOX operates at higher flux than other anabolic pathways and is modular, customizable, and programmable. The ability to use various primer and extender units, along with different termination mechanisms and the possibility of integrating the cycle products as intermediates for other pathways, make the rBOX a versatile platform for the synthesis of diverse product classes, including new-to-nature molecules. In addition, the ubiquitous nature of  $\beta$ -oxidation chemistries creates a potential for transferring these capabilities to other industrial organisms. Below we discuss some of the approaches that can be used to address remaining challenges, including the improvement of product titer, rate and yield and the implementation of the rBOX in other organisms.

One of the key considerations in tailoring the rBOX to generate products of a specific chain length and functionality as well as in optimizing pathway flux to improve rate and titer is the identification of enzyme candidate with optimal kinetic properties. Numerous rBOX enzymes have been characterized with some having narrow substrate range while others accepting wide range of carbon chain length and functional groups. Recent development in protein structure modeling tools such as AlphaFold (Jumper et al., 2021), provide a host of opportunities for protein structure analysis and engineering without the need for crystal structures. Learning from the structure analysis of the current database of rBOX enzymes aided by these tools can substantially enhance the knowledge of kinetic properties and the engineering of these enzymes for desired activity and specificity. In addition, low-cost protein engineering techniques in

conjunction with high throughput screening methods can reduce the time required for testing a large number of enzyme variants.

Finetuning the expression of enzymes to debottleneck pathway flux and maximize yield is another key consideration. Several dynamic and orthogonal regulation tools have been developed to help achieving this goal. Dynamic regulation of gene expression controlled by sensors such as light, pH and metabolites has been shown to improve biosynthesis of diverse molecules by limiting biomass, byproducts, and accumulation of toxic intermediates (Dinh & Prather, 2020; Shen et al., 2019). Moreover, decoupling growth and stationary phase via dynamic deregulation of key nodes in central metabolism during stationary phase can improve process robustness and scalability (Ye et al., 2021). In addition, the importance of relative enzyme availability for pathway optimization has been demonstrated both *in vivo* through orthogonal inducible promoter (Meyer et al., 2019) and *in vitro* through cell-free prototyping platforms (Karim et al., 2020). Altogether, these recently developed tools and techniques provide a great opportunity for improving the rBOX platform.

With increasing interests in using non-model organisms for direct utilization of greenhouse gases and organic wastes such as plastics to accomplish sustainability and combat climate change, it is imperative to develop host-agnostic, orthogonal platforms that can be transferred to different chassis organisms. rBOX is a great example of such platforms as it utilizes a universal metabolic precursor (acetyl-CoA) to produce a wide range of molecules with varying chain length and functionalities. Moreover, deletion of native  $\beta$ -oxidation genes as well as other competing pathways (e.g., thioesterases) for acetyl-CoA and other acyl-CoA pool can make the rBOX orthogonal to central metabolism. Advances in genome engineering tools, such as CRISPR/Cas9, have made engineering of non-model organisms very tractable. It has already

been demonstrated that organisms that do not produce high acetyl-CoA pool (notably organisms that mainly utilize pyruvate decarboxylase (PDC), such as *S. cerevisiae*) can be engineered to do so as well (Meadows et al., 2016). The future of rBOX will not just expand on the product diversity but also on the substrate diversity for more sustainable biomanufacturing that can support a circular bioeconomy.

## References

- Atsumi, S., Hanai, T., & Liao, J. C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*, 451(7174), 86-89.  
<https://doi.org/10.1038/nature06450>
- Berg, I. A. (2011). Ecological aspects of the distribution of different autotrophic CO<sub>2</sub> fixation pathways. *Appl Environ Microbiol*, 77(6), 1925-1936.  
<https://doi.org/10.1128/AEM.02473-10>
- Blaisse, M. R., Dong, H., Fu, B., & Chang, M. C. Y. (2017). Discovery and Engineering of Pathways for Production of  $\alpha$ -Branched Organic Acids. *Journal of the American Chemical Society*, 139(41), 14526-14532. <https://doi.org/10.1021/jacs.7b07400>
- Blaisse, M. R., Fu, B., & Chang, M. C. Y. (2018). Structural and Biochemical Studies of Substrate Selectivity in *Ascaris suum* Thiolases. *Biochemistry*, 57(22), 3155-3166.  
<https://doi.org/10.1021/acs.biochem.7b01123>

Bonk, B. M., Tarasova, Y., Hicks, M. A., Tidor, B., & Prather, K. L. J. (2018). Rational design of thiolase substrate specificity for metabolic engineering applications. *Biotechnol Bioeng*, 115(9), 2167-2182. <https://doi.org/10.1002/bit.26737>

Cai, W., & Zhang, W. (2018). Engineering modular polyketide synthases for production of biofuels and industrial chemicals. *Curr Opin Biotechnol*, 50, 32-38. <https://doi.org/10.1016/j.copbio.2017.08.017>

Cao, Y. X., Xiao, W. H., Liu, D., Zhang, J. L., Ding, M. Z., & Yuan, Y. J. (2015). Biosynthesis of odd-chain fatty alcohols in *Escherichia coli*. *Metab Eng*, 29, 113-123. <https://doi.org/10.1016/j.ymben.2015.03.005>

Chen, J. S., Colon, B., Dusel, B., Ziesack, M., Way, J. C., & Torella, J. P. (2015). Production of fatty acids in *Ralstonia eutropha* H16 by engineering beta-oxidation and carbon storage. *PeerJ*, 3, e1468. <https://doi.org/10.7717/peerj.1468>

Cheong, S., Clomburg, J. M., & Gonzalez, R. (2016). Energy- and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions. *Nat Biotechnol*, 34(5), 556-561. <https://doi.org/10.1038/nbt.3505>

Cho, I. J., Choi, K. R., & Lee, S. Y. (2020). Microbial production of fatty acids and derivative chemicals. *Curr Opin Biotechnol*, 65, 129-141. <https://doi.org/10.1016/j.copbio.2020.02.006>

Chou, A., Lee, S. H., Zhu, F., Clomburg, J. M., & Gonzalez, R. (2021). An orthogonal metabolic framework for one-carbon utilization. *Nat Metab*, 3(10), 1385-1399. <https://doi.org/10.1038/s42255-021-00453-0>

Cintolesi, A., Clomburg, J. M., & Gonzalez, R. (2014). In silico assessment of the metabolic capabilities of an engineered functional reversal of the beta-oxidation cycle for the

synthesis of longer-chain ( $C \geq 4$ ) products. *Metab Eng*, 23, 100-115.

<https://doi.org/10.1016/j.ymben.2014.02.011>

Clomburg, J. M., Blankschien, M. D., Vick, J. E., Chou, A., Kim, S., & Gonzalez, R. (2015).

Integrated engineering of beta-oxidation reversal and omega-oxidation pathways for the synthesis of medium chain omega-functionalized carboxylic acids. *Metab Eng*, 28, 202-212. <https://doi.org/10.1016/j.ymben.2015.01.007>

Clomburg, J. M., Contreras, S. C., Chou, A., Siegel, J. B., & Gonzalez, R. (2018). Combination of type II fatty acid biosynthesis enzymes and thiolases supports a functional beta-oxidation reversal. *Metab Eng*, 45, 11-19. <https://doi.org/10.1016/j.ymben.2017.11.003>

Clomburg, J. M., Vick, J. E., Blankschien, M. D., Rodriguez-Moya, M., & Gonzalez, R. (2012).

A synthetic biology approach to engineer a functional reversal of the beta-oxidation cycle. *ACS Synth Biol*, 1(11), 541-554. <https://doi.org/10.1021/sb3000782>

de Smit, S. M., de Leeuw, K. D., Buisman, C. J. N., & Strik, D. (2019). Continuous n-valerate formation from propionate and methanol in an anaerobic chain elongation open-culture bioreactor. *Biotechnol Biofuels*, 12, 132. <https://doi.org/10.1186/s13068-019-1468-x>

Dekishima, Y., Lan, E. I., Shen, C. R., Cho, K. M., & Liao, J. C. (2011). Extending carbon chain length of 1-butanol pathway for 1-hexanol synthesis from glucose by engineered *Escherichia coli*. *J Am Chem Soc*, 133(30), 11399-11401.

<https://doi.org/10.1021/ja203814d>

Dellomonaco, C., Clomburg, J. M., Miller, E. N., & Gonzalez, R. (2011). Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. *Nature*, 476(7360), 355-359. <https://doi.org/10.1038/nature10333>



Detman, A., Mielecki, D., Chojnacka, A., Salamon, A., Blaszczyk, M. K., & Sikora, A. (2019).

Cell factories converting lactate and acetate to butyrate: *Clostridium butyricum* and microbial communities from dark fermentation bioreactors. *Microb Cell Fact*, 18(1), 36.

<https://doi.org/10.1186/s12934-019-1085-1>

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

<https://doi.org/10.1016/j.copbio.2020.02.015>

Dong, H., Liffland, S., Hillmyer, M. A., & Chang, M. C. Y. (2019). Engineering in Vivo Production of alpha-Branched Polyesters. *J Am Chem Soc*, 141(42), 16877-16883.

<https://doi.org/10.1021/jacs.9b08585>

Garg, S., Wu, H., Clomburg, J. M., & Bennett, G. N. (2018). Bioconversion of methane to C-4 carboxylic acids using carbon flux through acetyl-CoA in engineered *Methylobacterium buryatense* 5GB1C. *Metab Eng*, 48, 175-183.

<https://doi.org/10.1016/j.ymben.2018.06.001>

Gulevich, A. Y., Skorokhodova, A. Y., Stasenko, A. A., Shakulov, R. S., & Debabov, V. G. (2016). Metabolic engineering of *Escherichia coli* for 1,3-butanediol biosynthesis through the inverted fatty acid  $\beta$ -oxidation cycle. *Applied Biochemistry and Microbiology*, 52(1),

15-22. <https://doi.org/10.1134/s0003683816010051>

Insomphun, C., Kobayashi, S., Fujiki, T., & Numata, K. (2016). Biosynthesis of polyhydroxyalkanoates containing hydroxyl group from glycolate in *Escherichia coli*.

*AMB Express*, 6(1), 29. <https://doi.org/10.1186/s13568-016-0200-5>

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A.,

- Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kaehne, F., Buchhaupt, M., & Schrader, J. (2011). A recombinant alpha-dioxygenase from rice to produce fatty aldehydes using *E. coli*. *Appl Microbiol Biotechnol*, 90(3), 989-995. <https://doi.org/10.1007/s00253-011-3165-y>
- Kallscheuer, N., Polen, T., Bott, M., & Marienhagen, J. (2017). Reversal of beta-oxidative pathways for the microbial production of chemicals and polymer building blocks. *Metab Eng*, 42, 33-42. <https://doi.org/10.1016/j.ymben.2017.05.004>
- Kang, M. K., & Nielsen, J. (2017). Biobased production of alkanes and alkenes through metabolic engineering of microorganisms. *J Ind Microbiol Biotechnol*, 44(4-5), 613-622. <https://doi.org/10.1007/s10295-016-1814-y>
- Karim, A. S., Dudley, Q. M., Juminaga, A., Yuan, Y., Crowe, S. A., Heggestad, J. T., Garg, S., Abdalla, T., Grubbe, W. S., Rasor, B. J., Coar, D. N., Torculas, M., Krein, M., Liew, F. E., Quattlebaum, A., Jensen, R. O., Stuart, J. A., Simpson, S. D., Kopke, M., & Jewett, M. C. (2020). In vitro prototyping and rapid optimization of biosynthetic enzymes for cell design. *Nat Chem Biol*, 16(8), 912-919. <https://doi.org/10.1038/s41589-020-0559-0>
- Kim, S., Cheong, S., Chou, A., & Gonzalez, R. (2016). Engineered fatty acid catabolism for fuel and chemical production. *Curr Opin Biotechnol*, 42, 206-215. <https://doi.org/10.1016/j.copbio.2016.07.004>
- Kim, S., Cheong, S., & Gonzalez, R. (2016). Engineering *Escherichia coli* for the synthesis of short- and medium-chain  $\alpha,\beta$ -unsaturated carboxylic acids. *Metabolic Engineering*, 36, 90-98. <https://doi.org/10.1016/j.ymben.2016.03.005>

- Kim, S., Clomburg, J. M., & Gonzalez, R. (2015). Synthesis of medium-chain length (C6-C10) fuels and chemicals via beta-oxidation reversal in *Escherichia coli*. *J Ind Microbiol Biotechnol*, 42(3), 465-475. <https://doi.org/10.1007/s10295-015-1589-6>
- Kim, S., & Gonzalez, R. (2018). Selective production of decanoic acid from iterative reversal of beta-oxidation pathway. *Biotechnol Bioeng*, 115(5), 1311-1320. <https://doi.org/10.1002/bit.26540>
- Kottenhahn, P., Philipps, G., & Jennewein, S. (2021). Hexanol biosynthesis from syngas by *Clostridium carboxidivorans* P7 - product toxicity, temperature dependence and in situ extraction. *Heliyon*, 7(8), e07732. <https://doi.org/10.1016/j.heliyon.2021.e07732>
- Kunjapur, A. M., & Prather, K. L. (2015). Microbial engineering for aldehyde synthesis. *Appl Environ Microbiol*, 81(6), 1892-1901. <https://doi.org/10.1128/AEM.03319-14>
- Lan, E. I., Dekishima, Y., Chuang, D. S., & Liao, J. C. (2013). Metabolic engineering of 2-pentanone synthesis in *Escherichia coli*. *AIChE Journal*, 59(9), 3167-3175. <https://doi.org/10.1002/aic.14086>
- Lange, B. M., Rujan, T., Martin, W., & Croteau, R. (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci U S A*, 97(24), 13172-13177. <https://doi.org/10.1073/pnas.240454797>
- Lennen, R. M., & Pfleger, B. F. (2013). Microbial production of fatty acid-derived fuels and chemicals. *Curr Opin Biotechnol*, 24(6), 1044-1053. <https://doi.org/10.1016/j.copbio.2013.02.028>
- Lian, J., & Zhao, H. (2015). Reversal of the beta-oxidation cycle in *Saccharomyces cerevisiae* for production of fuels and chemicals. *ACS Synth Biol*, 4(3), 332-341. <https://doi.org/10.1021/sb500243c>

- Liew, F., †, Nogle, R., †, Abdalla, T., †, Rasor, B. J., Canter, C., Jensen, R. O., Wang, L., Strutz, J., Chirania, P., Tissera, S. D., Mueller, A. P., Ruan, Z., Gao, A., Tran, L., Bromley, J. C., Daniell, J., Conrado, R., Tschaplinski, T. J., Giannone, R. J., . . . Köpke, M. (2021). Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nat Biotechnol*, *In Press*. <https://doi.org/10.1038/s41587-021-01195-w>
- Liu, B., Popp, D., Muller, N., Strauber, H., Harms, H., & Kleinstaub, S. (2020). Three Novel Clostridia Isolates Produce n-Caproate and iso-Butyrate from Lactate: Comparative Genomics of Chain-Elongating Bacteria. *Microorganisms*, 8(12). <https://doi.org/10.3390/microorganisms8121970>
- Liu, L., Zhou, S., & Deng, Y. (2020). The 3-ketoacyl-CoA thiolase: an engineered enzyme for carbon chain elongation of chemical compounds. *Appl Microbiol Biotechnol*, 104(19), 8117-8129. <https://doi.org/10.1007/s00253-020-10848-w>
- Machado, H. B., Dekishima, Y., Luo, H., Lan, E. I., & Liao, J. C. (2012). A selection platform for carbon chain elongation using the CoA-dependent pathway to produce linear higher alcohols. *Metab Eng*, 14(5), 504-511. <https://doi.org/10.1016/j.ymben.2012.07.002>
- Marcheschi, R. J., Li, H., Zhang, K., Noey, E. L., Kim, S., Chaubey, A., Houk, K. N., & Liao, J. C. (2012). A synthetic recursive "+1" pathway for carbon chain elongation. *ACS Chem Biol*, 7(4), 689-697. <https://doi.org/10.1021/cb200313e>
- Marella, E. R., Holkenbrink, C., Siewers, V., & Borodina, I. (2018). Engineering microbial fatty acid metabolism for biofuels and biochemicals. *Curr Opin Biotechnol*, 50, 39-46. <https://doi.org/10.1016/j.copbio.2017.10.002>

Martin, C. H., Dhamankar, H., Tseng, H. C., Sheppard, M. J., Reisch, C. R., & Prather, K. L. (2013). A platform pathway for production of 3-hydroxyacids provides a biosynthetic route to 3-hydroxy- $\gamma$ -butyrolactone. *Nat Commun*, 4, 1414.

<https://doi.org/10.1038/ncomms2418>

McMahon, M. D., & Prather, K. L. (2014). Functional screening and in vitro analysis reveal thioesterases with enhanced substrate specificity profiles that improve short-chain fatty acid production in *Escherichia coli*. *Appl Environ Microbiol*, 80(3), 1042-1050.

<https://doi.org/10.1128/AEM.03303-13>

Meadows, A. L., Hawkins, K. M., Tsegaye, Y., Antipov, E., Kim, Y., Raetz, L., Dahl, R. H., Tai, A., Mahatdejkul-Meadows, T., Xu, L., Zhao, L., Dasika, M. S., Murarka, A., Lenihan, J., Eng, D., Leng, J. S., Liu, C. L., Wenger, J. W., Jiang, H., . . . Tsong, A. E. (2016).

Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature*, 537(7622), 694-697. <https://doi.org/10.1038/nature19769>

Mehrer, C. R., Incha, M. R., Politz, M. C., & Pfleger, B. F. (2018). Anaerobic production of medium-chain fatty alcohols via a beta-reduction pathway. *Metab Eng*, 48, 63-71.

<https://doi.org/10.1016/j.ymben.2018.05.011>

Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J., & Voigt, C. A. (2019). *Escherichia coli* "Marionette" strains with 12 highly optimized small-molecule sensors. *Nat Chem Biol*, 15(2), 196-204. <https://doi.org/10.1038/s41589-018-0168-3>

Orita, I., Unno, G., Kato, R., & Fukui, T. (2022). Biosynthesis of Polyhydroxyalkanoate

Terpolymer from Methanol via the Reverse  $\beta$ -Oxidation Pathway in the Presence of

Lanthanide. *Microorganisms*, 10(1). <https://doi.org/10.3390/microorganisms10010184>

Ragsdale, S. W. (2008). Enzymology of the wood-Ljungdahl pathway of acetogenesis. *Ann N Y Acad Sci*, 1125, 129-136. <https://doi.org/10.1196/annals.1419.015>

Schwander, T., Schada von Borzyskowski, L., Burgener, S., Cortina, N. S., & Erb, T. J. (2016). A synthetic pathway for the fixation of carbon dioxide in vitro. *Science*, 354(6314), 900-904. <https://doi.org/10.1126/science.aah5237>

Shen, X., Wang, J., Li, C., Yuan, Q., & Yan, Y. (2019). Dynamic gene expression engineering as a tool in pathway engineering. *Curr Opin Biotechnol*, 59, 122-129. <https://doi.org/10.1016/j.copbio.2019.03.019>

Sheppard, M. J., Kunjapur, A. M., & Prather, K. L. J. (2016). Modular and selective biosynthesis of gasoline-range alkanes. *Metab Eng*, 33, 28-40. <https://doi.org/10.1016/j.ymben.2015.10.010>

Sheppard, M. J., Kunjapur, A. M., Wenck, S. J., & Prather, K. L. (2014). Retro-biosynthetic screening of a modular pathway design achieves selective route for microbial synthesis of 4-methyl-pentanol. *Nat Commun*, 5, 5031. <https://doi.org/10.1038/ncomms6031>

Shin, J. H., Andersen, A. J. C., Achterberg, P., & Olsson, L. (2021). Exploring functionality of the reverse beta-oxidation pathway in *Corynebacterium glutamicum* for production of adipic acid. *Microb Cell Fact*, 20(1), 155. <https://doi.org/10.1186/s12934-021-01647-7>

Slater, S., Houmiel, K. L., Tran, M., Mitsky, T. A., Taylor, N. B., Padgett, S. R., & Gruys, K. J. (1998). Multiple beta-ketothiolases mediate poly(beta-hydroxyalkanoate) copolymer synthesis in *Ralstonia eutropha*. *Journal of Bacteriology*, 180(8), 1979-1987. <https://doi.org/10.1128/JB.180.8.1979-1987.1998>

Staunton, J., & Weissman, K. J. (2001). Polyketide biosynthesis: a millennium review. *Nat Prod Rep*, 18(4), 380-416. <https://doi.org/10.1039/a909079g>

- Tan, Z., Clomburg, J. M., Cheong, S., Qian, S., & Gonzalez, R. (2020). A polyketoacyl-CoA thiolase-dependent pathway for the synthesis of polyketide backbones. *Nature Catalysis*, 3(7), 593-603. <https://doi.org/10.1038/s41929-020-0471-8>
- Tan, Z., Clomburg, J. M., & Gonzalez, R. (2018). Synthetic Pathway for the Production of Olivetolic Acid in *Escherichia coli*. *ACS Synth Biol*, 7(8), 1886-1896. <https://doi.org/10.1021/acssynbio.8b00075>
- Tseng, H.-C., Harwell, C. L., Martin, C. H., & Prather, K. L. (2010). Biosynthesis of chiral 3-hydroxyvalerate from single propionate-unrelated carbon sources in metabolically engineered *E. coli*. *Microbial Cell Factories*, 9(1), 96. <https://doi.org/10.1186/1475-2859-9-96>
- Tseng, H. C., Martin, C. H., Nielsen, D. R., & Prather, K. L. (2009). Metabolic engineering of *Escherichia coli* for enhanced production of (R)- and (S)-3-hydroxybutyrate. *Appl Environ Microbiol*, 75(10), 3137-3145. <https://doi.org/10.1128/AEM.02667-08>
- Tseng, H. C., & Prather, K. L. (2012). Controlled biosynthesis of odd-chain fuels and chemicals via engineered modular metabolic pathways. *Proc Natl Acad Sci U S A*, 109(44), 17925-17930. <https://doi.org/10.1073/pnas.1209002109>
- Vick, J. E., Clomburg, J. M., Blankschien, M. D., Chou, A., Kim, S., & Gonzalez, R. (2015). *Escherichia coli* Enoyl-Acyl Carrier Protein Reductase (FabI) Supports Efficient Operation of a Functional Reversal of the  $\beta$ -Oxidation Cycle. *Appl Environ Microbiol*, 81, 1406-1416. <https://doi.org/doi:10.1128/AEM.03521-14>
- Wang, Z. Q., Song, H., Koleski, E. J., Hara, N., Park, D. S., Kumar, G., Min, Y., Dauenhauer, P. J., & Chang, M. C. Y. (2021). A dual cellular-heterogeneous catalyst strategy for the



production of olefins from glucose. *Nature Chemistry*. <https://doi.org/10.1038/s41557-021-00820-0>

Wilbanks, B., & Trinh, C. T. (2017). Comprehensive characterization of toxicity of fermentative metabolites on microbial growth. *Biotechnol Biofuels*, 10, 262.

<https://doi.org/10.1186/s13068-017-0952-4>

Wu, J., Wang, Z., Duan, X., Zhou, P., Liu, P., Pang, Z., Wang, Y., Wang, X., Li, W., & Dong, M. (2019). Construction of artificial micro-aerobic metabolism for energy- and carbon-efficient synthesis of medium chain fatty acids in *Escherichia coli*. *Metab Eng*, 53, 1-13.

<https://doi.org/10.1016/j.ymben.2019.01.006>

Wu, J., Zhang, X., Xia, X., & Dong, M. (2017). A systematic optimization of medium chain fatty acid biosynthesis via the reverse beta-oxidation cycle in *Escherichia coli*. *Metab Eng*, 41,

115-124. <https://doi.org/10.1016/j.ymben.2017.03.012>

Ye, Z., Li, S., Hennigan, J. N., Lebeau, J., Moreb, E. A., Wolf, J., & Lynch, M. D. (2021). Two-stage dynamic deregulation of metabolism improves process robustness & scalability in engineered *E. coli*. *Metab Eng*, 68, 106-118. <https://doi.org/10.1016/j.ymben.2021.09.009>

Yu, J. L., Xia, X. X., Zhong, J. J., & Qian, Z. G. (2014). Direct biosynthesis of adipic acid from a synthetic pathway in recombinant *Escherichia coli*. *Biotechnol Bioeng*, 111(12), 2580-

2586. <https://doi.org/10.1002/bit.25293>

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

<https://doi.org/10.1128/AEM.00814-18>



Zhou, S., Hao, T., Xu, S., & Deng, Y. (2020). Coenzyme A thioester-mediated carbon chain elongation as a paintbrush to draw colorful chemical compounds. *Biotechnol Adv*, 43, 107575. <https://doi.org/10.1016/j.biotechadv.2020.107575>

ORIGINAL UNEDITED MANUSCRIPT

**End notes:**

**Supplementary Information** N/A

**Data Availability Statement** N/A

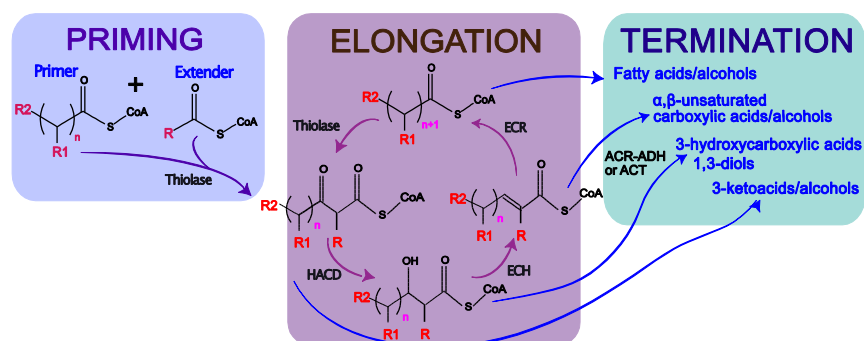
**Code Availability Statement** N/A

**Acknowledgements** This work has been funded by the U.S. Department of Energy (Award Number: DE-EE0008499 and DE-EE0008343) and the U.S. National Science Foundation (Award Number: CBET- 1805793).

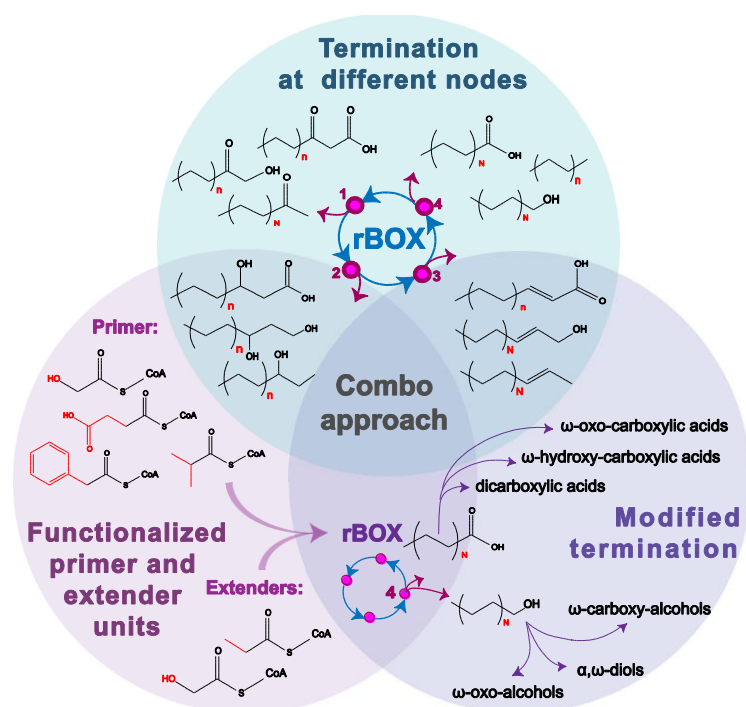
**Author Contributions** K.T., S.H.L., J.C., and R.G. prepared the manuscript. M.K. and M.C.J. edited the manuscript.

#### **Competing Financial Interests Statement**

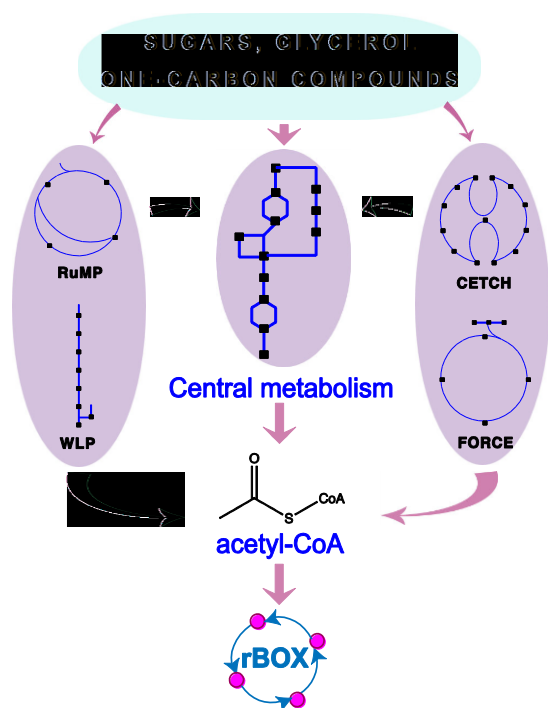
M.K. is a current employee of LanzaTech, a for-profit company pursuing commercialization of gas fermentation processes for a variety of sustainable chemicals. M.C.J. and R.G. consult for and have joint funding with LanzaTech. R. G. is the sole proprietor of RBN Biotech LLC, which holds rights to several rBOX IP. All other authors declare no competing interests.



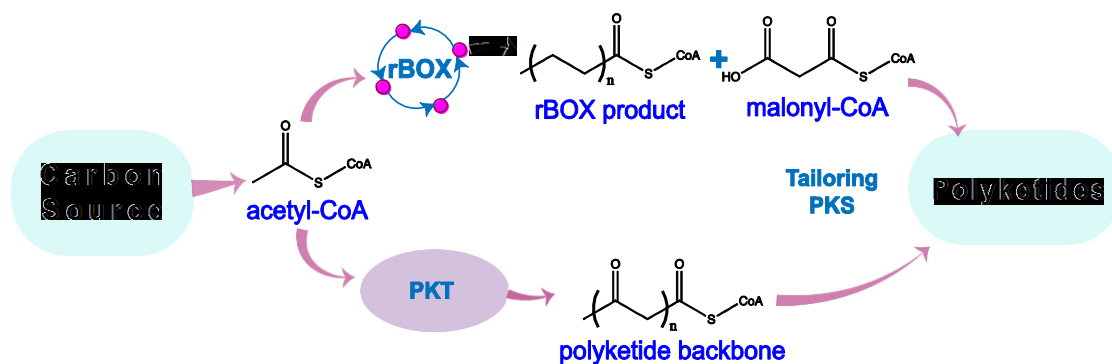
**Figure 1:** Modular structure of rBOX pathways and its use for production of alcohols and carboxylic acids with various functionalities. The three boxes represent conceptually discrete modules of the rBOX pathway, although they may share enzymes and substrates. Abbreviations: HACD - hydroxyacyl-CoA dehydrogenase; ECH - enoyl-CoA hydratase; ECR – enoyl-CoA reductase; ACR-ADH – acyl-CoA reductase-alcohol/aldehyde dehydrogenase for the alcohol termination pathway; ACT – acetyl-CoA transferase (thioesterase) for the carboxylic acid termination pathway.



**Figure 2:** Diversifying rBOX product classes through different strategies: 1) termination at different cycle nodes; 2) using alternative primers (glycolyl-CoA, succinyl-CoA, phenylacetyl-CoA, isopropyl-CoA and others not shown) and extender units (glycolyl-CoA, propionyl-CoA); 3) additional enzymatic modification at the  $\omega$  group. These approaches can also be combined to further diversify the product profile.



**Figure 3:** Interfacing central metabolism and upstream substrate utilization pathways with the rBOX platform. Methanotrophs and methylotrophs harboring Ribulose Monophosphate (RuMP) can assimilate methane, methanol and/or formaldehyde into central metabolites, which can then be converted to acetyl-CoA. Wood-Ljungdahl pathway (WLP) converts CO<sub>2</sub> or CO into acetyl-CoA. Synthetic CO<sub>2</sub>-fixing CETCH cycle (Schwander et al., 2016) operates with various CoA thioesters as pathway intermediates, which can be directly utilized as acyl-CoA primers or extenders for rBOX or further converted into acetyl-CoA via central metabolism. FORCE pathways (Chou et al., 2021) can utilize various one-carbon (C1) substrates to produce diverse acyl-CoAs, including acetyl-CoA.



**Figure 4:** Integrating elements of the rBOX pathway into polyketide biosynthesis in one of two approaches: 1) through condensation of rBOX pathway products with malonyl-CoA using polyketide synthases (PKSs) or 2) through synthesis of polyketide backbones in a PKS-independent manner by using polyketoacyl-CoA thiolase (PKTs).

**Table 1:** Selected examples of products synthesized using rBOX platforms.

Product class	Carbon source	Product	Titer	Enzymes			Reference
				Activation / Priming	Elongation	Termination	
carboxylic acids	glycerol	butyrate	3.4 g/L	AtoB	FadB, egTER/FabI	endogenous thioesterases	(Clomburg et al., 2012)
	glycerol	C6-C10 mixture	1.3 g/L	BktB	BktB, FadB, egTER	TesA	(Kim et al., 2015)
	glycerol	decanoic acid	2.1 g/L	BktB	BktB, FadB, egTER	FadM	(Kim & Gonzalez, 2018)
	glycerol	valerate	~398 mg/L	mePCTBktB	TdTER,hbd, crt	endogenous thioesterases	(Tseng et al., 2010)
	propionate and glucose	valerate	1.425 g/L	mePct, BktB	PhaB, PhaJ4, tdTER	Fs2108	(McMahon & Prather, 2014)
	glycerol	C6-C10 fatty acids	15.67 g/L (bioreactor)	BktB	BktB, fadB, egTER	YdiI	(Wu et al., 2019)
	glucose	octanoic acid	0.74 g/L	PaaJ9	FadJ, tdTER	thioesterase from <i>A. tetradium</i>	(Wang et al., 2021)
	acetate and lactate	butyrate	~200 mM carbon	endogenous rBOX enzymes from <i>Clostridium</i>			(Detman et al., 2019)
	lactate	n-caproate	18 mol%				(B. Liu et al., 2020)
	methanol and propionate	n-valerate	42.8 mM				(de Smit et al., 2019)
	methane	butyrate	0.08 g/L	AtoB	Hbd, Crt, endogenous Ter	YdiI	(Garg et al., 2018)
dicarboxylic acids	glucose	adipic acid	639 µg/L	PaaJ	PaaH1, Ech, egTER	Ptb and Buk1	(Yu et al., 2014)
	glycerol	adipic acid	2.5 g/L (bioreactor)	Cat1, PaaJ	PaaH, PaaF, tdTER	Acot8	(Cheong et al., 2016)
	glycerol	glutarate	36.5 mmol/L	Tfu_0875	Tfu_2399, Tfu_0067, Tfu_1647	Tfu_2576-7	(Zhao et al., 2018)
	glycerol	C6-C10 dicarboxylic acids	0.5 g/L	BktB	BktB, FadB and egTer	YdiI, AlkBGT, ChnD and ChnE	(Clomburg et al., 2015)
ω-hydroxyacids	glycerol	C6-10 ω-hydroxyacids	> 0.8 g/L			YdiI, AlkBGT	
alcohols	glycerol	C6-C10 mixture	0.3 g/L	AtoB and FadA	AtoB, fadA, FadB, egTER	<i>M. aquaeoli</i> Maqu2507 ACR	(Kim et al., 2015)
	glucose	C4-C16 mixture (anaerobic)	1.8 g/L	<i>V.fischeri</i> FadA	<i>V.fischeri</i> FadAB, tdTER	<i>M. aquaeoli</i> ACR	(Mehrer et al., 2018)
	propionate and glucose	pentanol	358 mg/L	BktB	Hbd, Crt, tdTER	<i>C. acetobutylicum</i> AdhE	(Tseng & Prather, 2012)
	syngas	hexanol	2.4 g/L (anaerobic)	endogenous Thl	Endogenous Hbd, Crt, Bcd, EtfAB	Endogenous AdhE	(Kottenhahn et al., 2021)
alkanes	glucose/glycerol	C3-C10 alkanes	0.2-4.3 mg/L	<i>Cupriavidus necator</i> BktB/ <i>Clostridium acetobutylicum</i> thl	BktB, PhaB, PhaJ, tdTer	endogenous thioesterases, <i>P. marinus</i> PMT1231, <i>Nocardia iowensis</i> CAR and <i>N. punctiforme</i> AD	(Sheppard et al., 2016)
methyl ketones	glucose	2-pentanone	0.24 g/L	BktB	Hbd, Crt and tdTer	PcaII from <i>Pseudomonas putida</i> , <i>Reut_1331_1332</i> and Adc from <i>Clostridium</i>	(Lan et al., 2013)
	glucose	acetone	0.7 g/L				

						<i>acetobutylicum</i>	
3-hydroxy-carboxylic acids	glucose	(R)-3-hydroxybutyrate	2.92 g/L	PhaA	PhaB	TesB	(Tseng et al., 2009)
	glucose	(S)-3-hydroxybutyrate	2.08 g/L	Thl	Hbd		
	butyrate	3-hydroxyhexanoate	22.5 mg/L	Pct, BktB	BktB, PhaB	TesB	(Martin et al., 2013)
	glycerol	(S)-3-hydroxybutyric acid	2.5 g/L	AtoB	FadB	endogenous thioesterases	(Clomburg et al., 2012)
	glucose	(S)-3-hydroxyvalerate	0.31 g/L	BktB, Ptb-Buk	Hbd	TesB	(Tseng et al., 2010)
		(R)-3-hydroxyvalerate	0.50 g/L		PhaB		
	propionate and glucose	3-hydroxyvalerate	2.16 g/L	mePct, BktB	PhaB	TesB	(McMahon & Prather, 2014)
	glucose	3-hydroxyoctanoate and 3-hydroxydecanoate	3.6 g/L	PaaJ9	FadJ, tdTER	thioesterase from <i>A. tetradium</i>	(Wang et al., 2021)
1,3-diols	syngas	(S)-3-hydroxybutyric acid	14.63 g/L (anaerobic bioreactor)	Thl	Hbd	native <i>C. autoethanogenum</i> thioesterase	(Karim et al., 2020)
		1,3-butanediol	0.5 g/l (anaerobic bioreactor)			native <i>C. autoethanogenum</i> aldehyde:ferredoxin oxidoreductase (AOR) and Adh	
	glucose	1,3-butanediol	27 mg/L	AtoB	FadB	MphF, FucO or YqhD	(Gulevich et al., 2016)
$\alpha,\beta$ -unsaturated carboxylic acids	glycerol	crotonic acid	3.2 g/L (bioreactor)	BktB	FadB	YdiI	(Kim, Cheong, & Gonzalez, 2016)
		2-hexenoic, 2-octenoic and 2-decenoic acid	189 mg/L		BktB, FadB, egTER		
	propionate and glucose	trans-2-pentenoate	695 mg/L	mePCT, BktB	PhaB, PhaJ4	YdiI	(McMahon & Prather, 2014)
	isobutyrate and glucose	4-methylvalerate	570 mg/L		PhaB, PhaJ4, tdTER	Fs2108	
	methane	crotonic acid	0.06 g/L	AtoB	Hbd, Crt	YdiI	(Garg et al., 2018)
branched-chain products	propionate and glycerol	tiglic acid	3.9 g/L	mePCT, FadAx	FadB2x, FadB1x	YdiI	(Cheong et al., 2016)
	isobutyrate and glycerol	4-methylpentanol	35 mg/L	Pct, BktB	FadB, egTER	Maqu2507	
	isobutyrate	3-hydroxy-4-methylvalerate	1.8 mg/L	mePCT, BktB	BktB, PhaB	TesB	(Martin et al., 2013)
	glycolate	3,4-dihydroxybutyric acid	0.3 mg/L				
	glucose and propionate	$\alpha$ -methyl enoic acid	1120 mg/L	Pct, <i>Ascaris suum</i> Acat5	AsHadH, EcH	YdiI	(Blaisse et al., 2017)



	glucose	4-methylpentanol	192 mg/L	<i>R. palustris</i> IbuA, <i>Cupriavidus necator</i> BktB	PhaB, PhaJ4, tdTER	Car ( <i>N. iowensis</i> ), ADH6 ( <i>Leifsonia</i> sp)	(Sheppard et al., 2014)
	glycerol	4-methylvalerate	34 mg/L	BktB	BktB, FabG, FabZ (R126W R121E), FabI	endogenous thioesterases	(Clomburg et al., 2018)
polyketides	glycerol and hexanoate	olivetolic acid	80 mg/L	BktB	BktB, FadB, egTER	OLS and OAC	(Tan et al., 2018)
	glycerol	triacetic acid lactone (TAL)	0.86 g/L	BktB	N/A	Spontaneous lactonization	(Tan et al., 2020)
	glycerol	alkylresorcinolic acid (ORA) and orcinol	~5 mg/L	BktB	N/A	<i>C. sativa</i> olivetolic acid cyclase (OAC)	(Tan et al., 2020)
polyester	glucose and propionate	$\alpha$ -methyl-branched PHA	18 mmol/L	AsAcat3, cpPCT	RephaB	CapPhaEC	(Dong et al., 2019)
	glycolate	PHA	1 wt%	mePCT, BktB	RephaB1	RephaC1	(Insomphun et al., 2016)
	methanol	PHA	5.4 mol% 3HV, 0.9% 3HHx	Emd, BktB	ReHbd, ReCrt2	AcPhaC <sub>NSDG</sub>	(Orita et al., 2022)