Expanding the repertoire of enzymatic C-C bond formation with one-carbon units

Dissertation

kumulativ

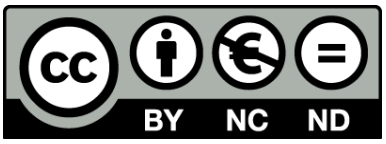
zur Erlangung des Grades eines
Doktor der Naturwissenschaften
(Dr. rer. nat.)

dem Fachbereich Biologie
der Philipps-Universität Marburg
vorgelegt von

Simon Burgener

Aus Visperterminen, Schweiz

Marburg, Oktober 2020



Die Untersuchungen zur vorliegenden Arbeit wurden von Oktober 2016 bis September 2020 unter der Betreuung von Herrn Prof. Dr. Tobias Jürgen Erb in Marburg am Max-Planck-Institut für terrestrische Mikrobiologie in der Abteilung Biochemistry and Synthetic Metabolism durchgeführt.

Vom Fachbereich Biologie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation angenommen am 03.12.2020

Erstgutachter: Prof. Dr. Tobias Jürgen Erb

Zweitgutachter: Prof. Dr. Johann Heider

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Lars-Oliver Essen

Prof. Dr. Hans-Ulrich Mösch

Tag der Disputation: 04.12.2020

Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel „Expanding the repertoire of enzymatic C-C bond formation with one-carbon units“ selbstständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfsmittel bedient habe.

Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 4. Oktober 2020

Simon Burgener

Table of Contents

Summary

C-C bonds are the basis for virtually all organic molecules on Earth. In nature, hundreds of different enzymes catalyze reactions in which C-C bonds are formed. A major task of these enzymes is the fixation of carbon, i.e. a C-C bond formation with at least one-carbon (C1) molecule. Most of these enzymes utilize electrophilic C1 species to fix carbon, while only very few use nucleophilic or radical C1 species. In this work the repertoire of enzymatic C-C bond formation was expanded by 4 new examples, 3 of which are based on a C1 nucleophile and one on a C1 radical.

The thiamine diphosphate (ThDP)-dependent enzyme oxalyl-CoA decarboxylase (OXC) generates a highly reactive carbanion/enamine intermediate that is protonated and released as formyl-CoA. Here it was shown that this intermediate can also undergo C-C bond formation with an electrophilic carbonyl center. This insight allowed to establish three novel C-C bond formation reactions.

First, it was demonstrated that benzaldehyde serves as an excellent electrophile, giving rise to mandelyl-CoA. In combination with oxalyl-CoA synthetase and a thioesterase this enabled the one-pot synthesis of aromatic (*S*)-α-hydroxy acids with enantiomeric excess up to 99%.

Second, it was found that OXC can also generate the carbanion/enamine intermediate from formyl-CoA. By coupling to exergonic reactions at high CO_2 concentrations, OXC was shown to be reversible, that is, it can carboxylate formyl-CoA to oxalyl-CoA.

Third, OXC was engineered to accept the C1 molecule formaldehyde as substrate, producing glycolyl-CoA. Through directed evolution the catalytic efficiency was improved by a factor of ~200 and the resulting variant was successfully employed in a whole-cell biocatalyst for the production of glycolate from formaldehyde.

The glycyl radical enzyme pyruvate formate-lyase (PFL) can abstract a hydrogen atom from formate, thereby generating a highly reactive formyl radical that undergoes C-C bond formation with an acetyl moiety stemming from acetyl-CoA. Here it was shown that PFL exhibits promiscuous activity with glycolyl-CoA. Based on this activity, a pathway was established *in vitro* that converts glycolate and formate to glycerate.

These additions to the toolbox of enzymatic C-C bond formation could contribute to achieve synthetic carbon fixation pathways in the future. Such pathways are thought to be instrumental in achieving a carbon neutral economy.

Zusammenfassung

C-C-Bindungen sind die Grundlage fast aller organischen Moleküle auf der Erde. In der Natur katalysieren hunderte verschiedener Enzyme Reaktionen, bei denen C-C-Bindungen gebildet werden. Eine Hauptaufgabe dieser Enzyme ist die Fixierung von Kohlenstoff, das heißt eine C-C-Bindungsbildung mit mindestens einem Ein-Kohlenstoff (C1) Molekül. Die meisten dieser Enzyme verwenden elektrophile C1-Spezies, wohingegen nur sehr wenige von nukleophilen oder radikalischen C1-Spezies Gebrauch machen. In dieser Arbeit wurde das Repertoire der enzymatischen C-C-Bindungsbildung um vier neue Beispiele erweitert, von denen drei auf einem C1-Nukleophil und eines auf einem C1-Radikal basieren.

Das Thiamin-Diphosphat (ThDP)-abhängige Enzym Oxalyl-CoA-Decarboxylase (OXC) erzeugt ein hochreaktives Carbanion/Enamin-Intermediat, das protoniert und als Formyl-CoA freigesetzt wird. In dieser Arbeit wurde gezeigt, dass dieses Zwischenprodukt auch eine C-C-Bindungsbildung mit einem elektrophilen Carbonylzentrum eingehen kann. Diese Erkenntnis ermöglichte die Entwicklung drei neuer C-C-bindungsbildender Reaktionen.
Erstens wurde gezeigt, dass Benzaldehyd als ausgezeichnetes Elektrophil dient und zu Mandelyl-CoA führt. In Kombination mit Oxalyl-CoA-Synthetase und einer Thioesterase ermöglichte dies die Eintopfsynthese von aromatischen (*S*)-α-Hydroxysäuren mit einem Enantiomerenüberschuss von bis zu 99%.
Zweitens wurde gefunden, dass OXC das Carbanion/Enamin-Intermediat auch aus Formyl-CoA erzeugen kann. Gekoppelt an exergonische Reaktionen und unter hohen CO_2-Konzentrationen wurde gezeigt, dass OXC reversibel ist, das heißt, Formyl-CoA zu Oxalyl-CoA carboxylieren kann.
Drittens wurde OXC dahingehend engineert, dass es aus dem C1-Molekül Formaldehyd Glykolyl-CoA erzeugt. Durch gerichtete Evolution konnte die katalytische Effizienz um einen Faktor von ~200 verbessert werden und die resultierende Enzymvariante wurde erfolgreich in einem Ganzzell-Biokatalysator zur Umsetzung von Formaldehyd zu Glykolat eingesetzt.

Das Glycylradikalenzym Pyruvatformiat-Lyase (PFL) ist in der Lage, ein Wasserstoffatom von Formiat zu abstrahieren, wodurch ein hochreaktives Formylradikal entsteht, das eine C-C-Bindungsbildung mit einer von Acetyl-CoA stammenden Acetylgruppe eingeht. Hier wurde gezeigt, dass PFL promiskuitive Aktivität mit Glykolyl-CoA aufweist, was *in vitro* einen Stoffwechselweg ermöglichte, der Glykolat und Formiat zu Glycerat umwandelt.

Diese 4 Ergänzungen zur Palette der enzymatischen Bildung von C-C-Bindungen könnten synthetische Kohlenstoff-Fixierungswege ermöglichen, und schlussendlich zu einer klimaneutralen Wirtschaft beitragen.

Parts of this thesis that have been published or are in preparation for publication:

Simon Burgener*, Niña Socorro Cortina, Tobias J. Erb (2020) Oxalyl-CoA Decarboxylase Enables Nucleophilic One-Carbon Extension of Aldehydes to Chiral α-Hydroxy Acids. *Angewandte Chemie International Edition*, 59: 5526-5530. * corresponding author

In this work, S.B. contributed to:
the conception of the project
the design and performance of experiments
the analysis of data
the writing of the manuscript

Maren Nattermann*, **Simon Burgener***, Luca Schulz, Pascal Pfister, Tobias J. Erb. Engineering a highly efficient acyloin condensation enzyme for synthetic one-carbon fixation. *Manuscript in preparation*. *equal contribution

In this work, S.B. contributed to:
the conception of the project
the design and performance of experiments
the analysis of data
the writing of the manuscript

Simon Burgener*, Luca Schulz*, Charles A. Cotton, Arren Bar-Even, Tobias J. Erb. Engineering synthetic formate fixation pathways based on pyruvate formate-lyase. *Manuscript in preparation*. *equal contribution.

In this work, S.B. contributed to:
the conception of the project
the design and performance of experiments
the analysis of data
the writing of the manuscript

The above stated contributions of S.B. to the publications or manuscripts in preparation, which are part of this thesis, are herewith endorsed by:

Simon Burgener Prof. Dr. Tobias J. Erb

Publications that are not discussed in this thesis:

Charles A. R. Cotton, Iria Bernhardsgrütter, Hai He, **Simon Burgener**, Luca Schulz, Nicole Paczia, Beau Dronsella, Alexander Erban, Stepan Toman, Marian Dempfle, Alberto De Maria, Joachim Kopka, Steffen N. Lindner, Tobias J. Erb, Arren Bar-Even (2020) Underground isoleucine biosynthesis pathways in *E. coli*. *eLife*, 9:e54207.

Simon Burgener, Shanshan Luo, Richard McLean, Tarryn E. Miller, Tobias J. Erb (2020) A roadmap towards integrated catalytic systems of the future. *Nature Catalysis*, 3:186-192.

Nico J. Claassens*, **Simon Burgener***, Bastian Vögeli, Tobias J. Erb, Arren Bar-Even (2019) A critical comparison of cellular and cell-free bioproduction systems. *Current Opinion in Biotechnology*, 60: 221-229. *equal contribution.

Simon Burgener, Thomas Schwander, Elvira Romero, Marco W. Fraaije, Tobias J. Erb (2018) Molecular basis for converting (2S)-methylsuccinyl-CoA dehydrogenase into an oxidase. *Molecules*, 3:68.

INTRODUCTION

All is beautiful and unceasing,

all is music and reason,

and all, like diamond,

is carbon first, then light.

- José Martí, Selected Writings

1 Introduction

1.1 C-C bond forming reactions

'Carbon first' certainly holds true for all life on earth. Carbon is an extremely versatile element with physical properties that allow it to form bonds with various elements, including other carbon atoms. Carbon-carbon (C-C) bonds are very stable and provide an excellent basis for large molecules – such as the molecules of life, which are all constructed around a core of C-C bonds. To build this "backbone of life", carbon atoms have to be covalently linked in C-C bond forming reactions, arguably the most important transformations in nature and chemical synthesis. Equally important is the breaking of C-C bonds, as it can provide carbon building blocks and energy.

In nature, most C-C bond forming and breaking reactions are catalyzed by enzymes. The importance of such enzymes is reflected in their abundance and diversity. For instance, the enzyme database BRENDA lists 251 unique enzymes in the Enzyme Commission (EC) subclass 4.1 (C-C lyases), and nine in the EC subclass 6.1 (C-C ligases). Many other subclasses also contain prominent C-C breaking or forming enzymes, such as crotonyl-CoA reductase/carboxylase (EC 1.3), glycine hydroxymethyltransferase (EC 2.1) acetyl-CoA C-acetyltransferase and pyruvate formate-lyase (both EC 2.3). Thus, there are at least 300 C-C breaking or forming enzymes, which amounts to approximately 4% of all 8083 BRENDA-listed enzymes. The distribution over several EC classes indicates that nature has found several unique strategies to create and break C-C bonds. Some of these are discussed in section 1.3.

C-C bonds are not only the cornerstone of life, but also of other important molecules such as fuels, plastics and pharmaceuticals. Given the high demand for such compounds, C-C bond forming reactions have garnered major attention in chemical synthesis. Chemists have developed at least 156 unique methods for C-C forming reactions,[1] some of which are unknown to nature, for example olefin metathesis and cross-coupling reactions, one of the most widely applied class of reactions in organic chemistry.[2] Inspired by nature, (bio)chemists recognized the potential of enzymes for chemical synthesis in the 20th century, and laid the foundation for biocatalysis, which takes advantage of the high selectivity that enzymes are able to achieve under mild and sustainable reaction conditions. Drawing on nature's vast repertoire of C-C bond formation, many enzymatic syntheses have been established,[3, 4] including industrial applications, such as the synthesis of chiral cyanohydrins by oxynitrilases.[5]

1.2 The one-carbon bioeconomy

In the light of the increasing atmospheric CO_2 concentration and the associated climate change, C-C bond forming reactions with one-carbon molecules (C1 fixation) are gaining particular interest. However, despite advances in recent years, chemical conversion of CO_2 into multicarbon compounds remains challenging. The main problem is that product selectivity is difficult to achieve with chemical synthesis. Additionally, many processes rely on expensive catalysts, toxic solvents and harsh reaction conditions and are therefore unsuitable for the goal of lowering atmospheric CO_2 concentrations. For instance, the industrial-scale Kolbe-Schmitt synthesis requires 100 bar CO_2 and 125 °C.[6] While industrial production of multicarbon compounds is challenging, CO_2 can be sustainably and efficiently converted into reduced C1 molecules such as CO, formic acid, methanol and methane, by means of hydrogenation, photochemistry, electrochemistry and biocatalysis.[7-11]

In contrast to the shortcomings of chemical CO_2 conversion, nature has faced the challenge of synthesizing multicarbon compounds from C1 molecules for billions of years and has evolved various efficient strategies. It was therefore proposed to harness the power of both worlds by combining chemical reduction of CO_2 with biocatalytic C1 fixation.[12-14] In this scenario, reduced C1 molecules are converted to value-added multicarbon compounds through microbial fermentation. Microbes can (re)oxidize the reduced C1 to CO_2, providing reducing equivalents and energy to the cell. Thus, the reduced C1 compounds serve not only as carbon source but also provide the energy required for growth and bioproduction. This paves the way towards a circular bioeconomy that uses atmospheric CO_2 as carbon source, and could eventually replace fossil carbon-dependent processes.

Even though carbon fixation has been optimized through evolution for billions of years, natural carbon fixation pathways have shortcomings, notably they are not operating at energetic optimum, or they are oxygen sensitive. Additionally, the complex architecture of most pathways (**Figure 1**) make it difficult to transplant them into production hosts. To overcome these shortcomings, biochemists have in recent years begun to design synthetic pathways. Based on the rapid progress in biology over the last decades, notably dropping costs for DNA sequencing and synthesis and ever-more profound understanding of enzymes and metabolism, it has become conceivable to invent and realize new metabolic pathways that do not exist in nature.[15-17] Similar to a chemist who strives for the most efficient route to transform a starting material into a product, biologists now aim for the pathway that allows for optimal growth and/or bioproduction. Synthetic pathways base on the notion that any given set of enzymes can be mixed and matched without

the constraints of the physiological context, enabling the design of the ‘perfect pathway’. Beyond the mix-and-match of natural enzymes, synthetic pathways can also be built around new-to-nature enzymes. This tremendously expands the solution space of metabolic pathways, in principle making any pathway possible that is thermodynamically feasible. In this vein, synthetic carbon fixation pathways can be tailored for minimal metabolic overlap (for example by reducing the number of reaction steps) or optimal energy efficiency and therefore, they hold high promise for address the challenges of establishing a sustainable bioeconomy.[18-20]

How to design a carbon fixation pathway? The key step of carbon fixation is a C-C bond forming reaction. As lined out in the following sections, the nature of the C-C bond forming reaction defines the architecture of a carbon assimilation pathway (**Figure 1**). In order to create novel, synthetic pathways it is therefore of paramount importance to understand the catalytic principles that underlie enzymatic C-C bond formation. The next section gives an overview on the knowledge about enzymatic C-C bond formation with C1 units, by looking at natural as well as repurposed or engineered carboligases.

1.3 Enzymatic one-carbon fixation

The basis of any carbon fixation is a C-C bond formation reaction in which a carbon atom of a C1 molecule is covalently bonded to a carbon atom of another molecule, creating a multicarbon compound. Enzymes use many different catalytic strategies to accomplish C1 fixation. Yet, these can be roughly divided into three categories based on the reactivity of the C1 species employed: C1 electrophiles, C1 nucleophiles and other reactive C1 species. Referred to here is the C1 species immediately before C-C bond formation. **Table 1** lists enzymes that occur in natural C1 fixation pathways, but also enzymes that have been repurposed or engineered for C1 fixation.

Table 1. List of C1-fixing enzymes, classified according to the reactivity of the C1 species.

C1 substrate (C oxidation state)	C1 electrophile	C1 nucleophile and other
CO_2 (+IV)	pyruvate synthase[21] 2-ketoglutarate synthase[22] isocitrate dehydrogenase[23] crotonyl-CoA carboxylase/reductase[24] 2-ketopropyl-CoM reductase[25] phosphoenol pyruvate carboxylase (HCO_3^-)[26] phosphoenol pyruvate carboxykinase[27] ribulose-bisphosphate carboxylase/oxygenase[28] phenylphosphate carboxylase[29] acetyl-CoA carboxylase (HCO_3^-)[30] propionyl-CoA carboxylase (HCO_3^-)[31] acetone carboxylase[32] methylcrotonyl-CoA carboxylase[33] acetophenone carboxylase[34] glycine cleavage system[35] *propionyl-CoA synthase*[36, 37] *>100 decarboxylases*	
CO (+II)		acetyl-CoA synthase[38] *V and Mo nitrogenase*[39, 40]
formate (+II)		*pyruvate formate-lyase*[41] *2-hydroxyacyl-CoA lyase (formyl-CoA)*[42]
cyanide (+II)		*aliphatic oxynitrilase*[43] *aromatic oxynitrilase*[5] *cyanoalanine synthase*[44]
formaldehyde (0)	glycine hydroxymethyltransferase (mTHF)[45] formaldehyde transketolase[46] 3-hexulose-6-phosphate synthase[48] *formolase*[47] *2-hydroxyacyl-CoA lyase* *threonine aldolase*[50] *2-keto-3-deoxy-L-rhamnonate aldolase*[50] *2-keto acid decarboxylase*[51]	glycine cleavage system (mTHF) *formolase*[47] *glyoxylate carboligase (glyoxylate)*[49]
methanol (-II)		acetyl-CoA synthase (MeCoFeSP)

The list contains all enzymes that are involved in natural C1 fixation as well as enzymes whose primary role is not C1 fixation, but which have been repurposed or engineered for C1 fixation; the latter are italicized. Enzymes that employ C1 species of both kinds are highlighted in blue. If the substrate is a derivatized C1 molecule, it is given in parenthesis after the enzyme name. In principle, more than 100 decarboxylases could be included in the section CO_2 as C1 electrophile; however, they were omitted for clarity (see also main text). Abbreviations: mTHF, 5,10-methylenetetrahydrofolate; MeCoFeSP, methyl-corrinoid iron-sulfur protein.

1.3.1 One-carbon electrophiles

As becomes evident from **Table 1**, the predominant strategies for natural carbon fixation rely on electrophilic C1 species, particularly CO_2 and formaldehyde. Presumably, the main reason for the success of this strategy is that both CO_2 and formaldehyde are intrinsically electrophilic, which makes them excellent acceptor units that do not require chemical activation. This is in contrast to all other C1 fixation reactions, which rely on the generation of a highly reactive C1 species prior to C-C bond formation (described section 1.3.2 and 1.3.3).

The most important carbon assimilation process is undoubtedly CO_2 fixation, as reflected in the diversity of carboxylases (**Table 1**). The evolutionary success of carboxylases is also evident from their abundance in metabolism: Carboxylases are not only part of six autotrophic pathways, but also assimilatory pathways, anaplerosis, biosynthetic pathways and redox balancing.[52] Rubisco probably being the most abundant enzyme on Earth further underscores the importance of carboxylases.[53] All carboxylases have in common that they first generate a nucleophile in the form of an enolate or enamine that then attacks CO_2. In virtually all cases, the nucleophile consists of more than one carbon atom and as a consequence thereof, the nucleophilic substrate has to be regenerated through several reaction steps before it can undergo the next C1 elongation. Herein lies the reason for the cyclic architecture of five out of seven natural CO_2 fixation pathways (**Figure 1**). The exceptions are the glycine cleavage system (GCS) and ACS/CODH, which are able to directly condense two C1 units, and thereby enable linear carbon fixation through the reductive glycine pathway and the Wood-Ljungdahl pathway, respectively (**Figure 1**).

In addition to their central role in natural pathways, carboxlyases have also been repurposed for synthetic carbon fixation pathways. Crotonyl-CoA carboxylase/reductase (CCR) is naturally occurring in the anaplerotic ethylmalonyl-CoA pathway and is the carboxylase with the highest turnover number reported so far.[24] Inspired by natural autotrophic pathways, our group envisioned a cyclic pathway based on two CCR-catalyzed carboxylations, the CETCH cycle.[54] To close the cycle, a regeneration module was established that replenishes the CO_2 acceptor crotonyl-CoA. The CETCH cycle was successfully assembled *in vitro* and assimilated CO_2 a rate comparable to natural autotrophic pathways. Another highly active carboxylase is phosphoenolpyruvate carboxylase (PPC) which is used in anaplerosis or to shuttle CO_2 in C4 plants.[26, 55] Combining it with malate dehydrogenase, malate thiokinase, malyl-CoA lyase and enzymes from glyoxylate assimilation gave rise to the synthetic malyl-CoA-glycerate cycle pathway that can assimilate C2 or C3 units into acetyl-CoA without net carbon loss.[56] These two examples demonstrate that

repurposing or engineering of carboxylases allows for tailor-made, cyclic carbon assimilation pathways.

The list of carboxylases could be extended by decarboxylases, since according to the principle of microscopic reversibility every reaction is reversible, including strongly exergonic decarboxylations. Under high concentrations of bicarbonate or pressurized CO_2, decarboxylases are able to catalyze thermodynamically unfavored carboxylation reactions.[57] (De)carboxylases have been recognized as a valuable alternative to chemical catalysts, because they do not require high reaction temperatures and exhibit high product selectivity. BRENDA lists more than 100 decarboxylases, with verified activities on a wide variety of substrates, indicating that this enzyme family bears great biocatalytic potential. One recent example is the use of 2-keto acid decarboxylase as methional carboxylase for the production of methionine.[58] Other examples include aromatic (de)carboxylases, such as ortho-benzoic acid decarboxylase, phenolic acid decarboxylases and prenylated FMN-dependent decarboxylases.[59] The results of CHAPTER II add another reverse decarboxylation to the biocatalytic repertoire: formyl-CoA carboxylation to oxalyl-CoA, catalyzed by oxalyl-CoA decarboxylase (OXC).

Although CO_2 can be directly used as electrophile by enzymes, C-C bond formation with CO_2 is thermodynamically challenging. Carboxylases overcome the thermodynamic cost of CO_2 fixation by coupling C-C bond formation to a highly exergonic reaction, for instance ATP hydrolysis, NAD(P)H oxidation or C-C bond cleavage. In contrast, formaldehyde is a very reactive C1 molecule that readily reacts with carbon nucleophiles. C-C bond formations with formaldehyde are generally exergonic and do not require coupling to another reaction. Nature has taken advantage of this reactivity and evolved three formaldehyde assimilation pathways: the serine cycle, the ribulose monophosphate (RuMP) cycle and the dihydroxyacetone (DHA) cycle. The latter two directly utilize formaldehyde, whereas the serine cycle relies on 5,10-methylene-THF as C1 substrate, which forms spontaneously from THF and formaldehyde.[60] Formaldehyde can be generated through the oxidation of methanol, thus these pathways enable growth on methanol as sole carbon source. In the serine cycle, glycine hydroxymethyltransferase synthesizes serine from glycine and 5,10-methylene-THF.[61] In the RuMP cycle, 3-hexulose-6-phosphate synthase (HPS) catalyzes the aldol condensation between D-ribulose 5-phosphate and formaldehyde, producing D-arabino-hex-3-ulose 6-phosphate.[62] In DHA cycle, thiamine diphosphate (ThDP)-dependent formaldehyde transketolase (FTK) extends D-xylulose 5-phosphate by formaldehyde and, after subsequent C-C bond cleavage, forms dihydroxyacetone and D-glyceraldehyde 3-phosphate.[63] All these pathways are cyclic and, like in the case of carboxylase-based autotrophic

pathways, the C1 acceptor molecule is a multicarbon compound that has to be regenerated by rearrangement of the carbon backbone (**Figure 1**).

Due to its inherently high reactivity, many other aldolases and ThDP-dependent lyases show promiscuous activity with formaldehyde. Recently, the promiscuous formaldehyde activity of threonine aldolase and 2-keto-3-deoxy-L-rhamnonate aldolase was exploited to enable methanol assimilation in *E. coli*.[50] The ThDP-dependent 2-keto acid decarboxylase has been repurposed for C-C bond formation, whereby aldehydes are elongated by one carbon atom derived from formaldehyde.[51] Two more ThDP-dependent enzymes are known to use formaldehyde as electrophile C1 substrate, formolase[47] and 2-hydroxyacyl-CoA lyase (HACL).[42] Since these two enzymes also generate a C1 nucleophile during catalysis, they are discussed in more detail in section 1.3.2. A novel example of synthetic C1 fixation with formaldehyde serving as the electrophile is described in CHAPTER II. Here, OXC was engineered to accept formaldehyde with affinity comparable to the naturally occurring formaldehyde-fixing enzymes FTK and HPS.

Overall, it can be concluded that C1 electrophiles are predominant in natural and engineered C1 fixation. Most of the underlying pathways are cyclic because a multicarbon acceptor molecule has to be regenerated (**Figure 1**). This strategy can be expressed in simple terms with the equation Cn+ C1 = C(n+1), where n > 1. In contrast, the next two sections describe enzymes that are able to directly condense two C1 units into a C2 compound and thus follow C1 + C1 = C2.

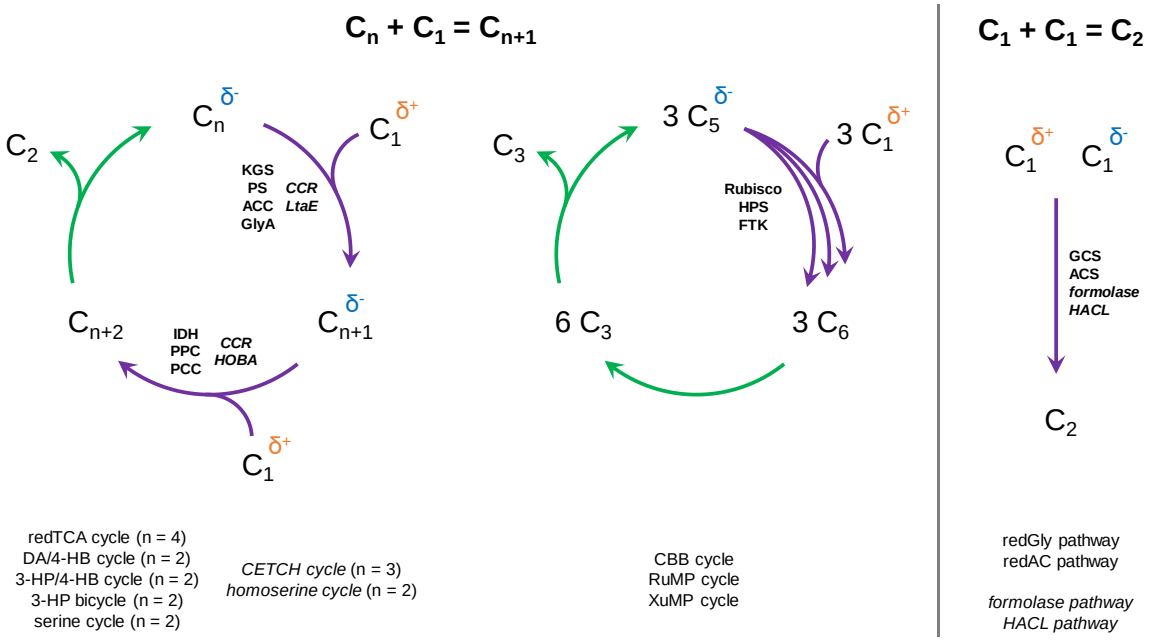


Figure 1. The architecture of C1 fixation pathways is defined by the C-C bond forming reaction(s). The pathways are simplified to C-C bond forming reactions (purple arrows) and carbon backbone rearrangements (green arrows). Enzymes for each C-C bond forming reactions are indicated. Synthetic pathways and the corresponding enzymes are italicized. The polarity of the carbon substrates is indicated with δ (+ for electrophilic and – fur nucleophilic). Reactions of the type Cn + C1 = C(n+1) require cleavage and rearrangement of the carbon backbone to regenerate the Cn nucleophile. There are two alternatives, (*Left*) two consecutive C1 extensions and cleavage of a C2 compound or (*Middle*) stoichiometric balancing of the C-C bond forming and splitting reactions to produce a C3 compound. Conversely, C1 + C1 = C2 reactions enable linear carbon fixation, without rearrangement reactions. Abbreviations: KGS, 2-ketoglutarate synthase; PS, pyruvate synthase; ACC, acetyl-CoA carboxylase; GlyA, glycine hydroxymethyltransferase; CCR, crotonyl-CoA carboxylase/reductase; LtaE, serine aldolase; IDH, isocitrate dehydrogenase; PPC, phosphoenol pyruvate carboxylase; PCC, propionyl-CoA carboxylase; HOBA, 4-hydroxy-2-ketobutanoate aldolase; HPS, 3-hexulose-6-phosphate synthase; FTK, formaldehyde transketolase; GCS, glycine cleavage system; ACS, acetyl-CoA synthase; HACL, 2-hydroxyacyl-CoA lyase; redTCA, reductive tricarboxylic acid; DA/4-HB, dicarboxylic acid; 4-HB, 4-hydroxybutyrate; 3-HP, 3-hydroxypropionate; CBB, Calvin-Benson-Bassham; RuMP, ribulose monophosphate; XuMP, dihydroxyacetone.

1.3.2 One-carbon nucleophiles

Creation of a C1 nucleophile requires chemical activation, as none of the carbon atoms in the C1 molecules methane, methanol, formaldehyde, formate, and CO_2 is intrinsically nucleophilic. Chemical activation is typically coupled to energy-consuming ATP hydrolysis. This could be a reason why in nature C1 fixation reactions with a C1 nucleophile are relatively scarce (**Table 1**). Merely four enzymes make use of this strategy: the GCS, glyoxylate carboligase (GCL), formolase, and HACL. To understand how these few enzymes achieve this seemingly difficult task, a brief summary of each mechanism is given.

In contrast to nature, C1 nucleophiles are very well-known to organic chemistry, especially cyanide is widely utilized in C-C bond forming reactions. Its use as C1 nucleophile goes back to the 19th century, when Hermann Kolbe, a founding father of organic synthesis and a graduate of the University of Marburg, first established it.[64] Enzymes, specifically, aliphatic and aromatic oxynitrilases,[5, 43] and cyanoalanine synthase[44] are also able to use cyanide as C1 nucleophile. In fact, oxynitrilase-catalyzed cyanide addition to aldehydes is one of the oldest and most successful examples of biocatalysis.[65] However, cyanide cannot be easily derived from CO_2, as this reaction requires nitrogen and C-N bond formation and therefore, cyanide-dependent C1 fixation is not directly relevant for the goal of lowering atmospheric CO_2 concentrations. The use of cyanide in synthesis is further hampered by its toxicity. Nevertheless, biocatalytic C-C bond formation offers the possibility to produce valuable multicarbon compounds with high selectivity and under environmental-friendly conditions. An example is described in CHAPTER I: OXC enabled the biocatalytic, cyanide-free synthesis of enantiopure mandelic acid by one-carbon extension of benzaldehyde under mild conditions.

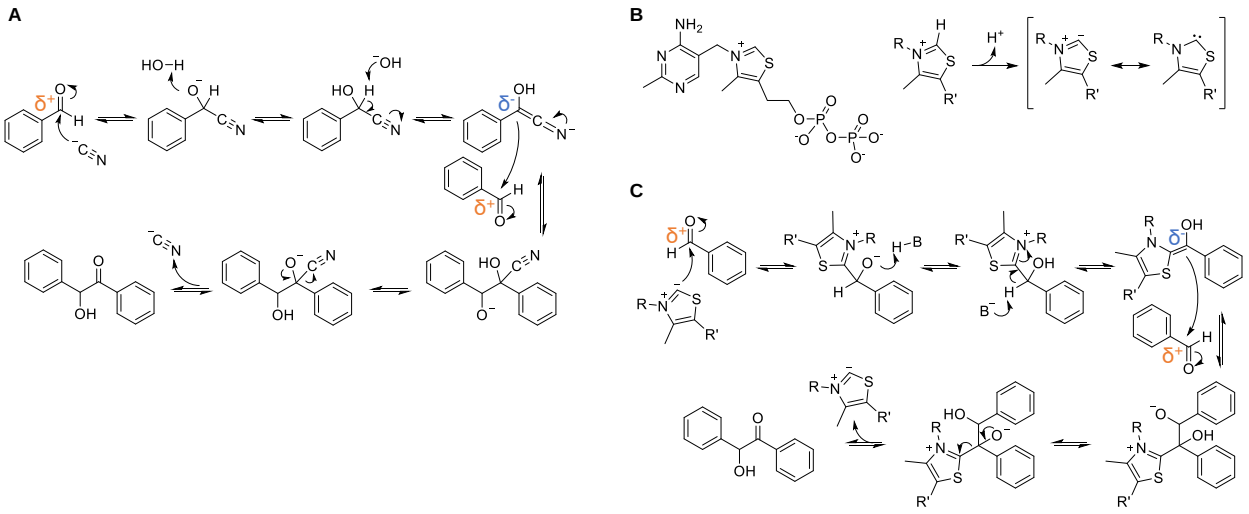


Figure 2. Umpolung of a carbonyl by cyanide and ThDP. The polarity of the relevant carbonyl centers is indicated; orange, electrophilic; blue, nucleophilic. **A)** Mechanism of the cyanide-catalyzed benzoin condensation. **B)** Structure of ThDP. ThDP is converted to its active carbene state by deprotonation. In ThDP-dependent enzymes this is typically mediated by an active site glutamate. **C)** Mechanism of the benzoin condensation, catalysed by the ThDP-dependent benzaldehyde lyase.

In chemistry, cyanide is not only a C1 source, but also a catalyst for C-C bond formation. In fact, the use as catalyst predates the Kolbe synthesis described above, as Liebig and Wöhler observed in 1832 that cyanide catalyzes the benzoin condensation.[66] According to the generally accepted mechanism,[67] the carbon atom of cyanide adds to an aldehyde, and thereby reverses the polarity of the carbonyl group (**Figure 2**). The newly generated nucleophilic center can then add to an

electrophilic carbonyl group, creating a C-C bond. This C-C bond formation is an extremely valuable tool for organic synthesis and was termed Umpolung to reflect the polarity reversal of the carbonyl group.[68] Enzymes are able to perform Umpolung with the help of the cofactor ThDP (**Figure 2**). This cofactor plays a central role in C-C bond breaking and forming reactions and a plethora of ThDP-dependent enzymes have been described.[69] However, only three ThDP-dependent enzymes reverse to polarity of a C1 carbonyl group to create a C-C bond: GCL, formolase and HACL.

GCL condenses two molecules of glyoxylate to tartronate semialdehyde.[49] In the first reaction step, glyoxylate binds to the ThDP and is decarboxylated. The resulting covalent intermediate, referred to as 'active formaldehyde', then performs a nucleophilic attack onto the second glyoxylate molecule and the product tartronate semialdehyde is finally released from the ThDP. Since both substrates are not C1 units, the GCL reaction is not strictly a C1 fixation reaction. However, if active formaldehyde could be produced from free formaldehyde (i.e. without decarboxylation of glyoxylate) this reaction would constitute a true C1 fixation. It is not known if GCL is able to activate formaldehyde, but other ThDP-dependent decarboxylases can generate the nucleophilic intermediate also from the respective aldehyde, without decarboxylation.[51, 70, 71] It is therefore intriguing to speculate that a formaldehyde assimilation pathway based on GCL or a similar carboligase could exist in nature.

Some evidence for this hypothesis was provided in 2015, when Siegel and colleagues designed a synthetic ThDP-dependent enzyme that generates 'active formaldehyde' and fuses it with formaldehyde.[47] The resulting C2 intermediate is then fused with another formaldehyde molecule to finally yield dihydroxyacetone. The condensation of three formaldehyde molecules into dihydroxyacetone is a formose reaction, hence this enzyme was termed formolase. Through computational design and several rounds of directed evolution, the catalytic efficiency of formolase was increased by a factor of ~100. This enabled a synthetic C1 fixation pathway that converts three molecules of formate into dihydroxyacetone - the first and so far only reported C1 + C1 + C1 = C3 strategy. Benzaldehyde lyase (BAL), which naturally catalyzes the condensation of two benzaldehyde molecules, was selected as scaffold for the design of formolase. Given that 'active formaldehyde' is an actual intermediate of GCL's catalytic cycle it could be argued that GCL would make for an excellent (and possibly better than BAL) blueprint for the design of a formolase.

The third ThDP-dependent enzyme using a C1 nucleophile is 2-hydroxyacyl-CoA lyase (HACL), which participates in fatty acid degradation in higher eukaryotes.[72] It cleaves 2-hydroxyacyl-CoA

into a fatty aldehyde and the C1 unit formyl-CoA. ThDP attacks on the carbonyl of the CoA ester substrate and promotes cleavage of the C-C_α bond. The remaining C1 nucleophile is then protonated and released as formyl-CoA. A similar ThDP-dependent enzyme is part of 2-hydroxyisobutyric acid degradation in Actinobacteria, where it cleaves 2-hydroxyisobutyryl-CoA into acetone and formyl-CoA.[73] Interestingly, this enzyme was proposed to run in the reverse, i.e. C-C bond forming direction, during acetone degradation in Desulfobacteraceae.[74] For HACL, reversibility was recently reported and it was shown that HACL has promiscuous activity with formaldehyde.[42] The condensation of formyl-CoA with formaldehyde yields glycolyl-CoA and constitutes a C1 + C1 = C2 reaction that was exploited to create a whole-cell biotransformation system converting formaldehyde into glycolate.[42] In this thesis, the closely related OXC was studied in more detail and its potential for C1 fixation was demonstrated. These results are presented in CHAPTER I and CHAPTER II.

Aside from the well-known Umpolung strategy, nature has only found one other way to generate and condense a C1 nucleophile. The GCS catalyzes the decarboxylation of glycine, producing 5,10-methylene-THF, CO_2, NH_3 and NADH.[35] As the name suggests, GCS is usually operative in the C-C bond cleaving direction, however, it can also run in reverse to produce glycine from CO_2 and 5,10-methylene-THF, a one-carbon carrier.[75] Thus, this enzyme complex is able to condense two C1 molecules, making it an excellent biocatalyst for C1 fixation. However, GCS-based C1 fixation seems to be rare in nature, considering that so far, there is only some indirect evidence that GCS may act as the main carbon fixation enzyme in an uncultured deltaproteobacterium.[76] The potential of GCS for C1 fixation was recently highlighted by the *in vivo* implementation of the reductive glycine pathway, a synthetic carbon fixation pathway that relies on the reverse action of GCS.[77-80] How does GCS achieve this remarkable catalysis? The biochemical study of GCS is very challenging, due to its multi subunit architecture, comprising of four different proteins. Nevertheless, the mechanism has been partially elucidated. The pyridoxal phosphate-dependent P protein catalyzes the reversible decarboxylation of glycine.[81] After decarboxylation, the remaining C1 unit is transferred to lipoate-containing H protein.[82] In the reverse, i.e. the C-C bond forming direction, a carbanion has to be generated that is able to attack CO_2. However, a profound understanding of the nature of this nucleophilic C1 species and how the P protein is able to generate it is lacking.

In conclusion, enzymes only rarely employ C1 nucleophiles, in fact, GCS is the only enzyme in natural carbon fixation doing so. However, as highlighted by the ThDP-dependent enzymes formolase and HACL, this strategy is very intriguing from a biotechnological perspective because

it enables direct condensation of two C1 molecules, and ultimately allows for linear carbon fixation pathways (**Figure 1**). Given their relatively simple structure, such pathways could facilitate *in vivo* implementation and eventually confer synthetic methylo- or formatotrophy on hetrotrophic organisms.

1.3.3 Other reactive one-carbon species

Aside from the abundant nucleophile/electrophile chemistry, some C-C bond forming enzymes make use of organometallic complexes or radicals to generate highly reactive carbon species. With respect to C1 fixation, four enzymes are of particular importance: ACS/CODH, vanadium and molybdenum nitrogenase (VNase and MoNase) and pyruvate formate-lyase (PFL).

ACS/CODH is the key enzyme of the Wood-Ljungdahl pathway found in acetogenic bacteria and methanogenic archaea.[38] This pathway enables synthesis of acetyl-CoA from two highly oxidized C1 molecules (CO_2, CO and formate) and is one of the six autotrophic pathways. In methanogens and sulfate-reducing bacteria, the Wood-Ljungdahl pathway operates in the reverse direction and provides the cell with energy by splitting the C-C bond of acetyl-CoA. As the name suggests, ACS/CODH is a bifunctional enzyme, catalyzing carbon reduction and C-C bond formation. CODH reduces CO_2 into CO at the expense of reduced ferredoxin. ACS condenses the in situ generated CO with CoA and a methyl group bound to the cobalt center of the corrinoid iron-sulfur protein (MeCoFeSP). First, the methyl group of MeCoFeSP and CO both bind to the same Ni center in ACS. The proximity of these two organometallic C1 species then promotes C-C bond formation. The resulting acetyl-Ni species undergoes thiolytic cleavage by CoA, regenerating the Ni center and producing acetyl-CoA.[38] The Wood-Ljungdahl pathway is the most energy efficient C1 pathway and therefore acetogens have been exploited for the bioproduction of multicarbon compounds from C1 units, notably CO-containing syngas.[83] However, ACS/CODH is highly O_2-sensitive, restricting the Wood-Ljungdahl pathway to strictly anaerobic conditions.

Another metalloenzyme capable of CO fixation is nitrogenase, specifically, VNase and MoNase. Nitrogenase's main task is to catalyze the 8-electron reduction of N_2 to two ammonia molecules.[84] A decade ago it was discovered that VNase and MoNase are also able to reduce CO into hydrocarbons, such as ethylene, ethane and propane.[39, 40, 85] Like nitrogen fixation, CO reduction is dependent on reduced ferredoxin or flavodoxin, however, the mechanism of C-C bond formation is not known. Currently, the use of VNase and MoNase in carbon fixation is limited by

low turnover numbers,[86] but the direct production of highly reduced multicarbon compounds from CO with one single enzyme harbors tremendous potential for a sustainable bioeconomy.

Next to organometallic complexes, enzymes can also make use of radical mechanisms to generate reactive C1 species. One notable example is PFL, which catalyzes reversible C-C bond cleavage of pyruvate, producing acetyl-CoA and formate.[41] PFL is a key enzyme of mixed acid fermentation and it is a strictly anaerobic glycyl radical enzyme that requires activation by PFL activase.[87] Catalysis is initiated by transfer of the radical onto the sulfur atom of an active site cysteine. In the reverse, i.e. C-C bond formation direction, the cysteyl radical then abstracts a hydrogen atom from formate, generating a highly reactive formyl radical that subsequently forms a C-C bond with the enzyme-bound acetyl moiety stemming from acetyl-CoA.[88, 89] Thus, PFL has the unique ability to create a C-C bond with the rather inert carbon atom of formate. This offers the possibility to assimilate formate without prior activation, which could be highly useful in engineering energy-efficient synthetic C1 pathways.[90] CHAPTER III describes the engineering of a synthetic formate assimilation pathway based on C-C bond formation by PFL.

Enzymes employing organometallic and radical C1 species are arguably the most remarkable carboligases. Their highly reactive intermediates enable energy efficient C-C bond formations of the type C1 + C1 = C2 or even higher numbered carbon compounds. However, the high oxygen sensitivity of these enzymes impedes engineering and therefore limits their biotechnological potential.

1.4 Aims of this thesis

As lined out above, carboligations with non-electrophilic C1 species are rare in nature and only a handful of enzymes utilize nucleophilic, radical or organometallic C1 species. Yet, as exemplified by several recent examples (**Table 1** and **Figure 1**), such enzymes hold high promise for synthetic metabolism and biocatalytic production of multicarbon compounds. With the exception of formolase, all these examples are based on repurposed wild-type enzymes. While this strategy has been successful, it is restricted to naturally occurring enzyme activity. It is to be expected that the creation of new-to-nature enzymes will open up a whole new realm of C1 fixation. Therefore, the central aim of this thesis is to expand the enzymatic repertoire of C1 carboligation reactions by means of protein engineering. In order to establish these, two of the most promising strategies

are explored in this thesis. CHAPTER I and CHAPTER II focus on C1 Umpolung via the ThDP-dependent enzyme OXC. In CHAPTER III the radical mechanism of PFL is exploited for direct fixation of formate.

1.5 References

[1] Wikipedia. (2020) Carbon-carbon bond forming reactions.
[2] Schultz, D., and Campeau, L. C. (2020) Harder, better, faster, *Nat Chem 12*, 661-664.
[3] Resch, V., Schrittwieser, J. H., Siirola, E., and Kroutil, W. (2011) Novel carbon-carbon bond formations for biocatalysis, *Curr Opin Biotech 22*, 793-799.
[4] Seoane, G. (2000) Enzymatic C-C bond-forming reactions in organic synthesis, *Curr Org Chem 4*, 283-304.
[5] Griengl, H., Schwab, H., and Fechter, M. (2000) The synthesis of chiral cyanohydrins by oxynitrilases, *Trends Biotechnol 18*, 252-256.
[6] Lindsey, A. S., and Jeskey, H. (1957) The Kolbe-Schmitt Reaction, *Chemical reviews 57*, 583-620.
[7] Baskaya, F. S., Zhao, X., Flickinger, M. C., and Wang, P. (2010) Thermodynamic feasibility of enzymatic reduction of carbon dioxide to methanol, *Appl Biochem Biotechnol 162*, 391-398.
[8] Schlager, S., Dibenedetto, A., Aresta, M., Apaydin, D. H., Dumitru, L. M., Neugebauer, H., and Sariciftci, N. S. (2017) Biocatalytic and Bioelectrocatalytic Approaches for the Reduction of Carbon Dioxide using Enzymes, *Energy Technol-Ger 5*, 812-821.
[9] Wang, Q., Warnan, J., Rodríguez-Jiménez, S., Leung, J. J., Kalathil, S., Andrei, V., Domen, K., and Reisner, E. (2020) Molecularly engineered photocatalyst sheet for scalable solar formate production from carbon dioxide and water, *Nature Energy*.
[10] Wang, W. H., Himeda, Y., Muckerman, J. T., Manbeck, G. F., and Fujita, E. (2015) CO2 Hydrogenation to Formate and Methanol as an Alternative to Photo- and Electrochemical CO2 Reduction, *Chemical reviews 115*, 12936-12973.
[11] Benson, E. E., Kubiak, C. P., Sathrum, A. J., and Smieja, J. M. (2009) Electrocatalytic and homogeneous approaches to conversion of CO2 to liquid fuels, *Chem Soc Rev 38*, 89-99.
[12] Pfeifenschneider, J., Brautaset, T., and Wendisch, V. F. (2017) Methanol as carbon substrate in the bio-economy: Metabolic engineering of aerobic methylotrophic bacteria for production of value-added chemicals, *Biofuel Bioprod Bior 11*, 719-731.
[13] Yishai, O., Lindner, S. N., de la Cruz, J. G., Tenenboim, H., and Bar-Even, A. (2016) The formate bio-economy, *Curr Opin Chem Biol 35*, 1-9.
[14] Pieja, A. J., Morse, M. C., and Cal, A. J. (2017) Methane to bioproducts: the future of the bioeconomy?, *Curr Opin Chem Biol 41*, 123-131.
[15] Meyer, A., Pellaux, R., and Panke, S. (2007) Bioengineering novel in vitro metabolic pathways using synthetic biology, *Curr Opin Microbiol 10*, 246-253.
[16] Martin, C. H., Nielsen, D. R., Solomon, K. V., and Prather, K. L. J. (2009) Synthetic Metabolism: Engineering Biology at the Protein and Pathway Scales, *Chem Biol 16*, 277-286.
[17] Erb, T. J., Jones, P. R., and Bar-Even, A. (2017) Synthetic metabolism: metabolic engineering meets enzyme design, *Curr Opin Chem Biol 37*, 56-62.
[18] Bar-Even, A., Noor, E., Lewis, N. E., and Milo, R. (2010) Design and analysis of synthetic carbon fixation pathways, *Proceedings of the National Academy of Sciences of the United States of America 107*, 8889-8894.
[19] Cotton, C. A. R., Claassens, N. J., Benito-Vaquerizo, S., and Bar-Even, A. (2020) Renewable methanol and formate as microbial feedstocks, *Curr Opin Biotech 62*, 168-180.
[20] Bar-Even, A., Noor, E., Flamholz, A., and Milo, R. (2013) Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes, *Biochim. Biophys. Acta 1827*, 1039-1047.
[21] Ragsdale, S. W. (2003) Pyruvate ferredoxin oxidoreductase and its radical intermediate, *Chemical reviews 103*, 2333-2346.
[22] Buchanan, B. B., and Evans, M. C. (1965) The synthesis of alpha-ketoglutarate from succinate and carbon dioxide by a subcellular preparation of a photosynthetic bacterium, *Proceedings of the National Academy of Sciences of the United States of America 54*, 1212-1218.
[23] Camacho, M. L., Brown, R. A., Bonete, M. J., Danson, M. J., and Hough, D. W. (1995) Isocitrate dehydrogenases from Haloferax volcanii and Sulfolobus solfataricus: enzyme purification, characterisation and N-terminal sequence, *FEMS Microbiol Lett 134*, 85-90.
[24] Erb, T. J., Berg, I. A., Brecht, V., Muller, M., Fuchs, G., and Alber, B. E. (2007) Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-

CoA pathway, *Proceedings of the National Academy of Sciences of the United States of America 104*, 10631-10636.
[25] Allen, J. R., Clark, D. D., Krum, J. G., and Ensign, S. A. (1999) A role for coenzyme M (2-mercaptoethanesulfonic acid) in a bacterial pathway of aliphatic epoxide carboxylation, *Proceedings of the National Academy of Sciences of the United States of America 96*, 8432-8437.
[26] Ashworth, J. M., Kornberg, H. L., and Ward, R. L. (1965) Role of Phosphopyruvate Carboxylase in Escherichia Coli, *Biochemical Journal 94*, P28-&.
[27] Ballard, F. J., and Hanson, R. W. (1967) Phosphoenolpyruvate Carboxykinase and Pyruvate Carboxylase in Developing Rat Liver, *Biochemical Journal 104*, 866-&.
[28] Hartman, F. C., and Harpel, M. R. (1994) Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase, *Annu Rev Biochem 63*, 197-234.
[29] Schuhle, K., and Fuchs, G. (2004) Phenylphosphate carboxylase: a new C-C lyase involved in anaerobic in phenol metabolism in Thauera aromatica, *Journal of bacteriology 186*, 4556-4567.
[30] Alberts, A. W., Gordon, S. G., and Vagelos, P. R. (1971) Acetyl CoA carboxylase: the purified transcarboxylase component, *Proceedings of the National Academy of Sciences of the United States of America 68*, 1259-1263.
[31] Kalousek, F., Darigo, M. D., and Rosenberg, L. E. (1980) Isolation and characterization of propionyl-CoA carboxylase from normal human liver. Evidence for a protomeric tetramer of nonidentical subunits, *The Journal of biological chemistry 255*, 60-65.
[32] Sluis, M. K., and Ensign, S. A. (1997) Purification and characterization of acetone carboxylase from Xanthobacter strain Py2, *Proceedings of the National Academy of Sciences of the United States of America 94*, 8456-8461.
[33] Chu, C. H., and Cheng, D. (2007) Expression, purification, characterization of human 3-methylcrotonyl-CoA carboxylase (MCCC), *Protein Expr Purif 53*, 421-427.
[34] Jobst, B., Schuhle, K., Linne, U., and Heider, J. (2010) ATP-dependent carboxylation of acetophenone by a novel type of carboxylase, *Journal of bacteriology 192*, 1387-1394.
[35] Kikuchi, G., Motokawa, Y., Yoshida, T., and Hiraga, K. (2008) Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia, *P Jpn Acad B-Phys 84*, 246-263.
[36] Bernhardsgrutter, I., Schell, K., Peter, D. M., Borjian, F., Saez, D. A., Vohringer-Martinez, E., and Erb, T. J. (2019) Awakening the Sleeping Carboxylase Function of Enzymes: Engineering the Natural CO2-Binding Potential of Reductases, *J Am Chem Soc 141*, 9778-9782.
[37] Bernhardsgrutter, I., Vogeli, B., Wagner, T., Peter, D. M., Cortina, N. S., Kahnt, J., Bange, G., Engilberge, S., Girard, E., Riobe, F., Maury, O., Shima, S., Zarzycki, J., and Erb, T. J. (2018) The multicatalytic compartment of propionyl-CoA synthase sequesters a toxic metabolite, *Nature chemical biology 14*, 1127-1132.
[38] Can, M., Armstrong, F. A., and Ragsdale, S. W. (2014) Structure, function, and mechanism of the nickel metalloenzymes, CO dehydrogenase, and acetyl-CoA synthase, *Chemical reviews 114*, 4149-4174.
[39] Lee, C. C., Hu, Y. L., and Ribbe, M. W. (2010) Vanadium Nitrogenase Reduces CO, *Science 329*, 642-642.
[40] Hu, Y. L., Lee, C. C., and Ribbe, M. W. (2011) Extending the Carbon Chain: Hydrocarbon Formation Catalyzed by Vanadium/Molybdenum Nitrogenases, *Science 333*, 753-755.
[41] Knappe, J., Blaschkowski, H. P., Grobner, P., and Schmitt, T. (1974) Pyruvate formate-lyase of Escherichia coli: the acetyl-enzyme intermediate, *European journal of biochemistry 50*, 253-263.
[42] Chou, A., Clomburg, J. M., Qian, S., and Gonzalez, R. (2019) 2-Hydroxyacyl-CoA lyase catalyzes acyloin condensation for one-carbon bioconversion, *Nature chemical biology 15*, 900-906.
[43] Fechter, M. H., Gruber, K., Avi, M., Skranc, W., Schuster, C., Pochlauer, P., Klepp, K. O., and Griengl, H. (2007) Stereoselective biocatalytic synthesis of (S)-2-hydroxy-2-methylbutyric acid via substrate engineering by using "thio-Disguised" precursors and oxynitrilase catalysis, *Chemistry-a European Journal 13*, 3369-3376.
[44] Warrilow, A. G., and Hawkesford, M. J. (2000) Cysteine synthase (O-acetylserine (thiol) lyase) substrate specificities classify the mitochondrial isoform as a cyanoalanine synthase, *J Exp Bot 51*, 985-993.
[45] Schirch, L. (1982) Serine hydroxymethyltransferase, *Adv Enzymol Relat Areas Mol Biol 53*, 83-112.

[46] Ro, Y. T., Eom, C. Y., Song, T. S., Cho, J. W., and Kim, Y. M. (1997) Dihydroxyacetone synthase from a methanol-utilizing carboxydobacterium, Acinetobacter sp. strain JC1 DSM 3803, *Journal of bacteriology 179*, 6041-6047.
[47] Siegel, J. B., Smith, A. L., Poust, S., Wargacki, A. J., Bar-Even, A., Louw, C., Shen, B. W., Eiben, C. B., Tran, H. M., Noor, E., Gallaher, J. L., Bale, J., Yoshikuni, Y., Gelb, M. H., Keasling, J. D., Stoddard, B. L., Lidstrom, M. E., and Baker, D. (2015) Computational protein design enables a novel one-carbon assimilation pathway, *Proceedings of the National Academy of Sciences of the United States of America 112*, 3704-3709.
[48] Rozova, O. N., But, S. Y., Khmelenina, V. N., Reshetnikov, A. S., Mustakhimov, I. I., and Trotsenko, Y. A. (2017) Characterization of two recombinant 3-hexulose-6-phosphate synthases from the halotolerant obligate methanotroph Methylomicrobium alcaliphilum 20Z, *Biochemistry-Moscow+ 82*, 176-185.
[49] Chang, Y. Y., Wang, A. Y., and Cronan, J. E., Jr. (1993) Molecular cloning, DNA sequencing, and biochemical analyses of Escherichia coli glyoxylate carboligase. An enzyme of the acetohydroxy acid synthase-pyruvate oxidase family, *The Journal of biological chemistry 268*, 3911-3919.
[50] He, H., Hoper, R., Dodenhoft, M., Marliere, P., and Bar-Even, A. (2020) An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in E. coli, *Metabolic Engineering 60*, 1-13.
[51] Germer, P., Gauchenova, E., Walter, L., and Muller, M. (2019) Thiamine Diphosphate Dependent KdcA-Catalysed Formyl Elongation of Aldehydes, *Chemcatchem 11*, 4276-4280.
[52] Erb, T. J. (2011) Carboxylases in Natural and Synthetic Microbial Pathways, *Applied and environmental microbiology 77*, 8466-8477.
[53] Bar-On, Y. M., and Milo, R. (2019) The global mass and average rate of rubisco, *Proceedings of the National Academy of Sciences of the United States of America 116*, 4738-4743.
[54] Schwander, T., Schada von Borzyskowski, L., Burgener, S., Cortina, N. S., and Erb, T. J. (2016) A synthetic pathway for the fixation of carbon dioxide in vitro, *Science 354*, 900-904.
[55] von Caemmerer, S., and Furbank, R. T. (2016) Strategies for improving C4 photosynthesis, *Curr Opin Plant Biol 31*, 125-134.
[56] Yu, H., Li, X. Q., Duchoud, F., Chuang, D. S., and Liao, J. C. (2018) Augmenting the Calvin-Benson-Bassham cycle by a synthetic malyl-CoA-glycerate carbon fixation pathway, *Nature communications 9*.
[57] Glueck, S. M., Gumus, S., Fabian, W. M., and Faber, K. (2010) Biocatalytic carboxylation, *Chem Soc Rev 39*, 313-328.
[58] Martin, J., Eisoldt, L., and Skerra, A. (2018) Fixation of gaseous CO2 by reversing a decarboxylase for the biocatalytic synthesis of the essential amino acid L-methionine, *Nat Catal 1*, 555-561.
[59] Payer, S. E., Faber, K., and Glueck, S. M. (2019) Non-Oxidative Enzymatic (De)Carboxylation of (Hetero)Aromatics and Acrylic Acid Derivatives, *Adv Synth Catal 361*, 2402-2420.
[60] He, H., Noor, E., Ramos-Parra, P. A., Garcia-Valencia, L. E., Patterson, J. A., de la Garza, R. I. D., Hanson, A. D., and Bar-Even, A. (2020) In Vivo Rate of Formaldehyde Condensation with Tetrahydrofolate, *Metabolites 10*.
[61] Heptinstall, J., and Quayle, J. R. (1970) Pathways leading to and from serine during growth of Pseudomonas AM1 on C1 compounds or succinate, *Biochem J 117*, 563-572.
[62] Johnson, P. A., and Quayle, J. R. (1965) Microbial Growth on C1 Compounds. Synthesis of Cell Constituents by Methane- and Methanol-Grown Pseudomonas Methanica, *Biochem J 95*, 859-867.
[63] Vandijken, J. P., Harder, W., Beardsmore, A. J., and Quayle, J. R. (1978) Dihydroxyacetone - Intermediate in Assimilation of Methanol by Yeasts, *Fems Microbiol Lett 4*, 97-102.
[64] Brooke, J. H. (1994) The Quiet Revolution - Hermann Kolbe and the Science of Organic-Chemistry - Rocke,Aj, *Isis 85*, 534-535.
[65] Rosenthaler, L. (1908) Enzyme effected asymmetrical synthesis, *Biochem Z 14*, 238-253.
[66] Wöhler, and Liebig. (1832) Untersuchungen über das Radikal der Benzoesäure, *Annalen der Pharmacie 3*, 249-282.
[67] Lapworth, A. (1904) Reactions involving the addition of hydrogen cyanide to carbon compounds Part II Cyanohydrins regarded as complex acids, *J Chem Soc 85*, 1206-1214.
[68] Seebach, D. (1979) Methods of Reactivity Umpolung, *Angewandte Chemie-International Edition in English 18*, 239-258.

[69] Kluger, R., and Tittmann, K. (2008) Thiamin diphosphate catalysis: enzymic and nonenzymic covalent intermediates, *Chemical reviews 108*, 1797-1833.
[70] Meyer, D., Walter, L., Kolter, G., Pohl, M., Müller, M., and Tittmann, K. (2011) Conversion of pyruvate decarboxylase into an enantioselective carboligase with biosynthetic potential, *Journal of the American Chemical Society 133*, 3609-3616.
[71] Dünkelmann, P., Kolter-Jung, D., Nitsche, A., Demir, A. S., Siegert, P., Lingen, B., Baumann, M., Pohl, M., and Müller, M. (2002) Development of a donor-acceptor concept for enzymatic cross-coupling reactions of aldehydes: the first asymmetric cross-benzoin condensation, *Journal of the American Chemical Society 124*, 12084-12085.
[72] Foulon, V., Sniekers, M., Huysmans, E., Asselberghs, S., Mahieu, V., Mannaerts, G. P., Van Veldhoven, P. P., and Casteels, M. (2005) Breakdown of 2-hydroxylated straight chain fatty acids via peroxisomal 2-hydroxyphytanoyl-CoA lyase: a revised pathway for the alpha-oxidation of straight chain fatty acids, *The Journal of biological chemistry 280*, 9802-9812.
[73] Rohwerder, T., Rohde, M. T., Jehmlich, N., and Purswani, J. (2020) Actinobacterial Degradation of 2-Hydroxyisobutyric Acid Proceeds via Acetone and Formyl-CoA by Employing a Thiamine-Dependent Lyase Reaction, *Front Microbiol 11*.
[74] Frey, J., Schneider, F., Huhn, T., Spiteller, D., Schink, B., and Schleheck, D. (2018) Two enzymes of the acetone degradation pathway of Desulfococcus biacutus: coenzyme B-12-dependent 2-hydroxyisobutyryl-CoA mutase and 3-hydroxybutyryl-CoA dehydrogenase, *Env Microbiol Rep 10*, 283-292.
[75] Gariboldi, R. T., and Drake, H. L. (1984) Glycine Synthase of the Purinolytic Bacterium, Clostridium-Acidiurici - Purification of the Glycine-Co2 Exchange System, *Journal of Biological Chemistry 259*, 6085-6089.
[76] Figueroa, I. A., Barnum, T. P., Somasekhar, P. Y., Carlstrom, C. I., Engelbrektson, A. L., and Coates, J. D. (2018) Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO2 fixation pathway, *Proceedings of the National Academy of Sciences of the United States of America 115*, E92-E101.
[77] Claassens, N. J., Bordanaba-Florit, G., Cotton, C. A. R., De Maria, A., Finger-Bou, M., Friedeheim, L., Giner-Laguarda, N., Munar-Palmer, M., Newell, W., Scarinci, G., Verbunt, J., de Vries, S. T., Yilmaz, S., and Bar-Even, A. (2020) Replacing the Calvin cycle with the reductive glycine pathway in Cupriavidus necator, *Metab Eng 62*, 30-41.
[78] Kim, S., Lindner, S. N., Aslan, S., Yishai, O., Wenk, S., Schann, K., and Bar-Even, A. (2020) Growth of E. coli on formate and methanol via the reductive glycine pathway, *Nature chemical biology 16*, 538-545.
[79] Bang, J., and Lee, S. Y. (2018) Assimilation of formic acid and CO2 by engineered Escherichia coli equipped with reconstructed one-carbon assimilation pathways, *Proc. Natl. Acad. Sci. U.S.A. 115*, E9271-E9279.
[80] Bang, J., Hwang, C. H., Ahn, J. H., Lee, J. A., and Lee, S. Y. (2020) Escherichia coli is engineered to grow on CO2 and formic acid, *Nature Microbiology*.
[81] Fujiwara, K., and Motokawa, Y. (1983) Mechanism of the Glycine Cleavage Reaction - Steady-State Kinetic-Studies of the P-Protein-Catalyzed Reaction, *Journal of Biological Chemistry 258*, 8156-8162.
[82] Fujiwara, K., Okamura, K., and Motokawa, Y. (1979) Hydrogen Carrier Protein from Chicken Liver - Purification, Characterization, and Role of Its Prosthetic Group, Lipoic Acid, in the Glycine Cleavage Reaction, *Arch Biochem Biophys 197*, 454-462.
[83] Henstra, A. M., Sipma, J., Rinzema, A., and Stams, A. J. (2007) Microbiology of synthesis gas fermentation for biofuel production, *Curr Opin Biotechnol 18*, 200-206.
[84] Winter, H. C., and Burris, R. H. (1976) Nitrogenase, *Annu Rev Biochem 45*, 409-426.
[85] Yang, Z. Y., Dean, D. R., and Seefeldt, L. C. (2011) Molybdenum nitrogenase catalyzes the reduction and coupling of CO to form hydrocarbons, *The Journal of biological chemistry 286*, 19417-19421.
[86] Rebelein, J. G., Hu, Y. L., and Ribbe, M. W. (2014) Differential Reduction of CO2 by Molybdenum and Vanadium Nitrogenases, *Angew Chem Int Edit 53*, 11543-11546.
[87] Broderick, J. B., Duderstadt, R. E., Fernandez, D. C., Wojtuszewski, K., Henshaw, T. F., and Johnson, M. K. (1997) Pyruvate formate-lyase activating enzyme is an iron-sulfur protein, *J Am Chem Soc 119*, 7396-7397.

[88] Becker, A., Fritz-Wolf, K., Kabsch, W., Knappe, J., Schultz, S., and Volker Wagner, A. F. (1999) Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase, *Nat. Struct. Biol. 6*, 969-975.
[89] Himo, F., and Eriksson, L. A. (1998) Catalytic mechanism of pyruvate formate-lyase (PFL). A theoretical study, *J. Am. Chem. Soc. 120*, 11449-11455.
[90] Bar-Even, A. (2016) Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways, *Biochemistry 55*, 3851-3863.

CHAPTER I

Oxalyl-CoA decarboxylase:
A Swiss Army knife for breaking
and making C-C bonds

2 Oxalyl-CoA Decarboxylase Enables Nucleophilic One-Carbon Extension of Aldehydes to Chiral α-Hydroxy Acids

Simon Burgener[1, *], Dr. Niña Socorro Cortina[1], Prof. Dr. Tobias J. Erb[1, 2, *]

[1] Department of Biochemistry & Synthetic Metabolism, Max-Planck-Institute for terrestrial Microbiology, Karl-von-Frisch-Str. 10, 35043 Marburg, Germany

[2] LOEWE Center for Synthetic Microbiology, 35043 Marburg, Germany

* Corresponding author

Angewandte Chemie International Edition, **2020**, 59: 5526-5530

Author contributions

S.B. and T.J.E. conceived of the project, and wrote the manuscript. S.B. designed and performed the experiments and analyzed the results. N.S.C. performed LC-MS measurements.

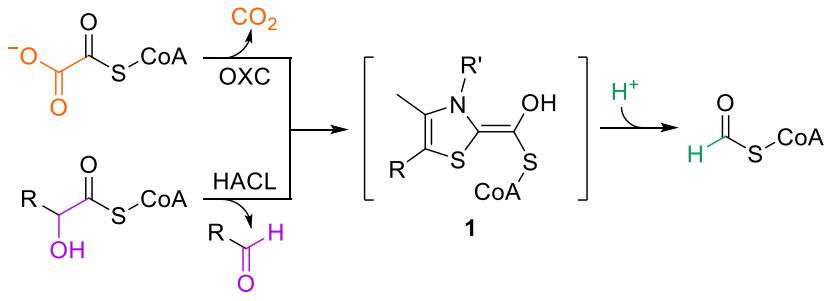
Abstract

The synthesis of complex molecules from simple, renewable carbon units is the goal of a sustainable economy. Here we explored the biocatalytic potential of the thiamine diphosphate-dependent (ThDP) oxalyl-CoA decarboxylase (OXC)/2-hydroxyacyl-CoA lyase (HACL) superfamily that naturally catalyze the shortening of acyl-CoA thioester substrates through the release of the C_1-unit formyl-CoA. We show that the OXC/HACL superfamily contains promiscuous members that can be reversed to perform nucleophilic C_1-extensions of various aldehydes to yield the corresponding 2-hydroxyacyl-CoA thioesters. We improved the catalytic properties of Methylorubrum extorquens OXC by rational enzyme engineering and combined it with two newly described enzymes – a specific oxalyl-CoA synthetase and a 2-hydroxacyl-CoA thioesterase. This enzyme cascade enabled continuous conversion of oxalate and aromatic aldehydes into valuable (*S*)-α-hydroxy acids with enantiomeric excess up to 99%. Altogether our study showcases the potential to develop ThDP-catalyzed nucleophilic C_1-extensions as sustainable production platform for chiral building blocks.

2.1 Introduction

One of biotechnology's central goals is the synthesis of multicarbon compounds under mild and sustainable conditions from renewable resources. This requires biocatalysts that enable selective C-C bond formation ('carboligation') between two carbon units. Thiamine diphosphate (ThDP)-dependent enzymes display high catalytic and substrate promiscuity with respect to C-C bond breaking and forming reactions and they catalyze carboligation reactions at high rates and with excellent stereo- and enantioselectivity.[1] Several biocatalytic applications have been developed that rely on ThDP-dependent carboligases; notable examples are pyruvate decarboxylase (PDC),[2] benzoylformate decarboxylase (BFD),[3] benzaldehyde lyase[4] and transketolase (TK).[5-7] Their broad catalytic repertoire makes ThDP-dependent enzymes promising starting points for the development of biocatalysts for C-C bond forming reactions.

In the context of a methanol and/or formate-based economy, carboligations with one-carbon units are of particular interest.[8, 9] The potential of ThDP-dependent enzymes for synthetic one-carbon fixation has been showcased by the computationally designed enzyme formolase, which condenses three formaldehyde molecules into dihydroxyacetone.[10] Here, we focused on the superfamily of ThDP-dependent oxalyl-CoA decarboxylase (OXC)/2-hydroxyacyl-CoA lyase (HACL). The OXC/HACL superfamily comprises of decarboxylating members (OXCs), as well as non-decarboxylating members (HACLs). Both catalyze the ThDP-dependent cleavage of formyl-CoA from their respective acyl-CoA thioester substrates (Scheme 1). OXCs catalyze the decarboxylation of the C_2-compound oxalyl-CoA to formyl-CoA,[11, 12] whereas HACLs cleave a 2-hydroxyacyl-CoA into a fatty aldehyde that is shortened by a C_1-unit.[13]



Scheme 1. OXC and HACL form the α-hydroxyl-CoA-ThDP carbanion/enamine intermediate (**1**) by decarboxylation of oxalyl-CoA and cleavage of a 2-hydroxyacyl-CoA, respectively. In the second half reaction **1** is protonated and released as formyl-CoA.

In OXC and HACL catalysis has been proposed to proceed through the same covalent intermediate **1** on the ThDP cofactor (Scheme 1).[13-15] After release of CO_2 or aldehyde, the remaining formyl-CoA moiety forms **1**. Analogous to other ThDP-dependent enzymes that form similar ThDP carbanion/enamine intermediates, we speculated that **1** can act as nucleophile in a carboligation reaction with an electrophilic carbon center, essentially reversing the native OXC/HACL reactions. This would enable nucleophilic C_1-extension reactions employing formyl-CoA or oxalyl-CoA as donor, which can in turn be produced from the cheap carbon sources formate[9] and oxalate, respectively. Recently, Chou et al. demonstrated that HACL catalyzes the acyloin condensation of formyl-CoA with aldehyde acceptor substrates.[16] However, the study focused on HACL and formyl-CoA, thus the carboligation potential of OXC with oxalyl-CoA as donor remains unknown.

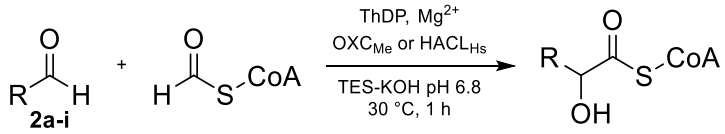
2.2 Results

To further explore the carboligase potential within the OXC/HACL superfamily, we recombinantly produced human HACL ($HACL_{Hs}$), as well as OXC from *Methylorubrum extorquens* (OXC_{Me}), $HACL_{Hs}$ expressed very poorly at 25 °C. Lowering temperature to 15 °C and using the strain ArcticExpress, protein production was improved, but was still rather low (~2 mg protein per L culture), especially compared to OXC (~30 mg/L). The low expression and/or stability limit the use of human HACL for biocatalytic applications. To test their carboligation activity, the enzymes were incubated with formyl-CoA and various aldehyde acceptors. 2-Hydroxyacyl-CoA thioesters were analyzed by LC-MS and the products verified by MS/MS fragmentation (Figure S1). Product formation was detected in the presence of formaldehyde, acetaldehyde, propionaldehyde, glycolaldehyde, succinic semialdehyde, benzaldehyde, and phenylacetaldehyde with both enzymes (Table 1). Glyceraldehyde, glyoxylate and acetone (not shown) were not accepted by OXC_{Me} and $HACL_{Hs}$. As expected from its physiological role as fatty acyl-CoA lyase, $HACL_{Hs}$ showed high activity with the aliphatic aldehydes acetaldehyde and propionaldehyde. On the other hand, OXC_{Me} showed very high activity with benzaldehyde, indicating a distinct catalytic spectrum between the two enzymes. In summary, $HACL_{Hs}$ and OXC_{Me} are able to perform nucleophilic C_1-extension reactions of different aldehydes with formyl-CoA as C_1-donor.

Next, we investigated whether both enzymes would also be able to perform decarboxylating carboligations with oxalyl-CoA as C_1-donor. We argued that the release of CO_2 should provide a strong thermodynamic driving force towards carboligation, analogous to the reactions reported

for glyoxylate carboligase,[17] PDC,[18] BFD,[19] TK,[5, 6] and branched chain ketoacid decarboxylase.[20] $HACL_{Hs}$ possessed only very low oxalyl-CoA decarboxylation activity (k_{cat} < 1 min^{-1}, see Figure S2), in contrast to OXC_{Me} (k_{cat} = 98 ± 3 s^{-1}, see Table 2 & Figure S3), which confirms OXC_{Me}'s physiological function as oxalyl-CoA decarboxylase.

Table 1. Comparison of the aldehyde substrate scope of OXC_{Me} and $HACL_{Hs}$.[a]



Aldehyde	R	Product name	OXC_{Me}[c]	$HACL_{Hs}$[c]
2a	H	glycolyl-CoA	100	11
2b	Me	lactyl-CoA	74	100
2c	CH_2Me	2-hydroxybutyryl-CoA	5	100
2d	CH_2OH	glyceryl-CoA	1	100
2e	$CHOHCH_2OH$	erythronyl-CoA	n.d.[b]	n.d.[b]
2f	COOH	tartronyl-CoA	n.d.[b]	n.d.[b]
2g	$(CH_2)_2COOH$	2-hydroxyglutaryl-CoA	1	100
2h	Ph	mandelyl-CoA	100	3
2i	CH_2Ph	3-phenyllactyl-CoA	22	100

[a] The reaction contained **2a-2i** (10 mM), formyl-CoA (1 mM), OXC_{Me} or $HACL_{Hs}$ (5 µM). Products were analyzed by LC-MS after 1 h reaction time. [b] Product not detected. [c] Relative activity in %. Relative activity refers to the comparison of OXC_{Me} and $HACL_{Hs}$ for each aldehyde substrate.

Given the high activity towards benzaldehyde, we chose it as model substrate to study the decarboxylating carboligation of OXC_{Me} in more detail. When started with oxalyl-CoA and benzaldehyde, OXC_{Me} as expected produced mandelyl-CoA at a rate of 4 min^{-1} (Figure S4). However, mandelyl-CoA formation was preceded by formation of formyl-CoA, suggesting that OXC_{Me} first rapidly decarboxylated oxalyl-CoA into formyl-CoA, followed by slow carboligation of formyl-CoA with benzaldehyde (Figure S4).

Decarboxylation of oxalyl-CoA proceeds via formation of **1**, which can be either protonated and released as formyl-CoA or undergo a nucleophilic attack on benzaldehyde to form mandelyl-CoA.

We speculated that mandelyl-CoA formation may be increased by suppressing the unwanted protonation reaction (and subsequent release of formyl-CoA) that competes with carboligation.

Table 2: Steady-state kinetic parameters of oxalyl-CoA decarboxylation catalyzed by OXC_{Me}.[a]

Mutation	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (s^{-1} M^{-1})
wild-type	98 ± 3	105 ± 11	9.3×10^5
Y497A	1.32 ± 0.04	180 ± 17	7.3×10^3
S568A	5.53 ± 0.11	23 ± 2	1.1×10^5
Y497A S568A	0.32 ± 0.01	103 ± 15	3.1×10^3

[a] Michaelis-Menten graphs are shown in Figure S3.

Structural, biochemical and theoretical studies on OXC from *Oxalobacter formigenes* (OXC_{Of}; 63% sequence identity to OXC_{Me}) demonstrated that protonation of **1** is mediated by a water molecule which is coordinated by several polar side chains, notably a tyrosine and a serine, which are conserved in OXC_{Me} (Tyr497 and Ser568).[14, 15] In OXC_{Me} variants Y497A and S568A, formyl-CoA formation was decreased 20- to 50-fold, while the K_M of both variants was largely unaffected (Table 2 & Figure S3). When starting with oxalyl-CoA and benzaldehyde $OXC_{Me\text{-}Y497A}$ showed 5-fold increased mandelyl-CoA production rate (20 min^{-1}; Figure S4). The mutation Y497A thus increased the ratio of carboligation to decarboxylation by a factor of approximately 400 compared to the wild-type, suggesting that we successfully redirected activity of the enzyme towards carboligation by suppressing protonation of **1**.

Hydrolysis of mandelyl-CoA leads to mandelic acid, which serves as building block for various drugs as well as a resolving agent in chiral resolution processes.[21] We set out to identify a thioesterase capable of hydrolyzing the non-natural metabolite mandelyl-CoA. We tested PaaI, TesB and YciA from *E. coli*, which were shown to hydrolyze a broad range of acyl-CoA thioesters.[22-24] YciA showed relatively high mandelyl-CoA thioesterase activity ($k_{obs} \approx 1.5$ s^{-1}), low activity with formyl-CoA ($k_{obs} \approx 0.06$ s^{-1}) and no activity with oxalyl-CoA (Figure S5 & 1B). When used in combination with OXC_{Me}, the two enzymes formed an enzymatic cascade for the conversion of oxalyl-CoA and benzaldehyde into mandelic acid and free CoA (data not shown).

We noticed that subsequent regeneration of free CoA into oxalyl-CoA would allow us to close a catalytic cycle for the continuous production of mandelic acid (Figure 1A). To establish such a catalytic cycle, we obtained an oxalyl-CoA synthetase AMP-forming (OXS) homologue from *M.*

extorquens,[25] and determined the enzyme's steady-state kinetic parameters (k_{cat} = 1.30 ± 0.02 s^{-1}; K_M(oxalate) = 9 ± 1 µM; Figure 1C). We then used OXS in combination with YciA and OXC_{Me} to continuously produce mandelic acid from oxalate and benzaldehyde. When we replaced OXC_{Me} by the Y497A variant, mandelic acid production rate increased 5-fold and the conversion increased 4-fold (Figure 1E). This was likely caused by decreased formation of the unwanted side product formyl-CoA (and its further hydrolysis by YciA). Chiral LC-MS revealed that (*S*)-mandelic acid was produced with enantiomeric excess of 97% (Figure 1D). Since YciA showed no stereospecificity in the hydrolysis of mandelyl-CoA (Figure 1B), the stereochemistry is exclusively determined by OXC_{Me}.

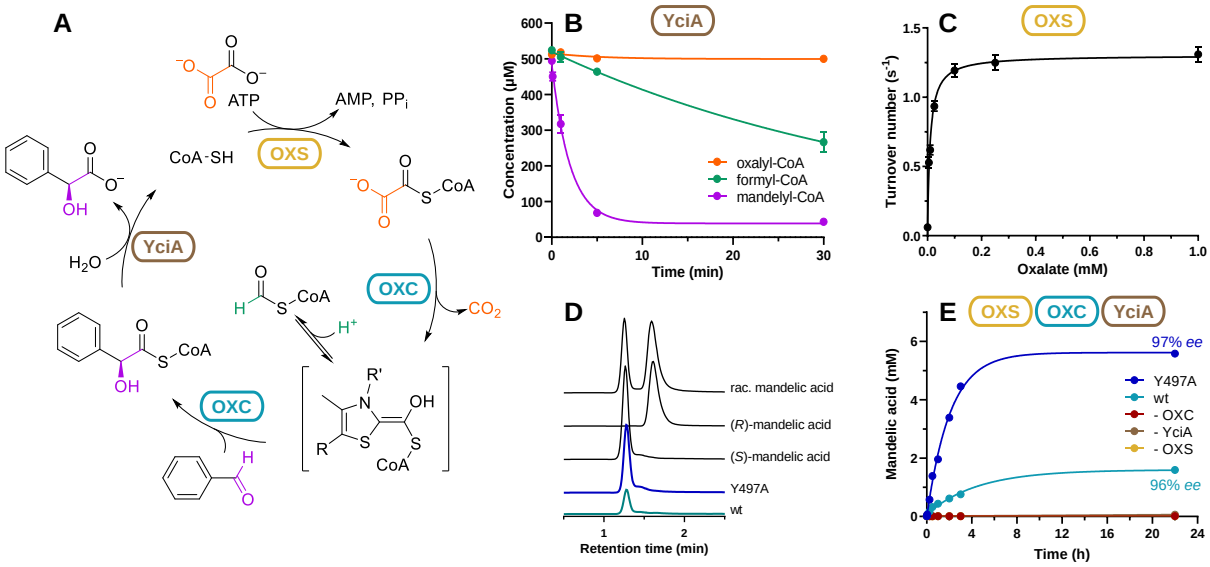


Figure 1. A) The OXS-OXC_{Me}-YciA cascade converts benzaldehyde and oxalate into mandelic acid and CO_2 under consumption of ATP. **B)** Thioesterase activity of YciA (2 µM) towards oxalyl-CoA, formyl-CoA and racemic mandelyl-CoA (0.5 mM each). The negative control without enzyme is shown in Figure S6. Error bars show standard deviation of two replicates. **C)** Michaelis-Menten graph of OXS with oxalate as substrate. **D)** Chromatograms of chiral HPLC analyzing the reaction products of the complete cascade with either wild-type OXC_{Me} (wt) or Y497A. On top are commercial mandelic acid standards. **E)** Mandelic acid formation of the cascade over time. The reaction contained **2h** (25 mM), disodium oxalate (10 mM), ATP (10 mM), CoA (0.5 mM), OXS (5 µM), OXC (5 µM), YciA (2 µM). *ee* of (S)-mandelic acid is indicated for the last time point (22 h).

Next, we tested the substrate scope of the catalytic cycle by replacing benzaldehyde with substituted variants **2i**-**2k** (Figure 2A). Under limiting ATP concentrations (10 mM), the expected products **3a**-**3d** were formed at varying yields (57-93%) and *ee* (44-99%) (Figure 2B). Notably, also sterically demanding **2k** was converted to **3d** with high yield, albeit with moderate *ee*. The

broad substrate scope of the OXS-OXC-YciA cascade was further confirmed by an extended screen, in which product formation was detected for ten other aromatic and three heteroaromatic aldehydes (**3e**-**3r**, Figure 2C). To test if the cascade can be scaled up, we performed the reaction on a semi-preparative scale (0.625 mmol **2h**). We added an ATP regeneration system comprising of creatine phosphate, creatine kinase and adenylate kinase. With catalytic amounts of ADP (0.5 mM) this five enzyme one-pot cascade produced **3a** with a yield of 53%. Taken together, these results indicate that the established enzyme cascade can be employed to produce various aromatic α-hydroxy acids with moderate to high (*S*)-selectivity.

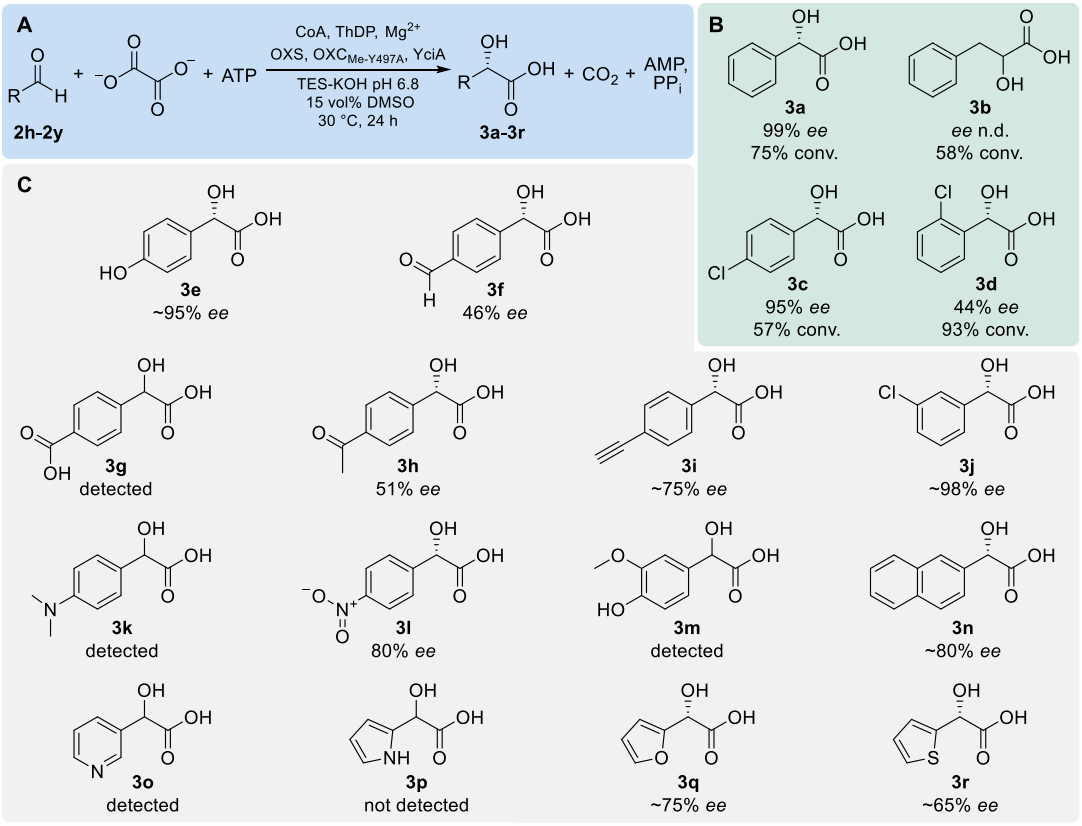


Figure 2. Scope of the OXS-OXC$_{Me-Y497A}$-YciA cascade for the synthesis of aromatic (*S*)-α-hydroxy acids. Chromatograms and time courses of the reactions are shown in Figure S7. **A)** The reactions contained **2h**-**2y** (25 mM), disodium oxalate (10 mM), ATP (10 mM), CoA (0.5 mM), OXS (5 µM), OXC$_{Me-Y497A}$ (5 µM), YciA (2 µM). Products were analyzed by chiral LC-MS after 24 h reaction time. ***ee's*** were estimated based on extracted ion counts. **B)** For **3a**-**3d** conversion and *ee* were quantified by comparison to a commercial standard. For **3b** separation of the enantiomers could not be achieved; n.d, not determined. **C) 3e**-**3r** were analyzed qualitatively. Where chromatographic separation allowed, the *ee* was estimated, assuming identical ionization of the two enantiomers.

2.3 Discussion

We demonstrated that members of the OXC/HACL superfamily are able to catalyze C_1-carboligation reactions between **1** – formed either through decarboxylation of oxalyl-CoA or deprotonation of formyl-CoA – and several aldehydes to yield chiral 2-hydroxyacyl-CoA thioesters. These nucleophilic C_1-extension reactions expand the spectrum of ThDP-dependent enzymes as versatile biocatalysts for C-C bond forming reactions.[26, 27]

What determines substrate specificity in the OXC/HACL superfamily? The observed variance in the aldehyde scope of OXC_{Me} and $HACL_{Hs}$ are likely caused by differences in the C-terminal domain, which forms a 'lid' on top of the active site.[14] While the bottom part of the active site is virtually identical between OXC and HACL (with exception of Tyr133 and Glu134 in OXC_{Me} that are replaced by Phe and Gln in $HACL_{Hs}$), the C-terminal lid-domain shows a high variability between individual superfamily members. Further characterization of the OXC/HACL superfamily could reveal more carboligases with aldehyde preference for a desired application.

We engineered OXC_{Me} towards improved carboligation rate at the expense of formyl-CoA formation rate by replacing Tyr497 with Ala. Interestingly, the mutation Y497A did not affect the enantioselectivity for (*S*)-mandelyl-CoA. This is reminiscent of engineering efforts on PDC from *Zymomonas mobilis*, where Glu473 positions a water molecule that acts as proton donor for the ThDP carbanion/enamine intermediate. Mutating this amino acid to glutamine led to a 100-fold preference of carboligation over aldehyde release, under full retention of enantioselectivity.[18] Considering the high demand of mandelic acid and its derivatives in the (*R*) configuration,[21] it would be interesting to engineer OXC_{Me} towards inverted enantioselectivity. This has been achieved for other ThDP-dependent carboligases.[28, 29]

The rational engineering of $OXC_{Me\text{-}Y497A}$ enabled a three enzyme cascade comprising of OXS, a newly identified oxalyl-CoA synthetase, and YciA, an efficient mandelyl-CoA thioesterase with only minor formyl- and oxalyl-CoA hydrolysis activities. The OXC_{Me}-mediated production of aromatic (*S*)-α-hydroxy acids in high *ee* from aldehydes and oxalate offers an alternative to hydrogen cyanide based syntheses catalyzed by nitrilases.[30] However, the requirements of CoA in catalytic amounts and an ATP regeneration system may limit the potential for a synthetic application on the larger scale. To this end, employing whole-cell catalysts may prove to be advantageous, providing not only the cascade enzymes but also ATP regeneration and a CoA pool without addition of purified enzymes and cofactors.[31]

Altogether, our study expands the spectrum of ThDP-dependent transformations by nucleophilic C_1-extensions, which gives access to α-hydroxy acids that are valuable chiral building blocks and showcases ways to establish *in vitro*- and *in vivo*-platforms for the continuous production of these compounds in the future.

Acknowledgements

We thank L. Schulz for cloning of *oxc* and B. Heinrich for NMR measurements. This work was supported by the German Ministry of Education and Research Grant FormatPlant (part of BioEconomy 2030, Plant Breeding Research for the Bioeconomy). A provisional patent application has been filed through the Max-Planck-Innovation based on the results presented here.

2.4 Materials and Methods

Chemicals

Unless stated otherwise, standard laboratory reagents were obtained from Sigma-Aldrich® (Steinheim, Germany) or Carl Roth GmbH & Co. KG (Karlsruhe, Germany) with the highest purity available. Propionaldehyde, Vanillin, (*R*)- and (*S*)-mandelic acid were obtained from Tokyo Chemical Industry (Zwijndrecht, Belgium). 4-chloromandelic acid was obtained from Alfa Aesar (ThermoFisher (Kandel) GmbH; Kandel, Germany).

Formyl-CoA synthesis

Formyl-CoA was synthesized as described previously.[32, 33] After Extraction with diethylether formyl-CoA was purified by preparative HPLC-MS with an acetonitrile gradient in 25 mM ammonium formate pH 4.2. The fractions containing the product were lyophilized and stored at -20 °C. Formyl-CoA was dissolved in aq. HCl (pH 4). The concentration was determined by enzymatic depletion with PduP,[34] following NADH consumption at 340 nm.

Oxalyl-CoA synthesis

Oxalyl-CoA was synthesized enzymatically with OXS. A 5 mL reaction containing 50 mg CoA (0.064 mmol, 1 eq.), 52 mg ATP (0.086 mmol, 1.3 eq.), 10 mg disodium oxalate (0.075 mmol, 1.2 eq.) in buffer (100 mM MES-KOH pH 6.8, 15 mM $MgCl_2$) was started by adding OXS to a final concentration of 0.4 mg/mL and incubated at 30 °C for 1 hour. The reaction was quenched with 250 µL formic acid and the enzyme removed by centrifugation (4,000 × g, 4 °C, 10 min). Oxalyl-CoA was purified by preparative HPLC-MS with an acetonitrile gradient in 25 mM ammonium formate pH 4.2. The fractions containing the product were lyophilized and stored at -20 °C. Oxalyl-CoA was dissolved in 10 mM acetate buffer pH 4.5. The concentration was determined by enzymatic depletion with PanE2, following NADPH consumption at 340 nm.

Mandelyl-CoA synthesis

Mandelyl-CoA and (*S*)-mandelyl-CoA were synthesized chemically with the carbonyldiimidazole (CDI) CoA-acylation method described previously.[35] 21 mg CDI (0.127 mmol, 4 eq.) was dissolved in 1 mL tetrahydrofuran, mandelic acid or (*S*)-mandelic acid was added (0.127 mmol, 4 eq.) and the mixture stirred at 22 °C for 15 min. 25 mg CoA (0.032 mmol, 1 eq.) was dissolved in 1 mL 0.5 M $NaHCO_3$ and added to the reaction mixture, followed by stirring at 22 °C for 30 min. THF was removed by applying vacuum (100 mbar) for 5 min. The mixture was then purified by preparative HPLC with a methanol gradient in 25 mM ammonium formate pH 8.0. The fractions containing the product were lyophilized and stored at -20 °C. Mandelyl-CoA was dissolved in 10 mM acetate buffer pH 4.5. The concentration was determined spectrophotometrically, using the extinction coefficient for saturated CoA esters (ε_{260nm} = 16.4 cm^{-1} mM^{-1}). The concentration was confirmed by enzymatic depletion with YciA, detecting the liberated CoA with Ellman's reagent (ε_{412nm} = 14.15 cm^{-1} mM^{-1}).[36]

Cloning and Mutagenesis

Oligonucleotides were obtained from Eurofins Genomics (Ebersbach, Germany). *oxc* (MexAM1_META1p0990), *oxs* (MexAM1_META1p2130) and *panE2* (MexAM1_META1p3141) were PCR-amplified from *Methylorubrum extorquens* chromosomal DNA using the corresponding primers (Table S1). The purified PCR products were digested with NdeI and BamHI and ligated into pET-16b. Correct cloning was confirmed by sequencing (Eurofins Genomics). Human *hacl1* (Gene ID: 26061) was obtained by gene synthesis (Table S2), performed by BaseClear (Leiden, The Netherlands). The gene was codon optimized for *E. coli* and sub-cloned into pET-16b. *adk*, *paaI*, *tesB* and *yciA* were obtained from the ASKA collection.[37] Point mutations were introduced into *oxc* by PCR using mismatch primers (Table S1). A 50 µL reaction contained 60 ng of pET-16b_OXC_{Me}, 0.25 µM forward and reverse primer, 200 µM dNTP, 5 µL 10x Reaction Buffer, 1 µL Phusion polymerase (2 U/µL). Template plasmid was removed by DpnI digest (10 U) at 37 °C immediately after PCR amplification. Mutations were confirmed by sequencing.

Table S1. Primers used for cloning of OXC and OXS and site-directed mutagenesis of OXC.

Primer name	**Nucleotide sequence (5‘ to 3‘)** (restriction sites or mismatches are underlined)
oxc_fw_NdeI	GTTCACATATGACCGTCCAGGCCCAG
oxc_rv_BamHI	CGCTGGATCCTCACTTCTTCTTCAAGGTGCTC
oxs_fw_NdeI	GTTCACATATGACGATGCTTCTGCC
oxs_rv_BamHI	CAATGGATCCTCAGACCAGCCCGAG
panE2_fw_NdeI	GCGCACATATGAGCATCGCGATCGTCG
panE2_rv_BamHI	CAGAGGATCCTCATGCTCCCTGGATCGC
oxc_Y497A_fw	CAACAACGGCATCGCTCGCGGCACCGAC
oxc_Y497A_rv	GTCGGTGCCGCGAGCGATGCCGTTGTTG
oxc_S568A_fw	CCGGCAGCGAGGCCGGCAATATCGG
oxc_S568A_rv	CCGATATTGCCGGCCTCGCTGCCGG

Table S2. Synthetic gene sequences

Gene name	**DNA sequence**
hacl1 (Gene ID 26061)	ATGCCGGACAGTAACTTCGCAGAGCGCAGCGAGGAGCAGGTGTCTGGTGCTAAAGTCATCGCTCAGGCCCTGAAA ACGCAAGATGTGGAGTACATATTTGGCATCGTAGGCATCCCAGTGACCGAAATCGCCATTGCTGCCCAGCAGCTAG GCATCAAGTACATCGGGATGAGGAATGAGCAAGCGGCTTGTTATGCTGCCTCCGCGATTGGATATCTGACAAGCAG GCCAGGAGTCTGCCTTGTTGTTTCTGGCCCAGGTCTCATCCATGCCTTGGGCGGTATGGCAAATGCAAACATGAAC TGCTGGCCCTTGCTTGTGATTGGTGGTTCCTCTGAAAGAAACCAAGAAACAATGGGAGCTTTCCAGGAGTTTCCTCA GGTTGAAGCTTGTAGATTATATACCAAGTTCTCTGCCCGCYCAAGCAGCATAGAAGCTATTCCTTTTGTTATTGAAAA GGCAGTGAGAAGCAGTATCTATGGTCGTCCAGGTGCTTGCTATGTTGACATACCAGCAGATTTTGTGAACCTTCAG GTGAATGTGAATTCTATAAAGTACATGGAACGCTGCATGTCACCTCCTATTAGCATGGCAGAAACCTCTGCTGTGTG CACGGCGGCTTCTGTTATTAGGAATGCCAAACAACCCCTTCTTATCATCGGGAAAGGTGCTGCTTACGCTCATGCA GAAGAGAGTATCAAGAAATTGGTGGAGCAATATAAACTGCCATTTTTGCCCACCCCTATGGGAAAGGGTGTTGTCC CTGACAACCATCCATACTGTGTAGGTGCAGCCAGATCCAGGGCTTTGCAATTTGCTGATGTAATTGTGTTATTTGGT GCCAGACTAAATTGGATTTTACATTTTGGACTGCCTCCAAGATATCAGCCAGATGTGAAGTTTATCCAGGTTGATATC TGTGCAGAAGAATTGGGGAATAATGTAAAGCCCGCTGTTACTTTGCTAGGAAACATACATGCTGTCACTAAGCAGCT TTTAGAGGAACTTGATAAAACACCATGGCAGTATCCTCCAGAGAGCAAGTGGTGGAAAACTCTGAGAGAAAAAATGA AGAGCAATGAAGCTGCATCCAAGGAACTAGCTTCTAAAAAATCCCTGCCTATGAATTATTACACAGTATTCTACCATG TTCAAGAACAACTACCTAGAGACTGTTTCGTGGTAAGTGAAGGAGCAAATACTATGGACATTGGACGGACTGTGCTT CAGAACTACCTTCCTCGTCACAGGCTTGATGCTGGTACTTTCGGAACAATGGGAGTTGGTTTGGGATTTGCTATTGC AGCTGCCGTGGTGGCTAAAGATAGAAGCCCTGGGCATTGGATCATCTGTGTGGAAGGAGACAGTGCATTTGGGTTT TCTGGCATGGAGGTAGAAACCATCTGCAGGTACAACTTGCCAATCATACTGTTGGTAGTGAATAACAATGGAATTTA CCAAGGTTTTGATACAGATACTTGGAAAGAAATGTTAAAATTTCAAGATGCTACTGCAGTGGTCCCTCCAATGTGTTT GCTGCCAAATTCACATTATGAGCAAGTCATGACTGCATTTGGAGGCAAAGGGTATTTTGTACAAACACCAGAAGAAC

	TCCAAAAATCCCTGGAGCAGAGCCTAGCAGACACAACTAAACCTTCTCTTATCAACATCATGATTGAGCCACAAGCC ACACGGAAGGCCCAGGATTTTCATTGGCTGACCCGCTCTAATATGTAA
hacl1 codon optimized for *E. coli*	ATGCCGGATTCAAACTTTGCTGAACGTTCAGAAGAACAGGTTTCTGGCGCAAAAGTGATTGCGCAAGCGCTAAAAA CCCAGGATGTCGAATATATCTTCGGTATAGTGGGTATTCCGGTGACGGAAATCGCCATCGCGGCTCAGCAACTCGG TATTAAATACATCGGTATGCGTAATGAGCAGGCGGCTTGTTACGCAGCAAGCGCGATTGGCTATCTGACCTCAAGA CCGGGGGTATGCCTTGTTGTCAGCGGCCCGGGCCTGATTCATGCCCTGGGTGGTATGGCGAATGCTAACATGAAC TGCTGGCCGCTGTTAGTTATAGGCGGCAGCAGTGAACGCAATCAGGAGACCATGGGCGCATTCCAAGAGTTTCCT CAGGTGGAAGCCTGCCGCCTGTATACCAAATTTTCTGCTAGACCTTCGTCAATAGAAGCCATCCCCTTCGTGATTGA AAAAGCGGTCCGTAGTTCTATCTATGGGCGCCCTGGGGCTTGCTATGTTGACATTCCCGCCGACTTCGTAAACCTC CAAGTGAATGTCAATTCCATTAAATACATGGAACGCTGTATGTCGCCGCCGATCAGTATGGCCGAGACCAGCGCCG TGTGCACAGCGGCGAGTGTTATCCGCAACGCCAAACAGCCATTACTAATAATTGGGAAAGGAGCCGCCTATGCTCA CGCTGAAGAGTCCATCAAAAAACTGGTGGAACAATACAAGTTGCCCTTTTTACCTACTCCCATGGGTAAAGGTGTTG TACCTGATAATCATCCATACTGCGTCGGTGCAGCCCGTTCCCGCGCACTACAGTTCGCTGACGTGATTGTGCTGTTT GGTGCGCGTTTAAACTGGATACTCCACTTCGGTTTACCGCCACGCTACCAGCCGGATGTGAAATTTATACAGGTTG ACATCTGTGCCGAAGAACTGGGCAATAATGTGAAACCAGCAGTCACTTTGCTGGGGAACATCCATGCAGTCACCAA ACAACTGCTTGAAGAGCTGGACAAGACGCCGTGGCAATATCCGCCAGAGTCAAAATGGTGGAAGACGCTACGTGA GAAAATGAAGAGCAACGAAGCGGCTTCGAAAGAGTTGGCCTCCAAAAAGTCATTGCCTATGAATTACTATACCGTCT TTTATCATGTCCAAGAGCAGCTGCCGCGGGATTGCTTTGTCGTGTCGGAGGGCGCCAACACCATGGATATAGGTCG CACTGTCCTTCAGAACTATCTGCCCCGGCATCGCCTGGACGCTGGTACCTTCGGAACTATGGGCGTAGGGCTGGG CTTTGCTATTGCGGCGGCAGTGGTAGCAAAGGATCGCAGCCCGGGGCAGTGGATTATTTGTGTAGAAGGGGACTC CGCATTTGGATTCTCTGGTATGGAGGTAGAGACAATATGCCGTTATAATTTGCCAATTATCCTTCTAGTGGTGAATAA CAATGGTATTTATCAGGGCTTTGATACAGATACGTGGAAGGAAATGCTGAAATTTCAAGATGCCACAGCCGTGGTAC CCCCCATGTGTTTGCTGCCGAACTCTCATTACGAGCAAGTGATGACAGCATTCGGCGGCAAGGGTTACTTTGTGCA GACACCGGAAGAACTTCAAAAGAGCCTTCGCCAGAGCTTGGCGGACACCACGAAACCGAGTCTGATTAATATAATG ATTGAACCACAGGCCACAAGGAAGGCACAAGACTTTCACTGGCTCACGCGTTCAAATATGTAA

Protein Production and Purification

All proteins except $HACL_{Hs}$ were heterologously produced in *E. coli* BL21 (DE3). 500 mL TB containing 100 µg/mL ampicillin (OXC_{Me}, OXS, PduP, and PanE2) or 34 µg/mL chloramphenicol (AdK, PaaI, TesB and YciA) was inoculated with freshly-transformed cells and incubated at 37 °C. After reaching an OD_{600} of 0.8 expression was induced by adding IPTG to a final concentration of 0.25 mM and the incubation temperature was lowered to 25 °C. $HACL_{Hs}$ was produced in *E. coli* ArcticExpress (DE3). 1 L LB containing 100 µg/mL ampicillin and 15 µg/mL gentamycin was inoculated with freshly-transformed cells and incubated at 37 °C. After reaching an OD_{600} of 0.8 the culture was cooled on ice for 15 min. Then expression was induced by adding IPTG to a final concentration of 0.1 mM and the incubation temperature was lowered to 15 °C. Cells were harvested after 16 h (24 h for $HACL_{Hs}$) by centrifugation (4500× g, 10 min) and resuspended in buffer A (500 mM KCl, 50 mM HEPES-KOH pH 7.6). If not used immediately, cell pellets were flash-frozen in liquid nitrogen and stored at −20 °C. The cell lysate obtained by sonication was clarified by centrifugation 75,000× g at 4 °C for 45 min. The supernatant was filtered through a 0.4 µm syringe tip filter (Sarstedt, Nümbrecht, Germany). Ni-affinity purification was performed with an Äkta FPLC system from GE Healthcare (GE Healthcare, Freiburg, Germany). The filtered soluble lysate was loaded onto a 1 mL Ni-Sepharose Fast Flow column (HisTrap FF, GE Healthcare, Little Chalfont, UK) that had been equilibrated with 10 mL buffer A. After washing with 20 mL 85% buffer A, 15% buffer B (500 mM KCl, 50 mM HEPES-KOH pH 7.6, 500 mM imidazole), the protein was eluted with 100% buffer B. Fractions containing purified protein were pooled and the buffer was exchanged to storage buffer (150 mM KCl, 50 mM HEPES-KOH pH 7.6) with a desalting column (HiTrap, GE Healthcare). Proteins were concentrated by ultrafiltration (Amicon Ultra). Concentration was determined on a NanoDrop 2000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA) using the extinction coefficient at 280 nm, as calculated by protparam (https://web.expasy.org/protparam/). Enzyme purity was confirmed by SDS-PAGE. The purified proteins were stored in 50 vol% glycerol at −20 °C. OXC_{Me} wild-type and mutants were flash-frozen in liquid nitrogen and stored at −80 °C.

LC-MS analses

Samples were prepared for LC-MS analysis by quenching an aliquot of a reaction with formic acid (final concentration 4%) and centrifuging for 10 min at 17,000 rcf, to remove precipitated proteins. LC-MS data were analyzed and quantified using MassHunter Qualitative Navigator and Quantitative Analysis software (Agilent, Waldbronn, Germany).

LC-MS detection of CoA esters

Samples were diluted 1:10 in H_2O. UPLC-high resolution MS of CoA-esters was performed as described previously.[38] CoA-esters were analyzed using an Agilent 6550 iFunnel Q-TOF LC-MS system equipped with an electrospray ionization source set to positive ionization mode. Compounds were separated on a RP-18 column (50 mm x 2.1 mm, particle size 1.7 µm, Kinetex EVO C18, Phenomenex) using a mobile phase system comprised of 50 mM ammonium formate pH 8.1 (A) and methanol (B). Chromatographic separation was carried out using the following gradient condition at a flow rate of 250 µL/min: 0 min 2.5% B; 2.5 min 2.5% B; 8 min 23% B; 10 min 80 %B; 11 min 80%; 12 min 2.5% B; 12.5 min 0% B. The column oven was set to 40 °C and autosampler was maintained at 10 °C. Standard injection volume was 1 µL. Capillary voltage was set at 3.5 kV and nitrogen gas was used as nebulizing (20 psig), drying (13 L/min, 225 °C) and sheath gas (12 L/min, 400°C). The TOF was calibrated using an ESI-L Low Concentration Tuning Mix (Agilent) before measurement (residuals less than 2 ppm for five reference ions) and was recalibrated during a run using 922.0908 *m/z* as reference mass. The scan range for MS and MS/MS data is 100-1000 *m/z* and 50-1000 *m/z* respectively. Collision energy used for MS/MS fragmentation was 35 eV.

LC-MS detection of mandelic acid derivatives

Samples were diluted 1:10 in H_2O. UPLC-high resolution MS analyses were performed on an Agilent 6550 iFunnel QTOF LC/MS system equipped with an electrospray ionization source to negative ionization mode. The analytes were isocratically chromatographed on a chiral column (100 mm x 2.1 mm, particle size 2.7 µm, Poroshell 120 Chiral-T, Agilent) kept at ambient tempature using a mobile phase system comprised of 30:70 20 mM ammonium formate pH 4 / methanol at a flow rate of 250 µL/min for 10 min. Samples were held at 10°C and injection volume was 1 µL. Capillary voltage was set at 3.5 kV and nitrogen gas was used as nebulizing (20 psig), drying (13 L/min, 225 °C) and sheath gas (12 L/min, 400°C). MS data were acquired with a scan range of 100-1100 *m/z*. For quantification the calculated *m/z* value of $[M-H]^-$ was used to obtain the extracted ion count from the total ion count.

Enzyme assays with LC-MS detection

Unless noted otherwise, all LC-MS assays were carried out at 30 °C in reaction buffer consisting of 50 mM TES-KOH pH 6.8, 10 mM $MgCl_2$. 0.5 mM ADP, 0.15 mM ThDP.

OXC_{Me} and $HACL_{Hs}$ aldehyde screen

In reaction buffer, 1 mM formyl-CoA and 10 mM aldehyde (formaldehyde, acetaldehyde, glycolaldehyde, propionaldehyde, glyceraldehyde, glyoxylate, succinic semialdehyde, benzaldehyde, and phenylacetaldehyde) was mixed with 5 µM OXC_{Me} and $HACL_{Hs}$, respectively. The reaction was stopped after 1 h and products analyzed with the CoA ester method described above.

YciA substrate screen

Thioesterase activity of YciA was determined by adding 2 µM YciA to the reaction buffer containing 25 mM benzaldehyde and 0.5 mM formyl-CoA, oxalyl-CoA and mandelyl-CoA, respectively. Samples were taken after 0, 1, 5 and 30 min and analyzed with the CoA ester method described above.

OXS- OXC_{Me}-YciA cascade prototyping

In a 1.5 mL microfuge tube, reaction buffer containing 0.5 mM CoA, 10 mM ATP, 25 mM benzaldehyde, 2 µM YciA, 5 µM OXC_{Me} and 5 µM OXS were mixed and the reaction was initiated by the addition of 10 mM disodium oxalate. For the negative controls each enzyme was omitted in a separate reaction. Samples were taken after 0, 3, 15, 40 min, 1, 2, 3 and 22 h and analyzed with the mandelic acid derivatives method described above. For quantification commercial, racemic mandelic acid was diluted in reaction buffer to appropriate concentrations to obtain a calibration curve (Figure S8).

OXS- OXC_{Me}-YciA cascade aldehyde scope

The aldehyde substrate screen of the OXS-$OXC_{Me\text{-}Y497A}$-YciA cascade was carried out as described above, except that the aromatic aldehydes were prepared as 33 mM stocks in 20 vol% DMSO and diluted to final concentration of 25 mM into the assay. Samples were analyzed with the mandelic acid derivatives method described above. Mandelic acid (**3a**), 4-chloromandelic acid (**3c**), 2-chloromandelic acid (**3d**) and 3-phenyllactic acid (**3b**) were quantified by comparison to commercial standards (Figure S8).

OXS- OXC_{Me}-YciA cascade on semi-preparative scale

In a glass vial (25 mL reaction volume) 50 mM TES-KOH pH 6.8, 10 mM $MgCl_2$, 0.5 mM ADP (0.0125 mmol, 5.3 mg), 0.15 mM ThDP (0.00375 mmol, 1.7 mg), 0.5 mM CoA (0.0125 mmol, 9.8 mg), 25 mM disodium oxalate (0.625 mmol, 84 mg), 25 mM benzaldehyde (0.625 mmol, 66 mg), 2 µM YciA, 10 µM $OXC_{Me\text{-}Y497A}$, 5 µM OXS, 1.3 µM adenylate kinase and 25 units creatine kinase (Roth) were mixed and the reaction was initiated by the addition of 10 mM creatine phosphate. The same amount of creatine phosphate was added after 1, 2, 4 and 6 h to reach a final concentration of 50 mM (1.25 mmol, 409 mg). The vials were incubated without shaking at 30 °C for 24 h. Then the pH was lowered to 3 by adding HCl. The quenched reaction was saturated with

NaCl and extracted with 25 mL diethylether. The organic phase was dried over $MgSO_4$ and filtered. After evaporation of the ether under vacuum, 50 mg (0.331 mmol, 53%) of a white solid remained, which was confirmed to be mandelic acid by UV (Figure S9), HPLC (Figure S10), and NMR (Figure S11). ^{1}H NMR (300 MHz, DMSO-d_6) δ 7.4 (m, 5H); 5.05 (s, 1H). ^{13}C NMR (300 MHz, DMSO-d_6) δ 174.5, 140.7, 128.6, 128.1, 127.1, 72.9.

Spectrophotometric enzyme assays

Assays were performed on a Cary-60 UV/Vis spectrophotometer (Agilent) at 30°C using quartz cuvettes (10 mm path length; Hellma, Müllheim, Germany). For the determination of steady-state kinetic parameters, each substrate concentration was measured in triplicates and the obtained curves were fit using GraphPad Prism 7. Hyperbolic curves were fit to the Michaelis-Menten equation to obtain apparent k_{cat} and K_M values.

Michaelis-Menten kinetics of OXS

Oxalyl-CoA production was followed by coupling OXS to purified PanE2, an NADPH-dependent oxalyl-CoA reductase from *M. extorquens*.[25] An assay containing 50 mM potassium phosphate pH 6.5, 0.3 mM NADPH, 10 mM $MgCl_2$, 1 mM ATP, 2.5 mM CoA, 1.2 µM PanE2, 176 nM OXS was preincubated for 2 min and the reaction started by adding disodium oxalate to a final concentration of 5, 10, 25, 100, 250, 1,000 µM, respectively. Reaction procedure was monitored by following the oxidation of NADPH at 340 nm.

Michaelis-Menten kinetics OXC_{Me}

Formyl-CoA production was followed by coupling OXC to purified PduP, a promiscuous CoA-dependent aldehyde dehydrogenase that reduces formyl-CoA to formaldehyde under NADH consumption.[34] An assay containing 50 mM MES-KOH pH 6.5, 0.3 mM NADH, 10 mM $MgCl_2$, 0.5 mM ADP, 0.15 mM ThDP, 5 µM PduP, and OXC_{Me} (concentration depending on the mutant) was preincubated for 2 min and the reaction started by adding oxalyl-CoA (concentrations depending on the mutant). Reaction procedure was monitored by following the oxidation of NADH at 340 nm. Michaelis-Menten graphs are shown in Figure S3.

2.5 References

[1] Kluger, R., and Tittmann, K. (2008) Thiamin diphosphate catalysis: enzymic and nonenzymic covalent intermediates, *Chem. Rev. 108*, 1797-1833.
[2] Schörken, U., and Sprenger, G. A. (1998) Thiamin-dependent enzymes as catalysts in chemoenzymatic syntheses, *Biochim. Biophys. Acta 1385*, 229-243.
[3] Demir, A. S., Dünnwald, T., Iding, H., Pohl, M., and Müller, M. (1999) Asymmetric benzoin reaction catalyzed by benzoylformate decarboxylase, *Tetrahedron Asymmetry 10*, 4769-4774.
[4] Demir, A. S., Sesenoglu, O., Eren, E., Hosrik, B., Pohl, M., Janzen, E., Kolter, D., Feldmann, R., Dünkelmann, P., and Müller, M. (2002) Enantioselective synthesis of alpha-hydroxy ketones via benzaldehyde lyase-catalyzed C-C bond formation reaction, *Adv. Synth. Catal. 344*, 96-103.
[5] Payongsri, P., Steadman, D., Strafford, J., MacMurray, A., Hailes, H. C., and Dalby, P. A. (2012) Rational substrate and enzyme engineering of transketolase for aromatics, *Org. Biomol. Chem. 10*, 9021-9029.
[6] L'enfant, M., Bruna, F., Lorilliere, M., Ocal, N., Fessner, W. D., Pollegioni, L., Charmantray, F., and Hecquet, L. (2019) One-Pot Cascade Synthesis of (3S)-Hydroxyketones Catalyzed by Transketolase via Hydroxypyruvate Generated in Situ from d-Serine by d-Amino Acid Oxidase, *Adv. Synth. Catal. 361*, 2550-2558.
[7] Ranoux, A., Karmee, S. K., Jin, J. F., Bhaduri, A., Caiazzo, A., Arends, I. W. C. E., and Hanefeld, U. (2012) Enhancement of the Substrate Scope of Transketolase, *ChemBioChem 13*, 1921-1931.
[8] Olah, G. A. (2013) Towards Oil Independence Through Renewable Methanol Chemistry, *Angew. Chem. Int. Ed. 52*, 104-107.
[9] Yishai, O., Lindner, S. N., de la Cruz, J. G., Tenenboim, H., and Bar-Even, A. (2016) The formate bio-economy, *Curr Opin Chem Biol 35*, 1-9.
[10] Siegel, J. B., Smith, A. L., Poust, S., Wargacki, A. J., Bar-Even, A., Louw, C., Shen, B. W., Eiben, C. B., Tran, H. M., Noor, E., Gallaher, J. L., Bale, J., Yoshikuni, Y., Gelb, M. H., Keasling, J. D., Stoddard, B. L., Lidstrom, M. E., and Baker, D. (2015) Computational protein design enables a novel one-carbon assimilation pathway, *P. Natl. Acad. Sci. U. S. A. 112*, 3704-3709.
[11] Quayle, J. R. (1963) Carbon Assimilation by Pseudomonas Oxalaticus (Ox1). 7. Decarboxylation of Oxalyl-Coenzyme a to Formyl-Coenzyme A, *Biochem. J. 89*, 492-503.
[12] Berthold, C. L., Moussatche, P., Richards, N. G., and Lindqvist, Y. (2005) Structural basis for activation of the thiamin diphosphate-dependent enzyme oxalyl-CoA decarboxylase by adenosine diphosphate, *J. Biol. Chem. 280*, 41645-41654.
[13] Foulon, V., Sniekers, M., Huysmans, E., Asselberghs, S., Mahieu, V., Mannaerts, G. P., Van Veldhoven, P. P., and Casteels, M. (2005) Breakdown of 2-hydroxylated straight chain fatty acids via peroxisomal 2-hydroxyphytanoyl-CoA lyase: a revised pathway for the alpha-oxidation of straight chain fatty acids, *J. Biol. Chem. 280*, 9802-9812.
[14] Berthold, C. L., Toyota, C. G., Moussatche, P., Wood, M. D., Leeper, F., Richards, N. G., and Lindqvist, Y. (2007) Crystallographic snapshots of oxalyl-CoA decarboxylase give insights into catalysis by nonoxidative ThDP-dependent decarboxylases, *Structure 15*, 853-861.
[15] Sheng, X., Liu, Y. J., and Zhang, R. (2014) A theoretical study of the catalytic mechanism of oxalyl-CoA decarboxylase, an enzyme for treating urolithiasis, *RSC Adv. 4*, 35777-35788.
[16] Chou, A., Clomburg, J. M., Qian, S., and Gonzalez, R. (2019) 2-Hydroxyacyl-CoA lyase catalyzes acyloin condensation for one-carbon bioconversion, *Nat. Chem. Biol. 15*, 900-906.
[17] Kaplun, A., Binshtein, E., Vyazmensky, M., Steinmetz, A., Barak, Z., Chipman, D. M., Tittmann, K., and Shaanan, B. (2008) Glyoxylate carboligase lacks the canonical active site glutamate of thiamine-dependent enzymes, *Nat. Chem. Biol. 4*, 113-118.
[18] Meyer, D., Walter, L., Kolter, G., Pohl, M., Müller, M., and Tittmann, K. (2011) Conversion of pyruvate decarboxylase into an enantioselective carboligase with biosynthetic potential, *J. Am. Chem. Soc. 133*, 3609-3616.
[19] Dünkelmann, P., Kolter-Jung, D., Nitsche, A., Demir, A. S., Siegert, P., Lingen, B., Baumann, M., Pohl, M., and Müller, M. (2002) Development of a donor-acceptor concept for enzymatic cross-coupling reactions of aldehydes: the first asymmetric cross-benzoin condensation, *J. Am. Chem. Soc. 124*, 12084-12085.
[20] Berthold, C. L., Gocke, D., Wood, M. D., Leeper, F. J., Pohl, M., and Schneider, G. (2007) Structure of the branched-chain keto acid decarboxylase (KdcA) from Lactococcus lactis provides insights into

the structural basis for the chemoselective and enantioselective carboligation reaction, *Acta Crystallogr. D 63*, 1217-1224.
[21] Gröger, H. (2001) Enzymatic Routes to Enantiomerically Pure Aromatic α-Hydroxy Carboxylic Acids: A Further Example for the Diversity of Biocatalysis, *Adv. Synth. Catal. 343*, 547-558.
[22] Sonntag, F., Buchhaupt, M., and Schrader, J. (2014) Thioesterases for ethylmalonyl-CoA pathway derived dicarboxylic acid production in Methylobacterium extorquens AM1, *App. Microbiol. Biotechnol. 98*, 4533-4544.
[23] Kuznetsova, E., Proudfoot, M., Sanders, S. A., Reinking, J., Savchenko, A., Arrowsmith, C. H., Edwards, A. M., and Yakunin, A. F. (2005) Enzyme genomics: Application of general enzymatic screens to discover new enzymes, *FEMS Microbiol. Rev. 29*, 263-279.
[24] Zhuang, Z., Song, F., Zhao, H., Li, L., Cao, J., Eisenstein, E., Herzberg, O., and Dunaway-Mariano, D. (2008) Divergence of function in the hot dog fold enzyme superfamily: the bacterial thioesterase YciA, *Biochemistry 47*, 2789-2796.
[25] Schneider, K., Skovran, E., and Vorholt, J. A. (2012) Oxalyl-coenzyme A reduction to glyoxylate is the preferred route of oxalate assimilation in Methylobacterium extorquens AM1, *J. Bacteriol. 194*, 3144-3155.
[26] Müller, M., Gocke, D., and Pohl, M. (2009) Thiamin diphosphate in biological chemistry: exploitation of diverse thiamin diphosphate-dependent enzymes for asymmetric chemoenzymatic synthesis, *FEBS J. 276*, 2894-2904.
[27] Giovannini, P. P., Bortolini, O., and Massi, A. (2016) Thiamine-Diphosphate-Dependent Enzymes as Catalytic Tools for the Asymmetric Benzoin-Type Reaction, *Eur. J. Org. Chem.*, 4441-4459.
[28] Westphal, R., Hahn, D., Mackfeld, U., Waltzer, S., Beigi, M., Widmann, M., Vogel, C., Pleiss, J., Muller, M., Rother, D., and Pohl, M. (2013) Tailoring the S-Selectivity of 2-Succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate Synthase (MenD) from Escherichia coli, *ChemCatChem 5*, 3587-3594.
[29] Rother, D., Kolter, G., Gerhards, T., Berthold, C. L., Gauchenova, E., Knoll, M., Pleiss, J., Muller, M., Schneider, G., and Pohl, M. (2011) S-Selective Mixed Carboligation by Structure-Based Design of the Pyruvate Decarboxylase from Acetobacter pasteurianus, *ChemCatChem 3*, 1587-1596.
[30] Gong, J. S., Lu, Z. M., Li, H., Shi, J. S., Zhou, Z. M., and Xu, Z. H. (2012) Nitrilases in nitrile biocatalysis: recent progress and forthcoming research, *Microb. Cell Fact. 11*, 142.
[31] Endo, T., and Koizumi, S. (2001) Microbial conversion with cofactor regeneration using genetically engineered bacteria, *Adv. Synth. Catal. 343*, 521-526.
[32] Sly, W. S., and Stadtman, E. R. (1963) Formate Metabolism. I. Formyl Coenzyme a, an Intermediate in the Formate-Dependent Decomposition of Acetyl Phosphate in Clostridium Kluyveri, *J. Biol. Chem. 238*, 2632-2638.
[33] Jonsson, S., Ricagno, S., Lindqvist, Y., and Richards, N. G. (2004) Kinetic and mechanistic characterization of the formyl-CoA transferase from Oxalobacter formigenes, *J. Biol. Chem. 279*, 36003-36012.
[34] Zarzycki, J., Sutter, M., Cortina, N. S., Erb, T. J., and Kerfeld, C. A. (2017) In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glycyl Radical Enzyme - Associated Bacterial Microcompartment, *Sci. Rep. 7*, 42757.
[35] Peter, D. M., Vögeli, B., Cortina, N. S., and Erb, T. J. (2016) A Chemo-Enzymatic Road Map to the Synthesis of CoA Esters, *Molecules 21*, 517.
[36] Riddles, P. W., Blakeley, R. L., and Zerner, B. (1983) Reassessment of Ellman's reagent, *Methods Enzymol 91*, 49-60.
[37] Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005) Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research, *DNA Res 12*, 291-299.
[38] Schwander, T., Schada von Borzyskowski, L., Burgener, S., Cortina, N. S., and Erb, T. J. (2016) A synthetic pathway for the fixation of carbon dioxide in vitro, *Science 354*, 900-904.

2.6 Supplementary Figures

Figure S1. MS/MS spectra of the 2-hydroxyacyl-CoA thioester products. Calculated m/z values of the fragmentation products are indicated on top. Spectra on the left show the parent ion. Spectra on the right show the fragmentation products of the parent ion (blue asterisk). The numbers in the spectrum indicate measured m/z values. Mandelyl-CoA was produced from deuterated benzaldehyde; the deuterium is retained on the α-carbon in the product.

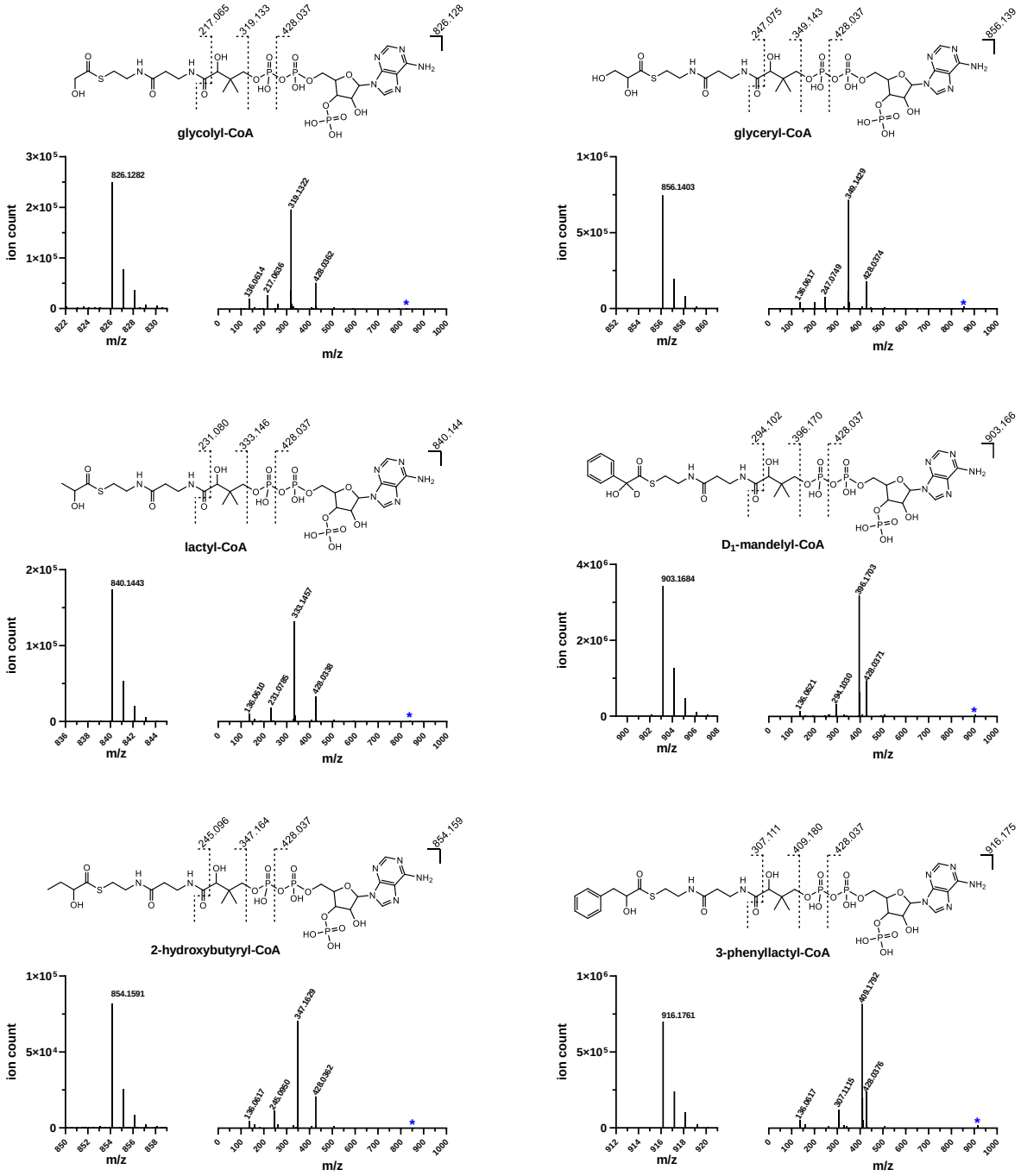


Figure S2. LC-MS analysis of oxalyl-CoA (1 mM) decarboxylation catalyzed by OXC_{Me}, $HACL_{Hs}$ and no enzyme control. The slope of HACL corresponds to a turnover number of approximately <1 min^{-1}.

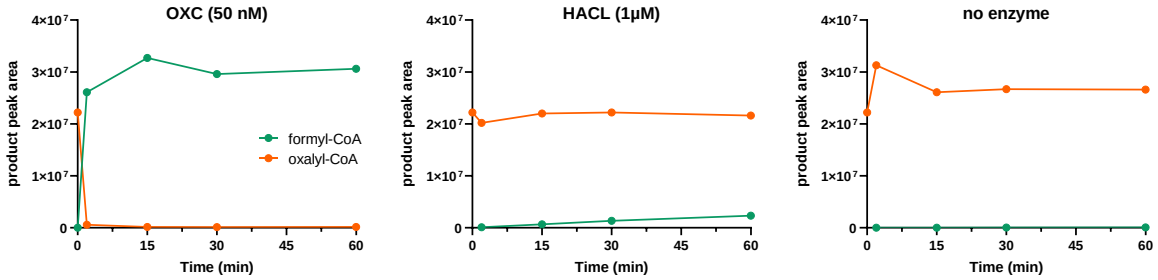


Figure S3. Michaelis-Menten graphs of OXC_{Me} and mutants thereof. Error bars show standard deviation of three replicates.

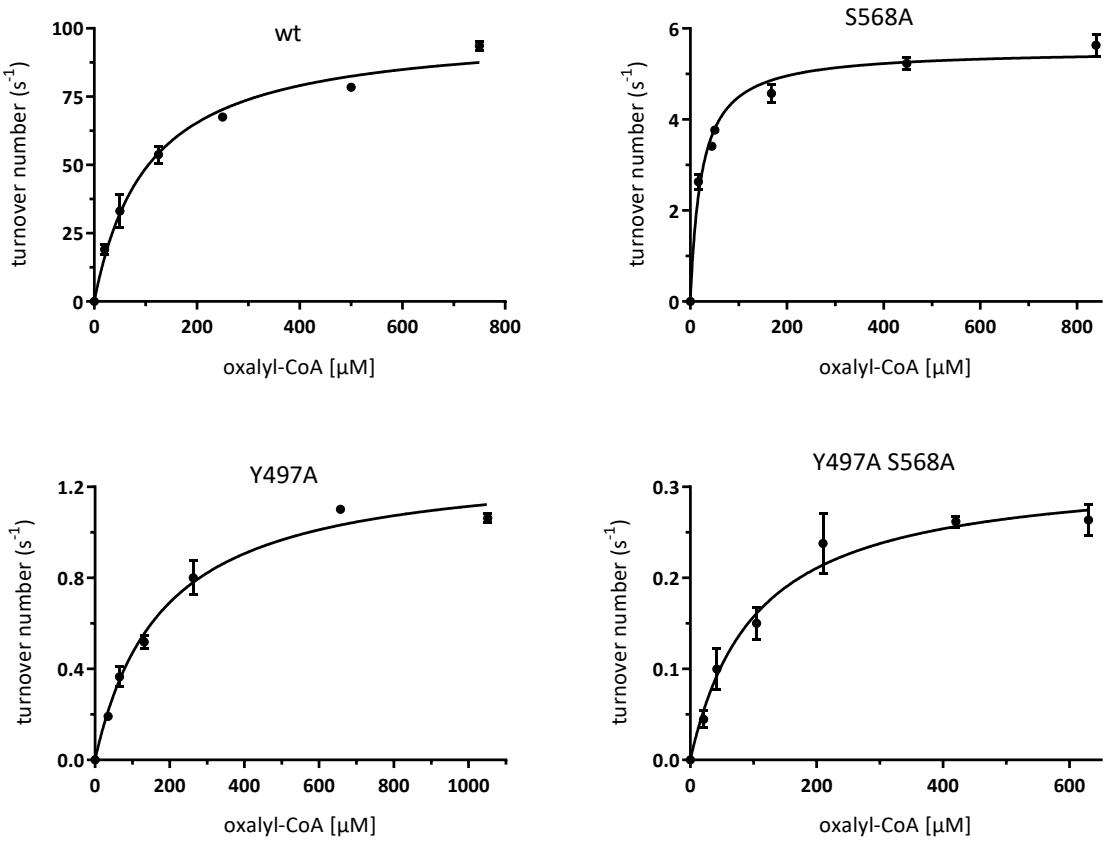
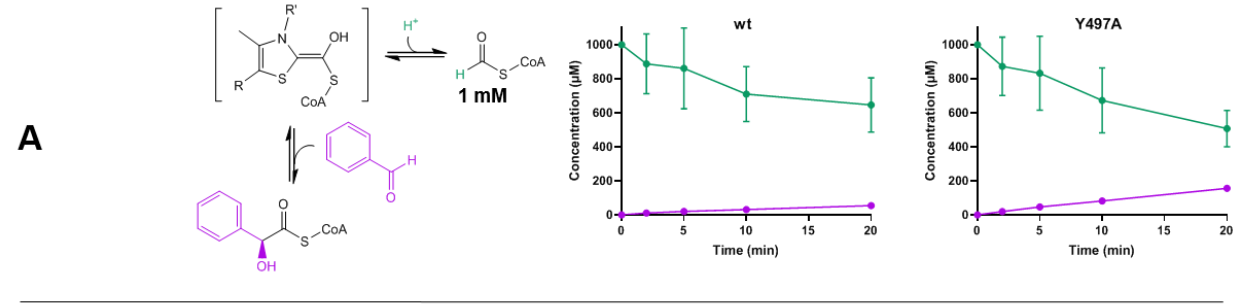
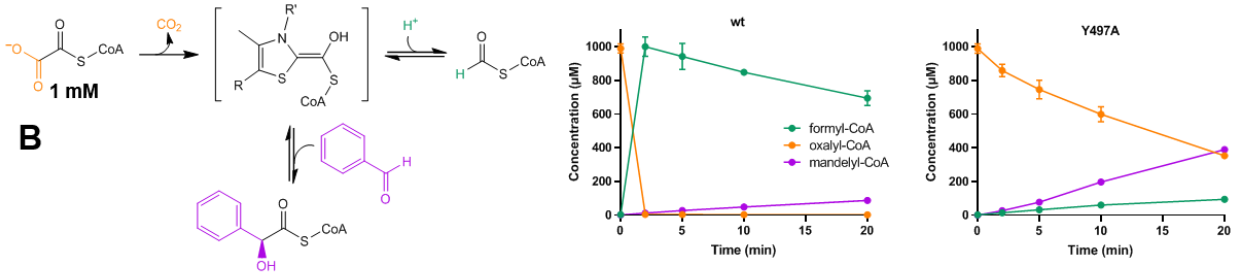


Figure S4. Competing reaction pathways of OXC. Reactions contained 1 µM OXC_{Me}, 25 mM benzaldehyde and were started by adding 1 mM formyl-CoA **(A)** or 1 mM oxalyl-CoA **(B)**. **C)** Apparent turnover numbers were determined by linear regression of the mandelyl-CoA formation rate over 20 min. Error bars show standard deviation of three replicates. Concentration of CoA-esters was determined by comparison to a standard curve obtained from chemically synthesized formyl-CoA, oxalyl-CoA and mandelyl-CoA, respectively.





C

	formyl-CoA	oxalyl-CoA
wt	3 min^{-1}	4 min^{-1}
Y497A	8 min^{-1}	20 min^{-1}

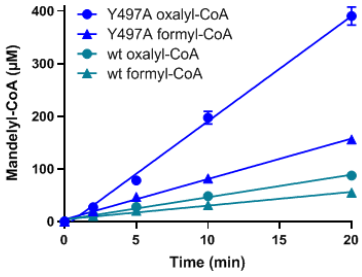


Figure S5. Screen for mandelyl-CoA thioesterase activity. CoA formation was detected with the Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid)), which reacts with free thiols under release of 2-nitro-5-thiobenzoate (ε_{412nm} = 14.15 mM^{-1} cm^{-1}). Assays were carried out at 30 °C in 50 mM MES-KOH pH 6.8 and contained 1 mM Ellman's reagent and 0.5 µM YciA or 0.5 µM TesB or 0.5 µM PaaI. The reaction was started by adding 0.1 mM mandelyl-CoA (time point indicated with the arrow).

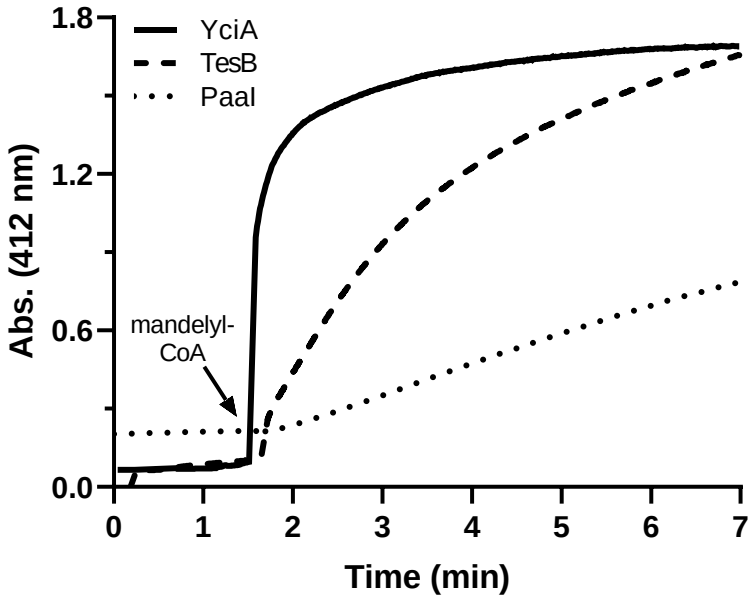


Figure S6. Thioesterase activity of YciA towards oxalyl-CoA, formyl-CoA and racemic mandelyl-CoA. Assays were carried out at 30 °C in reaction buffer consisting of 50 mM TES-KOH pH 6.8, 10 mM $MgCl_2$, 0.5 mM ADP, 0.15 mM ThDP and 25 mM benzaldehyde, 2 µM YciA (**A**) or no YciA (**B**). The reaction was started by adding either 0.5 mM formyl-CoA or 0.5 mM oxalyl-CoA or 0.5 mM mandelyl-CoA. Samples were taken after 0, 1, 5 and 30 minutes and analyzed with the LC-MS detection of CoA esters. CoA esters were quantified by comparison to synthetic standards of formyl-CoA, oxalyl-CoA and mandelyl-CoA, respectively. Error bars show standard deviation of two replicates.

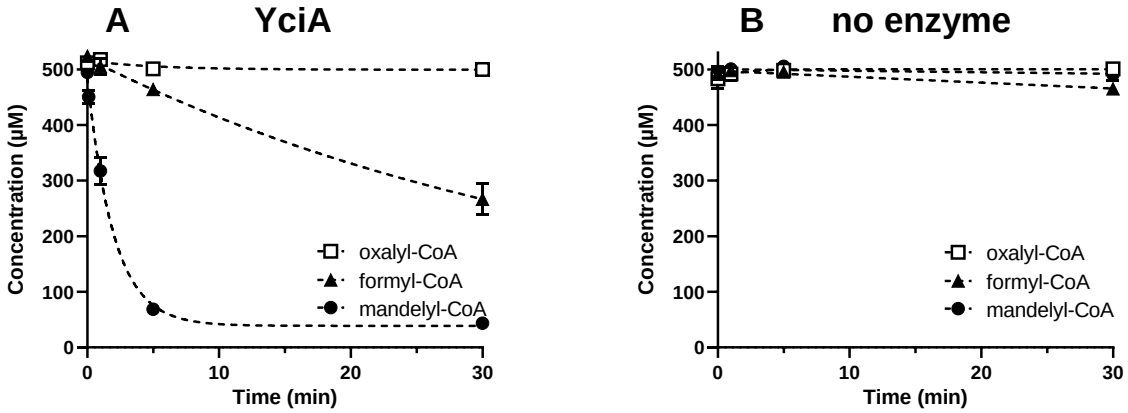


Figure S7. Chiral LC-MS product analysis of the aldehyde scope (Figure 2). On the left side are chromatograms of the last time point (24 h), showing extracted ion counts. The chemical structure of the product and its $[M\text{-}H]^-$ *m/z* used to obtain the extracted ion count are shown in each chromatogram. Dashed lines show commercial standards at a concentration of 10 mM. Plots on the right side show the time courses of the reactions. The lines show a one-phase association fit to the data. MA, mandelic acid. 4-CMA, 4-chloromandelic acid. (*S*)-2-CMA, (*S*)-2-chloromandelic acid. PLA, 3-phenyllactic acid.

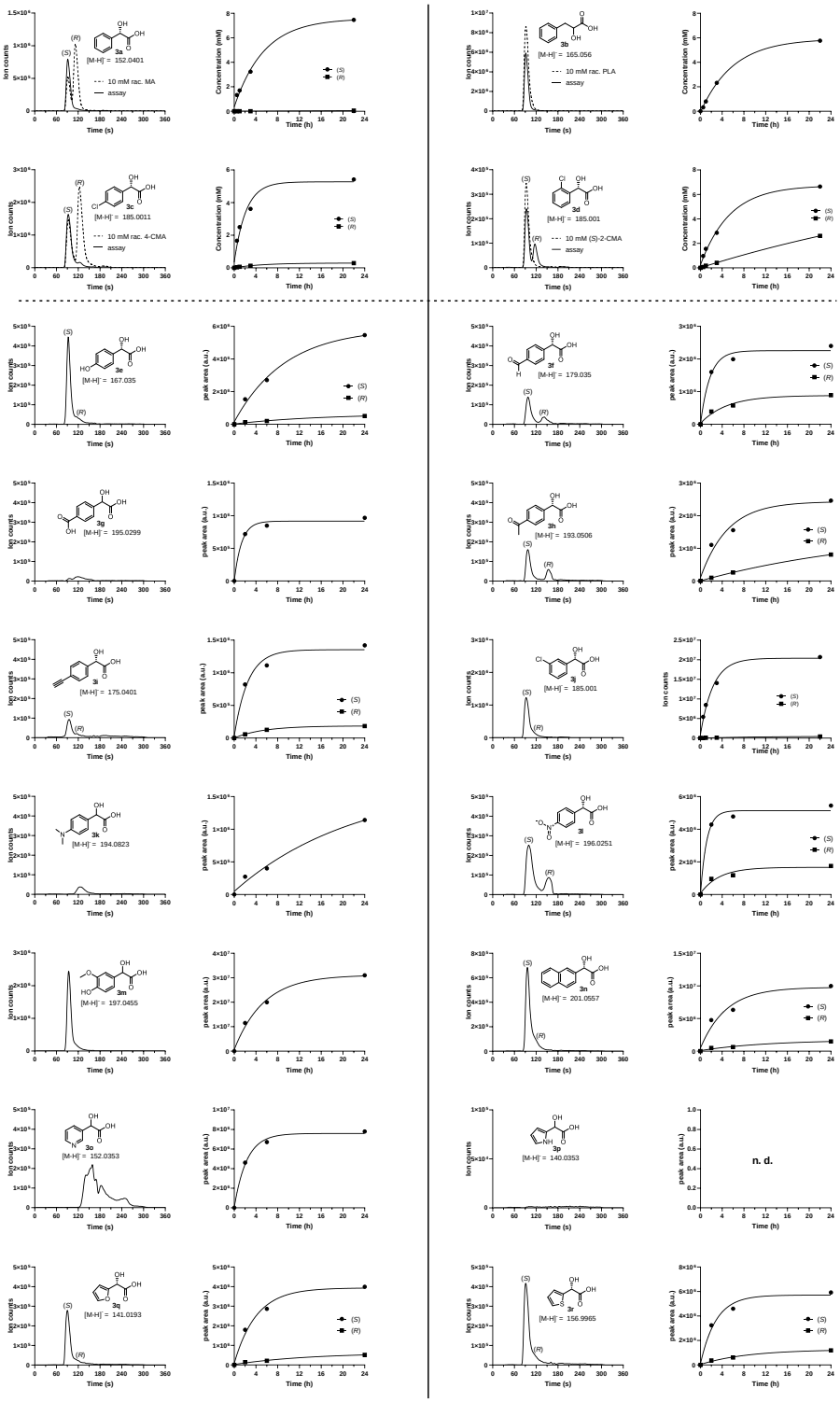
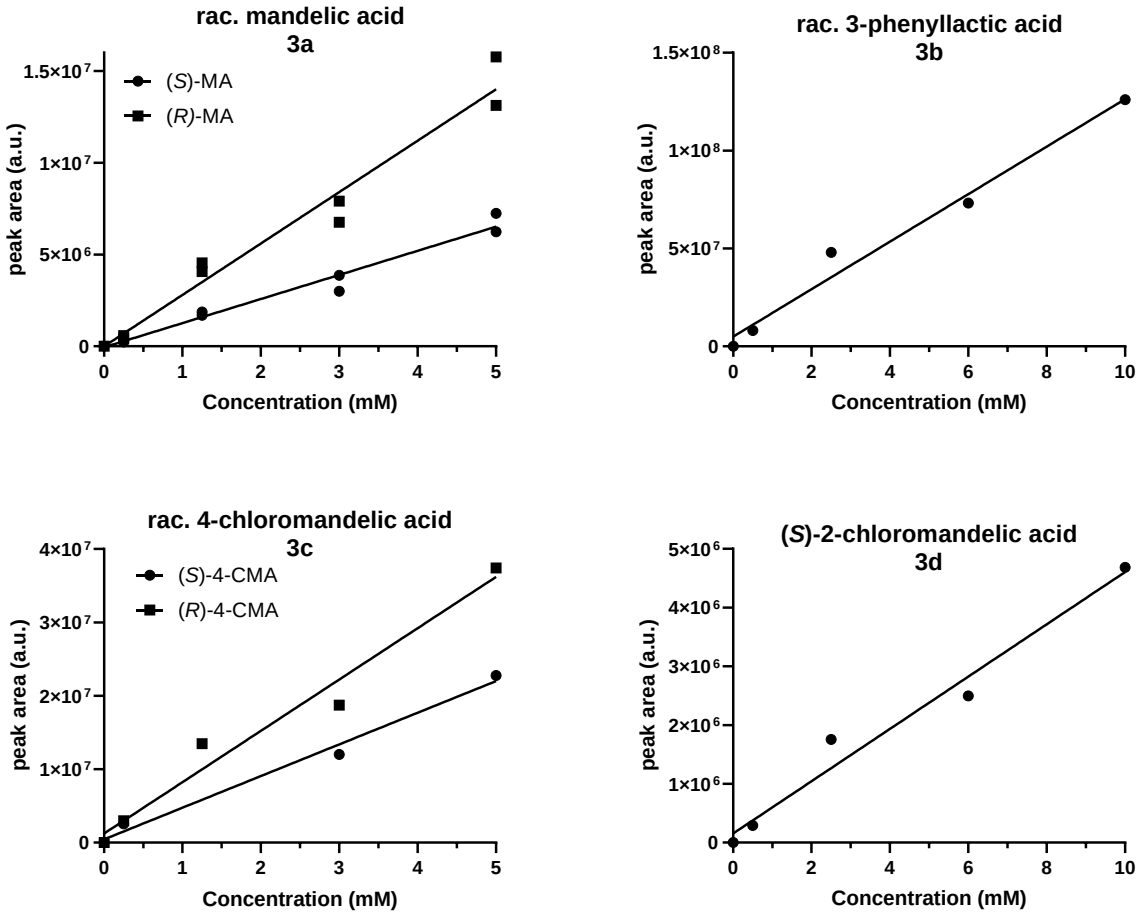


Figure S8. Calibration curves of commercially obtained racemic mandelic acid, racemic 3-phenyllactic acid, racemic 4-chloromandelic acid and (*S*)-2-chloromandelic acid. The lines show a linear fit to the data, slope (A) and Y-intercept (B) are indicated in the table below. Concentration of α-hydroxy acids in the reaction samples was calculated with the equation *concentration* = (*peak area* – B) / A.



Compound	A	B	R^2
(*S*)-**3a**	1313446	-51928	0.9760
(*R*)-**3a**	2801078	-3823	0.9707
3b	12133853	4946676	0.9797
(*S*)-**3c**	4312421	441083	0.9882
(*R*)-**3c**	6991962	1236119	0.9692
(*S*)-**3d**	445091	153684	0.9731

Figure S9. UV spectrum of the mandelic acid semi-preparative synthesis.

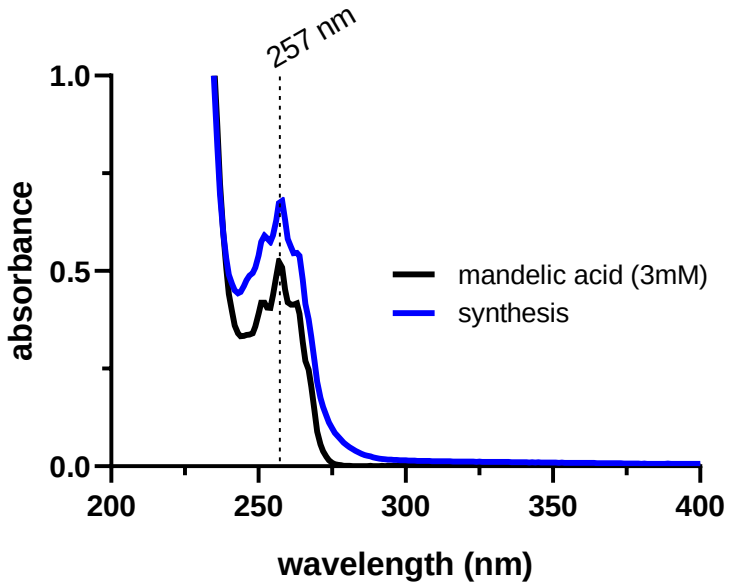


Figure S10. HPLC UV (210 nm) chromatogram of the mandelic acid semi-preparative synthesis.

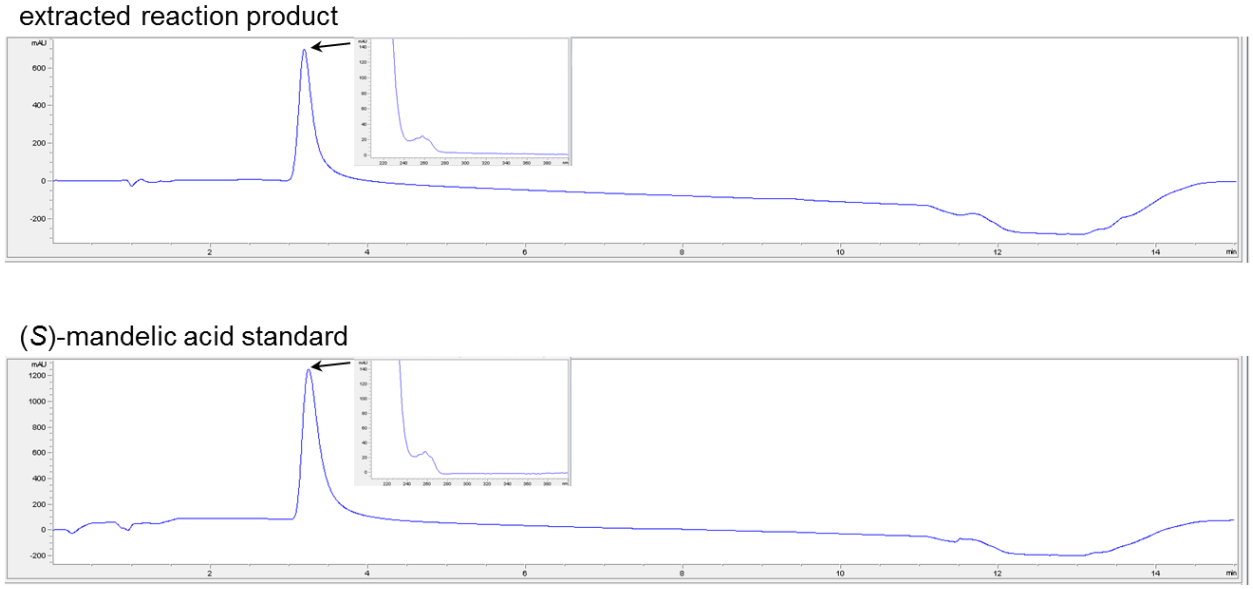
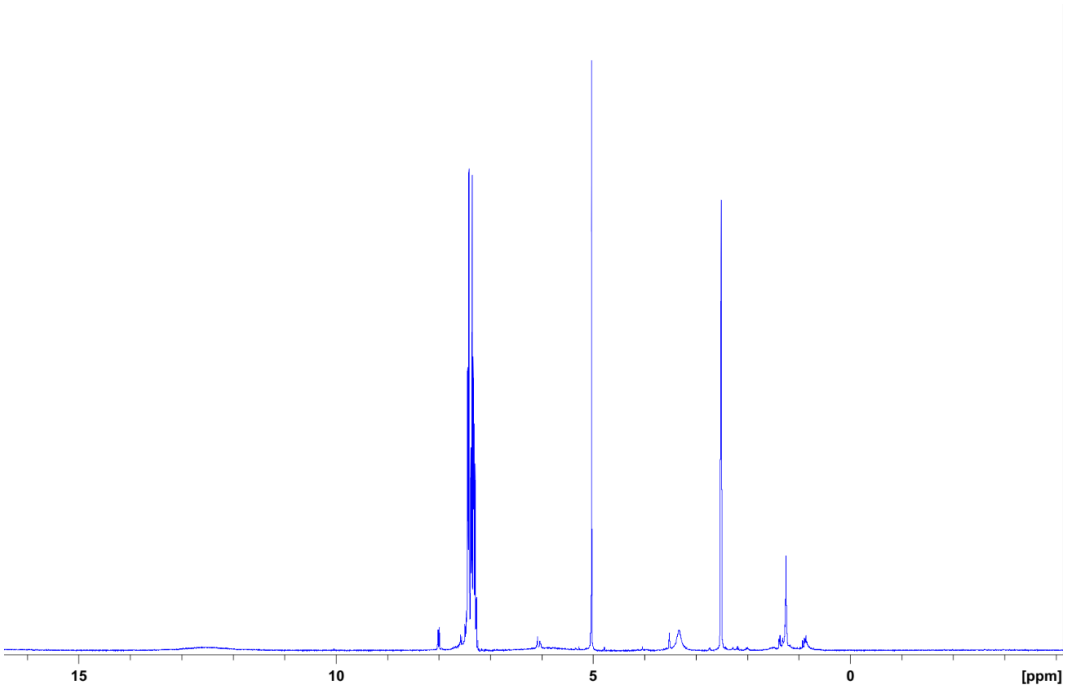
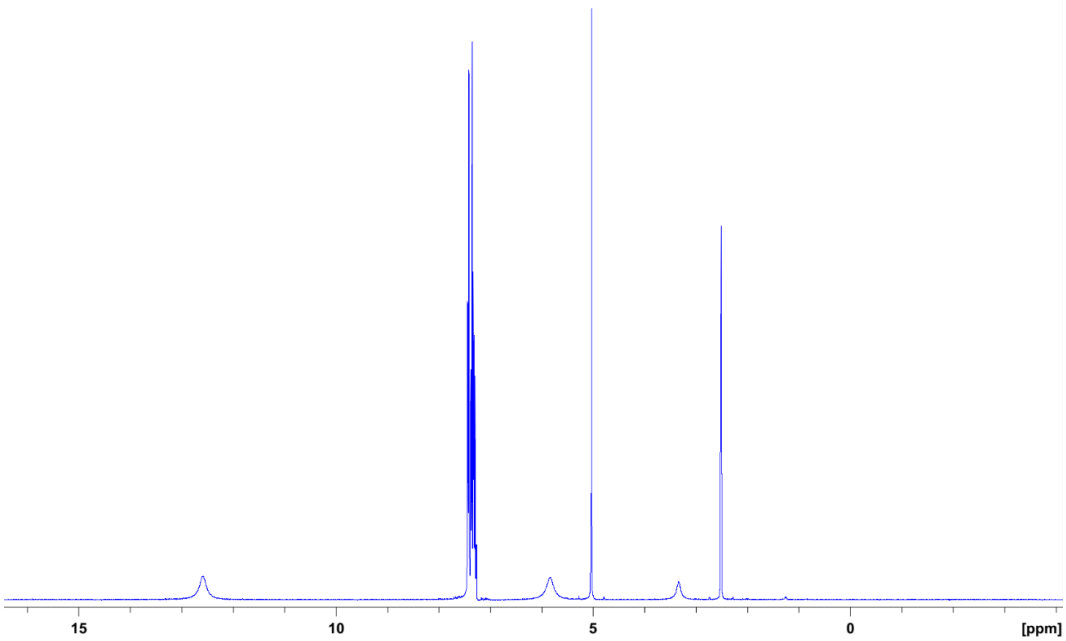


Figure S11. NMR spectra recorded in DMSO-d_6, 300 MHz. **A)** ^{1}H NMR of the extracted reaction product. **B)** ^{1}H NMR of commercial (*S*)-mandelic acid. **C)** ^{13}C NMR of the isolated reaction product. D) ^{13}C NMR of commercial (*S*)-mandelic acid.

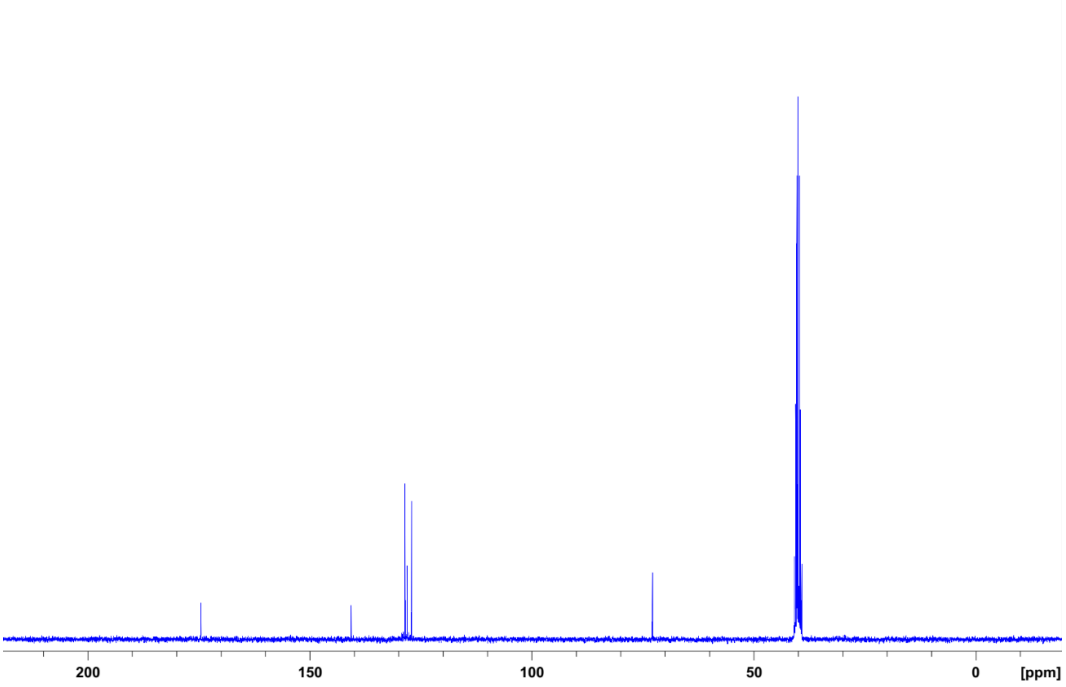
A



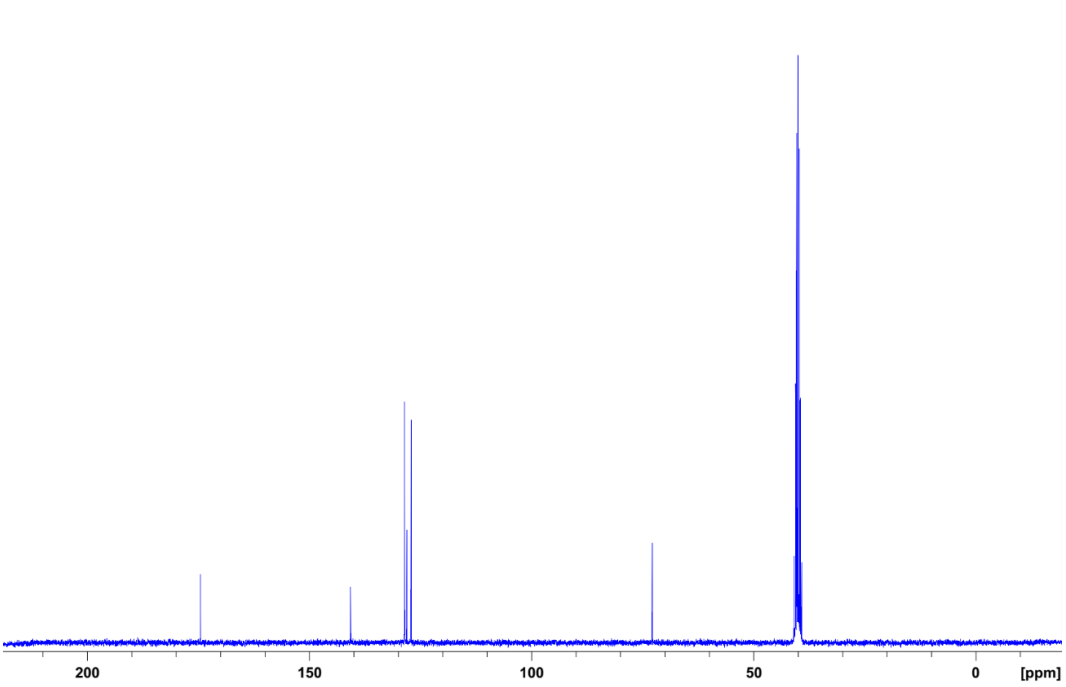
B



C



D



CHAPTER II

Swiss Army knife part 2:

C1 + C1 = C2

3 Engineering a highly efficient acyloin condensation enzyme for synthetic one-carbon fixation

Maren Nattermann[1,*], Simon Burgener[1,*], Pascal Pfister[1], Luca Schulz[1], Alexander Chou[2,3] Ramon Gonzalez[2,3], Tobias J. Erb[1,4]

* equal contribution

[1] Department of Biochemistry & Synthetic Metabolism, Max-Planck-Institute for terrestrial Microbiology, Karl-von-Frisch-Str. 10, 35043 Marburg, Germany

[2] Department of Chemical and Biomolecular Engineering, Rice University, Houston, TX, USA.

[3] Department of Chemical and Biomedical Engineering, University of South Florida, Tampa, FL, USA

[4] LOEWE Center for Synthetic Microbiology, 35043 Marburg, Germany

Manuscript in preparation

Author contributions

S.B. and T.J.E. conceived of the project. M.N. and S.B. designed and performed the experiments. P.P. solved the crystal structures. L.S. screened CoA transferases. A.C. designed and performed the whole-cell biotransformation experiments. T.J.E. and R.G. acquired funding. All authors analyzed the results and wrote the manuscript.

Abstract

Thiamine diphosphate-dependent enzymes have great potential for one-carbon fixation. The OXC/HACL superfamily has recently been discovered to catalyze acyloin condensations of formyl-CoA with a wide array of aldehydes. To enable synthetic one-carbon assimilation routes, we herein employed structure-guided iterative saturation mutagenesis to increase the activity of OXC for the condensation of formyl-CoA and formaldehyde. The resulting quadruple mutant MeOXC4 exhibits a 200-fold increased catalytic efficiency and formaldehyde affinity comparable to natural formaldehyde-fixing enzymes, enabling efficient production of glycolate in whole-cell biotransformations. Furthermore, in stark contrast to all HACLs characterized so far, MeOXC4 is highly soluble in *E. coli*, which paves the way for its use in synthetic one-carbon assimilation routes *in vivo*.

3.1 Introduction

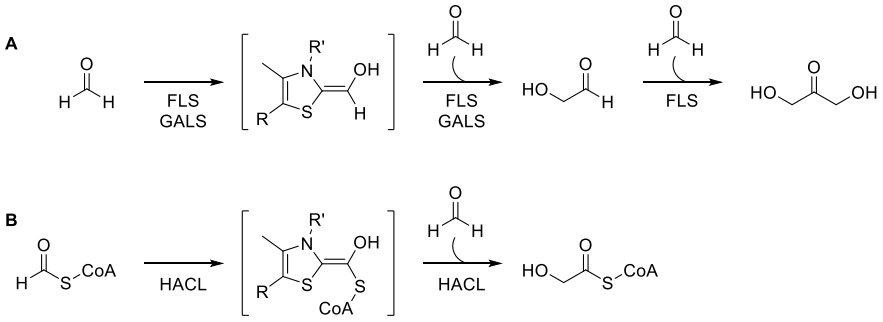
The synthesis of complex molecules from one-carbon (C1) compounds is key to a sustainable, circular economy. C1 compounds, such as formate and methanol, can be derived directly from CO_2 through several processes, including hydrogenation, photochemistry, electrochemistry or biocatalysis,[1-5] and thus serve as sustainable feedstock for the formation of value-added products via microbial fermentation.[6-8] Because natural methylo- and formatotrophs are not well-suited for large-scale biotechnological processes, current efforts aim at engineering well-established platform organisms for growth on C1 compounds.[9-11] Instead of transplanting natural existing pathways, the use of synthetic pathways might facilitate their realization in these organisms, by reducing metabolic complexity and minimizing overlap with the native metabolic network of the host.[12] Moreover, synthetic pathways also allow for carbon assimilation at higher energetic efficiencies compared to natural metabolism, which will ultimately translate into higher product yields.[13, 14]

The key step in all (one)-carbon assimilation pathways is a C-C bond-forming carboligation reaction between a nucleophilic and an electrophilic carbon atom. While many enzymes use electrophilic C1 species, such as CO_2 or formaldehyde, nucleophilic C1 species are rare. Most carboligases employ nucleophiles that contain more than one carbon atom, which makes the direct synthesis of C2 compounds from C1 units challenging.

One way to generate nucleophilic carbon centers is to invert the reactivity of a carbonyl species through Umpolung, which can be achieved through the cofactor thiamine diphosphate (ThDP).[15] In recent years, several ThDP-dependent enzymes have been engineered to catalyze C1-extension reactions, underscoring the potential of these enzymes for synthetic carbon fixation pathways. The most prominent example is the artificial enzyme formolase (FLS), which was derived from benzaldehyde lyase (BAL).[16] Similarly, glycolaldehyde synthase (GALS) was engineered from benzoylformate decarboxylase (BFD).[17] Both of these enzymes produce “active formaldehyde” in form of a nucleophilic ThDP-bound carbanion/enamine intermediate, which is condensed with another molecule of formaldehyde to form glycolaldehyde (**Scheme 1A**). In both cases, the very low initial activity was improved by directed evolution. However, despite ~100-fold increases, the final enzyme variants still exhibited poor catalytic efficiencies (k_{cat}/K_M < 10 M^{-1} s^{-1}; **Table 1**), as well as poor affinity for the highly toxic formaldehyde (K_M ≥ 90 mM), which has so far precluded an application of these enzymes *in vivo*.

We recently identified another ThPD-dependent enzyme class as formaldehyde carboligases. Members of the 2-hydroxyacyl-CoA lyase (HACL)/oxalyl-CoA decarboxylase (OXC) enzyme superfamily catalyze the condensation of formyl-CoA with formaldehyde to produce glycolyl-CoA (**Scheme 1B**).[18, 19] Notably, wild-type (WT) HACL showed more than ten-fold higher

catalytic efficiency compared to the engineered FLS and GALS, mainly due to a lower K_M for formaldehyde (29 mM; **Table 1**). We aimed at further improving HACL activity to eventually enable the production of glycolyl-CoA at physiologically relevant formaldehyde concentrations (<0.5 mM) and high rates. Using structure-guided enzyme engineering, we generated a bona fide glycolyl-CoA synthase (GCS), and demonstrate its use in *Escherichia coli*, which may ultimately enable synthetic C1 fixation pathways *in vivo*.



Scheme 1. ThDP-dependent C1 fixation reactions. FLS and GALS generate the nucleophilic carbanion/enamine intermediate from formaldehyde, whereas HACL uses formyl-CoA. In the second half reaction formaldehyde is condensed with the intermediate, producing glycolaldehyde and glycolyl-CoA, respectively.

Table 1. Steady-state parameters of formaldehyde-fixing enzymes.

Enzyme	k_{cat}(s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
BAL[16]	n.d.	n.d.	0.05
FLS[16]	n.d.	n.d.	5
BFD[17]	0.01	90	0.1
GALS[17]	1.6	170	9
HsHACL1[18]	3.8	160	24
RuHACL[18]	3.3	29	110

3.2 Results

MeOXC structure

Previously, we showed that MeOXC is capable of catalyzing the acyloin condensation of formyl-CoA with various aldehydes, including formaldehyde.[19] However, the catalytic efficiency of MeOXC (k_{cat}/K_M = 2 $M^{-1}s^{-1}$) is far below that of RuHACL (k_{cat}/K_M = 110 $M^{-1}s^{-1}$) and the K_M for both substrates is extremely high (formaldehyde: 100 mM and formyl-CoA: 3 mM; **Table 2**). We therefore sought to improve MeOXC towards a catalytically efficient GCS.

MeOXC is very closely related to OXC from *Oxalobacter formigenes* (OfOXC, 63% identity) and *E. coli* (EcOXC, 61%), for both of which crystal structures have been solved. Nevertheless, to gain detailed insight into the active site, we solved the crystal structure of MeOXC at a resolution of 1.9 Å (**Figure 1B**). Overall, the structure is very similar to that of OfOXC (rmsd = 0.425 for 7230 alinged atoms). Because we could not observe electron density after the residue Glu567, we modelled the C-terminal part based on the OfOXC structure with bound intermediate (PDB ID 2ji7) (**Figure 1B**). In OfOXC this part is flexible and electron density for the closed conformation was obtained only after soaking with substrate or product.[20]

During catalysis OXC and HACL both form an α-hydroxyl-CoA-ThDP carbanion/enamine intermediate that performs a nucleophilic attack onto formaldehyde (**Scheme 1**).[21] In HACL the intermediate is formed through proton abstraction from formyl-CoA. In contrast, OXC forms the intermediate by decarboxylation of oxalyl-CoA, and shows only little HACL-like activity with formyl-CoA.[19, 20] Because the overall reactions of HACL and OXC are very different and because the flexible C-terminal part shares little sequence similarity between OXC and HACL (**Figure 1A**), predicting mutations that would enhance HACL-activity in OXC are difficult. Instead of rationally designing mutants, we turned to a directed evolution approach, employing iterative saturation mutagenesis (ISM)[22] to exhaustively screen mutations in every position surrounding the active site. To this end, we identified 11 residues within 6 Å of Cα of the α-hydroxyl-CoA-ThDP carbanion/enamine intermediate as targets for ISM: Gly47, Ile48, Tyr134, Glu135, Gly414, Ala415, Met442, Tyr497, Glu567, Ser568, and Ile571 (**Figure 1B**).

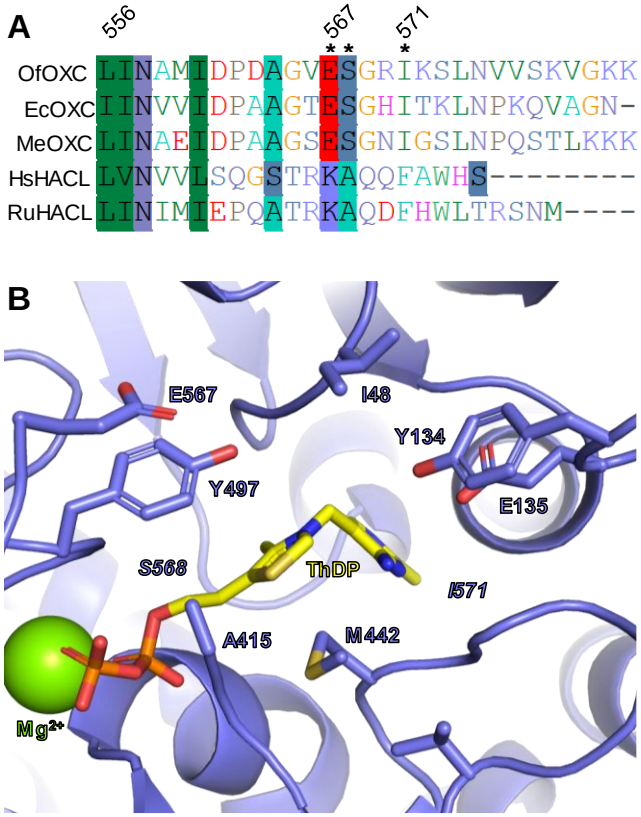


Figure 1. Sequence and structure comparison of OXC and HACL. A) MSA of the C-terminal amino acids of OXCs and HACLs. The high sequence disparity impedes homology modelling of this region, which is part of the active site in OXC. The residues marked with an asterisk are within 6 Å of Cα of the carbanion/enamine intermediate. **B)** Active site of MeOXC. 11 residues are within 6Å of Cα: Gly47, Ile48, Tyr134, Glu135, Gly414, Ala415, Met442, Tyr497, Glu567, Ser568, Ile571. The last three are part of the C-terminus that closes down over the active site upon substrate binding. Since we did not obtain electron density of this region, we modelled it based on the OfOXC structure.

High-throughput screen for GCS activity

ISM requires screening of thousands of variants. We conceived a high-throughput screen based on the hydrolysis of glycolyl-CoA to glycolate, which is subsequently oxidized by glycolate oxidase (GOX) to glyoxylate under stoichiometric production of H_2O_2 (**Figure 2A**). Horseradish peroxidase (HRP) consumes H_2O_2 to oxidize Ampliflu Red® to the fluorophore Resorufin. In summary, per molecule glycolyl-CoA formed by OXC from formyl-CoA and formaldehyde, one molecule Resorufin is produced.

Enzymes for the specific hydrolysis of glycolyl-CoA have not yet been described. We anticipated that it might be difficult to find a glycolyl-CoA thioesterase without activity towards formyl-CoA. Thus, we focused on identifying a glycolyl-CoA:formate transferase (GFT). This enzyme would not only cleave glycolyl-CoA but at the same time also regenerate formyl-CoA, creating a catalytic cycle that converts formate and formaldehyde into glycolate, the substrate for GOX. We screened FRC from *Oxalobacter formigenes*,[23] PCT from *Cupriavidus necactor*[24]

and *Clostridium propionicum*[25], as well as AbfT from *Clostridium aminobutyricum*[26], which showed the best GFT activity (k_{cat} = 0.40 ± 0.01 s^{-1}, app. K_M(glycolyl-CoA) = 12 ± 1 µM; **Figure 2B**).

We tested our screen by combining formyl-CoA, formaldehyde, formate and Ampliflu Red with purified MeOXC, AbfT, human GOX[27] and HRP (including all required cofactors). Fluorescence increase was only observed in the presence of all components (**Figure 2C**) and correlated with glyoxylate production, as determined by HPLC-MS. We made MeOXC the rate-limiting step in the assay and demonstrated that MeOXC activity could be quantified from *E. coli* lysates, which enabled us to screen MeOXC variants in high-throughput in cell-free extracts using 384-well plates.

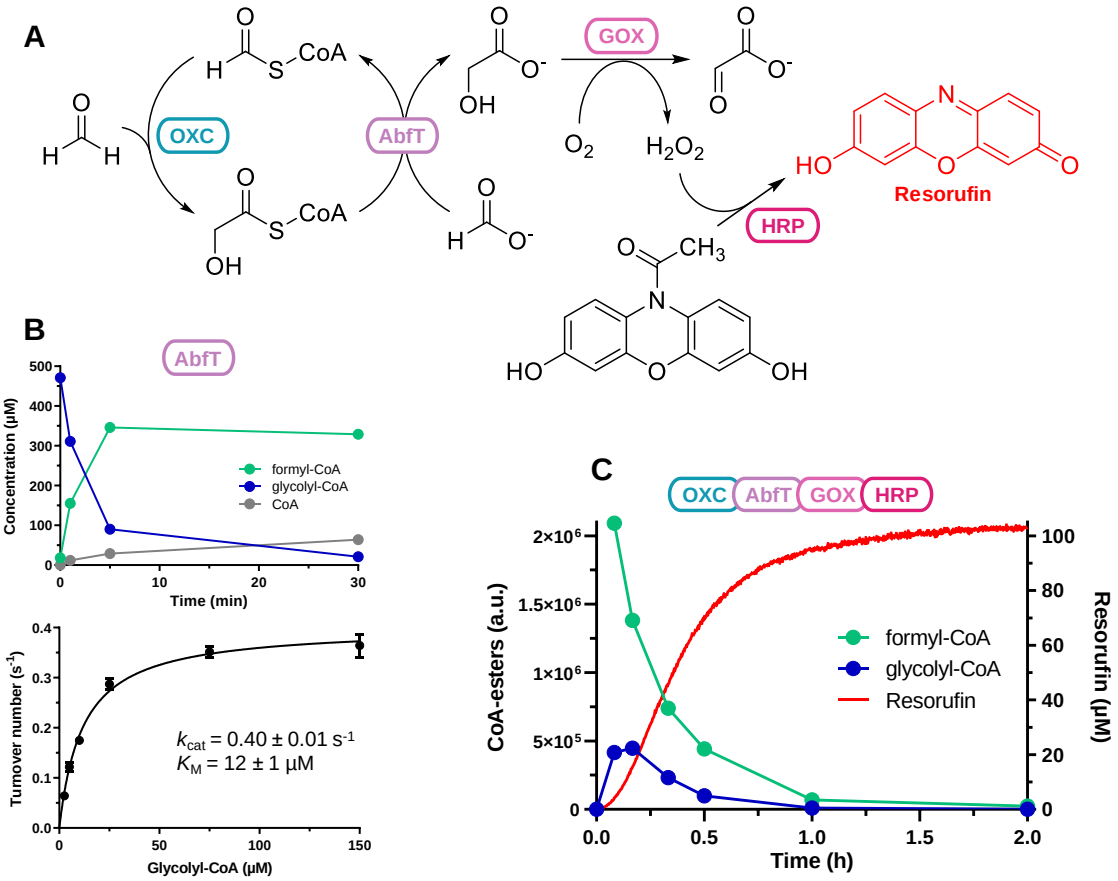


Figure 2. Establishing a high-throughput screen for GCS activity. A) Reaction scheme indicating enzymes for each step. Overall, glycolyl-CoA formation is detected via the red fluorophore resorufin. **B)** *Top*, The reaction progress of AbfT-catalyzed CoA transfer from glycolyl-CoA onto formate was monitored by LC-MS. *Bottom*, Michaelis-Menten plot of AbfT with glycolyl-CoA as donor and formate (25 mM) as acceptor. Error bars indicate the standard deviation of three replicates. **C)** Reaction progress of the OXC-AbfT-GOX-HRP cascade. Resorufin was detected via fluorescence in a plate-reader and CoA-esters were detected by LC-MS.

Iterative Saturation Mutagenesis of MeOXC

Having established a high-throughput screen, we created saturation mutagenesis libraries of the active site residues (see above) using the 22c trick to decrease the codon to amino acid ratio.[28] To reduce the number of variants to be screened in the first round, we excluded Gly47, Gly414 and Met442. For the screen, we also employed sub-saturating concentrations of formaldehyde and formyl-CoA (50 mM and 0.5 mM, respectively), to identify MeOXC variants with improved activity and/or higher affinity for both substrates.

In the first round of ISM (R1), libraries of position 48 and 571 contained only variants with reduced activity, suggesting that the isoleucine in both of these positions is critical to catalysis (**Figure 3**). Therefore, these residues were not screened in the following rounds R2 to R6. The libraries of the remaining six residues mostly contained variants with wild-type like activity but also some variants with significantly increased activity, indicating that these positions were generally good targets to improve enzyme activity (**Figure 3**). The best performing variant in R1 was an alanine to cysteine mutation in position 415. To confirm that the A415C variant was improved in catalysis, we purified this mutant and steady-state kinetics demonstrated that the mutation resulted in a decrease in the K_M for formaldehyde by 3-fold, and formyl-CoA by 19-fold (**Table 2**).

Based on the positive results of the first round, we continued with ISM by using the best performing variant of each round as template and saturating all other remaining sites step-wise. In R2, the mutation S568G conferred a ~3-fold improvement in formaldehyde affinity, prompting us to further decrease the formaldehyde concentration in R3 to 25 mM, where we identified E135G. For R4, we further lowered the concentration of formaldehyde to 10 mM to identify mutation Y497F. In R5, we screened the remaining residues 567 and 134, however, no more positive hits were found (**Figure 3**). We then screened a library in which position 134 and 135 were combinatorically saturated (R6; 400 variants), but also this library contained no positive variants. Thus, after screening a total of ~3,600 clones in 6 rounds, we obtained the final variant carrying the four mutations E135G, A415C, Y497F and S568G, which we termed MeOXC4.

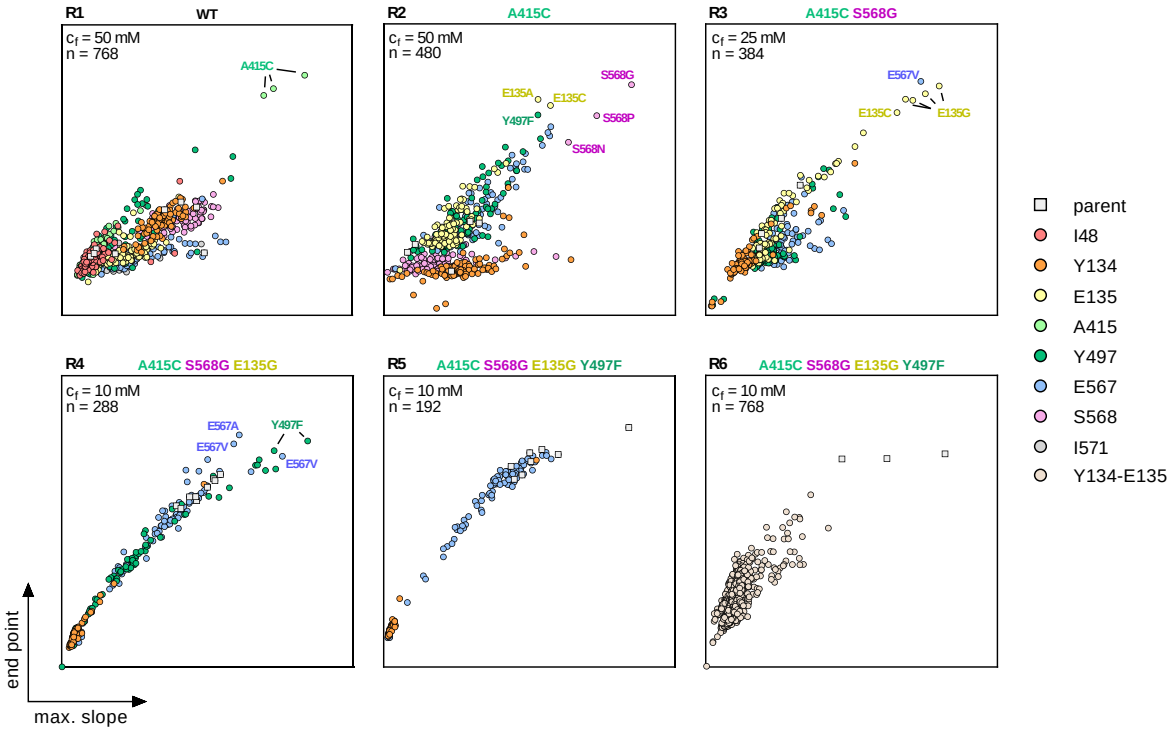


Figure 3. ISM of MeOXC for improved GCS activity. The saturation mutagenesis libraries were screened for GCS activity by the assay described in the main text. Plotted is the maximal slope versus end point of product formation. Each panel contains all libraries of the corresponding round. The best performing variants of each round were sequenced and their mutation is shown in the graphs with colored labels. As a reference, the parent (i.e. the template variant) was included in the screen. c_f denotes the formaldehyde concentration in the screen, n is the total number of screened clones per round.

Catalytic properties of MeOXC4

Next, we characterized MeOXC1-4 in more detail. All variants were produced at similarly high levels in *E. coli* (**Figure 4A**), indicating that their improvement was based on increased catalytic properties, and not on improved solubility and/or stability. In each round of our ISM, catalytic efficiency was increased, ultimately resulting in a ~200-fold improvement in k_{cat}/K_M for both substrates in MeOXC4 (**Figure 4B**). This was caused by a 20-fold lowered K_M for both substrates, in combination with a tenfold increased k_{cat} (**Table 2**).

We also tested whether MexOXC4 would promote the acyloin condensation of formyl-CoA with other acceptor substrates besides formaldehyde. MeOXC4 accepted a broad range of aldehyde substrates, including small hydrophilic and aliphatic aldehydes, bulky hydrophobic aldehydes and acetone, for which the WT had no detectable activity (**Figure 4C**).

While HACL-activity was strongly improved over the course of the ISM, the MeOXC variants gradually lost their native OXC-activity (**Figure 4B**). MeOXC4 retained less than 0.2% catalytic

efficiency for oxalyl-CoA decarboxylation (k_{cat} = 0.21 ± 0.01 s^{-1} and K_M = 122 ± 16 µM), indicating a tradeoff between OXC- and HACL-activity. This is exclusively due to a reduced k_{cat}, as the K_M remained virtually identical to wild-type MeOXC (105 µM).[19] Apparently, none of the mutations have an effect on oxalyl-CoA binding, which is in line with previous studies on OfOXC and MeOXC, where Tyr134, Glu135, Tyr497 and Ser568 were replaced by alanine without loss in affinity for oxalyl-CoA.[19, 20] Overall, the four mutations in MeOXC4 caused a speficity switch between HACL and native OXC activity of ~3.6 × 10^6.

Table 2. Steady-state parameters of MeOXC variants. Errors reflect the standard deviation of three independent measurements. Michaelis-Menten plots are shown in Figure S4.

	Formaldehyde			Formyl-CoA		
Enzyme	**k_{cat} (s^{-1})**	**app. K_M (mM)**	**k_{cat}/K_M (s^{-1} M^{-1})**	**k_{cat} (s^{-1})**	**app. K_M (mM)**	**k_{cat}/K_M (s^{-1} M^{-1})**
MeOXC	0.2 ± 0.01	100 ± 20	2	0.31 ± 0.08	3 ± 2	100
MeOXC1	0.34 ± 0.01	30 ± 2	10	0.23 ± 0.01	0.16 ± 0.05	1,400
MeOXC2	0.57 ± 0.05	12 ± 3	50	0.43 ± 0.02	0.09 ± 0.02	5,100
MeOXC3	0.22 ± 0.03	8 ± 5	30	2.0 ± 0.5	0.19 ± 0.08	11,000
MeOXC4	2.0 ± 0.2	5 ± 1	400	0.90 ± 0.07	0.05 ± 0.02	20,000

OXC is reversible – MeOXC4 not

To confirm that the switch was achieved by suppressing native OXC activity, we also sought to test the reverse reaction of OXC. We envisioned an enzymatic cascade in which the product of the reverse reaction, oxalyl-CoA, is removed from the equilibrium by reduction to glycolate (**Figure S5A**). This setup would render the overall reaction thermodynamically favorable under standard conditions ($\Delta G \approx$ -19 kJ mol^{-1})[29] and make the reaction, in contrast to earlier efforts,[30] reversible.

For the first reduction step we confirmed the NADPH-dependent oxalyl-CoA reductase activity of *M. extorquens* gene product PanE2[31] via LC-MS and determined its kinetic parameters photospectrometrically (**Figure S5B**). GhrB from *E. coli* catalyzes the NADPH-dependent reduction of glyoxylate to glycolate.[32] When tested in combination with PanE2 and GhrB, MeOXC WT was able to catalyze the carboxylation of formyl-CoA at a rate of ~0.5 min^{-1} (**Figure S5C-E**), while activity was completely abolished in MeOXC4 (**Figure S5D**), despite its higher formyl-CoA affinity, supporting that MeOXC4 lost its native OXC-activity.

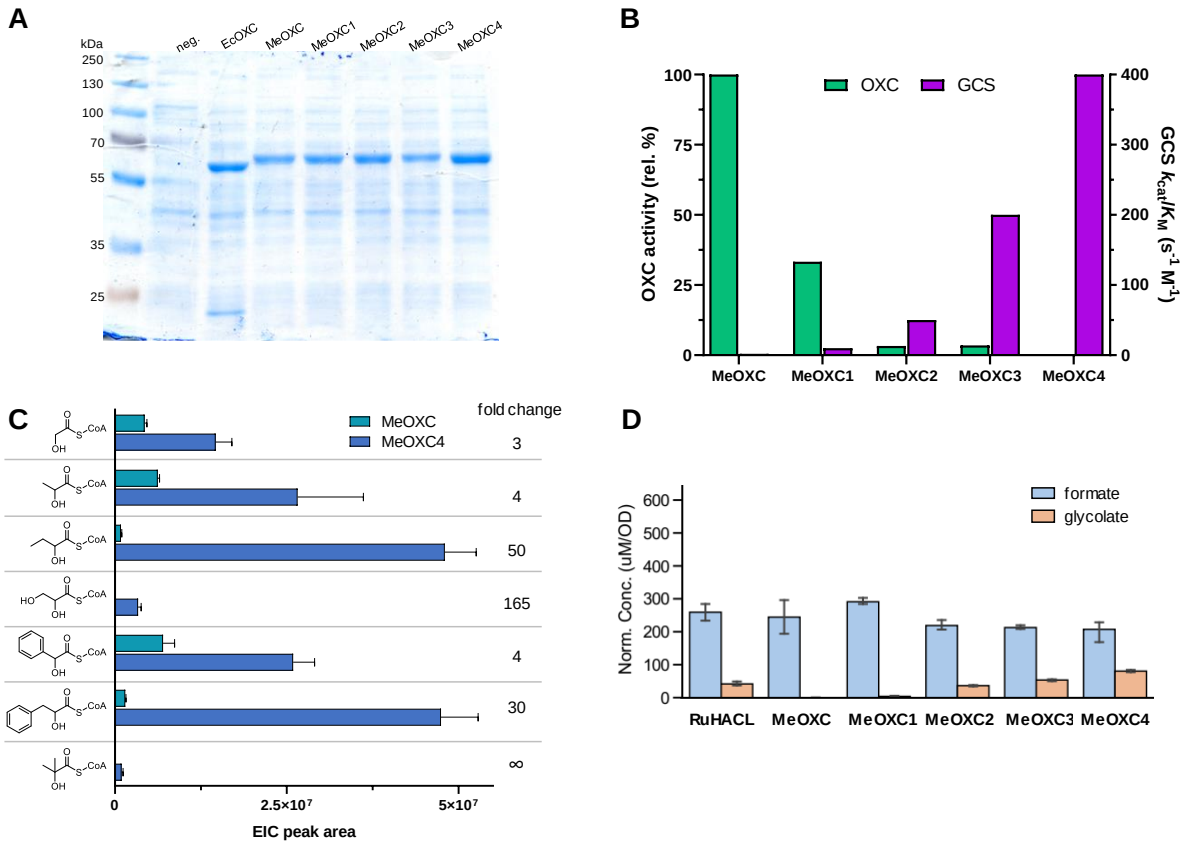


Figure 4. Characterization of MeOXC mutants from the ISM. A) SDS-PAGE analysis of the MeOXC mutants from each round of ISM, expressed in *E. coli* BL21. A total of 5 µg protein was loaded in each lane and empty pET-16b was used as negative control. **B)** Comparison of the OXC and GCS activities. OXC activity is the velocity of oxalyl-CoA decarboxylation at 1 mM oxalyl-CoA, relative to the wild-type. For GCS the absolute catalytic efficiency for formaldehyde is shown. **C)** Comparison of the aldehyde substrate scope of MeOXC and MeOXC4. Products were analyzed by LC-MS. Error bars show the standard deviation of three replicates. The numbers on the right indicate the fold change in activity of MeOXC4 compared to MeOXC. **D)** Formate and glycolate production in the *E. coli* whole-cell biotransformation system. Error bars show the standard deviation of three replicates.

Structure of MeOXC4

To better understand the effect of the mutations, we solved the crystal structure of MeOXC4 at a resolution of 2.4 Å. The overall structure is virtually identical to MeOXC with a rmsd = 0.190 Å for 8167 aligned atoms. As for MeOXC, we were not able to obtain interpretable electron density for the C-terminus, we therefore modelled the C-terminus based on OfOXC.

The newly introduced sulfur moiety of Cys415 is in close contact with residue 568. This mutation caused a decrease in the K_M for formyl-CoA by more than an order of magnitude (**Table 2**). No other amino acid substitution at this position had a beneficial effect on the activity (**Figure 3**), suggesting that the sulfide moiety of the cysteine enhances the affinity for formyl-CoA. Ala415 is fully conserved in the OXC/HACL superfamily. Interestingly, introducing this

mutation into HACL from *Syncephalastrum racemosum* completely abolished GCS activity (data not shown). The C-terminal sequence shares no similarity between OXC and HACL (**Figure 1A**) and therefore, mutations around or in this region could have vastly different effects.

Right next to A415C, the second mutation S568G opened up space to accommodate the larger side chain of Cys415. Notably, S568G was only beneficial in combination with A415C, as it was not detected in the first round of ISM (**Figure 3**). In OfOXC Tyr497 and Ser568 form hydrogen bonds with a water molecule (W3). W3 is in close contact (2.5 Å) with another water molecule (W2) that was proposed to protonate Cα of the carbanion/enamine intermediate.[20] In MeOXC4, the two mutations S568G and Y497F abolished hydrogen bonding to W3, and thereby presumably changed the environment of W2.

The third mutation E135G caused the most drastic structural change. Removal of the entire side chain of residue 135 created a cavity at the active site. As a consequence, the neighboring Tyr134 moved into this cavity (by ~0.8 Å at the hydroxyl group). In OfOXC, Tyr134 and Glu135 form hydrogen bonds to a water molecule (W1) that is also close to W2 (3.5 Å). Thus, E135G further changed the environment of W2.

Interestingly, none of the libraries at residue 134 contained a variant with improved activity, in fact, virtually all variants of these libraries performed worse than the corresponding parent (**Figure 3**). This suggests that Tyr135 is critical for GCS catalysis and the mutation E135G allowed for precise positioning of Tyr135. In HACL, C-C bond cleavage is initiated by proton abstraction from the 2-hydroxyl group of the carbanion/enamine intermediate. Conversely, the reverse, i.e. C-C bond forming reaction requires a proton donor. Given its positioning, Tyr135 is a potential proton donor, which would explain why any mutation lead to reduced or abolished activity.

Taken together the structural information indicate that the OXC and HACL activities are fine-tuned by modulating the structure of the three water molecules, particularly W2. A structure with trapped intermediate(s) could give detailed insight into the mechanism, particularly the nature of the proton donor.

In vivo application of MeOXC4

We previously established a HACL-based whole-cell bioconversion system for the synthesis of glycolate from formaldehyde.[18] Despite extensive optimization, product yield was limited by the low expression and activity of HACL. MeOXC4 on the other hand is produced in high titers in *E. coli* (**Figure 4A**) and has kinetic parameters that are comparable to the best performing

HACL, RuHACL (k_{cat}/K_M = 400 M^{-1} s^{-1} and 110 M^{-1} s^{-1}, respectively). Indeed, the use of MeOXC4 increased glycolate production by a factor of ~2 compared to RuHACL (**Figure 4D**), confirming that MeOXC4 can act as a GCS in a physiological context. Taken together, our results demonstrate that we successfully turned MeOXC into a highly soluble GCS that may pave the way towards novel, synthetic one-carbon assimilation metabolism.

3.3 Discussion

Through ISM we successfully evolved OXC into a *bona fide* GCS. None of the newly introduced amino acids of MeOXC4 is found in the entire OXC/HACL family, which indicates that a fitness maximum for GCS activity was found in the evolutionary trajectory of the ISM that is vastly different from native HACLs. This raises the question how HACL and MeOXC4 are able to catalyze the same reaction with two significantly different active sites.

It has been proposed that in OXC the enamine/α-carbanion intermediate is nonplanar, rendering the Cα more basic and facilitating the rate-limiting protonation step.[20] In contrast, in other ThDP–dependent enzymes that require a second acceptor substrate, the enamine-like planar structure was observed.[33-35] It is intriguing to speculate that MeOXC4's increased carboligation activity is caused by a higher enamine character of the intermediate, which suppresses protonation and thereby favors nucleophilic attack on the aldehyde. This is further supported by the fact that we observed a tradeoff between OXC- and HACL-activity during ISM.

Only two naturally occurring enzymes are known to catalyze formaldehyde fixation, 3-hexulose-6-phosphate synthase (HPS) in the ribulose monophosphate pathway and formaldehyde transketolase (FTK) in the dihydroxyacetone pathway. HPS has formaldehyde K_M in the range 0.15 to 3 mM[36, 37] and for FTK two values have been reported, 0.4 mM and 1.9 mM.[38, 39] Thus, with a K_M of 5 mM, MeOXC4 is comparable to natural formaldehyde-fixing enzymes and is an excellent basis for synthetic carbon fixation pathways.

We demonstrated that the active site of MeOXC is malleable and can be fine-tuned for activity towards carbonyl electrophiles. By exchanging formaldehyde with other aldehydes, our screen can be adapted to detect production of 2-hydroxy acids such as lactate, glycerate, 2-hydroxyglutarate, mandelate and 3-phenyllactate. Specific oxidases as well as promiscuous CoA transferases have been described for most of these acids.[40-44]

Establishing OXC/HACL as C-C bond formation platform may pave the way for biocatalytic production of valuable 2-hydroxy acids and novel artificial C1 assimilation pathways. Our results here showcase the great potential of enzyme engineering for synthetic biology.

Acknowledgements

We thank N. S. Cortina, P. Claus and N. Paczia for LC-MS measurements. We thank G. Santoni and D. von Stetten for help with x-ray data collection. This work was supported by the German Ministry of Education and Research Grant FormatPlant (part of BioEconomy 2030, Plant Breeding Research for the Bioeconomy).

3.4 Materials & Methods

Chemicals

Chemicals were obtained from Sigma-Aldrich (Munich, Germany) and Carl Roth GmbH (Karlsruhe, Germany). Coenzyme A was obtained from Roche Diagnostics Deutschland GmBH (Mannheim, Germany). Biochemicals, commercially available proteins and materials for cloning and protein production were obtained from Thermo Fisher Scientific (St. Leon-Rot, Germany), New England Biolabs GmbH (Frankfurt am Main, Germany) and Macherey-Nagel GmbH (Düren, Germany). Primers and synthesized genes were obtained from Eurofins MWG GmbH (Ebersberg, Germany).

Formyl-CoA and oxalyl-CoA were synthesized as described previously.[19] Glycolyl-CoA was synthesized via carbonyldiimidazole coupling, as described previously.[45] Glycolyl-CoA was purified by preparative HPLC-MS with a methanol gradient in 25 mM ammonium formate pH 4.2. The fractions containing the product were lyophilized and stored at -20 C.

For use in assays and standards, CoA esters were dissolved in 100 mM potassium phosphate, pH 5.5. The concentration was determined by enzymatic depletion with PduP[46], following NADH consumption at 340 nm.

Cloning and Mutagenesis

The overexpression plasmids for AbfT, FRC and PCTs were obtained from the internal lab collection. Human GOX was codon optimized and synthesized by Twist Bioscience. *panE2* was PCR-amplified from *M. extorquens* chromosomal DNA using the corresponding primers (Table S1). The purified PCR products were digested with NdeI and BamHI and ligated into pET-16b. Correct cloning was confirmed by sequencing (Eurofins Genomics, Ebersbach, Germany). The overexpression plasmid for GhrB was obtained from the ASKA collection.[47] Point mutations were introduced by PCR using mismatch primers (**Table S1**). A 50 µL reaction contained 60 ng of template DNA, 0.25 µM forward and reverse primer, 200 µM dNTP, 5 µL 10x Reaction Buffer, 1 µL Phusion polymerase (2 U/µL). Template plasmid was removed by DpnI digest (10 U) at 37 °C immediately after PCR amplification. Mutations were confirmed by sequencing.

Table S1. Primers used in this study. N = A, C, G, T. D = A, G, T. H = A, C, T. V = A, C, G.

Name	Nucleotide sequence
oxc_I48_NDT	CGATCTACAACGTGCCCGGC**NDT**CCGATCACCGATCTCGG
oxc_I48_VHG	CGATCTACAACGTGCCCGGC**VHG**CCGATCACCGATCTCGG
oxc_I48_TGG	CGATCTACAACGTGCCCGGC**TGG**CCGATCACCGATCTCGG
oxc_I48_rv	GCCGGGCACGTTGTAGATCGTCTCGATGCCGTTGAGCTTG
oxc_A415_NDT	CGATCCTCGTCAACGAGGGT**NDT**AACACCCTCGATCTGGC
oxc_A415_VHG	CGATCCTCGTCAACGAGGGT**VHG**AACACCCTCGATCTGGC
oxc_A415_TGG	CGATCCTCGTCAACGAGGGT**TGG**AACACCCTCGATCTGGC
oxc_A415_rv	CCCTCGTTGACGAGGATCGCGTCGGGCCGCTCCTTGATG
oxc_Y497_NDT	CTTCAACAACAACGGCATC**NDT**CGCGGCACCGACACCGATC
oxc_Y497_VHG	CTTCAACAACAACGGCATC**VHG**CGCGGCACCGACACCGATC
oxc_Y497_TGG	CTTCAACAACAACGGCATC**TGG**CGCGGCACCGACACCGATC
oxc_Y497_rv	GATGCCGTTGTTGTTGAAGATGACGATGCAGACCGGCAGC
oxc_I571_NDT	CCGGCAGCGAGAGCGGCAAT**NDT**GGCAGCCTCAACCCGCAG
oxc_I571_VHG	CCGGCAGCGAGAGCGGCAAT**VHG**GGCAGCCTCAACCCGCAG
oxc_I571_TGG	CCGGCAGCGAGAGCGGCAAT**TGG**GGCAGCCTCAACCCGCAG
oxc_I571_rv	GCCGCTCTCGCTGCCGGCGGCCGGGTCGATCTCCGCATTG
oxc_Y134_NDT	CGATCTGCAGCAGGGCGAC**NDT**GAGGAGATGGACCAGC
oxc_Y134_VHG	CGATCTGCAGCAGGGCGAC**VHG**GAGGAGATGGACCAGC
oxc_Y134_TGG	CGATCTGCAGCAGGGCGAC**TGG**GAGGAGATGGACCAGC
oxc_Y134_rv	GTCGCCCTGCTGCAGATCGACGATCTCGCGCTCGGAGGAG
oxc_E135_NDT	GATCTGCAGCAGGGCGACTAC**NDT**GAGATGGACCAGCTCGC
oxc_E135_VHG	GATCTGCAGCAGGGCGACTAC**VHG**GAGATGGACCAGCTCGC
oxc_E135_TGG	GATCTGCAGCAGGGCGACTAC**TGG**GAGATGGACCAGCTCGC
oxc_E135_rv	GTAGTCGCCCTGCTGCAGATCGACGATCTCGCGCTCGGAG
oxc_E567_NDT	CGACCCGGCCGCCGGCAGC**NDT**AGCGGCAATATCGGCAGC
oxc_E567_VHG	CGACCCGGCCGCCGGCAGC**VHG**AGCGGCAATATCGGCAGC
oxc_E567_TGG	CGACCCGGCCGCCGGCAGC**TGG**AGCGGCAATATCGGCAGC
oxc_E567_rv	GCTGCCGGCGGCCGGGTCGATCTCCGCATTGATCAGAGTC
oxc_S568_NDT	CCCGGCCGCCGGCAGCGAG**NDT**GGCAATATCGGCAGCCTC
oxc_S568_VHG	CCCGGCCGCCGGCAGCGAG**VHG**GGCAATATCGGCAGCCTC
oxc_S568_TGG	CCCGGCCGCCGGCAGCGAG**TGG**GGCAATATCGGCAGCCTC
oxc_S568_rv	CTCGCTGCCGGCGGCCGGGTCGATCTCCGCATTGATCAG
oxc_S568G_E567_NDT	CGACCCGGCCGCCGGCAGC**NDT**GGTGGCAATATCGGCAGC
oxc_S568G_E567_VHG	CGACCCGGCCGCCGGCAGC**VHG**GGTGGCAATATCGGCAGC
oxc_S568G_E567_TGG	CGACCCGGCCGCCGGCAGC**TGG**GGTGGCAATATCGGCAGC
oxc_E135G_Y134_NDT	CGATCTGCAGCAGGGCGAC**NDT**GGTGAGATGGACCAGC
oxc_E135G_Y134_VHG	CGATCTGCAGCAGGGCGAC**VHG**GGTGAGATGGACCAGC
oxc_E135G_Y134_TGG	CGATCTGCAGCAGGGCGAC**TGG**GGTGAGATGGACCAGC
oxc_Y134E135_NDTNDT	GATCTGCAGCAGGGCGAC**NDTNDT**GAGATGGACCAGCTCGC
oxc_Y134E135_NDTVHG	GATCTGCAGCAGGGCGAC**NDTVHG**GAGATGGACCAGCTCGC
oxc_Y134E135_NDTTGG	GATCTGCAGCAGGGCGAC**NDTTGG**GAGATGGACCAGCTCGC
oxc_Y134E135_VHGNDT	GATCTGCAGCAGGGCGAC**VHGNDT**GAGATGGACCAGCTCGC
oxc_Y134E135_VHGVHG	GATCTGCAGCAGGGCGAC**VHGVHG**GAGATGGACCAGCTCGC
oxc_Y134E135_VHGTGG	GATCTGCAGCAGGGCGAC**VHGTGG**GAGATGGACCAGCTCGC
oxc_Y134E135_TGGNDT	GATCTGCAGCAGGGCGAC**TGGNDT**GAGATGGACCAGCTCGC
oxc_Y134E135_TGGVHG	GATCTGCAGCAGGGCGAC**TGGVHG**GAGATGGACCAGCTCGC
oxc_E70A_fw	CATCTCCTTCCGCCACGCGCAGAATGCGGGCAAC
oxc_E70A_rv	GTTGCCCGCATTCTGCGCGTGGCGGAAGGAGATG
panE2_fw	GCGCACATATGAGCATCGCGATCGTCG
panE2_rv	CAGAGGATCCTCATGCTCCCTGGATCGC

Protein Production and Purification

All proteins were heterologously produced in *E. coli* BL21 (DE3). 500 mL TB containing 100 µg/mL ampicillin (OXC, GOX, FRC, AbfT, PCTs, PduP, PanE2) or 34 µg/mL chloramphenicol (GhrB) was inoculated with freshly-transformed cells and incubated at 37 °C. After reaching an OD_{600} of 0.8 expression was induced by adding IPTG to a final concentration of 0.5 mM and the incubation temperature was lowered to 25 °C. After reaching an OD_{600} of 0.8 the culture was cooled on ice for 15 min. Then expression was induced by adding IPTG to a final concentration of 0.1 mM and the incubation temperature was lowered to 15 °C. Cells were harvested after 16 h by centrifugation (4500× g, 10 min) and resuspended in buffer A (500 mM KCl, 50 mM HEPES-KOH pH 7.8). If not used immediately, cell pellets were stored at −20 °C. The cell lysate obtained by sonication was clarified by centrifugation 75,000× g at 4 °C for 45 min. The supernatant was filtered through a 0.4 µm syringe tip filter (Sarstedt, Nümbrecht, Germany). Ni-affinity purification was performed with an Äkta FPLC system from GE Healthcare (GE Healthcare, Freiburg, Germany). The filtered soluble lysate was loaded onto a 1 mL Ni-Sepharose Fast Flow column (HisTrap FF, GE Healthcare, Little Chalfont, UK) that had beenx equilibrated with 10 mL buffer A. After washing with 20 mL 85% buffer A, 15% buffer B (500 mM KCl, 50 mM HEPES-KOH pH 7.8, 500 mM imidazole), the protein was eluted with 100% buffer B. Fractions containing purified protein were pooled and the buffer was exchanged to storage buffer (150 mM KCl, 50 mM HEPES-KOH pH 7.8) with a desalting column (HiTrap, GE Healthcare). PduP was applied to a 1 ml StrepTrap column (GE Healthcare) that had been equilibrated with storage buffer. The column was washed with storage buffer and the recombinant enzyme was eluted with storage buffer containing 2.5 mM desthiobiotin. Proteins were concentrated by ultrafiltration (Amicon Ultra). Concentration was determined on a NanoDrop 2000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA) using the extinction coefficient at 280 nm, as calculated by protparam (https://web.expasy.org/protparam/). Enzyme purity was confirmed by SDS-PAGE. The purified proteins were stored in 50% glycerol at −20 °C. OXC wild-type and mutants were flash-frozen in liquid nitrogen and stored at −80 °C.

Crystallization & Structure Determination

MeOXC and MeOXC4 were purified as described above. Immediately after affinity purification, the eluate was loaded onto a Sepharose Gel Filtration column equilibrated in SEC Buffer (75 mM KCl, 25 mM HEPES-KOH pH 7.8). Fractions corresponding to tetrameric enzyme were collected, pooled and concentrated to 10 mg/mL on 30,000 MWCO filters (Amicon Ultra). Enzyme purity was evaluated via SDS page. The enzyme was supplemented with 10 mM $MgCl_2$, 2 mM TPP and 1 mM CoA and crystal plates were seeded using the sitting drop method, diluting equal volume of enzyme in reservoir solution. The reservoir solution contained for MeOXC 45 % w/v Pentaerythritol ethoxylate (3/4 EO/OH), 100 mM Sodium acetate pH 4.6, 400 mM KCl); and for MeOXC4 25 % w/v Pentaerythritol propoxylate (17/8 PO/OH), 100 mM TRIS pH 8.5, 50 mM $MgCl_2$). Substrate soaking with mandelyl-CoA or glycolyl-CoA impacted the final resolution of the crystals without revealing electron density for these ligands.

Data for MeOXC was collected at the Beamline ID23 -1 (European Synchrotron Radiation Facility, Grenoble, France), whereas Data for MeOXC4 was collected at the Beamline P13 (Deutsches Elektronen-Synchrotron, Hamburg, Germany). All images were processed using XDS.[48] The data set was scaled using the program suite ccp4.[49] The Phenix software package[50] was used to perform molecular replacement (PhaserMR) for phasing of the MeOXC

dataset by using the oxalyl-CoA decarboxylase from *Oxalobacter formigenes* (PDB 2C31) as search model. The refined structure of MeOXC was then used as search model for MeOXC4. Initial models were built with Phenix.AutoBuild and refined with the phenix.refine. Manual refinement and ligand modelling was done in COOT.[51] Final B-facor refinement and water positioning was also performed via phenix.refine.

Table S2. Data collection and refinement statistics.

	MeOXC	**MeOXC4**
Beam line	ESRF_ID23eh1, Gernoble, France	DESY P13, Hamburg, Germany
PDB ID	*tbd*	*tbd*
Ligands	TPP, ADP, Mg^{2+}	TPP, ADP, Mg^{2+}
Wavelength (Å)	0.972	0.976
Resolution range (Å)	39.11 - 1.9 (1.968 - 1.9)	29.97 - 2.401 (2.487 - 2.401)
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell dimensions a, b, c (Å)	160.176, 181.798, 202.366	161.112, 180.355, 202.032
α, β, γ (°)	90, 90, 90	90, 90, 90
Total reflections	2591098 (355003)	3098058 (454913)
Unique reflections	455609 (65577)	228390 (32969)
Multiplicity	5.7 (5.4)	13.6 (13.8)
Completeness (%)	98.64 (97.25)	99.41 (99.26)
Mean (I)/σ (I)	8.82 (1.99)	7.41 (2.09)
Wilson B-factor ($Å^2$)	23.99	25.58
R_{merge}	0.04624 (0.367)	0.06957 (0.3255)
R_{meas}	0.06539 (0.519)	0.09839 (0.4604)
CC1/2	0.997 (0.883)	0.997 (0.92)
Reflections used in refinement	454237 (44467)	227340 (22473)
Reflections used for R_{free}	1987 (195)	1991 (196)
R_{work}	0.1611 (0.3061)	0.2093 (0.2721)
R_{free}	0.1940 (0.3227)	0.2570 (0.3576)
Number of non-hydrogen atoms	36766	34358
macromolecules	32572	32361
ligands	432	432
solvent	3762	1565
Protein residues	4397	4385
RMS(bonds) (Å)	0.011	0.012
RMS(angles) (°)	1.07	1.45
Ramachandran favored (%)	97.59	96.92
Ramachandran allowed (%)	2.23	2.74
Ramachandran outliers (%)	0.18	0.35
Rotamer outliers (%)	0.79	2.72
Average B-factor	29.09	34.79
macromolecules	28.30	34.83
ligands	23.95	31.12
solvent	36.53	35.05

Statistics for the highest-resolution shell are shown in parentheses. tbd, to be determined.

LC-MS analysis of CoA-esters

Samples were prepared for LC-MS analysis by quenching an aliquot of a reaction with formic acid (final concentration 4%) and centrifuging for 10 min at 17,000 rcf, to remove precipitated proteins. Samples were diluted 1:10 in H_2O before analysis.

Quantitative determination of glycolyl-CoA was done by LC-MS/MS. The chromatographic separation was performed on an Agilent Infinity II 1290 HPLC system using a EVO C18 column (50 × 2.1 mm, 1.7 µm particle size, 100 Å pore size) connected to a guard column of the same specification (20 × 2.1 mm, 5 µm particle size) (Phenomoenx, Torrance, CA, USA) a constant

flow rate of 0.25 ml/min with mobile phase A being 50 mM Ammonium Formate (Sigma-Alrich, St. Lois, MO, USA) and phase B being Methanol 99.9% LC-MS (VWR, Darmstadt, Germany) at 25° C . The injection volume was 2 µl. The mobile phase profile consisted of the following steps and linear gradients: 0 – 4.5 min constant at 3 % B; 4.5 – 5.5 min from 3 to 80 % B; 5.5 – 6.5 min constant at 80 % B; 6.5 – 7.5 min from 80 to 3 % B; 7.5 – 8.5 min constant at 3 % B. An Agilent 6495 ion funnel mass spectrometer was used in positive mode with an electrospray ionization source and the following conditions: ESI spray voltage 1000 V, sheath gas 400° C at 12 l/min, nebulizer pressure 20 psig and drying gas 100° C at 11 l/min. Compounds were identified based on their mass transition and retention time compared to standards. Chromatograms were integrated using MassHunter software (Agilent, Santa Clara, CA, USA). Absolute concentrations were calculated based on an external calibration curve prepared in sample matrix after confirming that the analyte cannot be detected in the matrix prior to standard addition.

Qualitative analysis of CoA esters (**Figure 4C**) was performed using UPLC-high resolution MS as described previously.[14]

GFT screen

For the initial screen, 50 mM TES-KOH pH 6.8, 100 mM sodium formate and 1 µM transferase (FRC, AbfT, CnPCT, CpPCT) were mixed in a 1.5 mL microfuge tube at 30 °C. The reaction was started by addition of 1 mM glycolyl-CoA. Samples were taken after 5 and 60 min and analyzed with the CoA ester method described above. For the second screen, 50 mM TES-KOH pH 6.8, 50 mM sodium formate and 2.5 µM transferase (AbfT, CnPCT) were mixed in a 1.5 mL microfuge tube at 30 °C. The reaction was started by addition of 0.5 mM glycolyl-CoA. Samples were taken after 0, 1, 5, 30 and 120 min and analyzed with the CoA ester method described above.

GCS kinetics

GCS kinetics for formaldehyde were determined in reactions containing 100 mM potassium phosphate pH 6.9, 10 mM $MgCl_2$, 500 µM ADP, 150 µM TPP and 1 mM formyl-CoA. Enzyme concentration was varied for each mutant and initial velocity was determined for five formaldehyde concentrations, as shown in **Table S3**. Similarly, kinetics for formyl-CoA were determined using reactions containing 100 mM potassium phosphate pH 6.9, 10 mM $MgCl_2$, 500 µM ADP and 150 µM TPP. Formaldehyde and enzyme concentration were varied for each mutant to achieve saturating conditions and initial velocity was determined for six formyl-CoA concentrations as shown in **Table S3**. Reactions were incubated at 30 °C and samples were taken at 1, 2 and 5 min. Samples analysed by LC-MS as described above. Data was analysed in GraphPad Prism using a Michalis Menten fit. A gel of the purified MeOXC variants is shown in **Figure S3**.

Table S3. Concentrations of substrates and enzymes in determination of GCS steady-state parameters of MeOXC variants. Enz, enzyme; FALD, formaldehyde; F-CoA, formyl-CoA

	Formaldehyde kinetics			Formyl-CoA kinetics		
	Enz (µM)	**FALD (mM)**	**F-CoA (mM)**	**Enz. (µM)**	**FALD (mM)**	**F-CoA (mM)**
MeOXC	5	500, 200, 100, 50, 20	1	5	500	12.5, 7.5, 3.75, 2.5, 1.25, 0.5
MeOXC1	5	100, 50, 20, 10, 4	1	2	150	7.5, 4.5, 2.25, 1.5, 0.75, 0.3
MeOXC2	2.5	50, 20, 10, 5, 2	1	1	80	1, 0.6, 0.3, 0.2, 0.1, 0.05
MeOXC3	0.5	200, 100, 40, 20, 8	1	1	35	0.25, 0.15, 0.075, 0.05, 0.025, 0.01
MeOXC4	1	25, 10, 5, 2, 1	1	0.2	30	1.5, 1, 0.45, 0.3, 0.15, 0.05

Aldehyde scope of MeOXC vs MeOXC4

The aldehyde scope of MeOXC and MeOXC4 was evaluated via LC-MS. Reactions contained 100 mM potassium phosphate pH 6.9, 5 mM $MgCl_2$, 150 µM TPP, 1 mM formyl-CoA, 10 µM of MeOXC or MeOXC4 and either 100 mM formaldehyde, acetaldehyde, propionaldehyde or glycolaldehyde, or 20 mM benzaldehyde or phenylacetaldehyde, or 1 M acetone. Reactions were incubated at 30 °C for 10 min and then quenched for LC-MS analysis.

ISM of MeOXC

ISM was performed on pET16b_mexOXC using the 22c trick[28] and the primers listed in **Table S1**. Briefly, NDT, VHG and TGG forward primers were mixed at a ratio of 12:9:1 to create a forward primer mix. 50 µl PCR reactions contained 1x Q5 Buffer (NEB), 1x High GC enhancer (NEB), 0.4 mM dNTPs (Thermo Fisher), 0.8 µM forward primer mix, 0.8 µM reverse primer, 50 ng template DNA and 0.5 µl Q5 HF polymerase (NEB). Per target site, 4 50 µl PCR reactions were run in parallel. Two-step PCR Cycles were performed as follows: 30 s 98 °C, 32x [10 s 98 °C and 72 °C 6 min], final extension 72 °C 15 min, 4 °C hold. PCR product was purified using the Machery Nagel NucleoSpin Gel and PCR Clean-up kit and digested overnight at 37 °C using DpnI FD (Thermo Fisher) and the supplied buffer. Digested DNA was purified again and transformed into chemically competent DH5α cells (Invitrogen) by standard methods. After selection on LB plates supplied with ampicillin, colonies were washed from the plates, purified using the NucleoSpin Plasmid Miniprep kit (Machery Nagel) and send for sequencing (Microsynth). Library quality was evaluated from the Sanger sequencing traces (Figure S2). For the activity screen, libraries were transformed into competent *E.coli* BL21(DE3). Transformants were picked into 600 µl LB Amp_{100} in 2.0 ml 96 Deep Well Plates with V Bottom (Plate One). As a control, each plate was also inoculated with the corresponding parent variant in the *E.coli* BL21(DE3) background. The 96-well plates were sealed with Rotilabo Cell Culture Sealing Film (Carl Roth) and grown overnight at 37 °C, 180 rpm. The next day, fresh plates were inoculated using 30 µl of starter culture and again 600 µl LB Amp_{100}. The master plates were supplemented with 200 µl 80% glycerol per well and stored at -80 °C. The fresh plates were grown at 37 °C, 180 rpm to OD_{600} = 0.4-0.6 and then induced by addition of 250 µM IPTG. Overexpression occurred for 16 h at 25 °C, 180 rpm. Cells were harvested by centrifugation in a Multifuge X1R (Heraeus) at 2000 g, 4 °C for 30 min. LB was removed and cell pellets were lysed by addition of 60 µl CellLytic B Cell Lysis Reagent (Sigma Aldrich) and incubation for 10

min at 20 °C, 120 rpm. Cell debris was spun down by centrifugation at 2000 g, 4 °C for 5 min. Lysate screens were performed in an Infinite M PLEX plate reader (Tecan) using 100 mM K_2HPO_4, pH 7.5, 5 mM $MgCl_2$, 150 µM ThDP, 50 mM formate, 50 – 10 mM formaldehyde, 0.5 mM formyl-CoA, 0.5 mM Ampliflu Red (Sigma Aldrich), 1 U/mL horse-radish peroxidase (Sigma Aldrich), 2.5 µM purified GOX, 1 µM purified AbfT and 20% (v/v) OXC lysate. Product formation was followed for 2 h. The maximal slope was determined via the first derivative of the production formation. To identify positive clones, end point concentration was plotted versus maximal slope. Positive clones were sequenced using the frozen master plates for re-inoculation. Improved activity of the purified enzyme mutant was confirmed using the same reaction conditions as for the library screen.

Spectrophotometric enzyme assays

Spectrophotometric assays were performed on a Cary-60 UV/Vis spectrophotometer (Agilent) at 30°C using quartz cuvettes (10 mm path length; Hellma). For the determination of steady-state kinetic parameters, each concentration was measured in triplicates and the obtained curves were fit using GraphPad Prism 7. The data was fit to the Michaelis-Menten equation to obtain k_{cat} and K_M values.

AbfT. For the glycolyl-CoA:formate transferase activity of AbfT an assay containing 50 mM MES-KOH pH 6.8, 5 mM disodium oxalate, 50 mM sodium formate, 0.3 mM NADPH, 1.5 µM FRC, 0.2 µM PanE2 and 2 µM AbfT was preincubated for 2 min and the reaction started by adding glycolyl-CoA (2.5, 5, 10, 25, 100 and 150 µM). Reaction procedure was monitored by following the oxidation of NADPH at 340 nm.

OXC forward. The oxalyl-CoA decarboxylase activity was measured as described previously.[19] 50 mM MES-KOH pH 6.5, 0.3 mM NADH, 10 mM $MgCl_2$, 0.5 mM ADP, 0.15 mM TPP, 5 µM PduP, and OXC (concentration depending on the mutant) was preincubated for 2 min and the reaction started by adding oxalyl-CoA (concentrations depending on the mutant). Reaction procedure was monitored by following the oxidation of NADH at 340 nm.

OXC reverse. The reverse reaction of OXC was monitored by mixing 50 mM potassium phosphate pH 6.5, 100 mM $NaHCO_3$, 0.3 mM NADPH, 5 mM $MgCl_2$, 0.5 mM ADP, 0.15 mM TPP, 1 mM formyl-CoA, 1µg/ml carbonic anhydrase, 0.6 µM PanE2, 2 µM GhrB, 6.6 µM OXC. The reaction was started by adding OXC and formyl-CoA, respectively, and activity monitored by following the oxidation of NADPH at 340 nm. Activity only occurred after addition of the last component.

PanE2. For the oxalyl-CoA reductase activity of PanE2 an assay containing 100 mM potassium phosphate pH 7.5, 0.5 mM NADPH, 18 nM PanE2 was preincubated for 2 min and the reaction started by adding oxalyl-CoA to a final concentration of 5, 12.5, 25, 100 and 250 µM, respectively. Reaction procedure was monitored by following the oxidation of NADPH at 365 nm (ε_{365nm} = 3.4 mM^{-1} cm^{-1}).

E. coli whole-cell biotransformation

E. coli whole-cell biotransformations of formaldehyde to glycolate were performed as described previously.[18] MeOXC variants were tested by exchanging the HACL gene with the corresponding MeOXC gene.

3.5 References

[1] Baskaya, F. S., Zhao, X., Flickinger, M. C., and Wang, P. (2010) Thermodynamic feasibility of enzymatic reduction of carbon dioxide to methanol, *Appl Biochem Biotechnol 162*, 391-398.

[2] Schlager, S., Dibenedetto, A., Aresta, M., Apaydin, D. H., Dumitru, L. M., Neugebauer, H., and Sariciftci, N. S. (2017) Biocatalytic and Bioelectrocatalytic Approaches for the Reduction of Carbon Dioxide using Enzymes, *Energy Technol-Ger 5*, 812-821.

[3] Wang, Q., Warnan, J., Rodríguez-Jiménez, S., Leung, J. J., Kalathil, S., Andrei, V., Domen, K., and Reisner, E. (2020) Molecularly engineered photocatalyst sheet for scalable solar formate production from carbon dioxide and water, *Nature Energy*.

[4] Wang, W. H., Himeda, Y., Muckerman, J. T., Manbeck, G. F., and Fujita, E. (2015) CO2 Hydrogenation to Formate and Methanol as an Alternative to Photo- and Electrochemical CO2 Reduction, *Chemical reviews 115*, 12936-12973.

[5] Benson, E. E., Kubiak, C. P., Sathrum, A. J., and Smieja, J. M. (2009) Electrocatalytic and homogeneous approaches to conversion of CO2 to liquid fuels, *Chem Soc Rev 38*, 89-99.

[6] Yishai, O., Lindner, S. N., de la Cruz, J. G., Tenenboim, H., and Bar-Even, A. (2016) The formate bio-economy, *Curr Opin Chem Biol 35*, 1-9.

[7] Cotton, C. A. R., Claassens, N. J., Benito-Vaquerizo, S., and Bar-Even, A. (2020) Renewable methanol and formate as microbial feedstocks, *Curr Opin Biotech 62*, 168-180.

[8] Olah, G. A. (2005) Beyond oil and gas: The methanol economy, *Angew Chem Int Edit 44*, 2636-2639.

[9] Kim, S., Lindner, S. N., Aslan, S., Yishai, O., Wenk, S., Schann, K., and Bar-Even, A. (2020) Growth of E. coli on formate and methanol via the reductive glycine pathway, *Nat Chem Biol 16*, 538-545.

[10] Chen, F. Y. H., Jung, H. W., Tsuei, C. Y., and Liao, J. C. (2020) Converting Escherichia coli to a Synthetic Methylotroph Growing Solely on Methanol, *Cell 182*, 933-+.

[11] Claassens, N. J., Bordanaba-Florit, G., Cotton, C. A. R., De Maria, A., Finger-Bou, M., Friedeheim, L., Giner-Laguarda, N., Munar-Palmer, M., Newell, W., Scarinci, G., Verbunt, J., de Vries, S. T., Yilmaz, S., and Bar-Even, A. (2020) Replacing the Calvin cycle with the reductive glycine pathway in Cupriavidus necator, *Metab Eng 62*, 30-41.

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

[13] Bar-Even, A. (2016) Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways, *Biochemistry-Us 55*, 3851-3863.

[14] Schwander, T., von Borzyskowski, L. S., Burgener, S., Cortina, N. S., and Erb, T. J. (2016) A synthetic pathway for the fixation of carbon dioxide in vitro, *Science 354*, 900-904.

[15] Kluger, R., and Tittmann, K. (2008) Thiamin diphosphate catalysis: enzymic and nonenzymic covalent intermediates, *Chemical reviews 108*, 1797-1833.

[16] Siegel, J. B., Smith, A. L., Poust, S., Wargacki, A. J., Bar-Even, A., Louw, C., Shen, B. W., Eiben, C. B., Tran, H. M., Noor, E., Gallaher, J. L., Bale, J., Yoshikuni, Y., Gelb, M. H., Keasling, J. D., Stoddard, B. L., Lidstrom, M. E., and Baker, D. (2015) Computational protein design enables a novel one-carbon assimilation pathway, *Proceedings of the National Academy of Sciences of the United States of America 112*, 3704-3709.

[17] Lu, X., Liu, Y., Yang, Y., Wang, S., Wang, Q., Wang, X., Yan, Z., Cheng, J., Liu, C., Yang, X., Luo, H., Yang, S., Gou, J., Ye, L., Lu, L., Zhang, Z., Guo, Y., Nie, Y., Lin, J., Li, S., Tian, C., Cai, T., Zhuo, B., Ma, H., Wang, W., Ma, Y., Liu, Y., Li, Y., and Jiang, H. (2019) Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design, *Nature communications 10*, 1378.

[18] Chou, A., Clomburg, J. M., Qian, S., and Gonzalez, R. (2019) 2-Hydroxyacyl-CoA lyase catalyzes acyloin condensation for one-carbon bioconversion, *Nature chemical biology 15*, 900-906.

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

[20] Berthold, C. L., Toyota, C. G., Moussatche, P., Wood, M. D., Leeper, F., Richards, N. G., and Lindqvist, Y. (2007) Crystallographic snapshots of oxalyl-CoA decarboxylase give insights into catalysis by nonoxidative ThDP-dependent decarboxylases, *Structure 15*, 853-861.

[21] Foulon, V., Sniekers, M., Huysmans, E., Asselberghs, S., Mahieu, V., Mannaerts, G. P., Van Veldhoven, P. P., and Casteels, M. (2005) Breakdown of 2-hydroxylated straight chain fatty acids via peroxisomal 2-hydroxyphytanoyl-CoA lyase, *Journal of Biological Chemistry 280*, 9802-9812.

[22] Reetz, M. T., and Carballeira, J. D. (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes, *Nature protocols 2*, 891-903.

[23] Baetz, A. L., and Allison, M. J. (1990) Purification and characterization of formyl-coenzyme A transferase from Oxalobacter formigenes, *Journal of bacteriology 172*, 3537-3540.
[24] Volodina, E., Schurmann, M., Lindenkamp, N., and Steinbuchel, A. (2014) Characterization of propionate CoA-transferase from Ralstonia eutropha H16, *Applied microbiology and biotechnology 98*, 3579-3589.
[25] Selmer, T., Willanzheimer, A., and Hetzel, M. (2002) Propionate CoA-transferase from Clostridium propionicum. Cloning of gene and identification of glutamate 324 at the active site, *European journal of biochemistry 269*, 372-380.
[26] Scherf, U., and Buckel, W. (1991) Purification and properties of 4-hydroxybutyrate coenzyme A transferase from Clostridium aminobutyricum, *Applied and environmental microbiology 57*, 2699-2702.
[27] Schwam, H., Michelson, S., Randall, W. C., Sondey, J. M., and Hirschmann, R. (1979) Purification and Characterization of Human-Liver Glycolate Oxidase - Molecular-Weight, Subunit, and Kinetic-Properties, *Biochemistry 18*, 2828-2833.
[28] Kille, S., Acevedo-Rocha, C. G., Parra, L. P., Zhang, Z. G., Opperman, D. J., Reetz, M. T., and Acevedo, J. P. (2013) Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis, *ACS synthetic biology 2*, 83-92.
[29] Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012) eQuilibrator-the biochemical thermodynamics calculator, *Nucleic Acids Res 40*, D770-D775.
[30] Quayle, J. R. (1963) Carbon Assimilation by Pseudomonas Oxalaticus (Ox1). 7. Decarboxylation of Oxalyl-Coenzyme a to Formyl-Coenzyme A, *The Biochemical journal 89*, 492-503.
[31] Schneider, K., Skovran, E., and Vorholt, J. A. (2012) Oxalyl-coenzyme A reduction to glyoxylate is the preferred route of oxalate assimilation in Methylobacterium extorquens AM1, *Journal of bacteriology 194*, 3144-3155.
[32] Nunez, M. F., Pellicer, M. T., Badia, J., Aguilar, J., and Baldoma, L. (2001) Biochemical characterization of the 2-ketoacid reductases encoded by ycdW and yiaE genes in Escherichia coli, *Biochem J 354*, 707-715.
[33] Fiedler, E., Thorell, S., Sandalova, T., Golbik, R., Konig, S., and Schneider, G. (2002) Snapshot of a key intermediate in enzymatic thiamin catalysis: crystal structure of the alpha-carbanion of (alpha,beta-dihydroxyethyl)-thiamin diphosphate in the active site of transketolase from Saccharomyces cerevisiae, *Proceedings of the National Academy of Sciences of the United States of America 99*, 591-595.
[34] Nakai, T., Nakagawa, N., Maoka, N., Masui, R., Kuramitsu, S., and Kamiya, N. (2004) Ligand-induced conformational changes and a reaction intermediate in branched-chain 2-oxo acid dehydrogenase (E1) from Thermus thermophilus HB8, as revealed by X-ray crystallography, *J Mol Biol 337*, 1011-1033.
[35] Wille, G., Meyer, D., Steinmetz, A., Hinze, E., Golbik, R., and Tittmann, K. (2006) The catalytic cycle of a thiamin diphosphate enzyme examined by cryocrystallography, *Nature chemical biology 2*, 324-328.
[36] Kato, N., Yurimoto, H., and Thauer, R. K. (2006) The physiological role of the ribulose monophosphate pathway in bacteria and archaea, *Biosci Biotechnol Biochem 70*, 10-21.
[37] Orita, I., Sakamoto, N., Kato, N., Yurimoto, H., and Sakai, Y. (2007) Bifunctional enzyme fusion of 3-hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase, *Applied microbiology and biotechnology 76*, 439-445.
[38] Ro, Y. T., Eom, C. Y., Song, T. S., Cho, J. W., and Kim, Y. M. (1997) Dihydroxyacetone synthase from a methanol-utilizing carboxydobacterium, Acinetobacter sp. strain JC1 DSM 3803, *Journal of bacteriology 179*, 6041-6047.
[39] Kato, N., Higuchi, T., Sakazawa, C., Nishizawa, T., Tani, Y., and Yamada, H. (1982) Purification and Properties of a Transketolase Responsible for Formaldehyde Fixation in a Methanol-Utilizing Yeast, Candida-Boidinii (Kloeckera Sp) No 2201, *Biochim Biophys Acta 715*, 143-150.
[40] Yorita, K., Aki, K., OhkumaSoyejima, T., Kokubo, T., Misaki, H., and Massey, V. (1996) Conversion of L-lactate oxidase to a long chain alpha-hydroxyacid oxidase by site-directed mutagenesis of alanine 95 to glycine, *Journal of Biological Chemistry 271*, 28300-28305.
[41] Yorita, K., Janko, K., Aki, K., Ghisla, S., Palfey, B. A., and Massey, V. (1997) On the reaction mechanism of L-lactate oxidase: Quantitative structure-activity analysis of the reaction with para-substituted L-mandelates, *Proceedings of the National Academy of Sciences of the United States of America 94*, 9590-9595.
[42] Kalliri, E., Mulrooney, S. B., and Hausinger, R. P. (2008) Identification of Escherichia coli YgaF as an L-2-hydroxyglutarate oxidase, *Journal of bacteriology 190*, 3793-3798.

[43] Dewanti, A. R., Xu, Y., and Mitra, B. (2004) Role of glycine 81 in (S)-mandelate dehydrogenase from Pseudomonas putida in substrate specificity and oxidase activity, *Biochemistry 43*, 10692-10700.
[44] Yang, J. E., Park, S. J., Kim, W. J., Kim, H. J., Kim, B. J., Lee, H., Shin, J., and Lee, S. Y. (2018) One-step fermentative production of aromatic polyesters from glucose by metabolically engineered Escherichia coli strains, *Nature communications 9*.
[45] Trudeau, D. L., Edlich-Muth, C., Zarzycki, J., Scheffen, M., Goldsmith, M., Khersonsky, O., Avizemer, Z., Fleishman, S. J., Cotton, C. A. R., Erb, T. J., Tawfik, D. S., and Bar-Even, A. (2018) Design and in vitro realization of carbon-conserving photorespiration, *Proceedings of the National Academy of Sciences of the United States of America 115*, E11455-E11464.
[46] Zarzycki, J., Sutter, M., Cortina, N. S., Erb, T. J., and Kerfeld, C. A. (2017) In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glycyl Radical Enzyme - Associated Bacterial Microcompartment, *Scientific reports 7*, 42757.
[47] Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005) Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research, *DNA research : an international journal for rapid publication of reports on genes and genomes 12*, 291-299.
[48] Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement, *Acta Crystallogr D 66*, 133-144.
[49] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallographica Section D-Structural Biology 67*, 235-242.
[50] Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallographica Section D-Structural Biology 66*, 213-221.
[51] Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallographica Section D-Structural Biology 60*, 2126-2132.

3.6 Supplementary Figures

Figure S1. A) CoA transferases screen for GFT activity. Formyl-CoA was detected by LC-MS. Activity was normalized to the highest activity (AbfT). FRC is formyl-CoA:oxalate CoA-transferase from *Oxalobacter formigenes*, AbfT is 4-hydroxybutyrate CoA-transferase from *Clostridium aminobutyricum*, CnPCT is propionyl-CoA transferase from *Cupriavidus necator*, CpPCT is propionyl-CoA transferase from *Clostridium propionicum*. **B)** The reaction progress of AbfT-catalyzed CoA transfer from glycolyl-CoA onto formate was monitored by LC-MS. The build-up of CoA is due to hydrolysis of formyl-CoA.

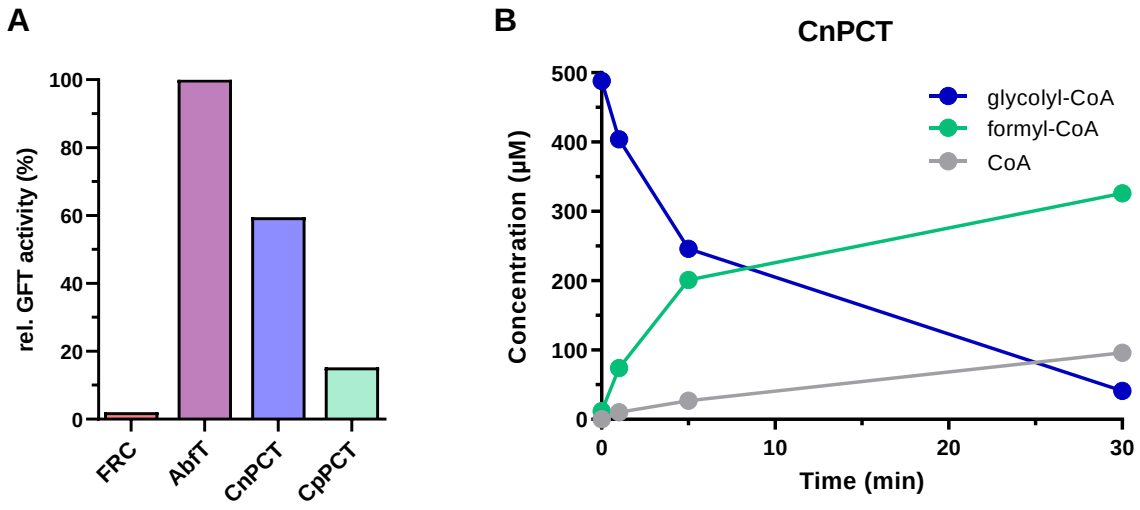


Figure S2. Sequencing chromatograms of the 22c-trick libraries. Green, A. Blue, C. Black, G. Red, T.

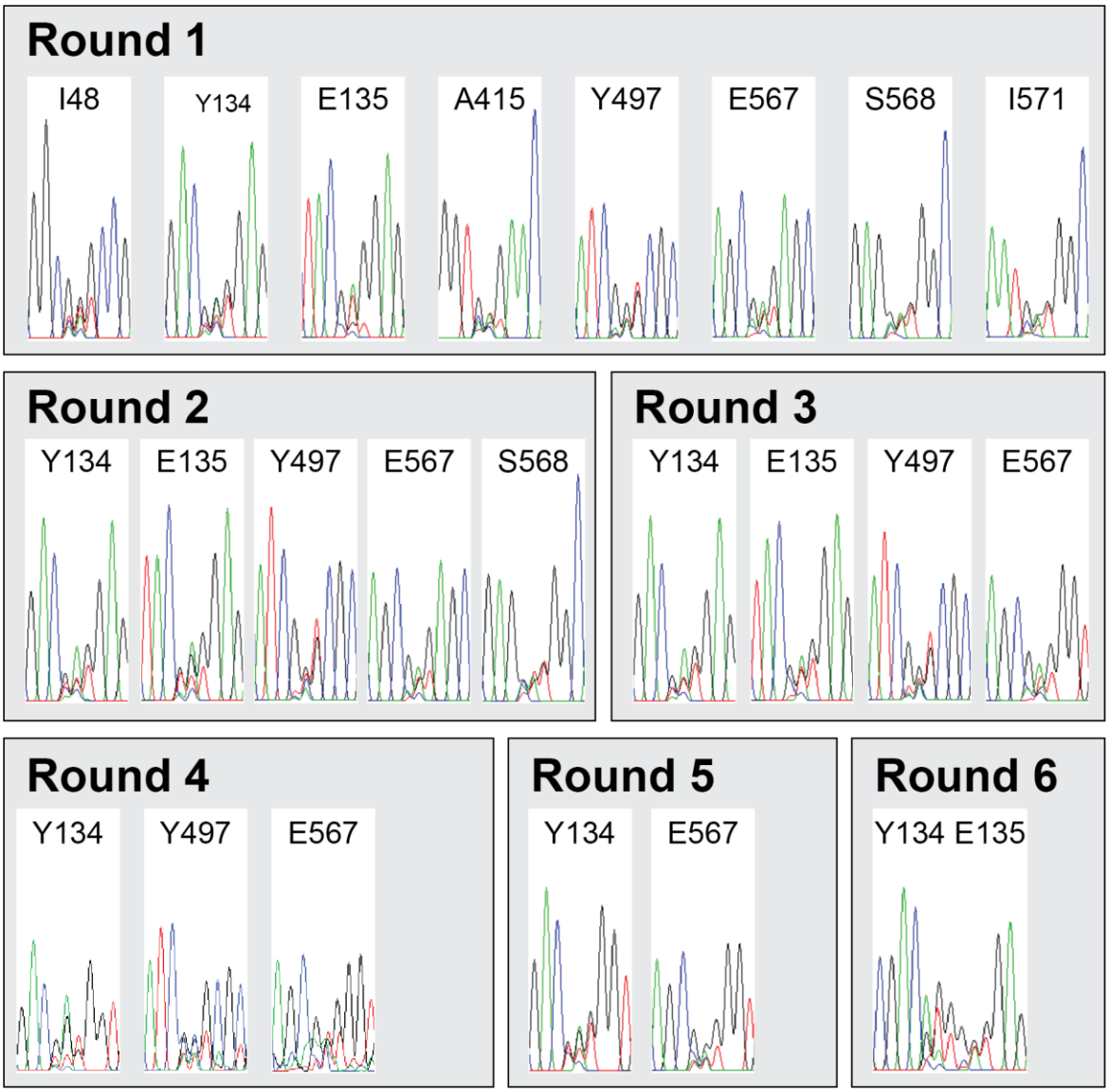


Figure S3. SDS-PAGE of purified MeOXC variants.

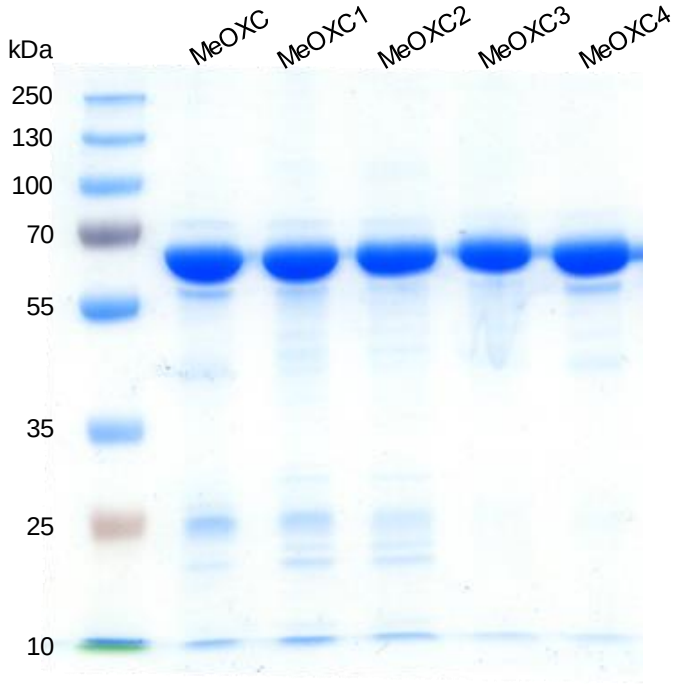


Figure S4. Michaelis-Menten graphs of MeOXC variants with formaldehyde (top) and formyl-CoA (bottom) as substrate. In the box is MeOXC4 with oxalyl-CoA as substrate.

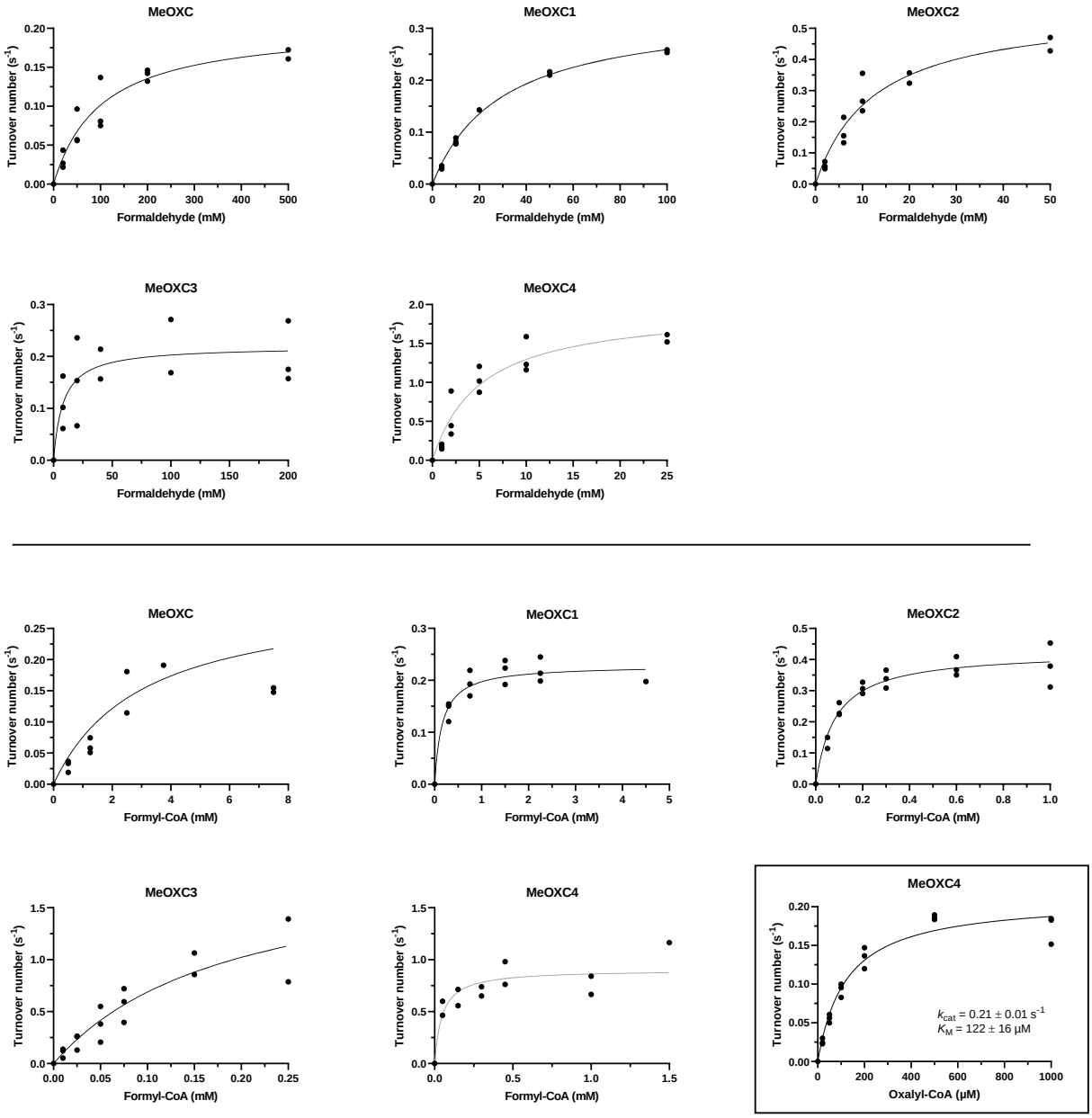
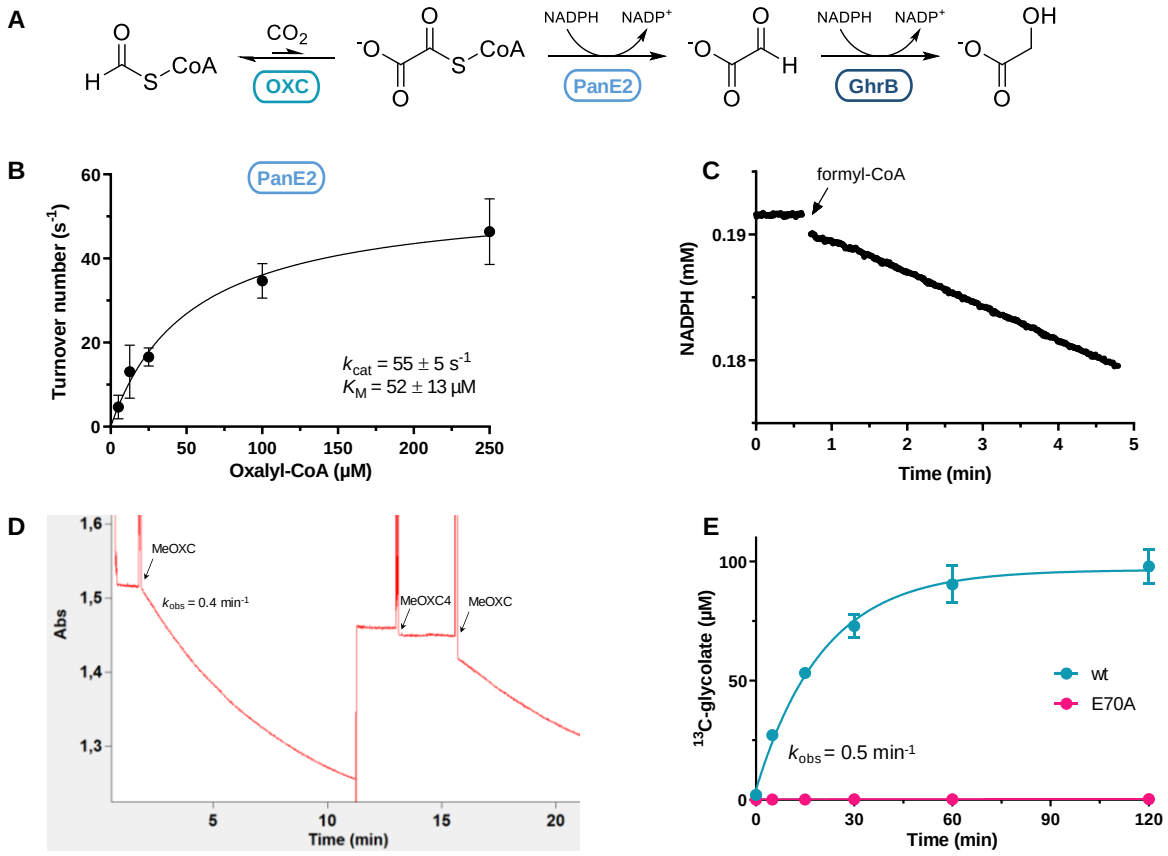


Figure S5. Reverse reaction of OXC. A) Reaction scheme of the OXC-PanE2-GhrB cascade, converting formyl-CoA and CO2 into glycolate under consumption of 2 NADPH. **B)** Michaelis-Menten plot of PanE2 catalyzing the NADPH-dependent reduction of oxalyl-CoA to glyoxylate. The NADPH concentration was 0.25 mM. Error bars indicate the standard deviation of three replicates. **C & D)** Operation of the full cascade was monitored by NADPH oxidation (340 nm). The arrows indicate the addition of formyl-CoA (C), or MeOXC (D). **E)** LC-MS detection of ^{13}C-glycolate produced in the cascade reaction with wild-type OXC or the mutant E70A, which is catalytically inactive. That E70A does not show any product formation confirms that the reaction is not catalyzed by free or unspecifically bound ThDP. The k_{obs} values in **D** & **E** were estimated based on the initial slope of product formation.



CHAPTER III

A radical strategy:

the C1 carboligase pyruvate formate-lyase

4 Engineering synthetic formate fixation pathways based on pyruvate formate-lyase

Simon Burgener[1,*], Luca Schulz[1,*], Charles A. R. Cotton[2], Arren Bar-Even[2], Tobias J. Erb[1,3]

* equal contribution

[1] Department of Biochemistry & Synthetic Metabolism, Max-Planck-Institute for terrestrial Microbiology, Karl-von-Frisch-Str. 10, 35043 Marburg, Germany

[2] Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany

[3] LOEWE Center for Synthetic Microbiology, 35043 Marburg, Germany

Manuscript in preparation

Author contributions

A.B.-E. and T.J.E. conceived of the project. S.B. and L.S. designed and performed the experiments and analyzed the results. C.A.C. created the *E. coli* PFL knockout strain. S.B., L.S. and T.J.E wrote the manuscript.

Abstract

Formate can be produced from CO_2 with renewable energy, making it a promising microbial feedstock for sustainable bioproduction. However, only very few natural formatotrophs are biotechnologically relevant or they employ formate fixation pathways that are not operating at energetic optimum. Here we propose two energy-efficient synthetic formate fixation pathways based on C-C bond formation by pyruvate formate-lyase. These pathways can use formate directly for C-C bond formation without prior ATP-dependent activation, which is common to natural formate assimilation pathways. We characterized and engineered enzymes for the three reaction steps activation, C-C bond formation and reduction. We then demonstrate *in vitro* that the glycolyl-CoA pathway is able to convert formate and glycolate into D-glycerate at the expense of 2 ATP and 1 NADPH. Implementing this three-enzyme pathway *in vivo* may pave the way for synthetic formate assimilation and ultimately enable bioproduction of value-added multicarbon compounds from formate.

4.1 Introduction

Engineered microbial growth on formate or methanol is a central goal of synthetic metabolism.[1-4] These one-carbon (C1) compounds can efficiently be derived from CO_2 with renewable energy[5] and are miscible in water, which avoids mass transfer limitations that apply to gaseous C1 compounds. Despite ongoing progress, the cultivation and engineering of natural methylo- and formatotrophs remains challenging.[6] Recent efforts have therefore sought to engineer biotechnologically relevant platform organisms such as *E. coli* and *S. cerevisiae* for growth on methanol or formate. This has been achieved by transferring the required genes of the naturally occurring C1 fixation pathways ribulose monophosphate (RuMP)[7-10], the dihydroxyacetone (DHA)[11] and the serine cycle[12] into these hosts.

While this approach enabled assimilation of C1 compounds, it is limited by the shortcomings of the natural metabolic pathways, such as non-optimal resource utilization or incompatibility with the host organism.[13] Synthetic C1-assimilating pathways are designed to be more energy efficient while limiting overlap with endogenous metabolism and several recent examples showcase their potential.[14-19] Another major drawback of utilizing naturally occurring methanol and formate fixation pathways is that carbon is almost exclusively assimilated in the form-aldehyde oxidation state. In a thermodynamically challenging reaction methanol or formate are first converted to formaldehyde, an intermediate that is highly cytotoxic to most organisms.[20] To circumvent this undesired intermediate, we sought to establish a C1 fixation pathway that directly assimilates the significantly less toxic formate.

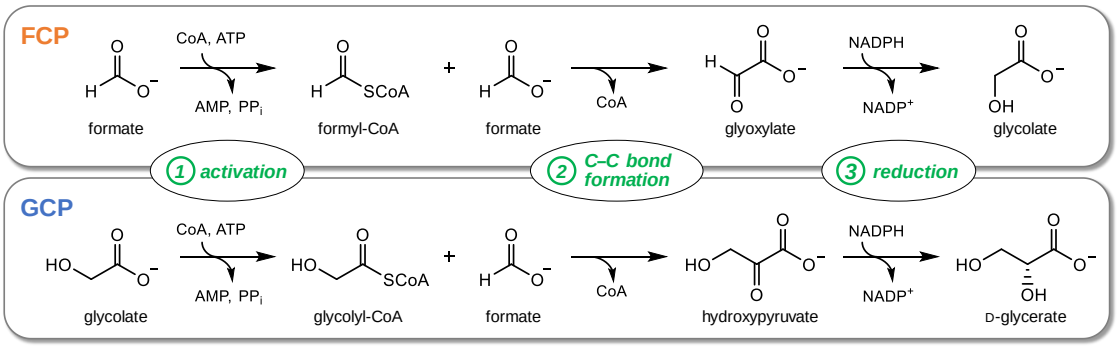
Only two natural pathways are known to support growth on formate as sole carbon source, the serine cycle, and the reductive acetyl-CoA pathway.[21] Recently, the reductive glycine pathway was proposed to support growth on formate, however, experimental evidence is lacking.[22] Interestingly, this pathway had previously been designed as synthetic pathway[23] and virtually simultaneously to its discovery in nature, it was implemented in *E. coli* and *Cupriavidus necator*.[24-26] In these organisms, synthetic methyl- and formatotrophy was established through adaptive laboratory evolution.[26, 27] Common to all three formate fixation pathways is that formate is first activated to formyl-THF at the expense of ATP.[21] This is necessary because the carbon atom of formate is neither nucleophilic nor electrophilic, which makes C-C bond formation difficult. The direct use of formate would avoid ATP consumption and may therefore enable more energy-efficient formate assimilation.[21]

One way to overcome the low reactivity of formate is to employ radical based C-C bond forming reactions. The glycyl radical enzyme pyruvate formate-lyase (PFL) catalyzes pyruvate cleavage into acetyl-CoA and formate during mixed acid fermentation. It has previously been demonstrated that PFL is able to catalyze the reverse reaction in vitro[28-30] and this reaction is reversible under physiological conditions ($\Delta_r G'^m$ = -21 kJ/mol).[31] While there is no evidence

that PFL operates in the pyruvate synthesis direction in nature, the potential of PFL-mediated synthetic formate assimilation was demonstrated in *E. coli*.[32]

Recently, PFL from *E. coli* (EcPFL) was shown to condense propionyl-CoA and formate to 2-ketobutyrate *in vivo*.[33] Additionally, *E coli* possesses another PFL-like enzyme (TdcE), which catalyzes the condensation of formate with acetyl-CoA and propionyl-CoA, respectively.[34] Based on these observations we reasoned that PFL or TdcE might be able to use other acyl-CoA donors. Promiscuity towards the accepted acyl-CoA substrates would lead to the production of different 2-keto acid products and may pave the way for synthetic formate fixation pathways.[21]

Here, we propose two synthetic formate pathways that rely on promiscuous PFL activity (**Scheme 1**). The formyl-CoA pathway (FCP) converts 2 molecules of formate into glycolate using one NADPH and two ATP equivalents with favorable thermodynamics ($\Delta_r G'^m$ = -53 kJ/mol).[31] The glycolyl-CoA pathway (GCP) converts formate and glycolate into glycerate under consumption of one NADPH and two ATP equivalents with weak thermodynamic driving force ($\Delta_r G'^m$ = -3 kJ/mol).[31] In the first step the acid is activated to the corresponding CoA ester by an acyl-CoA synthetase. PFL then condenses the CoA ester with formate, giving rise to a 2-keto acid, which is subsequently reduced to the corresponding 2-hydroxy acid. In principle the FCP and the GCP can be combined to synthesize glycerate from three molecules of formate with 4 ATP and 2 NADPH. This hypothetical pathway consisting of merely 6 reactions constitutes one of the most energy-efficient one-carbon fixation pathways.[21]



Scheme 1. The proposed synthetic formate fixation pathways. In step 1, formate or glycolate are activated to the corresponding CoA-ester, which are condensed with formate in step 2 to yield glyoxylate or hydroxypyruvate. In step 3 the 2-keto acids are reduced to 2-hydroxy acids glycolate and D-glycerate, respectively. Overall, both pathways extend a carboxylic acid by one carbon atom, derived from formate, at the expense of one ATP and one NADPH.

4.2 Results

To establish the FCP and GCP in vitro, we set out to identify enzymes for each reaction. The first two steps of both pathways are not known to occur in nature. Only for the last steps, i.e. the reduction of glyoxylate and hydroxypyruvate, enzymes are described. Conveniently, GhrB from *E. coli* catalyzes both reactions with high catalytic efficiency,[35] and was therefore the 2-keto acid reductase of choice for both pathways.

First, we focused on the activation of formate and glycolate to the corresponding CoA ester. We have previously engineered an acetyl-CoA synthetase (ACS) from *Erythrobacter* sp. NAP1 (EryACS) for improved glycolate activity, by introducing the active site mutation V379A.[36] The activity was further increased by producing the enzyme in an *E. coli* strain lacking the peptidyl-lysine N-acetyltransferase PatZ, which inhibits ACS.[37] The resulting glycolyl-CoA synthetase (GCS) showed good kinetic parameters with glycolate as substrate: k_{cat} = 11.1 ± 0.6 s^{-1} and app. K_M = 13 ± 3 mM.[36]

ACS from *E. coli* has been reported to accept formate as substrate, albeit with low catalytic efficiency.[38] EryACS showed slightly better kinetics (**Table 1**) and we therefore engineered this homologue to improve its formate activity. In GCS, Val379 was mutated to Ala, which opens up the active site to accommodate glycolate. Similarly, we replaced Val379 by Ile, which should improve formate binding. Indeed, the catalytic efficiency increased by a factor of ~3, mainly through an improved turnover number (**Table 1**). In contrast, V379A and V379N had lower activity and affinity on formate. We also targeted Thr304 for mutagenesis, because it is in close proximity of Val379 (**Figure S1**). T304N had a slightly higher activity, however, combining this mutation with V379I resulted in a very high apparent K_M. As is the case for GCS, the activity of EryACS V379I was further increased by expression in an *E. coli* strain lacking PatZ. Overall, our rational engineering efforts resulted in a formyl-CoA synthetase (FCS) with ~17-fold higher catalytic effiency compared to the wild-type (**Table 1**).

The mutation V379I led to an approximately ten-fold higher K_M for acetate (**Table 1**). Together with our previous efforts on GCS, these results underline the importance of Val379 in substrate recognition and demonstrate that mutations can enhance the activity of ACS with other carboxylic acids.

Table 1: Steady-state kinetic parameters of EryACS with formate and acetate as substrate.

	Formate			Acetate		
Enzyme	**k_{cat} (s^{-1})**	**app. K_M (mM)**	**k_{cat}/K_M (s^{-1} M^{-1})**	**k_{cat} (s^{-1})**	**app. K_M (µM)**	**k_{cat}/K_M (s^{-1} M^{-1})**
EcACS	0.10 ± 0.01	122 ± 23	0.8	n.d.	n.d.	n.d.
EryACS	0.67 ± 0.03	73 ± 8	9.2	0.95 ± 0.03	27 ± 4	3.5×10^4
T304N	1.5 ± 0.1	121 ± 18	12	n.d.	n.d.	n.d.
T304I	0.41 ± 0.03	39 ± 7	11	n.d.	n.d.	n.d.
V379A	0.16 ± 0.01	158 ± 20	1	n.d.	n.d.	n.d.
V379N	0.16 ± 0.01	366 ± 61	0.4	n.d.	n.d.	n.d.
V379I	2.5 ± 0.2	82 ± 13	30	3.1 ± 0.1	201 ± 16	1.5×10^4
T304N V379I	0.5 ± 0.1	328 ± 112	1.5	n.d.	n.d.	n.d.
FCS	17.4 ± 1.3	113 ± 17	153	16.9 ± 1.3	250 ± 46	6.7×10^4

Errors reflect standard deviation of three independent measurements. Michaelis-Menten graphs are shown in **Figure S2**. FCS is EryACS V379I produced in *E. coli* Δ*patZ*. n.d., not determined.

Having identified enzymes for the first and the third step, we focused on the C-C bond forming reaction between formate and formyl-CoA or glycolyl-CoA. As mentioned above, we hypothesized that PFL may exhibit promiscuous activity for these non-natural substrates. Despite the broad abundance of PFL among bacterial and eukaryotic species (**Figure 1C**), biochemical characterization is limited to very few homologues. So far, only three PFLs (from *E. coli*, *Streptococcus mutans* and *Chlamydomonas reinhardtii*) and TdcE have been characterized biochemically.[34, 39-41]

To better understand the activity and substrate scope within the PFL enzyme family, we performed a phylogenetic analysis of 500 PFL sequences, to inform the selection of candidates to study. All sequences not containing the active site motif Cys-Cys were discarded, as this has been shown to be strictly required for PFL activity.[42] The multiple sequence alignment (MSA) combined with the EcPFL structure showed that the active site is highly conserved throughout all PFLs (**Figure S3**). An exception is TdcE with the two substitutions A273G and Y323F (**Figure 1A & 1B**), which open up the active site to accept bulkier substrates, in line with TdcE's role as 2-ketobutyrate-cleaving enzyme.[34] Interestingly, TdcE is the enzyme in our MSA with these mutations, suggesting that the anaerobic threonine degradation pathway to which TdcE belongs is rare or that a "true" PFL may catalyze this reaction in other organisms. The second exception are *Streptococcus* PFLs, which carry the substitution F327Y and W333F (**Figure S3**). These PFLs differ from all others not only in the active site residues, but also in their overall sequence and form a distinct monophyletic clade (**Figure 1C**). Additionally, our analysis again highlighted that multiple phylogenetically distinct clades have no characterized

representative. Based on our analysis we selected PFL from *E. coli*, *S. mutans*, and *Clostridium butyricum* (two paralogues, abbreviated by their amino acid length as Cb743 and Cb750) for detailed biochemical characterization.

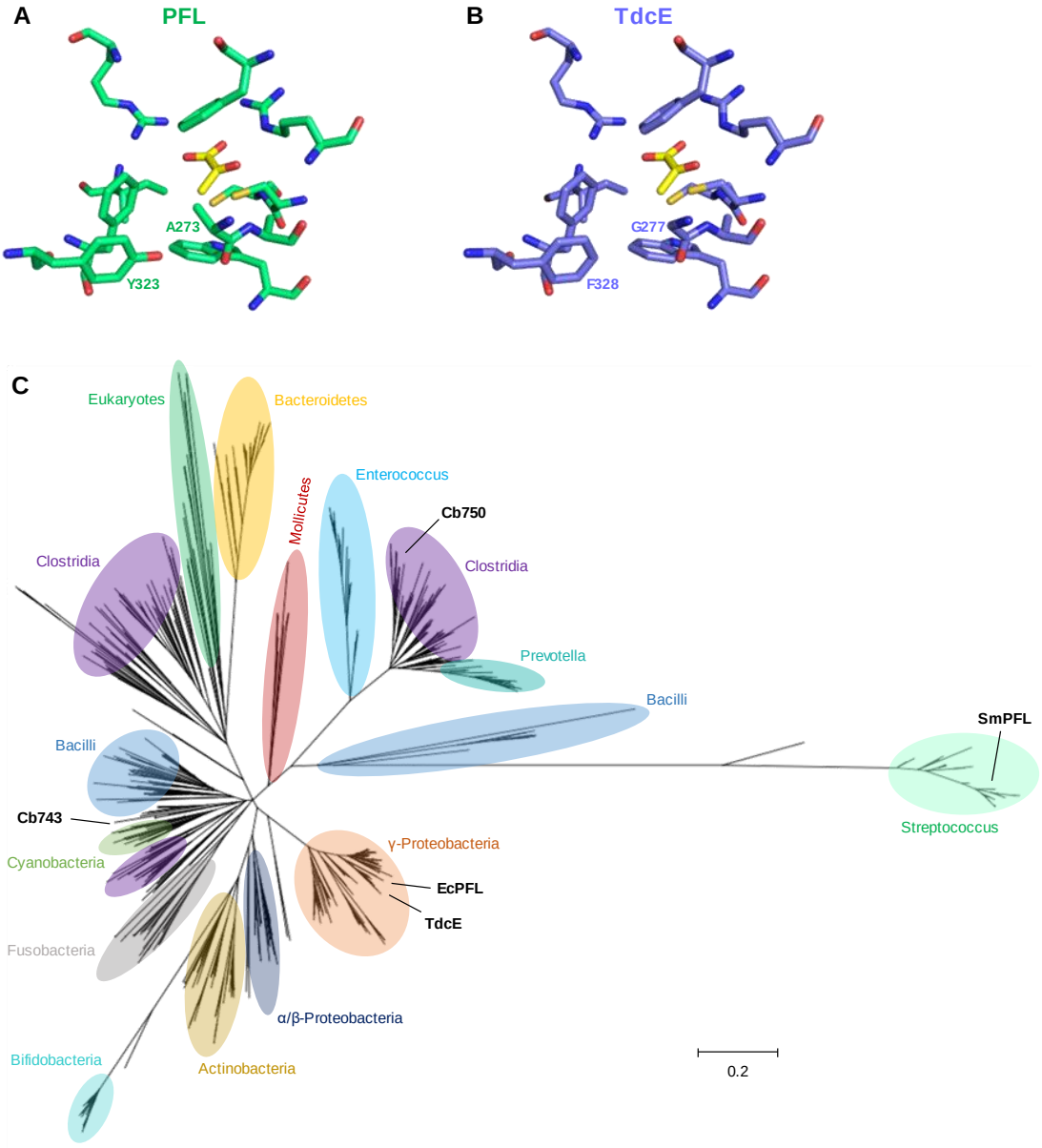


Figure 1. Structural and phylogenetic analysis of the PFL family. A) Active site of EcPFL (PDB code 1H16). Amino acids are shown in green, pyruvate in yellow. Ala273 and Tyr323 are in close contact with the methyl moiety of pyruvate. **B)** Homology model of TdcE based in EcPFL (PDB code 1H16) Amino acids are shown in purple, pyruvate in yellow. In comparison to EcPFL, Phe328 and Gly277 create space to accommodate the larger substrate 2-ketobutyrate. **C)** Phylogenetic tree of the PFL protein family. EcPFL, Cb743, Cb759 and SmPFL were selected for detailed characterization, based on their distribution over different clades and are highlighted on the phylogeny by black labels. The scale bar shows the average number of amino acid substitutions per site.

PFL requires activation by its activating enzyme (AE), a radical SAM enzyme containing an oxygen labile [4Fe-4S]-cluster.[43] Initially we used the AE from *E. coli* (EcAE) for activation of all PFLs. Activity was, however, only detected for EcPFL, suggesting that EcAE is not able to activate Cb743, Cb750 and SmPFL (**Table 3**). This prompted us to test AEs from *C. butyricum* (CbAE) and *S. mutans* (SmAE). After iron-sulfur cluster reconstitution, all AEs had increased absorbance in the 400 to 500 nm range, indicating a higher [4Fe-4S]-cluster content (**Figure 2**), and were able to activate their corresponding PFL (**Table 3**). Notably, CbAE was also able to activate all other PFLs to some extent (**Table 3**). These results confirm the activities of CbAE, SmAE, Cb743, Cb750 and SmPFL *in vitro* and indicate that AEs exert a degree of specificity for their cognate interaction partner. For further *in vitro* characterization we therefore used the corresponding AE for each PFL.

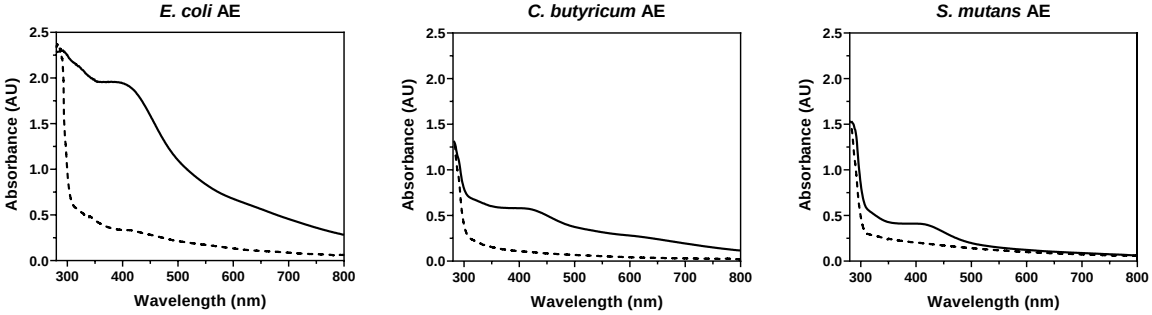


Figure 2. UV-Vis spectra of the AEs before (dashed line) and after reconstitution (solid line). The increased absorbance between 400 and 500 nm is indicative of a higher [4Fe-4S]-cluster content.

Table 3. Relative pyruvate cleavage activity of PFL homologues activated by EcAE, CbAE or SmAE.

	EcPFL	**Cb750**	**Cb743**	**SmPFL**
EcAE	100%	<1%	<1%	<1%
CbAE	70 ± 15 %	100%	100%	10 ± 4 %
SmAE	<1%	<1%	5 ± 2 %	100%

Activity is given relative to the cognate interaction and errors indicate standard deviation of three independent measurements.

With the aim of finding PFL variants with good kinetic traits, especially in the formate condensation direction, we determined steady-state kinetics for all PFLs in the pyruvate cleavage (forward) and pyruvate formation (reverse) direction. In the forward direction, turnover numbers were between 109 and 170 s^{-1}, and in the reverse direction between 30 and 70 s^{-1} (**Table 4**). In both cases the k_{cat} values are more than five times higher than the previously reported ones,[39] confirming that our activation protocol yields highly active PFL.

EcPFL has a comparably low K_M for pyruvate of 0.6 mM and the K_M for formate was approximately 50 mM for all PFLs (**Table 4**). Cb743 gave inconsistent results, presumably caused by incomplete and unreliable activation by CbAE. *C. butyricum* has other uncharacterized AEs, which could be the physiological partner of Cb743. Overall, these results indicate that PFLs from the selected, phylogenetically distinct clades have similar kinetic properties.

Table 4. Apparent steady-state kinetic parameters of PFL homologues.

	Pyruvate		CoA		Acetyl-CoA		Formate	
Enzyme	k_{cat} (s^{-1})	K_M (mM)	k_{cat} (s^{-1})	K_M (µM)	k_{cat} (s^{-1})	K_M (µM)	k_{cat} (s^{-1})	K_M (mM)
EcPFL	22 ± 1[a]	0.6 ± 0.1	32 ± 2[a]	79 ± 13	36 ± 2	176 ± 31	38 ± 3	49 ± 8
Cb750	170 ± 3	7.3 ± 0.4	168 ± 7	49 ± 6	67 ± 3	229 ± 20	57 ± 2	82 ± 6
Cb743	131 ± 6	5 ± 1	115 ± 10	156 ± 40	16 ± 3[a]	407 ± 158	45 ± 8	62 ± 29
SmPFL	153 ± 10	21 ± 3	109 ± 3	16 ± 2	31 ± 1	51 ± 8	30 ± 1	54 ± 8

Turnover numbers were calculated assuming full glycyl radical activation of PFL. Errors reflect standard deviation of three independent measurements. Michaelis-Menten graphs are shown in **Figure S4**. [a] These values were obtained with improperly activated PFL and may therefore appear too low.

Next, we screened the PFL variants for activity with formyl-CoA and glycolyl-CoA, respectively, by incubating activated PFLs with the respective CoA ester, formate and GhrB. GhrB was added to increase the driving force of the reaction and to enable quantification of the produced glycolate and glycerate by LC-MS. Initially, no or very low product formation was observed. LC-MS analysis revealed that formyl-CoA and, to a lesser extent, glycolyl-CoA were unstable under the assay conditions (data not shown). Low CoA ester stability was also observed in the presence of only buffer and 0.5 mM DTT (**Figure S5**). Presumably, the free thiols of DTT act as acyl acceptors and cleave formyl-CoA and glycolyl-CoA. As established during the optimization of the activation protocol, DTT was necessary to reach high PFL activity. However, we reasoned that other, non-thiol containing reducing agents could give high activity without causing CoA ester instability. To this end, we replaced DTT with the same concentration of tris-(2-carboxyethyl)-phosphin (TCEP) and measured pyruvate cleavage activity. Virtually identical activities were obtained when using TCEP instead of DTT (**Figure 3A & 3B**).

Using our improved assay conditions, we were able to detect activity with glycolyl-CoA (**Figure 3C**), but not formyl-CoA. As described above, the two residues Ala273 and Tyr323 are involved in substrate binding. We mutated Ala273 to Gly with the rationale that a larger binding pocket would improve binding of glycolyl-CoA. Analogously, we created the mutation A273V for formyl-CoA. These mutants retained PFL activity, with the exception of SmPFL A67V (**Figure 3B**), indicating that they can still be activated. The mutation A273G increased the glycolyl-CoA

activity in EcPFL, but not in CbPFL (**Figure 3C**). However, CbPFL A267G had also an 11-fold lower PFL activity, thus, overall this mutation caused a specificity switch towards glycolyl-CoA, which could be beneficial under physiological conditions.

Glycolyl-CoA is slightly smaller than propionyl-CoA, for which TdcE shows high activity.[33] With this rationale we narrowed TdcE's active site by mutating Phe328 to Tyr. This mutation caused decreased pyruvate cleavage and glycolyl-CoA condensation activity. Thus, despite the successful engineering of EcPFL, wild-type TdcE remained the best enzyme for glycolyl-CoA condensation. We were not able detect any activity with formyl-CoA, even with the engineered variants.

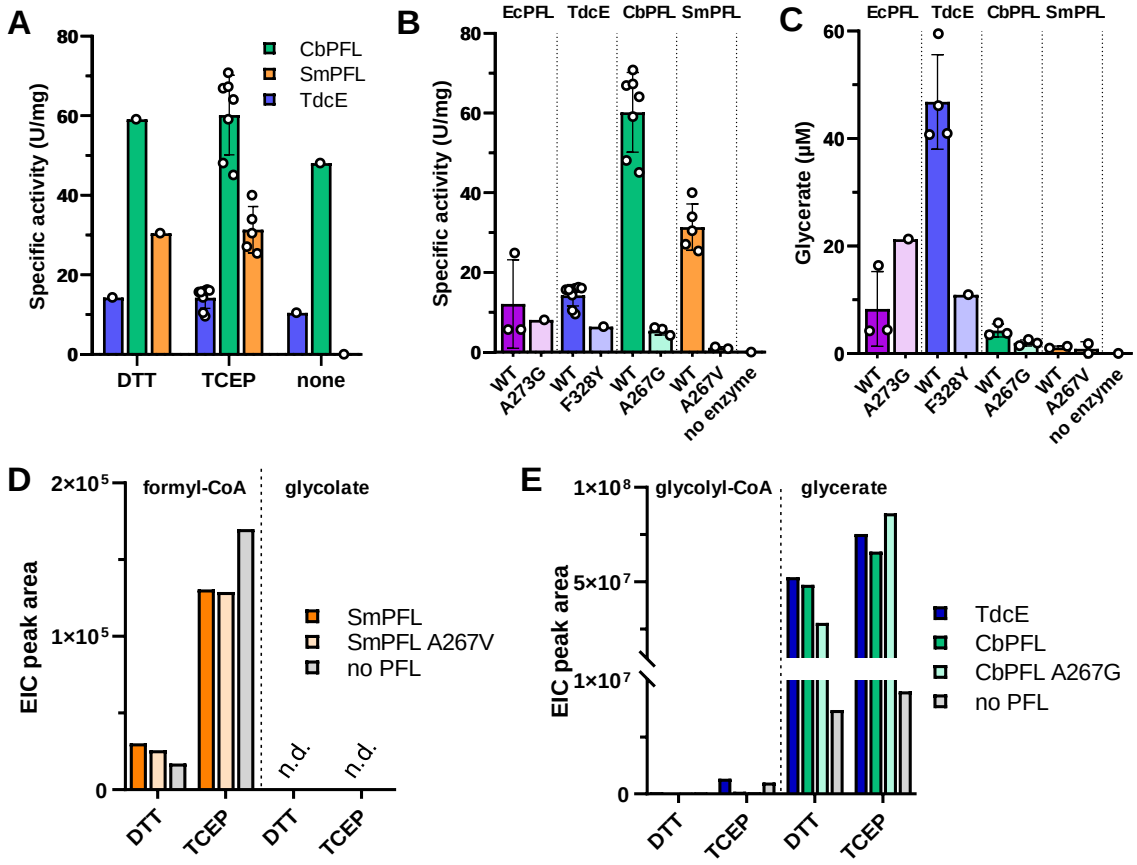


Figure 3. PFL engineering and assembly of the GCP. A) Pyruvate cleavage activity of PFL variants with DTT or TCEP as reducing agents in the activation reaction. **B)** Pyruvate cleavage activity of PFL variants with TCEP. **C)** Glycerate production from glycolyl-CoA and formate by the coupled reaction of PFL variants and GhrB. Reaction time was 1 h. **D)** Extracted ion counts of formyl-CoA and glycolate produced via the FCP pathway, comprising FCS, PFL and GhrB with DTT and TCEP, respectively. Glycolate was not detectable. **E)** Extracted ion counts of glycolyl-CoA and glycerate produced via the GCP pathway, comprising GCS, PFL and GhrB with DTT and TCEP, respectively.

Having identified enzymes for all three steps of the GCP, we tested the whole pathway by combining GCS and GhrB with PFL (TdcE, CbPFL or CbPFL A267G). Glycerate production was observed with the complete GCP, whereas no PFL-controls produced no product (**Figure**

3E). As expected, glycerate yields were higher when TCEP was used instead of DTT, which is likely due to higher glycolyl-CoA availability (**Figure 3E**). These results demonstrate that the GCP is able to produce glycerate from glycolate and formate. With further optimizations, the GCP could become a viable, energy-efficient route for synthetic formate fixation.

We also tested the FCP, but, while we observed the accumulation of formyl-CoA, no glycolate was detectable, owing to the inability of PFL to accept formyl-CoA. As expected, formyl-CoA production was ~4-fold higher in the presence of TCEP compared to DTT (**Figure 3D**), confirming that DTT drains CoA ester pools *in vitro*.

4.3 Discussion

We propose two energy-efficient synthetic formate fixation pathways: FCP and GCP. We screened and engineered enzymes for the missing steps, i.e. (1) activation of formate and glycolate, (2) C-C bond formation between formate and formyl-CoA/glycolyl-CoA. As a proof of principle, formate fixation through the GCP was demonstrated in vitro.

The GCP converts the C2 compound glycolate into the C3 compound D-glycerate with formate as C1 source. Similarly, the photorespiration pathway synthesizes D-glycerate from two glycolate molecules. However, the release of previously fixed CO_2 during photorespiration reduces the efficiency of photosynthesis and eventually crop yield. The GCP on the other hand is formate-fixing, i.e. carbon positive. Recently, a carbon positive, synthetic photorespiration bypass was established *in vitro*, the tartronyl-CoA pathway (TaCoP).[36] The TaCoP features an engineered glycolyl-CoA carboxylase (GCC) that extends glycolyl-CoA with CO_2 (coming from bicarbonate) to tartronyl-CoA, which is subsequently reduced to D-glycerate. Thus, the GCP and the TaCoP are very similar, essentially only differing in the C1 source. The use of formate could be advantageous because it prevents the competition between a carboxylase and Rubisco for CO_2 or bicarbonate. Note that formate is present in chloroplasts of higher plants.[44] However, oxygen-tolerance of PFL (or TdcE) would have to be evolved.

Previously, by examining underground isoleucine biosynthesis in *E. coli* we were able to show that TdcE synthesizes 2-ketobutyrate when propionate is present.[33] Similarly, TdcE could form hydroxypyruvate if glycolate becomes available. Yet, this requires conversion of glycolate into glycolyl-CoA and there is no evidence that glycolyl-CoA occurs in any natural metabolic pathway. Nevertheless, it is possible that PFL-based glycolate assimilation is naturally occurring.

Why is PFL unable to accept formyl-CoA as substrate? In the first reaction step, a formyl-cysteine intermediate is formed by transacylation of formyl-CoA. Formyl-CoA is inherently unstable, particularly in the presence of free thiols, such as DTT (**Figure S4**). Presumably, the

formyl-cysteine intermediate is of similar or even lower stability and degrades before subsequent C-C bond formation can take place.

Even though the FCP could not be established because no PFL variant with the required activity could be found, the development of formate-activating FCS might prove useful. FCS could enable other synthetic C1 fixation pathways, for instance the formolase pathway[38] or pathways based on 2-hydroxyacyl-CoA lyase (HACL). HACL catalyzes the acyloin condensation of formyl-CoA with a broad range of aldehydes, including formaldehyde.[45, 46] Combining FCS and HACL would create a synthetic pathway that converts formate and formaldehyde to glycolyl-CoA, very similar to the proposed FCP. Notably, this pathway does not require anaerobic conditions, because HACL is completely oxygen insensitive. Thus, FCS is a valuable addition to the growing toolbox of C1 fixation enzymes.

Our study highlights the potential of combining enzyme and metabolic engineering in creating artificial, tailor-made metabolic pathways. Synthetic formate fixation pathways could pave the way for sustainable production of value-added compounds with formate as sole feedstock.

Acknowledgements

We thank J. Heider for gently providing instrument time on the anaerobic spectrophotometer. This work was supported by the German Ministry of Education and Research Grant FormatPlant (part of BioEconomy 2030, Plant Breeding Research for the Bioeconomy).

4.4 Materials and Methods

Chemicals

Chemicals were obtained from Sigma-Aldrich (Munich, Germany) and Carl Roth GmbH (Karlsruhe, Germany). Coenzyme A was obtained from Roche Diagnostics Deutschland GmBH (Mannheim, Germany). Biochemicals, commercially available proteins and materials for cloning and protein production were obtained from Thermo Fisher Scientific (St. Leon-Rot, Germany), New England Biolabs GmbH (Frankfurt am Main, Germany) and Macherey-Nagel GmbH (Düren, Germany). Primers and synthesized genes were obtained from Eurofins MWG GmbH (Ebersberg, Germany). Materials and equipment for protein purification were obtained from GE Healthcare (Freiburg, Germany), BioRad (Munich, Germany) and Merck Millipore GmbH (Schwalbach, Germany).

Synthesis of CoA-thioesters

CoA esters were synthesized as described previously. In short, acetyl-CoA was synthesized via acetic anhydride,[47] formyl-CoA via formylthiophenol[46] and glycolyl-CoA using carbodiimidazole coupling.[47, 48]
The concentrations of CoA esters were quantified by absorption at 260 nm (ε = 16.4 mM^{-1} cm^{-1}) and depletion using 2 µM CoA-dependent propionaldehyde dehydrogenase PduP in an assay containing 200 µM NADH, 100 mM Tris-HCl (pH 7.0) and approximately 50 µM CoA ester.

Cloning and site-directed mutagenesis

Oligonucleotides were synthesized by Eurofins Genomics (Ebersbach, Germany). For purification, preparation, cloning, transformation and amplification of DNA, standard protocols by the supplier were used. The overexpression plasmids for EcAE, EcPFL, TdcE and GhrB were obtained from the ASKA collection.[49] The overexpression plasmids for CbAE, SmAE, Cb750, Cb743 and SmPFL were codon optimized (Table S2) and synthesized by Eurofins Genomics (Ebersberg, Germany). Point mutations were introduced by site-directed mutagenesis. Plasmids and primers used for site-directed mutagenesis are listed in Table S1. A 50 µL PCR reaction contained 60 ng of template plasmid, 0.25 µM forward and reverse primer, 200 µM dNTP, 5 µL 10x reaction buffer and 1 µL Phusion® High-Fidelity DNA Polymerase (2 U/µL). PCR was performed in 35 cycles, including denaturation for 30 s at 98 °C, annealing for 30 s at 65 °C and chain elongation for 4.5 min at 72 °C. Template plasmid was removed by DpnI digest (10 U) at 37 °C. Mutated plasmids were confirmed by sequencing.

Protein production and purification

EcPFL and TdcE and mutants thereof were produced in *E. coli* SIJ488 Δ*pflA* Δ*pflB* Δ*tdcE*, FCS was produced in *E. coli* BL21-AI Δ*patZ*,[36] all other proteins were produced in *E. coli* BL21(DE3). Cells were transformed with the expression plasmid and grown overnight on LB agar plates containing the appropriate antibiotic. Terrific broth (TB) containing 100 µg/mL ampicillin (CbAE, SmAE, Cb743, Cb750, SmPFL, EryACS) or 34 µg/mL chloramphenicol (EcAE, EcPFL, TdcE, GhrB) was inoculated with the freshly transformed cells and aerobically incubated at 37 °C.

After reaching an OD_{600} of 0.8 expression was induced by adding IPTG to a final concentration of 0.5 mM. The protein was produced aerobically at 25°C for 16 hours. Cells were pelleted by centrifugation at 4.500x g for 15 minutes and resuspended in Buffer A (50 mM HEPES-KOH pH 7.8, 500 mM KCl). Cells were lysed by sonication and the lysate was clarified by centrifugation at 37.000x g at 4°C for 45 minutes. The supernatant was filtered through a 0.45 µm syringe tip filter (Sarstedt) and loaded onto a 1 mL HisTrap FF column (GE Healthcare, Freiburg, Germany). After loading, the column was washed with Buffer A containing 75 mM imidazole and protein was then eluted with Buffer A containing 500 mM imidazole and desalted using a HiTrap 5 ml Desalting column (GE Healthcare, Freiburg, Germany) in 50 mM HEPES-KOH pH 7.8, 150 mM KCl. Elution fractions were concentrated on Amicon Ultra-4 centrifugal filters (Merck Millipore, Darmstadt, Germany). Purified proteins were flash frozen in liquid nitrogen and stored at -80°C. Protein concentration was determined on a Nanodrop 2000 (Thermo Scientific, St. Leon-Rot, Germany), using the calculated extinction coefficient (calculated with protparam; https://web.expasy.org/protparam).

Activity assay of acyl-CoA synthetases

Acyl-CoA synthetase activities were determined via a coupled spectrophotometric assay by measuring the production of AMP. Herein produced AMP is phosphorylated twice using adenylate kinase and pyruvate kinase. The created pyruvate is subsequently reduced to lactate by an NADH-dependent lactate dehydrogenase. Reaction procedure was monitored by following the oxidation of NADH at 340 nm. The assay was performed at 30 °C and contained 100 mM MOPS-KOH pH 7.8, 0.3 mM NADH, 5 mM $MgCl_2$, 2.5 mM ATP, 0.5 mM CoA, 2.5 mM phosphoenol pyruvate, 200 µg mL^{-1} adenylate kinase, 15 U mL^{-1} pyruvate kinase, 23 U mL^{-1} lactate dehydrogenase and the ACS variant. After 2 min preincubation, the reaction was initiated by adding formate or acetate.

General anaerobic stock handling and storage

All anaerobic steps were performed in an anaerobic glovebox (Coy Laboratories) under a N_2 atmosphere containing 3% to 3.5% H2. Residual O_2 was removed by palladium catalysts. O_2 concentration was monitored and maintained below 5 ppm at all times. All stocks were prepared under anaerobic conditions or prepared outside of the glovebox and equilibrated for at least 3 h inside the glovebox. Dithionite and DTT stocks were freshly prepared under anaerobic conditions each day.

Iron-sulfur cluster reconstitution

For all AEs, [4Fe-4S] cluster reconstitution was performed as described previously for EcAE.[33, 50] All steps were performed anaerobically. UV-Vis spectra were recorded before and after reconstitution with a Cary 4000 UV-Vis spectrometer (Agilent Technologies) using a FiberMate2™ Fiber Optic Coupler (Harrick Scientific Products) with a path length of 1 cm (Figure 2). The reconstituted AEs were aliquoted in rubber stoppered HPLC vials, flash frozen in liquid nitrogen and stored at -80 °C until further use.

Optimized activation assay of PFL

To activate PFL in vitro, 10 µM reconstituted activating enzyme was incubated with 100 mM Tris-HCl (pH 7.5), 50 mM KCl, 10 mM DTT, 50 mM sodium oxamate and 50 µM S-Adenosyl methionine for 10 minutes. 1 µM pyruvate formate-lyase and 25 µM sodium dithionite were added to start the activation. The activation was incubated at 30 °C for 30-60 minutes, and then placed on ice. To follow the activation state of PFL during the activation assay, PFL was diluted into the activity assay for the forward reaction as described below.

Activity assay of PFL forward reaction

The pyruvate cleavage activity was determined in an anaerobic chamber at 30°C in 460 µl of 100 mM Tris-HCl (pH 7.5) buffer containing 50 mM KCl, 10 mM malate, 1 mM NAD^+, 40 mM pyruvate, 4.5 U/ml citrate synthase from porcine heart (Sigma-Aldrich), 27.5 U/ml malate dehydrogenase from porcine hearth (Sigma-Aldrich), 300 µM CoA and 1:100 diluted activation assay. The activation assay was used to start the reaction. The reduction of NAD^+ was observed at 340 nm ($\Delta\varepsilon$ = 6.22 mM^{-1} cm^{-1}) with a Cary 4000 UV-Vis spectrometer (Agilent Technologies GmbH, Waldbronn, Germany) using a FiberMate2™ Fiber Optic Coupler (Harrick Scientific Products Inc., Pleasantville, NY, USA) with a path length of 1 cm.

Activity assay of PFL reverse reaction

The reverse reaction activity was determined in an anaerobic chamber at 30°C in 460 µl of 100 mM Tris-HCl (pH 7.5) buffer containing 50 mM KCl, 1 mM acetyl-CoA, 0.15 mM NADH, 300 mM sodium formate, 20 U/ml L-lactate dehydrogenase from rabbit muscle (Sigma-Aldrich) and 1:50 diluted activation assay. The activation assay was used to start the reaction. The consumption of NADH was observed at 340 nm ($\Delta\varepsilon$ = 6.22 mM^{-1} cm^{-1}) with a Cary 4000 UV-Vis spectrometer (Agilent Technologies GmbH, Waldbronn, Germany) using a FiberMate2™ Fiber Optic Coupler (Harrick Scientific Products Inc., Pleasantville, NY, USA) with a path length of 1 cm.

Cross-activation assays

To test for cross-activation between AE and PFL combinations from different organisms, the forward PFL reaction was quantified as described above, with slight adaptations. One PFL orthologue was investigated at a time. In three separate optimized activation assays 1 µM of the PFL of interest was activated by 10 µM AE from *E. coli*, *C. butyricum* and *S. mutans*, respectively. Activation time curves of all three reactions were done as described above. To quantify the degree of activation, the activity of the non-natural PFL – AE interactions was compared to the natural PFL – AE pair. Each activation time curve was done in triplicates.

Determination of steady-state kinetics

All Michaelis-Menten kinetic measurements were done in triplicate. GraphPad Prism 9 was used for Michaelis-Menten parameter calculation and fitting. PFL kinetic parameters take the enzymes half-of-the-sites reactivity into account.[51]

Activity assay of PFL with formyl-CoA and glycolyl-CoA

PFL was activated as described above, except that 10 mM TCEP was used instead of DTT and PFL and AE concentrations were doubled, i.e. final concentration in the activation reaction was 2 µM and 20 µM, respectively. Activity with formyl-CoA and glycolyl-CoA was determined at 30 °C in a reaction containing 100 mM Tris-HCl (pH 7.5), 50 mM KCl, 100 mM sodium formate, 2 mM NADPH, 5 µM GhrB and 1 mM formyl-CoA or glycolyl-CoA. The reactions were started by addition of the activation assay in two-fold dilution, i.e. the final PFL concentration was 1 µM. The reaction was quenched in a final concentration of 5% formic acid and analyzed for glycolate and glycerate using the method described below.

Testing full FCP and GCP

Functionality of the FCP and GCP was tested anaerobically at 30 °C. PFL was activated as described above, except that 10 mM TCEP was used instead of DTT and PFL and AE concentrations were doubled, i.e. final concentration in the activation reaction was 2 µM and 20 µM, respectively. The reactions contained 100 mM Tris-HCl (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, 100 mM sodium formate, 2 mM NADPH, 10 mM ATP, 0.5 mM CoA, 5 µM GhrB, 5 µM FCS or GCS, and 50 mM glycolate (only for the GCP). The reactions were started by addition of the activation assay in two-fold dilution, i.e. the final PFL concentration was 1 µM. The reaction was quenched in a final concentration of 5% formic acid and analyzed for glycolate, glycerate and CoA esters using the methods described below.

UPLC-high resolution MS of glycolate, glycerate and CoA esters

Quenched samples were spun down at 17,000x g, 4 °C for 10 minutes to remove precipitated proteins. For CoA ester detection, samples were diluted 1:10. LC-MS analysis of CoA esters was done using an Agilent 6550 iFunnel Q-TOF LC-MS system as described previously.[52]

For the detection of glycolate and glycerate, samples were diluted 1:1000. Quantitative determination of Glycolate and Glycerate was performed using a HRES-LC-MS. The chromatographic separation was performed on an Thermo Scientific Integrion HPLC System quipped with a EGC 500 KOH eluent generator, a Dionex CR-ATC 600 trap column, and a Dionex ADRS 600 suppressor operated at a constant currant of 95 mA, using a Dionex IonPac AS11-HC-4µm column (150 x 2mm, Thermo Scientific) equipped with a 50 X 2 mm guard column of similar specificity (Thermo Scientific) at a constant eluent flow rate of 0.38 ml/min and an constant regenerater flow rate of 0.76 ml/min, and a column temperature of 30 °C. The injection volume was 1 µl. Automatic sample injection was done via a Thermo Scientific ASAP sampler. The automatically generated elution profile consisted of the following steps and linear gradients: 0 – 5 min constant at 5 mM KOH; 5 – 7 min from 5 to 100 mM KOH; 7 – 8 min constant at 100 mM KOH; 8 – 8.1 min from 100 to 5 mM KOH; 8.1 – 10 min constant at 5 mM. A Thermo Scientific ID-X Orbitrap mass spectrometer was used in negative mode with an electrospray ionization source and the following conditions: ESI spray voltage 3300 V, sheath gas at 50 arbitrary units, auxiliary gas at 15 arbitrary units, sweep gas at 5 arbitrary units, ion transfer tube temperature at 300 °C and vaporizer temperature at 350 °C. Compounds were detected in a targeted mode and identified based on their exact mass with a mass window of 3 ppm and retention time compared to standards. Extracted ion chromatograms of the $[M-H]^-$

forms were integrated using Tracefinder software (Thermo Scientific). Absolute concentrations were calculated based on an external calibration curve.

Phlyogenetic analysis

PFL protein sequences were obtained by a BLAST search with EcPFL as input. Sequences not containing the Cys-Cys active site motif were excluded. Multiple sequence alignment and phylogenetic tree were generated using MEGA 6.[53]

Table S1. List of all site-directed mutagenesis primers used in this study

Primer name	Nucleotide sequence (5' to 3')
EcPFL_A273G_fw	GTCTCAGAACGGTGGTGCAATGTCCTTC
EcPFL_A273G_rv	GAAGGACATTGCACCACCGTTCTGAGAC
EcPFL_A273S_fw	GTTAAGTCTCAGAACGGTTCTGCAATGTCCTTCGGTC
EcPFL_A273S_rv	GACCGAAGGACATTGCAGAACCGTTCTGAGACTTAAC
EcPFL_A273T_fw	GTTAAGTCTCAGAACGGTACTGCAATGTCCTTCGGTC
EcPFL_A273T_rv	GACCGAAGGACATTGCAGTACCGTTCTGAGACTTAAC
EcPFL_A273V_fw	GTCTCAGAACGGTGTTGCAATGTCCTTCG
EcPFL_A273V_rv	CGAAGGACATTGCAACACCGTTCTGAGAC
Cb750_A267G_fw	GATAAAAACCCAGAATGGTGGTGCAATGTCAATTGGGCGG
Cb750_A267G_rv	CCGCCCAATTGACATTGCACCACCATTCTGGGTTTTTATC
Cb750_A267S_rv	GATAAAAACCCAGAATGGTTCTGCAATGTCAATTGGGC
Cb750_A267S_rv	GCCCAATTGACATTGCAGAACCATTCTGGGTTTTTATC
Cb743_A262G_fw	CAGCAAAACGGCGGTGCAATGTCGCTG
Cb743_A262G_rv	CAGCGACATTGCACCGCCGTTTTGCTG
Cb743_A262S_fw	GATTAAACAGCAAAACGGCTCTGCAATGTCGCTGGGAC
Cb743_A262S_rv	GTCCCAGCGACATTGCAGAGCCGTTTTGCTGTTTAATC
SmPFL_A267G_fw	GTGATTAATGGCGGTGCGACGTCGTTAG
SmPFL_A267G_rv	CTAACGACGTCGCACCGCCATTAATCAC
SmPFL_A267S_fw	GTGTGATTAATGGCTCTGCGACGTCGTTAG
SmPFL_A267S_rv	CTAACGACGTCGCAGAGCCATTAATCACAC
SmPFL_A267V_fw	GTGTGATTAATGGCGTTGCGACGTCGTTAG
SmPFL_A267V_rv	CTAACGACGTCGCAACGCCATTAATCACAC
TdcE_F328Y_fw	GCGTACACCGGAATATGATTCGCTGTTCTC
TdcE_F328Y_rv	GAGAACAGCGAATCATATTCCGGTGTACGC
EryACS_V379I_fw	CTGCTGGGCTCTATTGGTGAGCCGATC
EryACS_V379I_rv	GATCGGCTCACCAATAGAGCCCAGCAG

Table S3. Codon-Optimized gene sequences. Restriction sites are underlined.

CbAE (Locus tag DRB99_RS17320)

5'CATATGCTCTCCCTTGAACACGTAATGAAAGAATGCAACTTAGATAAAAACATTGCGCGAATCCATAGTATCGAGACGT
TTGGCTCGGTAGATGGACCCGGAGTTCGTTTCGTTACCTTCCTTAAAGGGTGTCACATGCGTTGCCAGTTTTGTCACAAT
CCAGACACATGGGACATCAATGGCGGGGAGACACGCACAGCGGATGAGCTTTTGAGCCAAGCCTTGCGGTACAAGACT
TATTGGAAAAAAGGCGGCGGCATAACAGTAAGTGGTGGCGAGCCGCTCCTCCAGATTGATTTTCTCATCGAATTTTTCAA
AAAGGCAAAAAGCAAGGGCGTCCATGTCACCTTAGACACCTCTGGTAACCCGTTCACTCGTGAAGAACCTTTCTTCAGC
AAGTTTAATGAACTCATGAAAGTTACCGATCTGGTTATGCTGGATATAAAACAGATTGATGAAGAGAAACATAAAATTCTG
ACTGGTTGGTCAAACAGCAACATCCTGGATATGGCTAAATTTTTATCGGAGATCAACAAGCCAGTGTGGATTCGTCACGT
CCTGGTGCCTGGTGGTAGCGACAATGACGAACAATTGATTAAGCTGGATGAATTTATCAAAACCCTGAAAAATGTGGATC
GCGTGGAAGTGCTGCCGTACCACTCCCTGGGCACGTTTAAATGGGAGGAACTGGGTATTGAATATCCGCTGAAAGATGT
GGAGCCGCCGACCAAAGAACGCATTGAAAATGCCAAAAAACTGCTGCATACCTCTGAGTACAATGGCTACCTTTCGCGC
TAAGGATCC-3'

SmAE (Locus tag DRB99_RS17320)

5'CATATGATTGAAAAAGTCGATTACGAAAAGGTGACGGGCCTGGTCAATAGCACTGAGTCGTTCGGTAGTGTAGATGGT
CCGGGCATCCGTTTTGTGGTCTTCATGCAAGGATGTCAGATGCGTTGCCAGTATTGTCATAACCCTGACACGTGGGCGA
TGAAAAATGATCGTGCGACTGAACGCACCGCCGGAGACGTCTTTAAAGAGGCGCTTCGTTTCAAAGACTTTTGGGGTGA
CACCGGCGGTATCACCGTAAGCGGTGGGGAGGCGACGCTCCAAATGGACTTTTTAATCGCACTCTTCTCATTAGCCAAA
GAAAAAGGCATTCACACCACCCTAGACACATGCGCACTAACCTTTCGCAACACCCCTAAGTATCTGGAGAAGTACGAAA
AACTGATGGCTGTTACTGATCTTGTACTGCTGGACATCAAAGAGATAAACCCGGATCAGCATAAGATAGTGACGGGGCA
TTCCAACAAGACTATTTTGGCGTGCGCCCGCTACCTGAGCGACATAGGCAAACCGGTCTGGATACGACACGTTTTGGTA
CCCGGCCTAACAGACCGCGACGAAGACCTCATCAAATTAGGCGAATATGTGAAAACCTTGAAAAATGTGCAACGATTCG
AAATCCTCCCATATCATACCATGGGCGAATTTAAGTGGAGAGAGTTGGGTATCCCTTACCCGTTAGAAGGCGTGAAACC
CCCAACTCCCGACCGCGTGCGAAATGCCAAGAAACTGATGCACACGGAAACCTATGAAGAATACAAGAAAAGAATTAAC
CATTAAGGATCC-3'

Cb750 (Locus tag DRB99_RS17315)

5'CATATGTTGGAAAACGAACAATGGGCAGGTTTCGAGGGGCGCATCTGGAAGGAAGAAATTAACGTTCGTGACTTCATT
CAGAACAACTATACCCCGTACGATGGCGACACCTCCTTTCTGGCAGACCCAACTGAGGCCACGAATAAACTTTGGTCCA
GGTTACAAGAACTGCAAAAAGAGGAGCGCGCTAAAGGTGGCGTGCTGGACATGGAAACAGAGGTAGTCTCCGGCTTAA
CCGCTTATGGACCAGGCTATATTGACGAATCACTGAAGGAGTTAGAGAAAGTCGTAGGGCTGCAAACCGACAAACCTCT
GAAACGCGCCTTTATGCCGTACGGTGGAATCAAAATGGCAGAACAATCATGCGAAAACTATGGGTATAAGCCAAACCCG
GAACTGCATAAAATATTTACAGAGTACCACAAGACGCACAACCAGGGGGTCTTTGACGCATACACACCGGAAATGCGGG
CCGCGAGATCGTCGCATATTATCACGGGGCTCCCGGATACCTATGGACGTGGGCGGATTGTGGGTGATTACCGTCGCG
TGGCTTTGTACGGTATAGATGGGCTGATTGAGGAGAAGTCCAAAGACTTGGCTAATTGTGGCGATGGTACCATGACCGA
CGAAGTCATACGCTTACGTGAAGAAATTACTGACCAAATCAAAGCCCTGAAAGATATGAAAAAAATGGCGGAGGGTTATG
GCTACGACATCTCTAAACCGGCAAAAGACGCGCGCGAAGCTATACAGTGGCTCTATTTTGGTTACCTGGCAGCGATAAA
AACCCAGAATGGTGCTGCAATGTCAATTGGGCGGGTTTCCACGTTTCTTGATATTTATATCCAACGCGACCTAGACAATG
GGGTGCTGACCGAGGATGAGGCCCAGGAATTGATTGATCATCTTACGATGAAGTGCAGGATGGTTAAATTTGCTCGTAT
TACTAGCTACAATGAACTGTTTAGTGGCGATCCCAGTTGGGTCACCATTGGATTGGGAGGAATCGGAGTCGATGGCCGC
CACATGGTGACCAAAAACGACTATCGATTTCTGCACACGTTGGAGAACATGGGCCCGTCCCCTGAACCGAACCTGACAG
TTTTTTATAGCAGCGCCCTGCCCGAGAATTTCAAGAAATATGCCGCTAAAATCTCGGTGGATACTTCCTCGATCCAGTAC
GAAAACGACGACGTCATGAAACCGGTTTGGGGCGATGACTATTCTATCTGTTGTTGTGTGTCGGCAACCCAAACTGGCA
AAGAAATGCAGTTTTTTGGTGCGCGAGCAAACTTAGCTAAATGCCTGCTGTACGCCATTAACGGCGGTATCGATGAAAA
GTCTGGGAAACAAGTCGGTCCAGAATACAAGGGCATTACGAGTGAGTATCTCGACTACGACGAAGTGATCGCGAAGTAT
GAGCAGATGTCGGACTGGCTGGCAGATCTGTATGTGAACATTCTTAATTTAATCCAGTATATGCACGACAAATACTACTA
TGAAGCGGCGGAAATGGCCCTAATCGATACCGAGGTACGTCGTACCTTTGCGACCGGCATTGCGGGTTTCAGTCATGTA
ATCGATAGTTTAAGCGCAATCAAATACGCGAAAGTGAAAACTATCAGGAATGAAGATGGTATCGTTACGGACTACGAAAT
CGAAGGCGATTTTCCGCGCTATGGTAACGATGATGATCGTGCGGATGAGATTGGTATTTACGTCTTAAAATCTTTCTTAG
ATAAGATTAAAAAACGTCATACGTATCGTAACAGCGAACCTACGACGAGTTTGCTCACCATTACGTCAAATGTTGTGTAC
GGTAAATACACCGGTAATATGCCCGATGGTCGTGAGGCATGGACTCCGTTGGCCCCAGGCGCGTCTCCATCATACGGC
GCGGAAAAAAATGGCCTTCTGGCGAGCCTTAATTCTGTTGCGAAATTGCCGTATGAATGGGCCCTGGATGGCATTAGCA
ATACACAGACCATTAACCCTGATGCGCTGGGGCACGGCGAGGAGGAACAGGTTGATAAATTAGTGCAGGTTCTGGATG
GCTATTTTGATCAAGGCCCGCACCATTTGAATGTAAACGTATTCGGCATTGAAAAACTGAAAGATGCGATGGAACATCCG
GAAAAAGAAGAATATGCCAATTTCACAATCCGCGTCAGCGGATATGCCGTAAAATTCATTGATCTTACGCGCGAGCAGCA
ACTGGACGTGATTAGCCGTACCTGCCATGAACGCATGTAAGGATCC-3'

Cb743 (Locus tag DRB99_RS03980)

5'CATATGTTTAAACAGTGGGATGGTTTTAAGTCGGGGCAGTGGCAGAAAAGCGTTGATGTACGCAACTTTATCCAGCTG
AACTATACTCCCTACGAAGGGGACGACTCCTTCTTGGCCGGGGCCTCGGAGAACACCACCAAATTATGGGACGAGGTA
TCAGACCTGTTTAAAAAAGAACGTGACAACGGTGGTGTTCTTGATGTTGACACTGATACGGCGTCCCGCATCAATGCTTA
CGCACCGGGCTATATTGATAAAGCCATAGAGAAGATCGTGGGCGTGCAAACCGACGCGCCGCTGAAAAGAGCCGTCAT
GGCTGAGGGGGGAATACGTATGGCTGAAAACGCGGCGAAAGCGTATGGCTATGAGGTTAATCCGAAAATTTCAGAAATC
TTTACAAAATACCGTAAAACACATAACCAGGGAGTCTTTGATGCCTATACTGACGAGATGCGCCTCGCACGAAAATCGGG
GATCGTGACCGGGTTACCTGACGCATACGGCCGTGGCCGCATTATAGGCGATTACCGACGTGTGGCTCTTTATGGTGTA
GACAAACTGATTGAGGACAAAATCGCGCAGAAAAAAACCTTGGAGGTAACTACAATCGACGAGGAAGTAATCCGACTTC
GCGAGGAACTGTCTGATCAGATTGTAGCCTTAAAAGAGCTTAAAGAGATGGCGCTGTCCTATGGTATAGATATTAGCGCA
CCGGCGACCAATGCACAAGAAGCGGTGCAGTGGCTGTATTTCGGGTACTTAGCTGCGATTAAACAGCAAAACGGCGCT
GCAATGTCGCTGGGACGTACAAGCACTTTCCTGGATATTTACATCGAACGCGACCTTCAAAATGGGGTCATCACGGAGG
AAGAAGCGCAGGAGATCATCGATCACTTCGTTATGAAATTGCGACTGGTGAAATTCCTGCGTACACCGGAATACAATGA
CCTGTTTAGCGGAGATCCTACCTGGGTCACTGAGTCTATTGCCGGAATGGGCCTTGATGGTCGCACGCTGGTTACCAAA
AACTCCTTCCGTATCCTGAACACGCTGTATACGATAGGTCCTTCGCCGGAACCTAACCTGACCGTACTATGGTCAACGC
GTCTGCCCCAGGGCTTTAAAGACTTTTGCAGCAAAGTTTCTATTGATACCAGTTCCGTCCAATATGAAAATGACGACCTA
ATGTTAGAATACTGGGGAGATGACTACGCCATCGCCTGTTGCGTCAGTGCGATGAAAGTTGGCAAACAGATGCAATTTTT
TGGTGCGCGTGTCAACTTAGCTAAAACCCTACTATATACGATTAATGGAGGCAAGGACGAAAAGTACGGTATGCAGGTC
GGACCGAAAATGGAACCGATAACCAGTGAATATCTGGATTACAACGAAGTGATGGAAAAGTTTGAAGTCATGACGGACT
GGCTGGCGAATCTGTACGTTAATACGCTGAACTGTATCCATTACATGCACGATAAATATAGTTACGAATCTCTTCAGATG
GCGCTCCATGATCGGGACGTATTTCGTACCATGGCCTGTGGTATTGCCGGGTTATCAGTTTGTGCTGATAGCCTGTCTG
CGATCAAATATGCGAAGGTCAAGCCGATTCGCAACGAAGAAGGCATTGCAGTTGATTTCGAAGTAGAAGGTGATTTCCC
TAAGTACGGCAATGATGATGATCGTGTTGATGATATTGCAGTTTTCTTGGTGGAAAACATGATGAATAAAATCCGTAAAAA
CAAGACCTATCGGAATGCGTACCATACGCAGTCAGTGCTCACTATTACATCGAACGTGGTGTATGGTAAGAAAACAGGC
ACGACGCCATGCGGTCGGAAGGCCGGTGAACCGTTTGCTCCGGGCGCGAACCCAATGCACGGCCGCGATAACAGTGG
TAGCCTTGCGAGTCTGAACAGTGTGGCCAAGTTACCTTACGAACACAGTCAGGATGGCATCTCTAATACCTTTAGCATTG
TGCCGGATGCACTGGGCAAAACCAAAGAAGATCAGATTACCAACCTCTCAGCTATGATGGATGGCTACTTCGGCCAAAA
GGCGCACCACCTCAATGTTAACGTGTTTAATCGTGCTACCCTCTTAGATGCAATGGATCATCCTGAAGAATATCCCCAGC

TCACCATTCGCGTGTCAGGCTACGCCGTCAACTTTATTAAATTGACTCGGGAACAACAGCTGGATGTGATTAATCGCACG
TTTCACGAAAAAATGTAAGGATCC-3‘

SmPFL (Locus tag SMU_RS01955)

5‘CATATGGCAACCGTTAAAACTAATACCGATGTTTTCGAGAAAGCGTGGGAAGGATTCAAAGGCACGGATTGGAAGGAT
CGGGCGAGCATTTCTCGATTCGTCCAAGATAATTATACGCCGTACGATGGAGATGAGTCGTTTCTAGCTGGCCCAACCG
AAAGAAGCCTACATATTAAGAAGGTCGTCGAAGAAACCAAGGCCCACTACGAGGAGACGCGATTTCCTATGGATACTCG
TATCACCTCTATTGCTGACATTCCCGCAGGATATATCGATAAAGAGAATGAACTTATATTCGGTATCCAAAATGACGAGCT
GTTTAAGCTCAACTTTATGCCAAAAGGGGGGATTCGGATGGCCGAAACCGCGCTGAAAGAGCATGGCTACGAGCCCGA
CCCCGCGGTGCATGAGATCTTTACCAAATATGCTACTACTGTAAACGATGGGATCTTCCGGGCCTATACGTCGAACATTC
GGCGTGCACGTCATGCGCATACCGTTACGGGCTTACCAGATGCTTATTCGCGTGGGCGTATAATAGGTGTATATGCCCG
TTTGGCACTCTATGGGGCGGACTATTTAATGCAGGAAAAAGTTAATGACTGGAACTCCATTGCTGAGATTGACGAGGAAT
CTATTCGCTTACGTGAAGAAATCAACCTTCAATATCAGGCGCTTGGGGAAGTGGTCCGCTTGGGAGACCTTTACGGCCT
CGACGTTCGTAAACCTGCGATGAATGTTAAAGAAGCCATCCAGTGGATTAACATCGCGTTCATGGCAGTCTGCCGTGTG
ATTAATGGCGCTGCGACGTCGTTAGGTCGTGTACCAATCGTGTTGGACATTTTTGCAGAACGTGACCTTGCCCGTGGGA
CTTTTACGGAGTCGGAAATTCAGGAGTTTGTGGATGATTTTGTTATGAAACTTCGCACCGTTAAATTTGCACGTACCAAG
GCTTACGACGAGCTGTACTCCGGAGATCCTACTTTCATTACCACCTCCATGGCGGGAATGGGTGCCGATGGACGCCATC
GCGTAACAAAAATGGACTACCGCTTTTTGAACACGTTAGACAACATAGGTAACGCACCAGAGCCTAACCTGACCGTCCT
CTGGTCTTCCAAATTGCCGTATCCGTTTCGCCACTATTGTATGAGCATGAGTCATAAGCATAGTTCTATTCAGTACGAAG
GTGTGACAACCATGGCCAAAGAAGGCTACGGGGAAATGAGCTGCATCTCCTGCTGTGTGAGCCCGTTGGACCCGGAAA
ATGAAGACCGTCGTCATAATCTTCAATACTTTGGCGCCCGTGTCAATGTACTCAAGGCGTTATTGACCGGTCTGAACGGT
GGCTATGATGACGTACACAAAGACTATAAAGTGTTTGATGTAGAACCGATTCGCGATGAAGTGCTGGACTTCGAAACCGT
GAAGGCCAATTTCGAAAAAGCGCTGGATTGGCTGACCGACACATACGTCGATGCAATGAATATTATCCACTATATGACG
GACAAATATAACTACGAAGCAGTTCAGATGGCCTTCTTACCGACACGCGTGAAAGCGAATATGGGATTTGGCATCTGCG
GCTTTAGCAATACAGTTGATTCTTTAAGCGCGATTAAGTACGCGACTGTCAAACCGATCCGCGATGAAGATGGCTATATT
TACGACTATGAAACGGTTGGCAATTTCCCGCGCTACGGCGAAGACGATGACCGCGTTGATAGTATCGCGGAGTGGCTG
CTGGAAGCCTTTCACACCCGCCTGGCACGCCATAAACTGTACAAAGATTCAGAAGCCACTGTGAGTCTGCTGACGATCA
CCAGCAACGTGGCGTATTCAAAACAAACAGGCAATAGCCCGGTCCATAAAGGGGTGTATCTTAACGAAGATGGTAGCGT
TAATCTGTCAAAAGTTGAGTTTTTCAGCCCGGGTGCTAACCCGTCGAATAAGGCATCAGGCGGTTGGCTACAAAACCTG
AATAGCCTGAAAAAGCTGGATTTCGCACATGCGAATGATGGCATTAGTCTGACAACCCAGGTCTCCCCGAAAGCGCTGG
GCAAAACCTTTGATGAGCAGGTAGCGAACCTGGTCACGATTCTGGATGGTTATTTTGAAGGTGGCGGCCAGCACGTGAA
TCTGAACGTAATGGATCTGAAAGATGTGTATGATAAAATCATGAACGGTGAAGATGTGATCGTACGCATCTCGGGCTACT
GTGTCAACACCAAATACCTGACGAAAGAGCAAAAAACGGAACTGACCCAGCGCGTATTCCACGAAGTGCTGAGCATGGA
TGATGCCGCCACTGATCTGGTGAACAACAAATAAGGATCC-3‘

4.5 References

[1] Olah, G. A. (2013) Towards oil independence through renewable methanol chemistry, *Angew. Chem. Int. Ed. Engl. 52*, 104-107.
[2] Pfeifenschneider, J., Brautaset, T., and Wendisch, V. F. (2017) Methanol as carbon substrate in the bio-economy: Metabolic engineering of aerobic methylotrophic bacteria for production of value-added chemicals, *Biofuel. Bioprod. Biorefin. 11*, 719-731.
[3] Yishai, O., Lindner, S. N., Gonzalez de la Cruz, J., Tenenboim, H., and Bar-Even, A. (2016) The formate bio-economy, *Curr. Opin. Chem. Biol. 35*, 1-9.
[4] Cotton, C. A., Claassens, N. J., Benito-Vaquerizo, S., and Bar-Even, A. (2019) Renewable methanol and formate as microbial feedstocks, *Curr. Opin. Biotechnol. 62*, 168-180.
[5] Jouny, M., Luc, W., and Jiao, F. (2018) General Techno-Economic Analysis of CO2 Electrolysis Systems, *Ind. Eng. Chem. Res. 57*, 2165-2177.
[6] Bertsch, J., and Muller, V. (2015) Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria, *Biotechnol. Biofuels. 8*, 210.
[7] Meyer, F., Keller, P., Hartl, J., Groninger, O. G., Kiefer, P., and Vorholt, J. A. (2018) Methanol-essential growth of Escherichia coli, *Nat. Commun. 9*, 1508.
[8] Chen, C. T., Chen, F. Y., Bogorad, I. W., Wu, T. Y., Zhang, R., Lee, A. S., and Liao, J. C. (2018) Synthetic methanol auxotrophy of Escherichia coli for methanol-dependent growth and production, *Metab. Eng. 49*, 257-266.
[9] He, H., Edlich-Muth, C., Lindner, S. N., and Bar-Even, A. (2018) Ribulose Monophosphate Shunt Provides Nearly All Biomass and Energy Required for Growth of E. coli, *ACS Synth. Biol. 7*, 1601-1611.
[10] Chen, F. Y. H., Jung, H. W., Tsuei, C. Y., and Liao, J. C. (2020) Converting Escherichia coli to a Synthetic Methylotroph Growing Solely on Methanol, *Cell 182*, 933-+.
[11] Dai, Z., Gu, H., Zhang, S., Xin, F., Zhang, W., Dong, W., Ma, J., Jia, H., and Jiang, M. (2017) Metabolic construction strategies for direct methanol utilization in Saccharomyces cerevisiae, *Bioresour. Technol. 245*, 1407-1412.
[12] Yu, H., and Liao, J. C. (2018) A modified serine cycle in Escherichia coli coverts methanol and CO2 to two-carbon compounds, *Nat. Commun. 9*, 3992.
[13] Cotton, C. A. R., Claassens, N. J., Benito-Vaquerizo, S., and Bar-Even, A. (2020) Renewable methanol and formate as microbial feedstocks, *Curr Opin Biotech 62*, 168-180.
[14] Siegel, J. B., Smith, A. L., Poust, S., Wargacki, A. J., Bar-Even, A., Louw, C., Shen, B. W., Eiben, C. B., Tran, H. M., Noor, E., Gallaher, J. L., Bale, J., Yoshikuni, Y., Gelb, M. H., Keasling, J. D., Stoddard, B. L., Lidstrom, M. E., and Baker, D. (2015) Computational protein design enables a novel one-carbon assimilation pathway, *Proc. Natl. Acad. Sci. U.S.A. 112*, 3704-3709.
[15] Yang, X., Yuan, Q., Luo, H., Li, F., Mao, Y., Zhao, X., Du, J., Li, P., Ju, X., Zheng, Y., Chen, Y., Liu, Y., Jiang, H., Yao, Y., Ma, H., and Ma, Y. (2019) Systematic design and in vitro validation of novel one-carbon assimilation pathways, *Metab. Eng. 56*, 142-153.
[16] Wang, C., Ren, J., Zhou, L., Li, Z., Chen, L., and Zeng, A. P. (2019) An Aldolase-Catalyzed New Metabolic Pathway for the Assimilation of Formaldehyde and Methanol To Synthesize 2-Keto-4-hydroxybutyrate and 1,3-Propanediol in Escherichia coli, *ACS Synth. Biol. 8*, 2483-2493.
[17] Lu, X., Liu, Y., Yang, Y., Wang, S., Wang, Q., Wang, X., Yan, Z., Cheng, J., Liu, C., Yang, X., Luo, H., Yang, S., Gou, J., Ye, L., Lu, L., Zhang, Z., Guo, Y., Nie, Y., Lin, J., Li, S., Tian, C., Cai, T., Zhuo, B., Ma, H., Wang, W., Ma, Y., Liu, Y., Li, Y., and Jiang, H. (2019) Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design, *Nat. Commun. 10*, 1378.
[18] Chou, A., Clomburg, J. M., Qian, S., and Gonzalez, R. (2019) 2-Hydroxyacyl-CoA lyase catalyzes acyloin condensation for one-carbon bioconversion, *Nat. Chem. Biol. 15*, 900-906.
[19] He, H., Hoper, R., Dodenhoft, M., Marliere, P., and Bar-Even, A. (2020) An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in E. coli, *Metab. Eng.*
[20] Patterson, J. A., He, H., Folz, J. S., Li, Q., Wilson, M. A., Fiehn, O., Bruner, S. D., Bar-Even, A., and Hanson, A. D. (2020) Thioproline formation as a driver of formaldehyde toxicity in Escherichia coli, *Biochemical Journal 477*, 1745-1757.
[21] Bar-Even, A. (2016) Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways, *Biochemistry 55*, 3851-3863.
[22] Figueroa, I. A., Barnum, T. P., Somasekhar, P. Y., Carlstrom, C. I., Engelbrektson, A. L., and Coates, J. D. (2018) Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO2 fixation pathway, *Proc. Natl. Acad. Sci. U.S.A. 115*, E92-E101.

[23] Bar-Even, A., Noor, E., Flamholz, A., and Milo, R. (2013) Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes, *Biochim. Biophys. Acta 1827*, 1039-1047.
[24] Yishai, O., Bouzon, M., Doring, V., and Bar-Even, A. (2018) In Vivo Assimilation of One-Carbon via a Synthetic Reductive Glycine Pathway in Escherichia coli, *ACS Synth. Biol. 7*, 2023-2028.
[25] Bang, J., and Lee, S. Y. (2018) Assimilation of formic acid and CO2 by engineered Escherichia coli equipped with reconstructed one-carbon assimilation pathways, *Proc. Natl. Acad. Sci. U.S.A. 115*, E9271-E9279.
[26] Claassens, N. J., Bordanaba-Florit, G., Cotton, C. A. R., De Maria, A., Finger-Bou, M., Friedeheim, L., Giner-Laguarda, N., Munar-Palmer, M., Newell, W., Scarinci, G., Verbunt, J., de Vries, S. T., Yilmaz, S., and Bar-Even, A. (2020) Replacing the Calvin cycle with the reductive glycine pathway in Cupriavidus necator, *Metab Eng 62*, 30-41.
[27] Kim, S., Lindner, S. N., Aslan, S., Yishai, O., Wenk, S., Schann, K., and Bar-Even, A. (2020) Growth of E. coli on formate and methanol via the reductive glycine pathway, *Nat. Chem. Biol.*
[28] Thauer, R. K., Kirchniawy, F. H., and Jungermann, K. A. (1972) Properties and function of the pyruvate-formate-lyase reaction in clostridiae, *Eur. J. Biochem. 27*, 282-290.
[29] Knappe, J., Blaschkowski, H. P., Grobner, P., and Schmitt, T. (1974) Pyruvate formate-lyase of Escherichia coli: the acetyl-enzyme intermediate, *Eur. J. Biochem. 50*, 253-263.
[30] Himo, F., and Eriksson, L. A. (1998) Catalytic mechanism of pyruvate formate-lyase (PFL). A theoretical study, *J. Am. Chem. Soc. 120*, 11449-11455.
[31] Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012) eQuilibrator-the biochemical thermodynamics calculator, *Nucleic Acids Res 40*, D770-D775.
[32] Zelcbuch, L., Lindner, S. N., Zegman, Y., Vainberg Slutskin, I., Antonovsky, N., Gleizer, S., Milo, R., and Bar-Even, A. (2016) Pyruvate Formate-Lyase Enables Efficient Growth of Escherichia coli on Acetate and Formate, *Biochemistry 55*, 2423-2426.
[33] Cotton, C. A., Bernhardsgrutter, I., He, H., Burgener, S., Schulz, L., Paczia, N., Dronsella, B., Erban, A., Toman, S., Dempfle, M., De Maria, A., Kopka, J., Lindner, S. N., Erb, T. J., and Bar-Even, A. (2020) Underground isoleucine biosynthesis pathways in E. coli, *Elife 9*.
[34] Hesslinger, C., Fairhurst, S. A., and Sawers, G. (1998) Novel keto acid formate-lyase and propionate kinase enzymes are components of an anaerobic pathway in Escherichia coli that degrades L-threonine to propionate, *Mol Microbiol 27*, 477-492.
[35] Nunez, M. F., Pellicer, M. T., Badia, J., Aguilar, J., and Baldoma, L. (2001) Biochemical characterization of the 2-ketoacid reductases encoded by ycdW and yiaE genes in Escherichia coli, *Biochem. J. 354*, 707-715.
[36] Scheffen, M., Marchal, D. G., Beneyton, T., Schuller, S. K., Klose, M., Diehl, C., Lehmann, J., Pfister, P., Carrillo, M., He, H., Aslan, S., Cortina, N. S., Claus, P., Bollschweiler, D., Baret, J.-C., Schuller, J. M., Zarzycki, J., Bar-Even, A., and Erb, T. J. (2020) A new-to-nature carboxylation module to improve natural and synthetic CO2 fixation, *Nat Catal*, accepted.
[37] de Diego Puente, T., Gallego-Jara, J., Castano-Cerezo, S., Bernal Sanchez, V., Fernandez Espin, V., Garcia de la Torre, J., Manjon Rubio, A., and Canovas Diaz, M. (2015) The Protein Acetyltransferase PatZ from Escherichia coli Is Regulated by Autoacetylation-induced Oligomerization, *The Journal of biological chemistry 290*, 23077-23093.
[38] Siegel, J. B., Smith, A. L., Poust, S., Wargacki, A. J., Bar-Even, A., Louw, C., Shen, B. W., Eiben, C. B., Tran, H. M., Noor, E., Gallaher, J. L., Bale, J., Yoshikuni, Y., Gelb, M. H., Keasling, J. D., Stoddard, B. L., Lidstrom, M. E., and Baker, D. (2015) Computational protein design enables a novel one-carbon assimilation pathway, *Proceedings of the National Academy of Sciences of the United States of America 112*, 3704-3709.
[39] Knappe, J., Blaschkowski, H. P., Grobner, P., and Schmitt, T. (1974) Pyruvate formate-lyase of Escherichia coli: the acetyl-enzyme intermediate, *European journal of biochemistry 50*, 253-263.
[40] Takahashi, S., Abbe, K., and Yamada, T. (1982) Purification of pyruvate formate-lyase from Streptococcus mutans and its regulatory properties, *Journal of bacteriology 149*, 1034-1040.
[41] Hemschemeier, A., Jacobs, J., and Happe, T. (2008) Biochemical and physiological characterization of the pyruvate formate-lyase Pfl1 of Chlamydomonas reinhardtii, a typically bacterial enzyme in a eukaryotic alga, *Eukaryot Cell 7*, 518-526.
[42] Becker, A., Fritz-Wolf, K., Kabsch, W., Knappe, J., Schultz, S., and Volker Wagner, A. F. (1999) Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase, *Nat. Struct. Biol. 6*, 969-975.
[43] Broderick, J. B., Duderstadt, R. E., Fernandez, D. C., Wojtuszewski, K., Henshaw, T. F., and Johnson, M. K. (1997) Pyruvate formate-lyase activating enzyme is an iron-sulfur protein, *J Am Chem Soc 119*, 7396-7397.

[44] Igamberdiev, A. U., Bykova, N. V., and Kleczkowski, L. A. (1999) Origins and metabolism of formate in higher plants, *Plant Physiol Bioch 37*, 503-513.
[45] Chou, A., Clomburg, J. M., Qian, S., and Gonzalez, R. (2019) 2-Hydroxyacyl-CoA lyase catalyzes acyloin condensation for one-carbon bioconversion, *Nature chemical biology 15*, 900-906.
[46] Burgener, S., Cortina, N. S., and Erb, T. J. (2020) Oxalyl-CoA Decarboxylase Enables Nucleophilic One-Carbon Extension of Aldehydes to Chiral alpha-Hydroxy Acids, *Angew Chem Int Ed Engl 59*, 5526-5530.
[47] Peter, D. M., Vogeli, B., Cortina, N. S., and Erb, T. J. (2016) A Chemo-Enzymatic Road Map to the Synthesis of CoA Esters, *Molecules 21*.
[48] Trudeau, D. L., Edlich-Muth, C., Zarzycki, J., Scheffen, M., Goldsmith, M., Khersonsky, O., Avizemer, Z., Fleishman, S. J., Cotton, C. A. R., Erb, T. J., Tawfik, D. S., and Bar-Even, A. (2018) Design and in vitro realization of carbon-conserving photorespiration, *Proceedings of the National Academy of Sciences of the United States of America 115*, E11455-E11464.
[49] Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005) Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research, *DNA research : an international journal for rapid publication of reports on genes and genomes 12*, 291-299.
[50] Byer, A. S., McDaniel, E. C., Impano, S., Broderick, W. E., and Broderick, J. B. (2018) Mechanistic Studies of Radical SAM Enzymes: Pyruvate Formate-Lyase Activating Enzyme and Lysine 2,3-Aminomutase Case Studies, *Methods Enzymol 606*, 269-318.
[51] Unkrig, V., Neugebauer, F. A., and Knappe, J. (1989) The free radical of pyruvate formate-lyase. Characterization by EPR spectroscopy and involvement in catalysis as studied with the substrate-analogue hypophosphite, *Eur J Biochem 184*, 723-728.
[52] Schwander, T., Schada von Borzyskowski, L., Burgener, S., Cortina, N. S., and Erb, T. J. (2016) A synthetic pathway for the fixation of carbon dioxide in vitro, *Science 354*, 900-904.
[53] Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0, *Mol Biol Evol 30*, 2725-2729.

4.6 Supplementary Figures

Figure S1. Active site homology model for acetyl-CoA synthetase 1 of *Erythrobacter* sp. NAP1 (EryACS) based on the crystal structure of an acetyl-CoA synthetase from *Cryptococcus neoformans* (PDB ID 5K8F). The acetyl-AMP is shown in yellow. Distances are given in Å. Replacing Thr304 or Val379 with bulkier amino acids closes the binding pocket for the acetyl-moiety, and presumably enhances activity with formate.

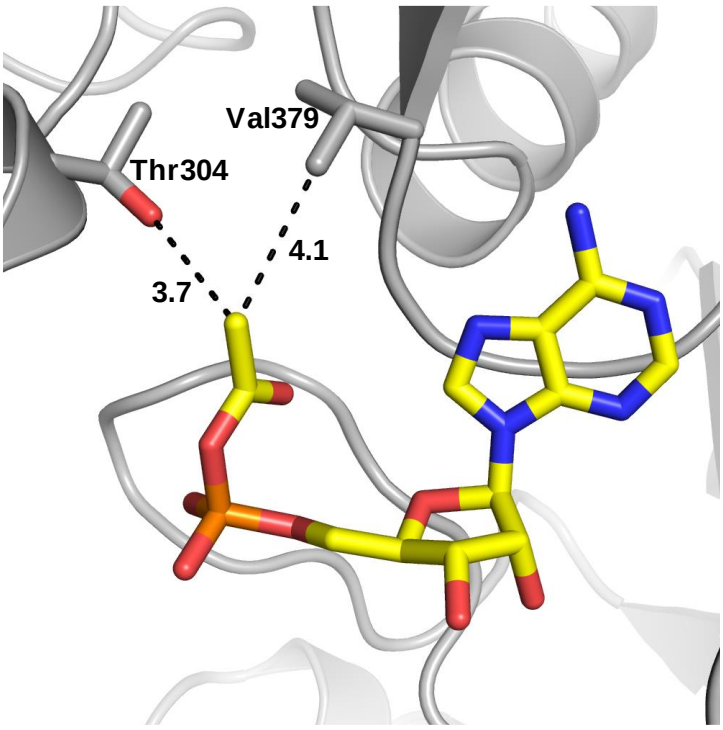


Figure S2. Michaelis-Menten graphs of EcACS, EryACS and mutants thereof with formate and acetate, respectively, as substrate. Error bars show standard deviation of three replicates.

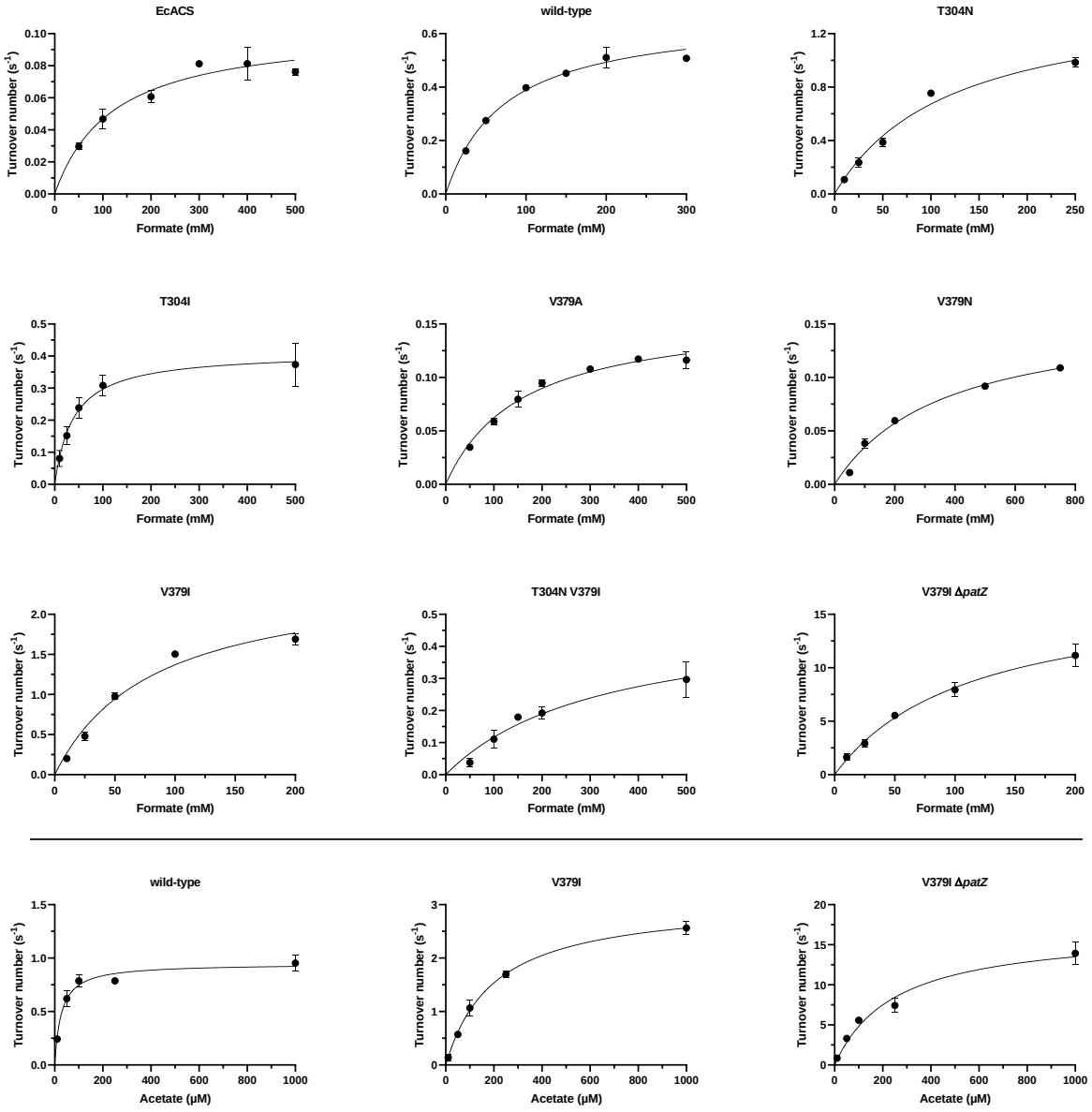


Figure S3. Multiple sequence alignment of PFLs. Relevant active site residues are labelled (numbering corresponds to EcPFL). Cys-Cys motif in yellow, the radical harboring glycine in red and the four non-conserved residues in orange. Note the substitutions, TdcE A273G and Y323F; SmPFL F327Y and W333F.

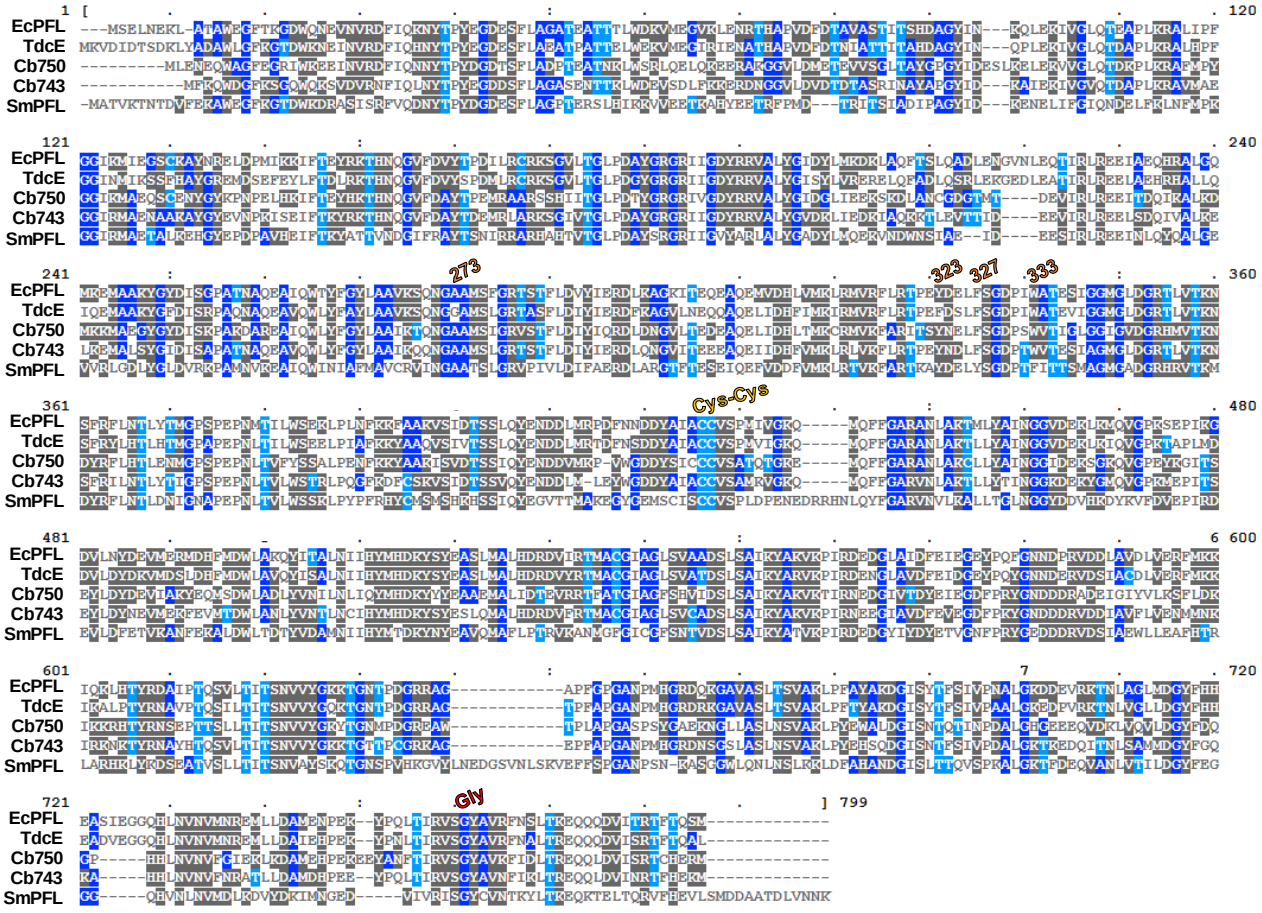


Figure S4. Michaelis-Menten graphs of PFLs with pyruvate, CoA, acetyl-CoA amd formate. Error bars show standard deviation of three replicates.

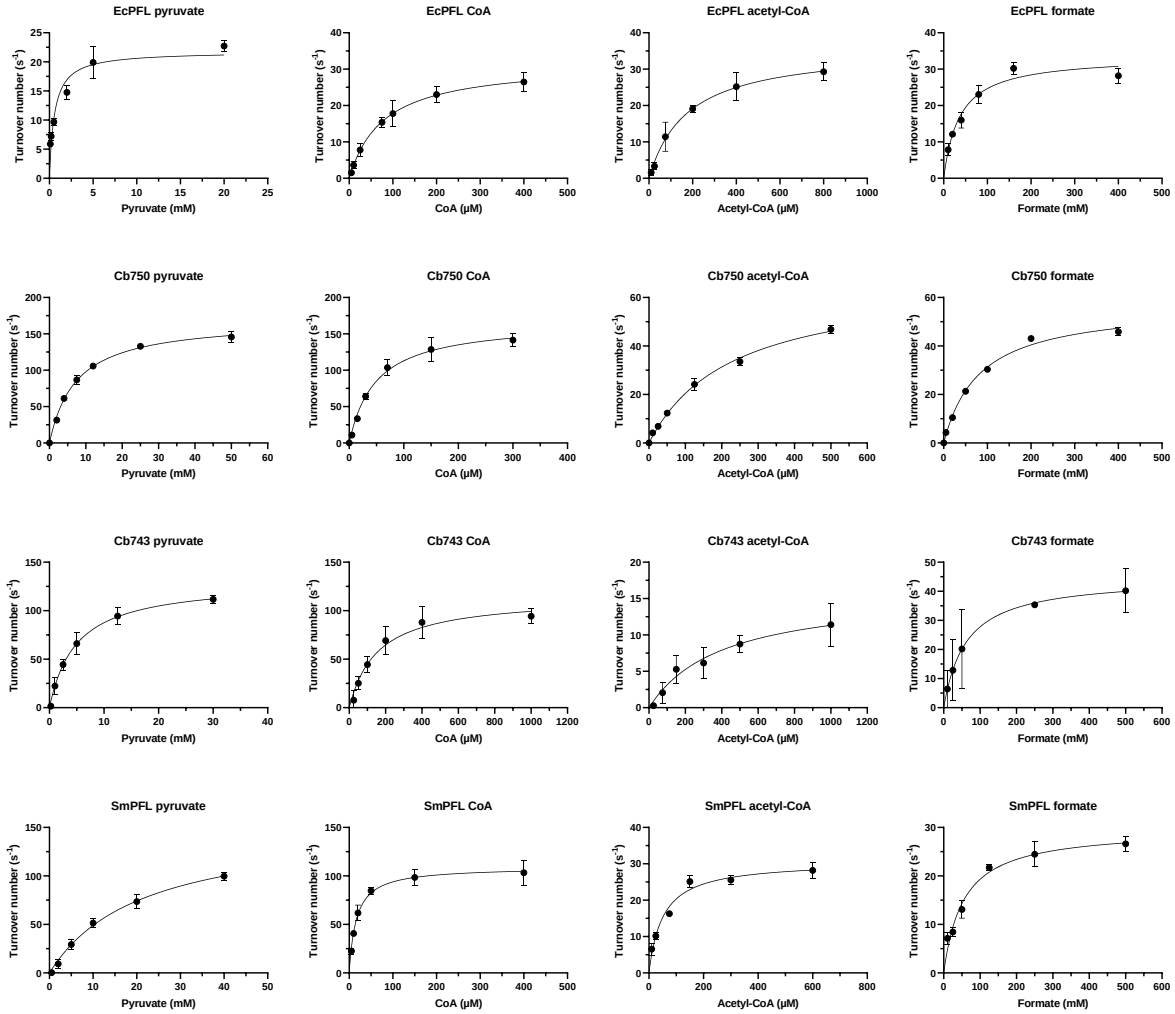
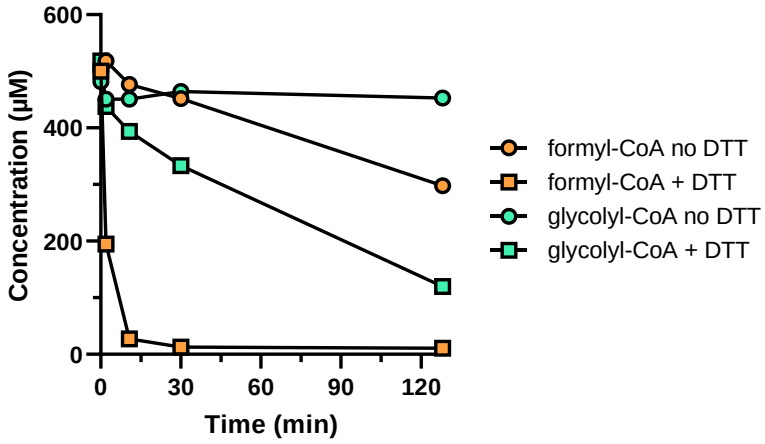


Figure S5. Stability of formyl-CoA and glycolyl-CoA in the presence of 0.5 mM dithiothreitol (DTT).



DISCUSSION

Even diamonds start as coal.

- Incubus

5 General discussion

The central aim of this thesis was to expand the repertoire of enzymatic C1 carboligation reactions, in particular C-C bond formations with a C1 nucleophile or radical. As lined out in the INTRODUCTION these reactivities are rarely employed by enzymes and filling in this gap could give rise to new biocatalytic transformations and synthetic carbon fixation pathways. Ultimately, novel C1 fixation enzymes could contribute to a sustainable, circular bioeconomy. The results presented in this thesis add 4 new examples of enzymatic C-C bond formation. These are briefly discussed below, with a special focus on their potential contribution to a sustainable bioeconomy.

1. **C1 elongation of aldehydes** $C_n + C_1 = C_{n+1}$

   CHAPTER I describes the repurposing and engineering of oxalyl-CoA decarboxylase (OXC) as a carboligase. OXC, with the help of the cofactor thiamine diphosphate (ThDP), usually breaks a C-C bond by decarboxylating oxalyl-CoA.[1] During catalysis, a nucleophilic carbanion/enamine intermediate forms that is protonated and released as formyl-CoA.[2] Here, it was shown that this intermediate can also undergo C-C bond formation with an aldehyde, producing a 2-hydroxyacyl-CoA ester. A single point mutation in the active site shifted the competing reactions of protonation and carboligation in favor of the latter by a factor of 400. This variant in combination with an oxalyl-CoA ligase and a thioesterase enabled the C1 elongation of aldehydes, giving rise to enantiopure aromatic α-hydroxy acids.

   Clearly, this example is not directly relevant to the goal of lowering atmospheric CO_2 because the C1 unit comes from oxalate and the reaction actually releases CO_2. It nevertheless is a valuable addition to the C-C bond formation toolbox, as it enables the synthesis of important chiral building blocks under mild conditions with cheap starting materials. Moreover, unlike most other mandelic acid syntheses, OXC does not rely on toxic cyanide. By adapting this method to a whole-cell biocatalyst (with glucose providing ATP), this could become a sustainable and safe option for the asymmetric synthesis of α-hydroxy acids.

2. **Carboxylation of formyl-CoA** $C_1 + C_1 = C_2$

   CHAPTER II demonstrates that OXC can catalyze the reverse reaction, that is, the carboxylation of formyl-CoA. The carboxylation of an activated C1 compound resembles the reactions catalyzed by acetyl-CoA synthase/CO dehydrogenase (ACS/CODH)[3] and the glycine cleavage system (GCS),[4] thus, reverse OXC provides a promising CO_2-utilizing C1-C1 condensation reaction. However, the low turnover number (~0.5 min^{-1}) and the

requirement for high concentrations of CO_2 and formyl-CoA, as well as two NADPH-consuming coupling enzymes make this a rather academic example. Still, if the CO_2 affinity could be improved and the carboxylation reaction could directly be coupled to formyl-CoA production, this could become a structurally simple alternative to the complex enzymes ACS/CODH and GCS.

Our group has recently elucidated the principles of CO_2 binding in CCR.[5] This knowledge enabled rational engineering of a related reductase into a carboxylase.[6] It would be interesting to see if these principles could also serve as a guideline to build a CO_2-binding pocket into the unrelated OXC. Not only would this increase OXC's carboxylation activity, it would also show that the principles of CO_2 binding are broadly applicable to create novel carboxylases.

3. **Condensation of formyl-CoA and formaldehyde** $C_1 + C_1 = C_2$

CHAPTER II describes the OXC-catalyzed C1-C1 condensation of formyl-CoA and formaldehyde to glycolyl-CoA. Directed evolution improved the catalytic efficiency by a factor of ~200, resulting in an artificial formaldehyde-fixing enzyme (MeOXC4) that is comparable to natural ones. The high productivity in a whole-cell context indicates that MeOXC4 is a promising enzyme for synthetic C1 fixation pathways.

Notably, MeOXC4 also extends glycolaldehyde into glyceryl-CoA. As previously reported,[7] the CoA-acylating aldehyde dehydrogenase PduP reduces glycolyl-CoA to glycolaldehyde and preliminary results from this work showed that formyl-CoA and glyceryl-CoA are also reduced to formaldehyde and glyceraldehyde, respectively. The combination of MeOXC4 and PduP could therefore enable an iterative pathway to synthesize a three-carbon compound from three formaldehyde molecules, the iterative OXC/PduP pathway (iOP; **Figure 1**). Addition of formyl-CoA synthetase (FCS; see CHAPTER III) could further extend the pathway to utilize formate as C1 source. Overall, this pathway is very similar to the formolase pathway, which compares favorably with the nine naturally occurring C1 fixation pathways, in particular with respect to the number of reaction steps and chemical driving force.[8] The only differences between the iOP and the formolase pathway are the final product and the use of C1 units; the latter uses three formaldehydes to synthesize dihydroxyacetone, whereas the iOP converts two formyl-CoA and one formaldehyde to glyceraldehyde (**Figure 1**). Glyceraldehyde can be phosphorylated and enter glycolysis as D-glyceraldehyde 3-phosphate. The stereochemical outcome of the MeOXC4-catalyzed condensation is currently

not known, however, kinases for both stereoisomers of glyceraldehyde have been described.[9, 10]

The lower dependence of the iOP on toxic formaldehyde could prove advantageous *in vivo*. In addition, MeOXC4 exhibits an 80-fold higher k_{cat}/K_M with formaldehyde than formolase. In conclusion, the successful engineering of MeOXC4 could pave the way for a new-to-nature C1 + C1 + C1 = C3 pathway, and eventually synthetic formatotrophic growth, which has not yet been achieved with the formolase pathway, despite extensive efforts.[8]

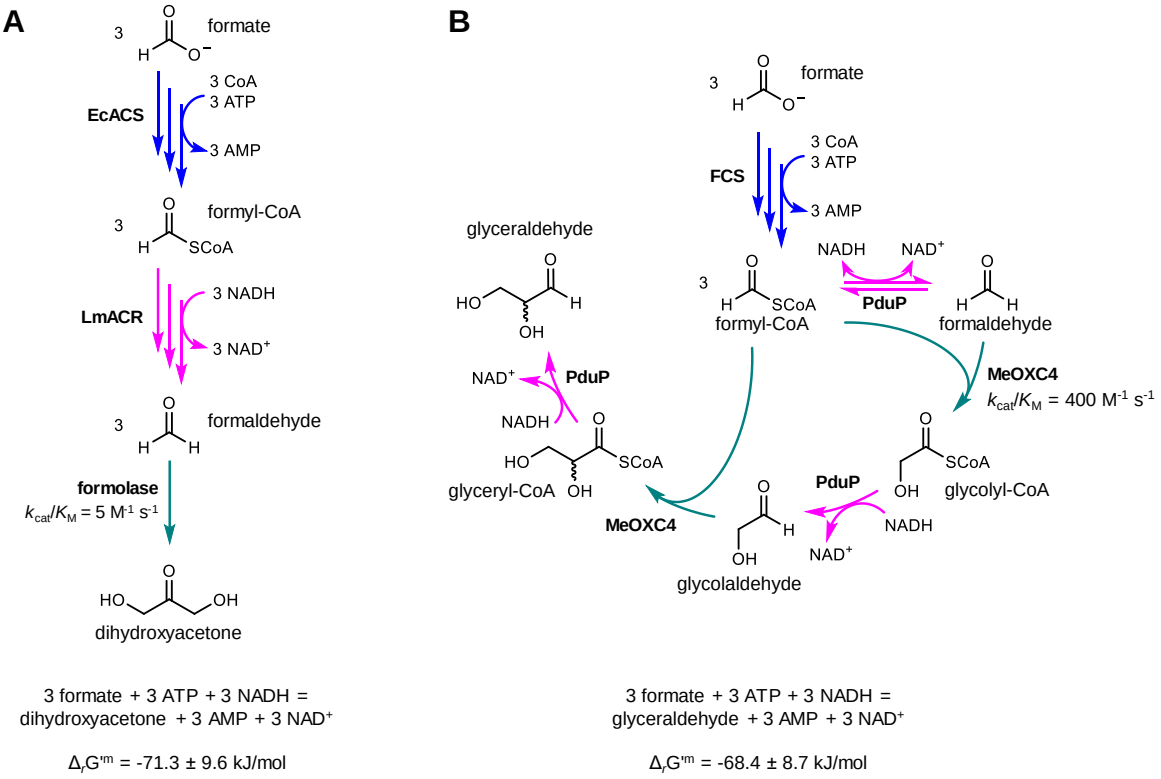


Figure 1. Comparison of the formolase and the iOP pathway. Enzymes for each step are given in bold. Arrows are colour coded: blue, formate activation; pink, CoA ester reduction; green, C-C bond formation. The overall reaction equation and Gibbs free energy under physiological conditions (calculated using eQuilibrator[11]) are shown below. **A)** The linear formolase pathway condenses three formaldehyde molecules into dihydroxyacetone.[8] **B)** The proposed iterative OXC/PduP pathway is based on two consecutive C1 extensions by MeOXC4 and three CoA ester reductions by PduP and produces glyceraldehyde.

4. **Condensation of glycolyl-CoA and formate** $C_n + C_1 = C_{n+1}$

In CHAPTER III it was demonstrated that pyruvate formate-lyase (PFL) catalyzes the condensation of glycolyl-CoA and formate to hydroxypyruvate. In combination with glycolyl-CoA synthetase and hydroxypyruvate reductase, this enabled a pathway for the conversion of glycolate and formate to D-glycerate. As discussed in CHAPTER III, this reaction sequence is a potential carbon-positive photorespiration bypass that utilizes formate instead of CO_2. However, the oxygen sensitivity of PFL will make the implementation into photosynthetic organisms, especially plants, challenging. One way this problem could be tackled is to localize PFL in the mitochondria, where oxygen concentration can be as low as 0.1 µM, owing to the high respiration rate.[12] There is also some evidence that PFL is active in yeast under aerobic conditions when coexpressed with flavodoxin.[13] Also beneficial could be the coexpression of YfiD, which is a small protein that can replace the C-terminal domain of oxygen-damaged PFL.[14, 15]

Another potential application of this reaction is to combine it with the glycolyl-CoA synthase MeOXC4 (see Chapter II). Similarly to the iOP described above, this would constitute a pathway that converts three molecules of formate to a three carbon compound. Only 2 NAD(P)H are required, because the final product glycerate is more oxidized than glyceraldehyde. D-glycerate can be phosphorylated and enter glycolysis as 2- or 3-phosphoglycerate.

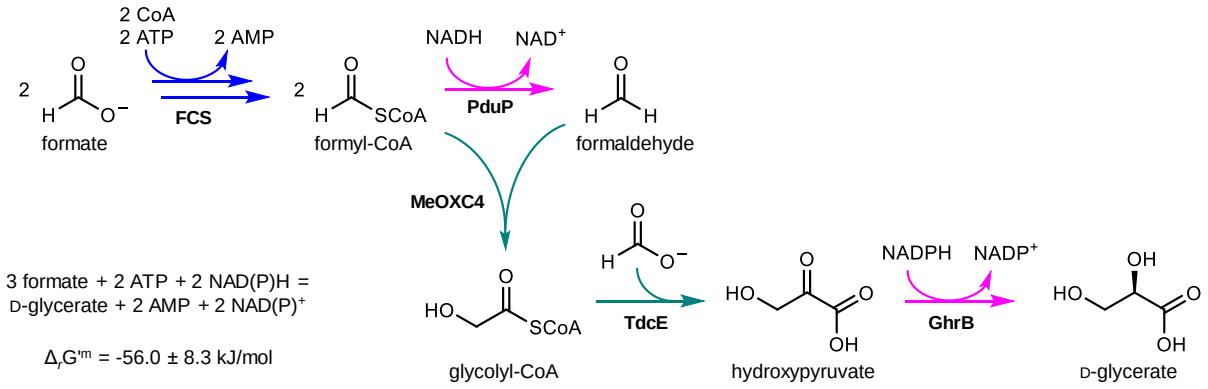


Figure 2. A proposed a pathway comprising of enzymes from all three CHAPTERS of this thesis. Enzymes for each step are given in bold. Arrows are colour coded: blue, formate activation; pink, reduction; green, C-C bond formation. The overall reaction equation and Gibbs free energy under physiological conditions (calculated using eQuilibrator[11]) are shown one the left.

Overall, in this thesis more than 20 enzymes from 10 different organisms were screened, characterized or engineered. Combinations of these enzymes to *in vitro* cascades gave rise to the four above-mentioned novel C1 fixation routes. The physiological background of the enzymes used are very diverse and not all are intuitively linked to carbon fixation. For instance, OXC has mostly been studied due its medical relevance. It is a key enzyme of the oxalate degradation by *Oxalobacter formigenes* in the human gastrointestinal tract[16] and patients with deficiency of this bacterium are more likely to develop kidney stones.[17] This exemplifies the power of mix-and-match biochemistry: by taking enzymes out of their natural context and combining them it is possible to create new biocatalytic syntheses and synthetic metabolism.[18] Moreover, this is a testament to the importance of studying enzymes from any origin and metabolic context - they may be of unexpected use in the future. The fact that new enzymes are still being discovered with high frequency suggests that we have only seen the tip of the iceberg in enzymatic activity. Thus, the perfect C-C bond-forming enzyme might exist, we just haven't found it yet.

Conversely, engineering new metabolism can also assist in identifying naturally existing, yet elusive metabolic pathways. The history of the reductive glycine pathway nicely highlights this. In 2013, Bar-Even and colleagues proposed that this (then considered synthetic) pathway could support efficient growth on C1 substrates.[19] Four years later, some evidence was found that the reductive glycine pathway could be the seventh natural CO_2 fixation pathway.[20] Even though biochemical validation is still lacking, this shows that synthetic pathways can guide the identification of yet unknown natural pathways. A few months after this discovery, two groups (one of them led by A. Bar-Even) independently reported the engineering of *E. coli* to assimilate carbon via the reductive glycine pathway[21, 22] Along these lines, it seems possible that natural C1 fixation pathways exist that are based on Umpolung of formyl-CoA by a HACL-like enzyme or on generation of a formyl radical by a PFL-like enzyme. The existence of a 2-hydroxyisobutyryl-CoA lyase that condenses formyl-CoA with acetone in the acetone degradation pathway of Desulfobacteraceae lends support to this hypothesis.[23]

This work is entirely based on cell-free *in vitro* studies. While this approach offers many advantages, notably the rapid testing of enzyme combinations and reaction conditions,[24] the ultimate goal of the synthetic metabolism community is to establish growth-supporting carbon fixation pathways in a living cell. Despite great advances in metabolic engineering, the implementation of synthetic metabolism into living organisms is still challenging. This is in particular true for carbon assimilation pathways, since it requires rewiring of the existing network

of carbon metabolism.[25] Nevertheless, the last two years have witnessed a veritable surge in engineered growth on C1 feedstocks. Six seminal studies reported successfully established growth on a C1 compound as sole carbon and energy source: *E. coli* and *Pichia pastoris* on CO_2 via the Calvin-Benson-Bassham cycle,[26, 27] *E. coli* and *Cupriavidus necator* on formate and CO_2 via the reductive glycine pathway,[28-30] and *E. coli* on methanol via the ribulose monophosphate pathway.[31] In all these examples, introducing the genes of the naturally occurring C1 pathways initially failed to create the carbon fixation phenotype. Only after extensive laboratory evolution the strains were able to grow on C1 compounds as sole carbon source.

Considering that transplantation of natural pathways is still difficult, the implementation of synthetic pathways based on promiscuous and/or engineered enzymes will likely require even greater efforts. However, living organisms also offer the eminent advantage that enzyme activity can directly be linked to growth. A recent example demonstrates how synthetic carbon assimilation pathways could be established *in vivo*.[32] Auxotrophic *E. coli* strains were developed that depend on the activity of the enzyme of interest for growth. These strains were then used to screen enzyme variants and two aldolases were found to have promiscuous activity with formaldehyde.[32] Strains like these could also be used to evolve engineered or designed enzymes towards improved activity. Once sufficient activity is reached, the whole pathway can be assembled.[33] By smart strain design, the amount of carbon that has to fixed via a pathway can be precisely controlled. In this way the selection stringency can be fine-tuned so that growth at low initial enzyme or pathway activity is possible. Incrementally increasing the selection pressure could ultimately lead to engineered growth via a synthetic pathway.

Clearly, there are major challenges left on the way to a carbon-neutral economy fueled by biological carbon fixation. However, recent progress and the work presented in this thesis indicate that synthetic metabolism and enzyme engineering harbor great potential in achieving this goal. It can therefore be anticipated that enzymes, and C-C bond forming ones in particular may become as valuable as diamonds in addressing the challenges of the 21st century. They may need some polishing, but after all, even diamonds start as coal.

5.1 References

[1] Quayle, J. R. (1963) Carbon Assimilation by Pseudomonas Oxalaticus (Ox1). 7. Decarboxylation of Oxalyl-Coenzyme a to Formyl-Coenzyme A, *Biochemical Journal 89*, 492-503.

[2] Berthold, C. L., Toyota, C. G., Moussatche, P., Wood, M. D., Leeper, F., Richards, N. G., and Lindqvist, Y. (2007) Crystallographic snapshots of oxalyl-CoA decarboxylase give insights into catalysis by nonoxidative ThDP-dependent decarboxylases, *Structure 15*, 853-861.

[3] Can, M., Armstrong, F. A., and Ragsdale, S. W. (2014) Structure, function, and mechanism of the nickel metalloenzymes, CO dehydrogenase, and acetyl-CoA synthase, *Chem Rev 114*, 4149-4174.

[4] Kikuchi, G., Motokawa, Y., Yoshida, T., and Hiraga, K. (2008) Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia, *Proceedings of the Japan Academy Series B-Physical and Biological Sciences 84*, 246-263.

[5] Stoffel, G. M. M., Saez, D. A., DeMirci, H., Vogeli, B., Rao, Y., Zarzycki, J., Yoshikuni, Y., Wakatsuki, S., Vohringer-Martinez, E., and Erb, T. J. (2019) Four amino acids define the CO2 binding pocket of enoyl-CoA carboxylases/reductases, *Proceedings of the National Academy of Sciences of the United States of America 116*, 13964-13969.

[6] Bernhardsgrutter, I., Schell, K., Peter, D. M., Borjian, F., Saez, D. A., Vohringer-Martinez, E., and Erb, T. J. (2019) Awakening the Sleeping Carboxylase Function of Enzymes: Engineering the Natural CO2-Binding Potential of Reductases, *J Am Chem Soc 141*, 9778-9782.

[7] Trudeau, D. L., Edlich-Muth, C., Zarzycki, J., Scheffen, M., Goldsmith, M., Khersonsky, O., Avizemer, Z., Fleishman, S. J., Cotton, C. A. R., Erb, T. J., Tawfik, D. S., and Bar-Even, A. (2018) Design and in vitro realization of carbon-conserving photorespiration, *Proceedings of the National Academy of Sciences of the United States of America 115*, E11455-E11464.

[8] Siegel, J. B., Smith, A. L., Poust, S., Wargacki, A. J., Bar-Even, A., Louw, C., Shen, B. W., Eiben, C. B., Tran, H. M., Noor, E., Gallaher, J. L., Bale, J., Yoshikuni, Y., Gelb, M. H., Keasling, J. D., Stoddard, B. L., Lidstrom, M. E., and Baker, D. (2015) Computational protein design enables a novel one-carbon assimilation pathway, *Proceedings of the National Academy of Sciences of the United States of America 112*, 3704-3709.

[9] Hers, H. G., and Kusaka, T. (1953) [The metabolism of fructose-1-phosphate in the liver], *Biochim Biophys Acta 11*, 427-437.

[10] Hayashi, S. I., and Lin, E. C. (1967) Purification and properties of glycerol kinase from Escherichia coli, *The Journal of biological chemistry 242*, 1030-1035.

[11] Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012) eQuilibrator-the biochemical thermodynamics calculator, *Nucleic Acids Res 40*, D770-D775.

[12] Wilson, D. F., Rumsey, W. L., Green, T. J., and Vanderkooi, J. M. (1988) The oxygen dependence of mitochondrial oxidative phosphorylation measured by a new optical method for measuring oxygen concentration, *The Journal of biological chemistry 263*, 2712-2718.

[13] Zhang, Y. M., Dai, Z. J., Krivoruchko, A., Chen, Y., Siewers, V., and Nielsen, J. (2015) Functional pyruvate formate lyase pathway expressed with two different electron donors in Saccharomyces cerevisiae at aerobic growth, *Fems Yeast Res 15*.

[14] Wagner, A. F., Schultz, S., Bomke, J., Pils, T., Lehmann, W. D., and Knappe, J. (2001) YfiD of Escherichia coli and Y06I of bacteriophage T4 as autonomous glycyl radical cofactors reconstituting the catalytic center of oxygen-fragmented pyruvate formate-lyase, *Biochem Biophys Res Commun 285*, 456-462.

[15] Bowman, S. E. J., Backman, L. R. F., Bjork, R. E., Andorfer, M. C., Yori, S., Caruso, A., Stultz, C. M., and Drennan, C. L. (2019) Solution structure and biochemical characterization of a spare part protein that restores activity to an oxygen-damaged glycyl radical enzyme, *J Biol Inorg Chem 24*, 817-829.

[16] Siener, R., Bangen, U., Sidhu, H., Honow, R., von Unruh, G., and Hesse, A. (2013) The role of Oxalobacter formigenes colonization in calcium oxalate stone disease, *Kidney Int 83*, 1144-1149.

[17] Kaufman, D. W., Kelly, J. P., Curhan, G. C., Anderson, T. E., Dretler, S. P., Preminger, G. M., and Cave, D. R. (2008) Oxalobacter formigenes may reduce the risk of calcium oxalate kidney stones, *J Am Soc Nephrol 19*, 1197-1203.

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

[19] Bar-Even, A., Noor, E., Flamholz, A., and Milo, R. (2013) Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes, *Biochim Biophys Acta 1827*, 1039-1047.
[20] Figueroa, I. A., Barnum, T. P., Somasekhar, P. Y., Carlstrom, C. I., Engelbrektson, A. L., and Coates, J. D. (2018) Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO2 fixation pathway, *Proc Natl Acad Sci U S A 115*, E92-E101.
[21] Bang, J., and Lee, S. Y. (2018) Assimilation of formic acid and CO2 by engineered Escherichia coli equipped with reconstructed one-carbon assimilation pathways, *Proc. Natl. Acad. Sci. U.S.A. 115*, E9271-E9279.
[22] Yishai, O., Bouzon, M., Doring, V., and Bar-Even, A. (2018) In Vivo Assimilation of One-Carbon via a Synthetic Reductive Glycine Pathway in Escherichia coli, *ACS Synth. Biol. 7*, 2023-2028.
[23] Frey, J., Schneider, F., Huhn, T., Spiteller, D., Schink, B., and Schleheck, D. (2018) Two enzymes of the acetone degradation pathway of Desulfococcus biacutus: coenzyme B-12-dependent 2-hydroxyisobutyryl-CoA mutase and 3-hydroxybutyryl-CoA dehydrogenase, *Env Microbiol Rep 10*, 283-292.
[24] Claassens, N. J., Burgener, S., Vogeli, B., Erb, T. J., and Bar-Even, A. (2019) A critical comparison of cellular and cell-free bioproduction systems, *Curr Opin Biotech 60*, 221-229.
[25] Aslan, S., Noor, E., and Bar-Even, A. (2017) Holistic bioengineering: rewiring central metabolism for enhanced bioproduction, *Biochemical Journal 474*, 3935-3950.
[26] Gassler, T., Sauer, M., Gasser, B., Egermeier, M., Troyer, C., Causon, T., Hann, S., Mattanovich, D., and Steiger, M. G. (2020) The industrial yeast Pichia pastoris is converted from a heterotroph into an autotroph capable of growth on CO2, *Nat Biotechnol 38*, 210-+.
[27] Gleizer, S., Ben-Nissan, R., Bar-On, Y. M., Antonovsky, N., Noor, E., Zohar, Y., Jona, G., Krieger, E., Shamshoum, M., Bar-Even, A., and Milo, R. (2019) Conversion of Escherichia coli to Generate All Biomass Carbon from CO2, *Cell 179*, 1255-+.
[28] Kim, S., Lindner, S. N., Aslan, S., Yishai, O., Wenk, S., Schann, K., and Bar-Even, A. (2020) Growth of E. coli on formate and methanol via the reductive glycine pathway, *Nat Chem Biol 16*, 538-545.
[29] Bang, J., Hwang, C. H., Ahn, J. H., Lee, J. A., and Lee, S. Y. (2020) Escherichia coli is engineered to grow on CO2 and formic acid, *Nature Microbiology*.
[30] Claassens, N. J., Bordanaba-Florit, G., Cotton, C. A. R., De Maria, A., Finger-Bou, M., Friedeheim, L., Giner-Laguarda, N., Munar-Palmer, M., Newell, W., Scarinci, G., Verbunt, J., de Vries, S. T., Yilmaz, S., and Bar-Even, A. (2020) Replacing the Calvin cycle with the reductive glycine pathway in Cupriavidus necator, *Metab Eng 62*, 30-41.
[31] Chen, F. Y. H., Jung, H. W., Tsuei, C. Y., and Liao, J. C. (2020) Converting Escherichia coli to a Synthetic Methylotroph Growing Solely on Methanol, *Cell 182*, 933-+.
[32] He, H., Hoper, R., Dodenhoft, M., Marliere, P., and Bar-Even, A. (2020) An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in E. coli, *Metabolic Engineering 60*, 1-13.
[33] Claassens, N. J., He, H., and Bar-Even, A. (2019) Synthetic Methanol and Formate Assimilation Via Modular Engineering and Selection Strategies, *Curr Issues Mol Biol 33*, 237-248.

Acknowledgements

The journey of a PhD is long and rocky, but it is not a lonely one. I would like to thank everyone who accompanied me on this adventure.

First of all **Tobi**. As a supervisor he gave me virtually boundless freedom, which allowed me to grow by pursuing my own ideas – the good ones and the bad ones. As a mentor he would always find the right buttons to push me further than I knew I could go. And finally, as a group leader he creates an atmosphere of trust and great scientific spirit.

I knew right from my fulminant debut that this lab was not only going to be a stimulating work place but that it would also become a second home. I was not proven wrong, as I have found many new friends that will make the time in Marburg unforgettable. Special thanks go to **Tarryn** for countless Quodversations and for sharing her enthusiasm about science in general, but sparking my interest in photosynthesis in particular. Special thanks also go to **Gabo**, for not only providing me with a place to sleep in the first few months but also for being a next-level neighbor in and out of the lab, who is always in for some banter but also to admire the wonderful world of enzymes. A big Merci to **Iria**, for responding unconditionally and whole-heartedly to all the Juuus of the AJB lab, and also for, together with **Wikus**, making sure my mother wouldn't miss my gigs. Thanks **Basti** & **Thomi** for teaching me how things are done in the Erb lab, in terms of techniques but also in terms of having a good time while working hard.

I had the great privilege to supervise two highly motivated and skilled students in **Luca** and **Maren**. Their contribution to this thesis cannot be stressed enough. I learnt a lot from you guys and it makes me proud to see that you continue to do (and enjoy) science.

I am deeply indebted to our close collaborator **Arren Bar-Even**, who tragically passed away while this thesis was being written. His enthusiastic and visionary view of metabolism will not only be missed by me, but the whole scientific community. This thesis shall be proof that his ideas will live on. Many thanks go to **Charlie** and **Ari** from Arren's lab for the fruitful collaborations.

I would like to thank my thesis advisory committee, **Prof. Johann Heider** and **Dr. Sabrina Höbenreich,** for providing valuable feedback on my work.

I am immensely grateful for my wonderful parents for letting me go my way, while making me keep my home in my heart. And last, but certainly not least, thank you **Lea** for always supporting me, even when it meant to regularly face the perils of travelling with Deutsche Bahn. This means the world to me.