Enzymatic systems for synthetic formate assimilation

Dissertation

kumulativ

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich Biologie

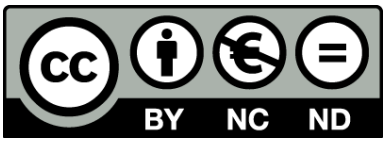
der Philipps-Universität Marburg

vorgelegt von

Maren Nattermann

aus Clausthal-Zellerfeld

Marburg, 2022



Die Arbeit zur vorliegenden Dissertation wurde von September 2019 bis Dezember 2022 unter der Betreuung von Herrn Prof. Dr. Tobias J. Erb am Max-Planck-Institut für terrestrische Mikrobiologie in der Abteilung für Biochemie und Synthetischen Metabolismus angefertigt.

Vom Fachbereich Biologie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation angenommen am 20.02.2023

Erstgutachter: Prof. Dr. Tobias J. Erb

Zweitgutachter: Prof. Dr. Johann Heider

Tag der Disputation: 15.03.2023

Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel *Enzymatic systems for synthetic formate assimilation* 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 01.12.2022

Maren Nattermann

Statement on the contribution to publications in this dissertation

I declare that this thesis was composed by myself and without external assistance. I confirm that the work presented has not been submitted for any other degree or professional qualification. I also declare that I have not made any prior doctoral attempt.

Furthermore, I confirm that no sources other than those indicated have been used and that appropriate credit has been given within this thesis where reference has been made the work of others.

I confirm that the work submitted is my own, except where explicitly stated otherwise in the text or where work that has formed part of jointly authored publications has been included. My contribution to this work has been explicitly stated below.

Marburg, 01.12.2022

Maren Nattermann

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

Maren Nattermann*, Simon Burgener*, Pascal Pfister, Alexander Chou, Luca Schulz, Seung Hwan Lee, Nicole Paczia, Jan Zarzycki, Ramon Gonzalez, and Tobias J. Erb (2021) Engineering a highly efficient carboligase for synthetic one-carbon metabolism. *ACS Catalysis*, 11(9): 5396–5404. https://doi.org/10.1021/acscatal.1c01237 *equal contribution

In this work, M.N. contributed to	the design and execution of experiments the analysis of data the writing of the manuscript

Maren Nattermann, Sebastian Wenk, Pascal Pfister, Nils Guntermann, Lennart Nickel, Charlotte Wallner, Jan Zarzycki, Nicole Paczia, Giancarlo Franciò, Walter Leitner, Arren Bar-Even and Tobias J. Erb (2022) A new-to-nature cascade for phosphate-dependent formate reduction at the thermodynamic limit. *Manuscript in revision*.

In this work, M.N. contributed to	the design and execution of experiments the organization of collaborations the analysis of data the writing of the manuscript

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

_____ _____

Maren Nattermann Prof. Dr. Tobias J. Erb

Publications that are not discussed in this thesis:

Amir Pandi, Christoph Diehl, Ali Yazdizadeh Kharrazi, Scott A. Scholz, Elizaveta Bobkova, Léon Faure, **Maren Nattermann**, David Adam, Nils Chapin, Yeganeh Foroughijabbari, Charles Moritz, Nicole Paczia, Niña Socorro Cortina, Jean-Loup Faulon and Tobias J. Erb (2022) A versatile active learning workflow for optimization of genetic and metabolic networks. *Nature Communications* 13: 3876. https://doi.org/10.1038/s41467-022-31245-z

Sarah Bierbaumer, **Maren Nattermann***, Luca Schulz*, Reinhard Zschoche, Tobias J. Erb, Christoph K. Winkler, Matthias Tinzl and Silvia M. Glueck (2022) Enzymatic CO_2 Fixation: from Nature to Application. *Manuscript in revision*. *equal contribution

Table of contents

Table of contents IX

Summary XII

Zusammenfassung XIV

General Introduction 1

Carbon at the intersection of biology and chemistry 1

One-carbon assimilation as the basis for synthetic metabolism 2

The formate bio-economy 7

Synthetic formate to formaldehyde conversion 7

Thiamine diphosphate as the basis of C1-C1 additions 9

Aims of the thesis 12

References 14

A new-to-nature cascade for phosphate-dependent formate reduction at the thermodynamic limit 24

Abstract 26

Introduction 27

Materials & Methods 29

Results 43

Discussion 55

References 56

Acknowledgements 60

Competing Interests 61

Supplementary Figures 62

Supplementary Tables 80

Supplementary References 81

Engineering a highly efficient carboligase for synthetic one-carbon metabolism 86

Abstract 90
Introduction 91
Results 92
Discussion 104
Acknowledgements 105
References 105
Materials & Methods 109
Supporting Tables 119
Supporting Figures 123
Supporting References 134
General Discussion 139
FOK-FPR: an energy-efficient path from formate t formaldehyde 139
GCS: an efficient C1-C1 carboligase 141
Outlook in the context of the formate bio-economy 144
References 145
Acknowledgements 149

Summary

A circular economy requires the re-valorization of waste products. Since the beginning of the industrial age, one of the major emissions of human civilization has been CO_2. While it could be regarded simply as a pollutant, it is, intrinsically, a resource – not only a carbon source, but also the main carbon source supporting all life on the planet. While it is sensible to use it as such, its chemical conversion into useful compounds is challenging – while nature has evolved pathways for its assimilation for millions of years. Biological systems for CO_2 valorization have therefore come into focus. In addition to CO_2, other one-carbon (C1) compounds, in particular methanol and formate, have been suggested for electro-biochemical applications. In these two-step processes, C1 compounds are generated by electrochemical hydrogenation of CO_2 and subsequently utilized by biological platform organisms as a source of both carbon and reductive power.

Out of the C1 compounds, formate is biologically assimilated at the highest energetic efficiency. Natural formate assimilations, however, is limited to specialist microbes that can be challenging to cultivate or genetically modify. The generation of synthetic formatotrophs, therefore, offers the potential of broadening the scope to more commonly used platform organisms such as *E. coli.* Here, two modes of action can be followed – one, the transposition of natural formate assimilation pathways and two, the generation of fully synthetic pathways for formate assimilation. In this thesis, two modules for synthetic formate assimilation are developed through rational enzyme engineering. First, a module for the reduction of formate to formaldehyde via formyl phosphate is created, permitting integration of formate into otherwise methanol-assimilating pathways. Second, a novel carboligase is designed that is able to condense formaldehyde with formyl-CoA, an activated formate species, into the C2 compound glycolyl-CoA. The combination of these two modules displays a potential for formate assimilation into a C2 compound at higher efficiency than previously published routes. In summary, this thesis enriches the scope of synthetic formatotrophy by two powerful building blocks that enable new paths in carbon metabolism.

Zusammenfassung

Eine wahrlich zirkuläre Wirtschaft bedingt die Aufwertung von Abfallstoffen. Seit Beginn der Industrialisierung ist CO_2 eines der größten Abfallprodukte der Menschheit. Während man dieses Gas als reine Verunreinigung betrachten kann, ist es doch hauptsächlich eine Ressource – nicht bloß eine Kohlenstoffquelle, but die Hauptquelle alles organischen Kohlenstoffs, ohne die Leben auf der Erde in dieser Form nicht entstanden wäre. Obwohl es also sinnvoll ist, es als solche zu nutzen, ist die chemische Konversion von CO_2 in industriell relevante Produkte schwierig. Im Kontrast dazu hat die Natur über Jahrmillionen an der Kohlenstoffassimilierung gefeilt. Aus diesem Grund rücken biologische Systeme der CO_2-Aufwertung immer mehr in den Fokus. Abgesehen von CO_2 können auch andere Einkohlenstoffe, vor allem Formiat und Methanol, in bioelektrischen Hybridsystemen verwertet werden. In solchen werden die Stoffe über die elektrochemische Hydrogenation von CO_2 hergestellt und im Anschluss von biologischen Platformorganismen als Kohlenstoff- und Energiequelle genutzt.

Formiat sticht hier hervor, da es die höchste energetische Effizienz für die biologische Assimilierung bietet. Allerdings ist die natürliche Formiatassimilierung auf mikrobielle Spezialisten beschränkt, die schwierig zu kultivieren oder genetisch unverfügbar sind. Aus diesem Grund bietet die Etablierung synthetischer Formatotrophie eine Gelegenheit, besser etablierte biologische Plattformen wie E. coli zu nutzen. Hierfür gibt es zwei verschiedene Optionen – zum einen die Transplantation natürlicher Formiatassimilierung, und auf der anderen Seite die Konstruktion vollsynthetischer Stoffwechselwege. Diese Arbeit beschäftigt sich mit der Generation zweier synthetischer Formiatassimilierungsmodule. Zuerst wird eine Reaktionskaskade zur Reduktion von Formiat zu Formaldehyd über Formylphosphat etabliert, welche die Integration von Formiat in methylotrophe Stoffwechselwege ermöglicht. Im Anschluss wird eine neuartige Carboligase entworfen, die Formaldehyd und Formyl-CoA in den Zweikohlenstoff Glycolyl-CoA umwandelt. Eine Kombination dieser Module zeigt deren Potential für die effiziente Formiatassimilierung zu einem Zweikohlenstoff auf. Zusammengefasst bereichert diese Arbeit das Feld der synthetischen Formatotrophie um zwei potente Bausteine, die dem Kohlenstoffstoffwechsel neue Wege ermöglichen.

General Introduction

General Introduction

Carbon at the intersection of biology and chemistry

Life on earth is built upon a backbone of carbon. Organic matter contains approximately 50% carbon[1, 2]. Two large fields of natural sciences are almost exclusively concerned with carbon-based molecules - organic chemistry and biochemistry. Organic chemists mostly deal with compounds useful to humans, such as plastics, fuel, pharmaceuticals or dyes. Biochemists, on the other hand, have long been concerned exclusively with the understanding of natural processes – the chemistry of metabolism, of biosynthesis or of cellular mechanisms. In recent years, organic chemistry and biochemistry have begun to merge – biochemical processes are used to produce pharmaceuticals[3], to generate fuel[4], or to manufacture and degrade plastic[5-8]. This intersection of the two fields is enabled by the efficiency of natural catalysts - enzymes. Where chemical synthesis commonly requires heat and high pressure, enzymes can perform identical reactions at ambient temperature and atmospheric pressure. In addition, where a chemist may control two or three reactions at a time, potentially having to extract and purify intermediary products, a biochemist may mix-and-match dozens of enzymes to create long production chains – as nature has done for millions of years.

A completely new field has emerged based around this idea: synthetic biology. Natural compounds have been extracted from organisms dating as far back as 3500 BC, used mostly for cosmetics and pharmaceuticals, which are still their major applications today[9]. This approach is intrinsically limited: it is strictly defined by the course evolution has taken. Products that conceptually could exist do not, simply because organisms did not evolve in an environment conducive to them. They may be produced at low yield or require difficult cultivation conditions. In addition, organisms are limited to a defined amount of energy and carbon sources and are unable to escape this limitation unless high selective pressure is applied, meaning cultivation costs may be un-economical. Synthetic biology aims to expand biological systems past the limits of evolution in a tailor-made approach that heavily relies on modularity – an incremental alteration of multiple (rate-limiting) components of a given production route.

On the product side, targets can either be existing chemicals or "designer" products created by the incorporation of non-natural building blocks[10]. Biological production of these

compounds often relies on enzyme promiscuity - the ability of an enzyme to recognize non-native substrate[11] - and enzyme engineering – altering enzymes for improved activity on the desired substrate[12, 13]. On the production side, metabolic engineering[13-15] is the most powerful tool. In the classical sense, this approach alters existing pathways by introduction of small production modules and regulation of genetic elements to optimize flux towards the desired production route. Advances in systems biology have added a host of computational tools for flux balance analysis to the toolkit[16-18], while progress in synthetic biology allows the introduction of more complex production routes, better regulation thereof, and, importantly for this thesis, enables metabolic utilization of non-native carbon sources.

Given the staggering rate of greenhouse gas emissions and the rapid progression of man-made climate change, the re-design and improvement of natural carbon assimilation has come into focus: aiming to extract CO_2 from the atmosphere and convert it into value-added products[19]. Broadly, this approach divides into two core modules: one, designing new pathways for carbon assimilation into central metabolism and two, implementing pathways for producing value-added molecules downstream of the assimilation. These modules may be combined in a single platform organism, split between a carbon fixing and a product-generating organism, or carried out fully *in vitro*.

One-carbon assimilation as the basis for synthetic metabolism

Nature derives a vast majority of its biomass directly or indirectly from photosynthesis[20]. Seven natural carbon assimilation pathways have been discovered, the Calvin-Benson-Bassham (CBB) cycle[21], the reverse tricarboxylic acid (rTCA) cycle[22], the dicarboxylate/ 4-hydroxybutyrate (DH/4HB) cycle[23], the 3-hydroxyproprionate/ 4-hydroxybutryrate (3HP/4HB) cycle[24] and the 3-hydroxyproprionate (3HP) bicycle[25], the Wood-Ljungdahl pathway (WLP)[26], the reductive glycine pathway (rGlyP)[27, 28],. These are joined by three natural one-carbon (C1) assimilation pathways, the ribulose monophosphate (RuMP) pathway[29], xylulose monophosphate (XuMP) pathway,[30] and the serine cycle,[31] which assimilate reduced C1 species like CO, formate, formaldehyde, methanol and methane (**Figure 1, Table 1**).

Thus, nature has found metabolic solutions to valorize C1 compounds with different oxidation states (**Figure 1**). Here, assimilation of CO_2 is the least efficient as it is fully oxidized. In

contrast, all other C1 species can be oxidized, generating energy for the cell in the form of reducing equivalents, usually NAD(P)H. With exception of the WLP, carbon is assimilated at oxidative states +IV (CO_2, HCO_3^-) and 0 (formaldehyde/ 5,10-mT_4F), representing the most electrophilic C1 species (**Figure 1, Table 1**). Formate, methanol and methane are usually oxidized to a more reactive species before assimilation.

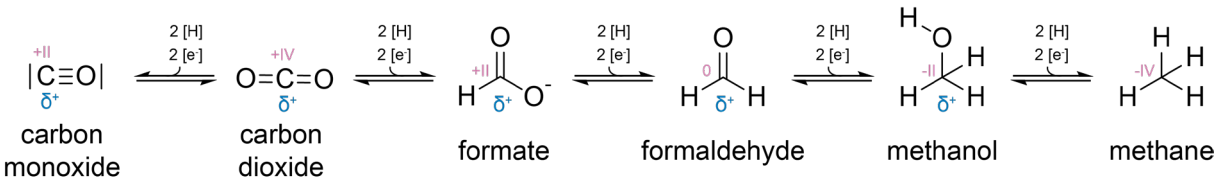


Figure 1 Overview of relevant one-carbon species. Oxidation numbers are indicated in red roman numerals. Partial charges are indicated in blue.

Synthetic C1 metabolism aims to expand beyond natural carbon assimilation to valorize excess carbon emissions. This includes transplantation of natural C1 assimilation pathways to different, better-understood platform organisms or introduction of individual enzymes to existing pathways to either improve or enable carbon assimilation[32-37]. In addition, fully synthetic pathways can be constructed by combining existing enzymatic functions and may contain enzymes with altered functions that catalyze non-native, promiscuous reactions[38, 39].

Synthetic pathways are commonly generated via design-build-test cycles. In the design phase, hypothetical pathways are constructed based on known (and sometimes, unknown) enzymatic functions. These designs are then evaluated for feasibility. Here, kinetic, energetic and thermodynamic parameters as well as implementation limitations are evaluated. Thermodynamics are considered based on computational modelling of the Gibbs free energy profile (ΔrG') of partial reactions or whole pathways using tools such as eQuilibrator[16, 17]. A second commonly used metric is the Max-min Driving Force (MDF) that identifies the reactions of the pathway that are most thermodynamically constrained[18]. Energetic considerations take into account the amount of reductive (NAD(P)H, FADH2 etc.) or energy carrying (NTPs, CoA-esters, etc.) equivalents consumed per turn of the cycle or per assimilated carbon. Finally, implementation feasibility is considered – taking into account the number of enzymes required, interactions with native metabolism, side-product formation and oxygen

tolerance. Often, these considerations are in opposition to one another, i.e. an energetically efficient pathway may be kinetically inefficient. After the design phase, the most promising pathways are assembled and characterized. This is usually done *in vitro*, where reaction parameters are more easily controlled and complexity is minimized. If activity is observed, optimizations may be implemented, such as replacements of the most limiting steps, removal of toxic side-product, etc.

Crucially, design-build-test cycles are not only applied to whole pathways, but also to individual enzymes, where kinetic parameters may be improved or selectivity increased by targeted or untargeted mutagenesis. These enzyme engineering efforts are often employed to switch an enzyme's activity from its native to a promiscuous one. The resulting variants can be core components to synthetic pathways, if one of the desired reactions is not naturally occurring.

Table 2 lists the current landscape of synthetic carbon assimilation. For synthetic C1 assimilation pathways, three benchmarks exist: (I), functionality *in vitro*, (II), functionality *in vivo* and (III), sustained growth. Some of the listed pathways, such as the ASAP[40] and rGPS-MCG[41], do not aim at *in vivo* implementation and therefore are not concerned with benchmarks (II) and (III). Others, such as the CETCH or the homoserine cycle, were aimed at sustained growth, but so far either were only active *in vitro* or showed activity *in vivo* when co-fed on other substrates[42-44]. Notably, fully orthogonal carbon assimilation that is assisted by a second substrate can be desired, as it allows a decoupling of cellular growth from product synthesis[45]. Similarly, whole-cell biocatalysts may be used, where cells are first grown and enzymes produced before C1 substrates are added. This is mostly applicable for pathways requiring the addition of high titers of formaldehyde, which is highly cytotoxic, such as the FORCE pathway[46]. Finally, very few synthetic pathways have been shown to achieve sustained growth on a C1 source, such as the serine threonine cycle[47].

Table 1 Natural C1 assimilation pathways. CBB: Calvin-Benson-Bassham cycle, rTCA: reverse tricarboxylic acid cycle, DC/4HB: dicarboxylate/ 4-hydroxybutyrate cycle, WLP: Wood-Ljungdahl pathway, rGlyP: reductive glycine pathway, 3HP/4HB: 3-hydroxyproprionate/ 4-hydroxybutryrate cycle, 3HP: 3-hydroxyproprionate bicycle, RuMP: ribulose monophosphate pathway, XuMP: xylulose monophosphate pathway, SerC: serine cycle, 5-mT_4F: 5-methyl-T_4F; 5,10-mT_4F: 5,10-methylene-T_4F, Me-CoFeSP: Methyl-cobalt- and iron-containing corrinoid iron–sulfur protein

	metabolic carbon source/ pathway intermediate						assimilating enzyme	C1 substrate	acceptor molecule
	CO	CO_2/ HCO_3^-	CHOOH	CH_2O	CH_3OH	CH_4			
CBB[21]		■					ribulose-1,5-bisphosphate carboxylase/ oxygenase	CO_2	ribulose-1,5-bisphosphate (C5)
rTCA[22]		■					2-ketogluterate synthase	CO_2	succinyl-CoA (C4)
							isocitrate dehydrogenase	CO_2	α-ketoglutarate (C5)
DC/4HB[23]		■					pyruvate:ferredoxin oxidoreductase	CO_2	acetyl-CoA (C2)
							phosphoenolpyruvate carboxylase	HCO_3^-	phosphoenole pyruvate (C3)
3HP/4HB[24]		■					acetyl-CoA/propionyl-CoA carboxylase	HCO_3^-	acetyl-CoA (C2)
							acetyl-CoA/propionyl-CoA carboxylase	HCO_3^-	propionyl-CoA (C3)
3HP[25]		■					acetyl-CoA/propionyl-CoA carboxylase	HCO_3^-	acetyl-CoA (C2)
							acetyl-CoA/propionyl-CoA carboxylase	HCO_3^-	propionyl-CoA (C3)
WLP[26],	■	■					CO dehydrogenase/ acetyl-CoA synthase	CO/ Me-CoFeSP	C1-C1 condensation
rGlyP[27, 28]		■	■				glycine cleavage/ synthase system	CO_2/ 5,10-mT_4F	C1-C1 condensation
RuMP[29]				■	■	■	3-hexulose-6-phosphate synthase	CH_2O	ribulose-5-phosphate (C5)
XuMP[30]				■	■		dihydroxyacetone synthase	CH_2O	xylulose-5-phosphate (C5)
SerC[31]		■	■	■	■	■	serine hydroxymethyl transferase	5,10-mT_4F	glycine (C2)
							phosphoenolpyruvate carboxylase	HCO_3^-	phosphoenole pyruvate (C3)

Table 2 Synthetic C1 assimilation pathways. [a] designed for *in vitro* application [b] *in vitro* activity confirmed, *in vivo* activity desired [c] *in vivo* activity confirmed [d] sustained growth on C1 source observed. Back: native activity, purple: promiscuous activity, blue: engineered activity. 5,10-mT$_4$F: 5,10-methylene-T$_4$F

	substrate/ pathway intermediate					assimilating enzyme	C1 substrate	acceptor molecule
	CO_2/ HCO_3^-	CHOOH	CH_2O	CH_3OH	CH_4			
CETCH[42, 43],[b]						crotonyl-CoA carboxylase/reductase	CO_2	crotonyl-CoA (C4)
						crotonyl-CoA carboxylase/reductase	CO_2	acryloyl-CoA (C3)
rGPS-MCG[41],[a]						phosphoenolpyruvate carboxylase	HCO_3^-	phosphoenole pyruvate (C3)
						crotonyl-CoA carboxylase/reductase	CO_2	crotonyl-CoA (C4)
POAP[48],[b]						pyruvate carboxylase	HCO_3^-	pyruvate (C3)
						pyruvate synthase	CO_2	acetyl-CoA (C2)
MCG[49],[c]						phosphoenolpyruvate carboxylase	HCO_3^-	phosphoenole pyruvate (C3)
GED[37],[c]						6-phosphogluconate dehydrogenase	CO_2	ribulose-5-phosphate (C5)
serine threonine[47],[d]						serine hydroxymethyl transferase	5,10-mT$_4$F	glycine (C2)
						phosphoenolpyruvate carboxylase	HCO_3^-	phosphoenole pyruvate (C3)
mod. serine[50],[c]						serine hydroxymethyl transferase	5,10-mT$_4$F	glycine (C2)
						phosphoenolpyruvate carboxylase	HCO_3^-	phosphoenole pyruvate (C3)
homoserine[44],[c]						serine aldolase	CH_2O	glycine (C2)
						4-hydroxy-2-oxobutanoate aldolase	CH_2O	pyruvate (C3)
FORCE[46],[c]						2-hydroxyacyl-CoA lyase	CH_2O, fCoA	C1-C1 condensation
ethylene glycole[51],[c]						glyoxylate carboligase	CH_2O	
GAA, GAPA[52],[b]								
ASAP[40],[a]						formolase		
HWLS[53],[c],								
SMGF[54],[c]								
C4 sugar[55],[a]								
lin. Met.[56],[c]								
SACA[57],[c]						benzoylformate decarboxylase		
Reverse PFL[58],[c]						pyruvate formate lyase	HCOOH	Acetyl-CoA (C2)

The formate bio-economy

Out of the C1 compounds, formate is possibly the easiest to handle. CO, CO_2 and CH_4 are gaseous and have low solubility in water, therefore they are challenging to make available to microorganisms in high concentration. Common workarounds for CO_2 assimilation are an increase of CO_2 content in the atmosphere or the use of bicarbonate, which has much higher solubility. Methanol is fully miscible with water, but evaporates quickly, is flammable and highly toxic to human workers. Formaldehyde, while gaseous and highly soluble in water, is cytotoxic at micromolar concentrations (see *CHAPTER 1*, **Supplementary Figure 11a**) and the least suitable C1 feedstock. Formate is highly soluble in water, poses no risk to humans and is tolerated at milllimolar titers by microorganisms (see *CHAPTER 1*, **Supplementary Figure 11b**). In addition, the energetic efficiency of formate assimilation has been calculated to be higher than that of any other C1 compound[59, 60]. Therefore, formate has been proposed as the core compound of a circular bioeconomy[61]. Herein, it is proposed that formate can be generated with relative ease from CO_2 using electrochemical hydrogenation[62-64]. Alternatively, CO_2 can be reduced to formate by the highly oxygen-sensitive formate dehydrogenase of the WLP[65]. Formate valorization can then occur via natural formatotrophs using the serine cycle, the WLP or the rGlyP. However, native formatotrophs are usually unsuited for industrial applications as their genetic availability is limited. Therefore, natural formate assimilation pathways have been transplanted into common platform organisms. The rGlyP in particular is of high interest here, as it is short, linear and aero-tolerant and requires the transplantation of only a few enzymes[66]. In addition, multiple synthetic formate assimilation pathways have been constructed, some based on the existing serine cycle or WLP[44, 53, 54], another employing the reverse reaction of pyruvate formate lyase[58, 67] and finally, a host of pathways based on formaldehyde-assimilating lyases have been demonstrated[40, 51-57, 68].

Synthetic formate to formaldehyde conversion

In formatotrophs, formate is exclusively assimilated via tetrahydrofolate (T_4F, in bacteria) or tetrahydromethanopterine (H_4MPT, in archaea). In **Figure 2**, the activation of formate to T_4F bound C1 species is shown. Metabolic integration occurs at the levels of 5,10-methylene-T_4F (5,10-mT_4F, serine cycle and rGlyP) and 5-methyl-T_4F (5-m T_4F, WLP), corresponding to the

oxidative states of formaldehyde and methanol. 5,10-mT$_4$F can spontaneously hydrolyze to formaldehyde and T$_4$F[69]. This feature was employed in a synthetic pathway to produce relevant levels of free formaldehyde that could be assimilated by formolase[68, 70]. This engineered benzaldehyde lyase converts 3 molecules of formaldehyde to dihydroxyacetone, which can be channeled into central metabolism via phosphorylation. In a similar fashion, it can be imagined that strictly methylotrophic pathways like the RuMP, XuMP or homoserine pathway – all of which use free formaldehyde in their assimilating reactions – could be turned into formatotrophic pathways.

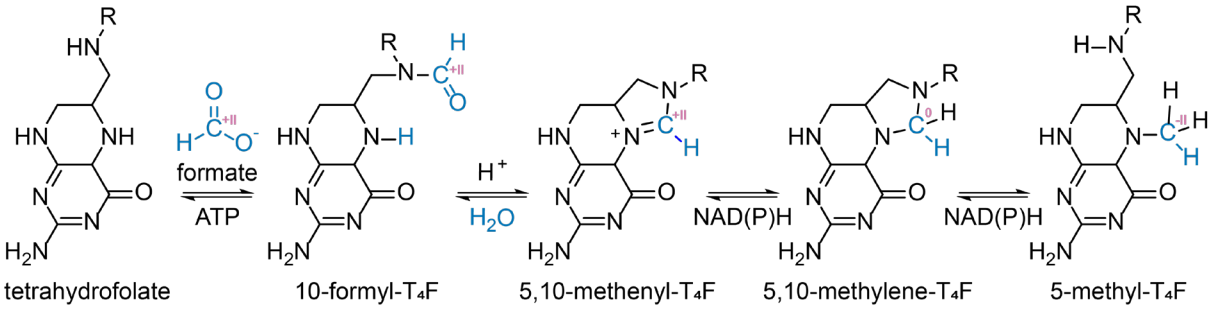


Figure 2 Activation of formate using the natural T$_4$F cascade. Oxidation numbers are indicated in red roman numerals. Adapted from Bierbaumer *et al.*, 2022[71].

When applied for formaldehyde production, the T$_4$F cascade suffers from a few drawbacks. While T$_4$F is used by aerobic organisms, it degrades oxidatively *in vitro*[72], making it difficult to implement T$_4$F-based cascades in cell-free systems. Up to 5,10-mT$_4$F, the cascade employs three enzymes, (see *CHAPTER 1*, **Figure 1**), resulting in a relatively high burden on cells (over)expressing it. Finally, the system is limited by spontaneous hydrolysis of the relatively stable 5,10-m T$_4$F[72]. As an alternative, a two enzyme system for the activation of formate to formaldehyde was designed based on the promiscuous activities of acetyl-CoA synthetase (ACS) and acyl-CoA reductase (ACR), employing two ATP equivalents to activate formate to formyl-CoA, followed by a reduction to formaldehyde[54, 68, 73] (see *CHAPTER 1*, **Figure 1**). While this cascade is both shorter and not limited by a spontaneous reaction, it also requires one additional ATP equivalent compared to the T$_4$F route. To circumvent this additional cost, the activation of formate to formyl phosphate by the formate kinase (FOK) activity of acetate

kinase (ACK)[74], followed by transfer to formyl-CoA by *E. coli* phosphate acetyl transferase (PTA) was proposed[46], although PTA activity on formyl phosphate was never demonstrated. This alternative requires the same amount of energy and enzymes as the T_4F cascade, while actively producing formaldehyde. Yet another, more elegant solution reducing both energy demand and the number of enzymes is the activation of formate to formyl phosphate by FOK followed by a reduction to formaldehyde by a reductase. However, such a formyl phosphate reductase (FPR) has never been described. In *CHAPTER 1*, FPR activity is demonstrated in members of four reductase families and an enzyme with improved *in vivo* activity is produced by targeted mutagenesis. The resulting FOK-FPR module is proposed as a suitable candidate for implementation in existing pathways, replacing the CoA or T_4F modules used therein. A particular benefit of FOK-FPR compared to a CoA-based route is a lack of interaction with the CoA pool of the cell. CoA reductases are notoriously promiscuous, with most formyl-CoA reductases reducing acetyl-CoA just as readily. Only limited success has been achieved trying to limit acetyl-CoA reduction while maintaining formyl-CoA reduction[54]. FOK-FPR does not have a direct influence on the CoA pool of the cell, circumventing this limitation.

Thiamine diphosphate as the basis of C1-C1 additions

Carbon assimilation can broadly be divided into two modes of action: direct C1-C1 additions and C1-Cn additions utilizing longer-chained acceptor molecules, the latter being much more common (**Table 1 & 2**). As shown in **Figure 1**, C1 substrates either have electrophilic character (CO_2, CO, formate, formaldehyde and methanol) or are inert (methane). Therefore, the easiest way of performing a C1 addition is the generation of a nucleophile on an acceptor molecule. In most reaction cascades, these nucleophiles are generated by proton abstraction proximal to carbonyl groups, generating enolates with high nucleophilic character (**Figure 3**). The main drawback of C1-Cn additions is the requirement for acceptor substrate regeneration at the expense of energy. This results in a cyclic structure of pathways using C1-Cn additions, which are usually long and relatively complex.

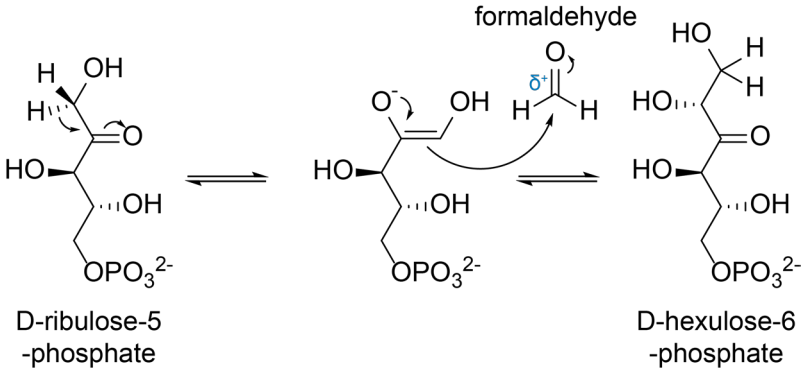


Figure 3 Enolate generation as the basis of nucleophilic attacks in longer chain acceptor molecules. Shown is the mechanism of 3-Hexulose-6-phosphate synthase. Adapted from Bierbaumer et al., 2022[71].

In contrast, C1-C1 additions do not require acceptor molecules, thus allowing a linear funneling of the C1 substrates into central metabolism. The only two natural pathways capable of direct C1-C1 addition, the WLP and rGlyP, utilize complex mechanisms to perform these additions. The CO dehydrogenase/ acetyl-CoA synthase of the WLP pathway contains a highly oxygen sensitive metal cluster that enables the condensation of CO with Me-CoFeSP[75], while the glycine cleavage system of the rGlyP generates a lipoamide carrier protein-bound NH2-CH2-S-R intermediate, which is subsequently transferred to PLP and attacks CO_2[76]. Therefore, the discovery of a low-complexity solution was highly desirable and has been found in the form of thiamine diphosphate (ThDP) dependent enzymes[40, 46, 51-57]. ThDP is a unique cofactor capable of performing Umpolung of carbons bound to its thiazolium ring (**Figure 4**), inverting electrophilic C1 species to nucleophiles. This chemical trick allows ThDP enzymes to remain structurally and mechanistically simple. ThDP is employed in nature by transketolases[77], decarboxylases[78] and even in C1 metabolism by dihydroxyacetone synthase[79]. Members of the benzaldehyde lyase[68, 70], benzoylformate decarboxylase[57], glyoxylate carboligase[56] and 2-hydroxy-acyl-CoA lyase[46, 80] (HACL) families have been shown to catalyze ThDP-dependent C1-C1 additions.

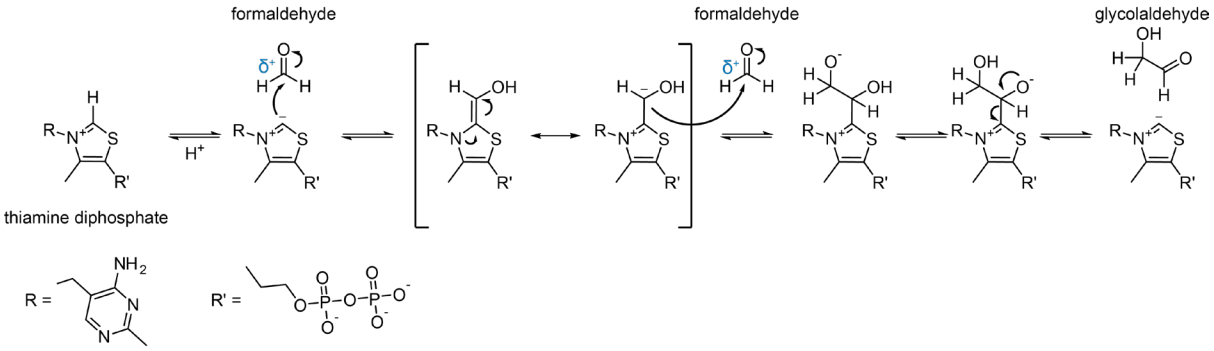


Figure 4 Postulated ThDP-dependent mechanism of the enzymatic formose reaction. ThDP is activated to a carbanion by deprotonation and attacks formaldehyde. The formation of a carbanion intermediate reverses the positive partial charge of the carbon atom into a negative charge, permitting nucleophilic attack on a second formaldehyde molecule. Glycolaldehyde is released and ThDP is regenerated. Adapted from previous publications[70, 81].

Notably, the enzymatic efficiency of HACL outperforms that of all other mentioned carboligases by one to two orders of magnitude (**Table 3**). In contrast to the other enzymes that catalyze the formose reaction to produce glycolaldehyde (C2) or dihydroxyacetone (C3) (**Figure 4**), HACL requires two substrates, formaldehyde and formyl-CoA, creating glycolyl-CoA (C2) (*CHAPTER 2*, **Figure SI 2**). The beneficial kinetic parameters of this glycolyl-CoA synthase (GCS) reaction make it an attractive target for further engineering efforts. However, engineering of HACL has yielded no improvement thus far[80], and as its structure has not been solved, targeted mutagenesis is difficult. Fortunately, within the related and structurally well-characterized oxalyl-CoA decarboxylase (OXC) family, C1-Cn addition of formyl-CoA to longer-chained aldehydes was observed[82]. Reasoning that OXC therefore may also be capable of performing the GCS reaction and be a good target for structure-based enzyme engineering, *CHAPTER 2* is concerned with generating a bona fide GCS out of *Methylorubrum extorquens* OXC (MeOXC). A use-case for GCS is well established in the FORCE system[46], where it serves as the core module for the generation of a variety of C2 compounds. Improvements to the efficiency of the enzyme has direct impact on yield and contributes to the viability of bioconversion systems compared to traditional production routes.

Table 3 kinetic parameters of C1-C1 carboligases.

family	enzyme	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
benzaldehyde lyase	formolase[68, 70] (to DHA)	n.d.	n.d.	4.7
glyoxylate carboligase	EcGCL R484M N283Q L478M[56]	0.1	18	5.2
benzoylformate decarboxylase	glycolaldehyde synthase[57]	1.6	170	9.6
2-hydroxy-acyl-CoA lyase	RuHACL G390N (formaldehyde)[80]	3.1	30	100
2-hydroxy-acyl-CoA lyase	RuHACL G390N (formyl-CoA)[80]	1.4	0.2	6600

Aims of the thesis

A sustainable C1 bio-economy relies on the establishment of (energy)-efficient pathways for C1 assimilation and product generation. This need is driven in large part by the low energy (electron) content of C1 feedstocks compared to traditional feedstocks such as sugars[83, 84]. Accounting for oxidative phosphorylation, glucose assimilation (via glycolysis) produces 6 ATP equivalents, whereas its production from CO_2 via the CBB cycle consumes 17 ATP equivalents. Even CH_4 assimilation – being the most reduced C1 source – via the RuMP pathway produces only 4 ATP equivalents. While *in vitro* cascades and photoautotrophic organisms can gain their energy from "limitless" electron sources like electricity or light, other organisms are dependent on alternative electron sources such as hydrogen or the C1 compound itself[85, 86]. Hence, it is of high importance for such systems to conserve the energy they expend. Therefore, generating synthetic modules with either higher enzymatic activity (lowering the requirement for enzyme production) or lower energy expense (lowering the depletion of ATP/ NAD(P)H pools) can improve the metabolic capacity of C1-consuming organisms.

In this thesis, enzyme engineering is employed to achieve this aim. *CHAPTER 1* seeks a kinetically, energetically and thermodynamically favorable formate to formaldehyde conversion cascade. This hypothetical route – FOK-FPR – requires the discovery of a novel non-natural reaction. FPR activity can be probed for in different phosphate reductase families and improved by homologue screening and targeted enzyme engineering. The chapter strives for improved formaldehyde production *in vivo* to enable future implementation in whole-cell

bioconversion systems[46] or even as a module enabling full formatotrophic growth, i.e. in conjunction with the RuMP pathway that has already been established in *E.coli*[34].

CHAPTER 2 lines up downstream of *CHAPTER 1* with its aim of finding an efficient enzyme for C1-C1 addition. OXC was previously identified as a highly efficient and promiscuous C1-elongating carboligase[82], suggesting it as a good template for an engineered GCS activity. After establishing basal GCS activity, the well-characterized active site of OXC can be subjected to structure-based enzyme engineering to improve efficiency. An improved GCS can then be implemented into existing whole-cell bioconversion systems[80], acting as a funnel for C1 compounds into value-added product or central metabolism.

In conjunction, the two chapters provide an efficient, low-energy option for synthetic C1 assimilation using formate as feedstock.

References

[1] Williams, C. L., Rachel, M. E., and Jaya Shankar, T. (2017) Biomass Compositional Analysis for Conversion to Renewable Fuels and Chemicals, In *Biomass Volume Estimation and Valorization for Energy* (Jaya Shankar, T., Ed.), IntechOpen, Rijeka.

[2] Zhang, D., Duan, D., Huang, Y., Yang, Y., and Ran, Y. (2017) Composition and structure of natural organic matter through advanced nuclear magnetic resonance techniques, *Chem. Biol. Technol. Agric. 4*, 8.

[3] Jozala, A. F., Geraldes, D. C., Tundisi, L. L., Feitosa, V. A., Breyer, C. A., Cardoso, S. L., Mazzola, P. G., Oliveira-Nascimento, L., Rangel-Yagui, C. O., Magalhaes, P. O., Oliveira, M. A., and Pessoa, A., Jr. (2016) Biopharmaceuticals from microorganisms: from production to purification, *Braz. J. Microbiol. 47 Suppl 1*, 51-63.

[4] Kumar, R., and Kumar, P. (2017) Future Microbial Applications for Bioenergy Production: A Perspective, *Front. Microbiol. 8*, 450.

[5] Urtuvia, V., Villegas, P., Gonzalez, M., and Seeger, M. (2014) Bacterial production of the biodegradable plastics polyhydroxyalkanoates, *Int. J. Biol. Macromol. 70*, 208-213.

[6] Ru, J., Huo, Y., and Yang, Y. (2020) Microbial Degradation and Valorization of Plastic Wastes, *Front. Microbiol. 11*, 442.

[7] Nduko, J. M., and Taguchi, S. (2020) Microbial Production of Biodegradable Lactate-Based Polymers and Oligomeric Building Blocks From Renewable and Waste Resources, *Front. Bioeng. Biotechnol. 8*, 618077.

[8] Zeenat, Elahi, A., Bukhari, D. A., Shamim, S., and Rehman, A. (2021) Plastics degradation by microbes: A sustainable approach, *Journal of King Saud University - Science 33*, 101538.

[9] Bart, H.-J. (2011) Extraction of Natural Products from Plants – An Introduction, In *Industrial Scale Natural Products Extraction*, 1-25.

[10] Breitling, R., and Takano, E. (2016) Synthetic Biology of Natural Products, *Cold Spring Harb. Perspect. Biol. 8*, a023994.

[11] Wu, Q., Liu, B.-K., and Lin, X.-F. (2010) Enzymatic Promiscuity for Organic Synthesis and Cascade Process, *Curr. Org. Chem. 14*, 1966-1988.

[12] Dalby, P. A. (2011) Strategy and success for the directed evolution of enzymes, *Curr. Opin. Struct. Biol. 21*, 473-480.

[13] Fisher, A. K., Freedman, B. G., Bevan, D. R., and Senger, R. S. (2014) A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories, *Comput. Struct. Biotechnol. J. 11*, 91-99.

[14] Nielsen, J. (2001) Metabolic engineering, *Appl. Microbiol. Biotechnol. 55*, 263-283.

[15] Woolston, B. M., Edgar, S., and Stephanopoulos, G. (2013) Metabolic engineering: past and future, *Annu. Rev. Chem. Biomol. Eng. 4*, 259-288.

[16] Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012) eQuilibrator--the biochemical thermodynamics calculator, *Nucleic Acids Res. 40*, D770-775.

[17] Beber, M. E., Gollub, M. G., Mozaffari, D., Shebek, K. M., Flamholz, A. I., Milo, R., and Noor, E. (2022) eQuilibrator 3.0: a database solution for thermodynamic constant estimation, *Nucleic Acids Res. 50*, D603-D609.

[18] Noor, E., Bar-Even, A., Flamholz, A., Reznik, E., Liebermeister, W., and Milo, R. (2014) Pathway thermodynamics highlights kinetic obstacles in central metabolism, *PLoS Comput. Biol. 10*, e1003483.

[19] Santos Correa, S., Schultz, J., Lauersen, K. J., and Soares Rosado, A. (2022) Natural carbon fixation and advances in synthetic engineering for redesigning and creating new fixation pathways, *J. Adv. Res.*

[20] Bar-On, Y. M., Phillips, R., and Milo, R. (2018) The biomass distribution on Earth, *Proc. Natl. Acad. Sci. U.S.A. 115*, 6506-6511.

[21] Calvin, M., and Benson, A. A. (1948) The Path of Carbon in Photosynthesis, *Science 107*, 476-480.

[22] Evans, M. C., Buchanan, B. B., and Arnon, D. I. (1966) A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium, *Proc. Natl. Acad. Sci. U.S.A. 55*, 928-934.

[23] Huber, H., Gallenberger, M., Jahn, U., Eylert, E., Berg, I. A., Kockelkorn, D., Eisenreich, W., and Fuchs, G. (2008) A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum Ignicoccus hospitalis, *Proc. Natl. Acad. Sci. U. S. A. 105*, 7851-7856.

[24] Berg, I. A., Kockelkorn, D., Buckel, W., and Fuchs, G. (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea, *Science 318*, 1782-1786.

[25] Holo, H. (1989) Chloroflexus aurantiacus secretes 3-hydroxypropionate, a possible intermediate in the assimilation of CO2 and acetate, *Arch. Microbiol. 151*, 252-256.

[26] Ljungdahl, L. G., and Wood, H. G. (1969) Total synthesis of acetate from CO2 by heterotrophic bacteria, *Annu. Rev. Microbiol. 23*, 515-538.

[27] 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.

[28] Sanchez-Andrea, I., Guedes, I. A., Hornung, B., Boeren, S., Lawson, C. E., Sousa, D. Z., Bar-Even, A., Claassens, N. J., and Stams, A. J. M. (2020) The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans, *Nat. Commun. 11*, 5090.

[29] Kemp, M. B., and Quayle, J. R. (1966) Microbial growth on C1 compounds. Incorporation of C1 units into allulose phosphate by extracts of Pseudomonas methanica, *Biochem. J. 99*, 41-48.

[30] Van Dijken, J. P., Harder, W., Beardsmore, A. J., and Quayle, J. R. (1978) Dihydroxyacetone: An intermediate in the assimilation of methanol by yeasts?, *FEMS Microbiol. Lett. 4*, 97-102.

[31] Large, P. J., Peel, D., and Quayle, J. R. (1962) Microbial growth on C(1) compounds. 3. Distribution of radioactivity in metabolites of methanol-grown Pseudomonas AM1 after incubation with [C]methanol and [C]bicarbonate, *Biochem. J. 82*, 483-488.

[32] Yishai, O., Goldbach, L., Tenenboim, H., Lindner, S. N., and Bar-Even, A. (2017) Engineered Assimilation of Exogenous and Endogenous Formate in Escherichia coli, *ACS Synth. Biol. 6*, 1722-1731.

[33] 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.

[34] 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.

[35] 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.

[36] Flamholz, A. I., Dugan, E., Blikstad, C., Gleizer, S., Ben-Nissan, R., Amram, S., Antonovsky, N., Ravishankar, S., Noor, E., Bar-Even, A., Milo, R., and Savage, D. F. (2020) Functional reconstitution of a bacterial CO2 concentrating mechanism in Escherichia coli, *eLife 9*, e59882.

[37] Satanowski, A., Dronsella, B., Noor, E., Vogeli, B., He, H., Wichmann, P., Erb, T. J., Lindner, S. N., and Bar-Even, A. (2020) Awakening a latent carbon fixation cycle in Escherichia coli, *Nat. Commun. 11*, 5812.

[38] Hult, K., and Berglund, P. (2007) Enzyme promiscuity: mechanism and applications, *Trends Biotechnol. 25*, 231-238.

[39] 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.

[40] Cai, T., Sun, H., Qiao, J., Zhu, L., Zhang, F., Zhang, J., Tang, Z., Wei, X., Yang, J., Yuan, Q., Wang, W., Yang, X., Chu, H., Wang, Q., You, C., Ma, H., Sun, Y., Li, Y., Li, C., Jiang, H., Wang, Q., and Ma, Y. (2021) Cell-free chemoenzymatic starch synthesis from carbon dioxide, *Science 373*, 1523-1527.

[41] Luo, S., Lin, P. P., Nieh, L.-Y., Liao, G.-B., Tang, P.-W., Chen, C., and Liao, J. C. (2022) A cell-free self-replenishing CO2-fixing system, *Nat. Catal. 5*, 154-162.

[42] 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.

[43] Miller, T. E., Beneyton, T., Schwander, T., Diehl, C., Girault, M., McLean, R., Chotel, T., Claus, P., Cortina, N. S., Baret, J. C., and Erb, T. J. (2020) Light-powered CO2 fixation in a chloroplast mimic with natural and synthetic parts, *Science 368*, 649-654.

[44] 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. 60*, 1-13.

[45] Pandit, A. V., Srinivasan, S., and Mahadevan, R. (2017) Redesigning metabolism based on orthogonality principles, *Nat. Commun. 8*, 15188.

[46] Chou, A., Lee, S. H., Zhu, F., Clomburg, J. M., and Gonzalez, R. (2021) An orthogonal metabolic framework for one-carbon utilization, *Nat. Metab. 3*, 1385-1399.

[47] Wenk, S., Rainaldi, V., He, H., Schann, K., Bouzon, M., Döring, V., Lindner, S. N., and Bar-Even, A. (2022) Synthetic carbon fixation via the autocatalytic serine threonine cycle, *bioRxiv*, 2022.2009.2028.509898.

[48] Xiao, L., Liu, G., Gong, F., Zhu, H., Zhang, Y., Cai, Z., and Li, Y. (2022) A Minimized Synthetic Carbon Fixation Cycle, *ACS Catal. 12*, 799-808.

[49] Yu, H., Li, X., 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, *Nat. Commun. 9*, 2008.

[50] 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.

[51] Jo, H.-J., Kim, J.-H., Kim, Y.-N., Seo, P.-W., Kim, C.-Y., Kim, J.-W., Yu, H.-n., Cheon, H., Lee, E. Y., Kim, J.-S., and Park, J.-B. (2022) Glyoxylate carboligase-based whole-cell biotransformation of formaldehyde into ethylene glycol via glycolaldehyde, *Green Chem. 24*, 218-226.

[52] 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.

[53] Hu, G., Li, Z., Ma, D., Ye, C., Zhang, L., Gao, C., Liu, L., and Chen, X. (2021) Light-driven CO2 sequestration in Escherichia coli to achieve theoretical yield of chemicals, *Nat. Catal. 4*, 395-406.

[54] Hu, G., Guo, L., Gao, C., Song, W., Liu, L., and Chen, X. (2022) Synergistic Metabolism of Glucose and Formate Increases the Yield of Short-Chain Organic Acids in Escherichia coli, *ACS Synth. Biol. 11*, 135-143.

[55] Güner, S., Wegat, V., Pick, A., and Sieber, V. (2021) Design of a synthetic enzyme cascade for the in vitro fixation of a C1 carbon source to a functional C4 sugar, *Green Chem. 23*, 6583-6590.

[56] Wang, X., Wang, Y., Liu, J., Li, Q., Zhang, Z., Zheng, P., Lu, F., and Sun, J. (2017) Biological conversion of methanol by evolved Escherichia coli carrying a linear methanol assimilation pathway, *Bioresour. Bioprocess. 4*.

[57] 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.

[58] 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.

[59] Claassens, N. J., Cotton, C. A. R., Kopljar, D., and Bar-Even, A. (2019) Making quantitative sense of electromicrobial production, *Nat. Catal. 2*, 437-447.

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

[61] 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.

[62] 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, *Chem. Rev. 115*, 12936-12973.

[63] Artz, J., Muller, T. E., Thenert, K., Kleinekorte, J., Meys, R., Sternberg, A., Bardow, A., and Leitner, W. (2018) Sustainable Conversion of Carbon Dioxide: An Integrated Review of Catalysis and Life Cycle Assessment, *Chem. Rev. 118*, 434-504.

[64] Schaub, T. (2018) CO2-based hydrogen storage: CO2 hydrogenation to formic acid, formaldehyde and methanol, *Phys. Sci. Rev. 3*, 20170015.

[65] Ljungdahl, L. G., and Andreesen, J. R. (1978) Formate dehydrogenase, a selenium-tungsten enzyme from Clostridium thermoaceticum, In *Meth. Enzymol.* (Fleischer, S., and Packer, L., Eds.), 360-372, Academic Press.

[66] 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.

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

[68] 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.

[69] He, H., Noor, E., Ramos-Parra, P. A., Garcia-Valencia, L. E., Patterson, J. A., Diaz de la Garza, R. I., Hanson, A. D., and Bar-Even, A. (2020) In Vivo Rate of Formaldehyde Condensation with Tetrahydrofolate, *Metabolites 10*, 65.

[70] Poust, S., Piety, J., Bar-Even, A., Louw, C., Baker, D., Keasling, J. D., and Siegel, J. B. (2015) Mechanistic Analysis of an Engineered Enzyme that Catalyzes the Formose Reaction, *Chembiochem 16*, 1950-1954.

[71] Bierbaumer, S., Nattermann, M., Schulz, L., Zschoche, R., Erb, T. J., Winkler, C. K., Tinzl, M., and Glueck, S. M. (2022) Enzymatic CO2 Fixation: from Nature to Application, *Manuscript in revision*.

[72] Blakley, R. L. (1960) Spectrophotometric studies on the combination of formaldehyde with tetrahydropteroylglutamic acid and other hydropteridines, *Biochem J 74*, 71-82.

[73] Wang, J., Anderson, K., Yang, E., He, L., and Lidstrom, M. E. (2021) Enzyme engineering and in vivo testing of a formate reduction pathway, *Synth. Biol. 6*, ysab020.

[74] Chittori, S., Savithri, H. S., and Murthy, M. R. (2012) Structural and mechanistic investigations on Salmonella typhimurium acetate kinase (AckA): identification of a putative ligand binding pocket at the dimeric interface, *BMC Struct Biol 12*, 24.

[75] Evans, D. J. (2005) Chemistry relating to the nickel enzymes CODH and ACS, *Coord. Chem. Rev. 249*, 1582-1595.

[76] Kikuchi, G. (1973) The glycine cleavage system: composition, reaction mechanism, and physiological significance, *Mol. Cell Biochem. 1*, 169-187.

[77] Kochetov, G. A., and Solovjeva, O. N. (2014) Structure and functioning mechanism of transketolase, *Biochim. Biophys. Acta 1844*, 1608-1618.

[78] Schellenberger, A. (1967) Structure and Mechanism of Action of the Active Center of Yeast Pyruvate Decarboxylase, *Angew. Chem. Int. Ed. 6*, 1024-1035.

[79] Bystrykh, L. V., Sokolov, A. P., and Trotsenko, Y. A. (1981) Purification and properties of dihydroxyacetone synthase from the methylotrophic yeast Candida boidinii, *FEBS Lett. 132*, 324-328.

[80] 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.

[81] Nattermann, M., Burgener, S., Pfister, P., Chou, A., Schulz, L., Lee, S. H., Paczia, N., Zarzycki, J., Gonzalez, R., and Erb, T. J. (2021) Engineering a Highly Efficient Carboligase for Synthetic One-Carbon Metabolism, *ACS Catal. 11*, 5396-5404.

[82] 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. 59*, 5526-5530.

[83] Comer, A. D., Long, M. R., Reed, J. L., and Pfleger, B. F. (2017) Flux Balance Analysis Indicates that Methane Is the Lowest Cost Feedstock for Microbial Cell Factories, *Metab. Eng. Commun. 5*, 26-33.

[84] Zhang, C., Ottenheim, C., Weingarten, M., and Ji, L. (2022) Microbial Utilization of Next-Generation Feedstocks for the Biomanufacturing of Value-Added Chemicals and Food Ingredients, *Front. Bioeng. Biotechnol. 10*, 874612.

[85] Lemaire, O. N., Jespersen, M., and Wagner, T. (2020) CO(2)-Fixation Strategies in Energy Extremophiles: What Can We Learn From Acetogens?, *Front. Microbiol. 11*, 486.

[86] Johnson, Z. J., Krutkin, D. D., Bohutskyi, P., and Kalyuzhnaya, M. G. (2021) Chapter Eight - Metals and methylotrophy: Via global gene expression studies, In *Meth. Enzymol.* (Cotruvo, J. A., Ed.), 185-213, Academic Press.

Chapter 1

A new-to-nature cascade for phosphate-dependent formate reduction at the thermodynamic limit

Maren Nattermann[1], Sebastian Wenk[2], Pascal Pfister[1], Nils Guntermann[3], Lennart Nickel[4], Charlotte Wallner[5], Jan Zarzycki[1], Nicole Pazcia[1], Giancarlo Franciò[3], Walter Leitner[3,6], Arren Bar-Even[2] and Tobias J. Erb[1,7]*

[1] Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

[2] Max Planck Institute of Molecular Plant Physiology, Golm, Germany

[3] Institute of Technical and Macromolecular Chemistry, RWTH Aachen University, Aachen, Germany

[4] Ruprecht Karl University, Heidelberg, Germany

[5] Philipps University, Marburg, Germany

[6] Max Planck Institute for Chemical Energy Conversion, Mülheim an der Ruhr, Germany

[7] Center for Synthetic Microbiology (SYNMIKRO), Marburg, Germany

In revision, Nature Communications 2022

Author contributions

M.N., A.B.E. and T.J.E. conceptualized the project and wrote the manuscript with contributions from all co-authors. M.N. performed the majority of the experiments. S. W. designed strains and experiments. P.P. and J.Z. collected crystallographic data, solved, and refined the structures. N.G., G.F. and W.L. established and carried out formyl phosphate synthesis. W.L. advised on related chemical pathways. N.P. performed LC-MS measurements. C.W. and L.N. carried out LC-MS and preliminary experiments.

Abstract

To realize a carbon-neutral bio-economy, the concept of electrobiocatalysis has been developed, in which CO2 is first chemically reduced to formate and subsequently assimilated by engineered microbes. A key step in the assimilation of formate is its reduction into formaldehyde, which is chemically challenging. Here, we developed a two-enzyme route in which formate is activated into formyl phosphate and reduced by NAD(P)H. The formyl phosphate route operates at the thermodynamically lowest limit compared to any other C1-activation pathway. Exploiting promiscuity of acetate kinase and N-acetyl-γ-glutamyl phosphate reductase, we demonstrate the phosphate route *in vitro* and *in vivo*. Through enzyme engineering, we created a reductase variant with improved formyl phosphate conversion *in vivo* and only little cross-talk with native metabolism. We further show that our route can be interfaced with recently developed pathways to provide the kinetically and thermodynamically most efficient route from formate into C2-compounds demonstrated to date.

Introduction

One of the major goals for establishing a carbon-neutral (or carbon-positive) bio-economy is the valorization of carbon dioxide as a feedstock. While a variety of novel CO_2-fixing cycles and microorganisms have been described recently [1-3], the direct use of CO_2 in biotechnology still poses several challenges and limitations, among others its low solubility in water. An alternative is the chemical conversion of CO_2 into soluble one-carbon (C1) compounds that can be further converted by biological platform organisms. One of the most promising C1-molecules in such chemo-bio hybrid approaches is formate. Formate can be produced electrochemically or by catalytic hydrogenation[4-6], is fully miscible with water, non-flammable and of little toxicity to the human worker[7].

To develop a formate-based bio-economy, recent efforts have focused on converting common platform organisms, such as *Escherichia coli* into synthetic formatotrophs that accept formate as both carbon and energy source [8, 9]. Formatotrophy naturally relies on the activation of formate to the tetrahydrofolate (T_4F) cofactor and subsequent reduction to methylene- T_4F (T_4F cascade, **Fig. 1**). Methylene-T_4F can be channeled into central metabolism directly via the serine cycle[10] or the reductive glycine pathway[11]. Alternatively, methylene-T_4F can spontaneously hydrolyze, releasing formaldehyde, which can enter central carbon metabolism through the ribulose monophosphate pathway[12], xylulose monophosphate pathway[13], homoserine cycle[14], or formolase pathways[15-19], although the hydrolysis rate is too slow to fully sustain growth. Aside biological application, formaldehyde is an attractive target for chemical synthesis that is difficult to access via purely chemical routes due to an over-reduction to methanol[20, 21].

Direct activation of formate to formaldehyde does not occur in nature, but would be highly desirable, as pathways would no longer depend on a spontaneous hydrolysis step. For that reason, different new-to-nature pathways have been proposed and/or developed. One option is the activation of formate to formyl-CoA, followed by the reduction to formaldehyde through the promiscuous activities of acetyl-CoA synthase (ACS) and acyl-CoA reductase (ACR) [15, 19, 22, 23]. Compared to the natural T_4F cascade, no hydrolysis is required and one enzyme less is needed. However, the CoA-based cascade requires hydrolysis of one additional phosphoanhydride bond for activation (ATP → AMP + PP_i, **Fig. 1**). While the CoA-based route

has been demonstrated in a proof-of-principle, residual activities of ACS with acetate and ACR with the native CoA metabolite pool (e.g., acetyl-CoA) currently limit the application of the CoA-based route *in vivo*[19].

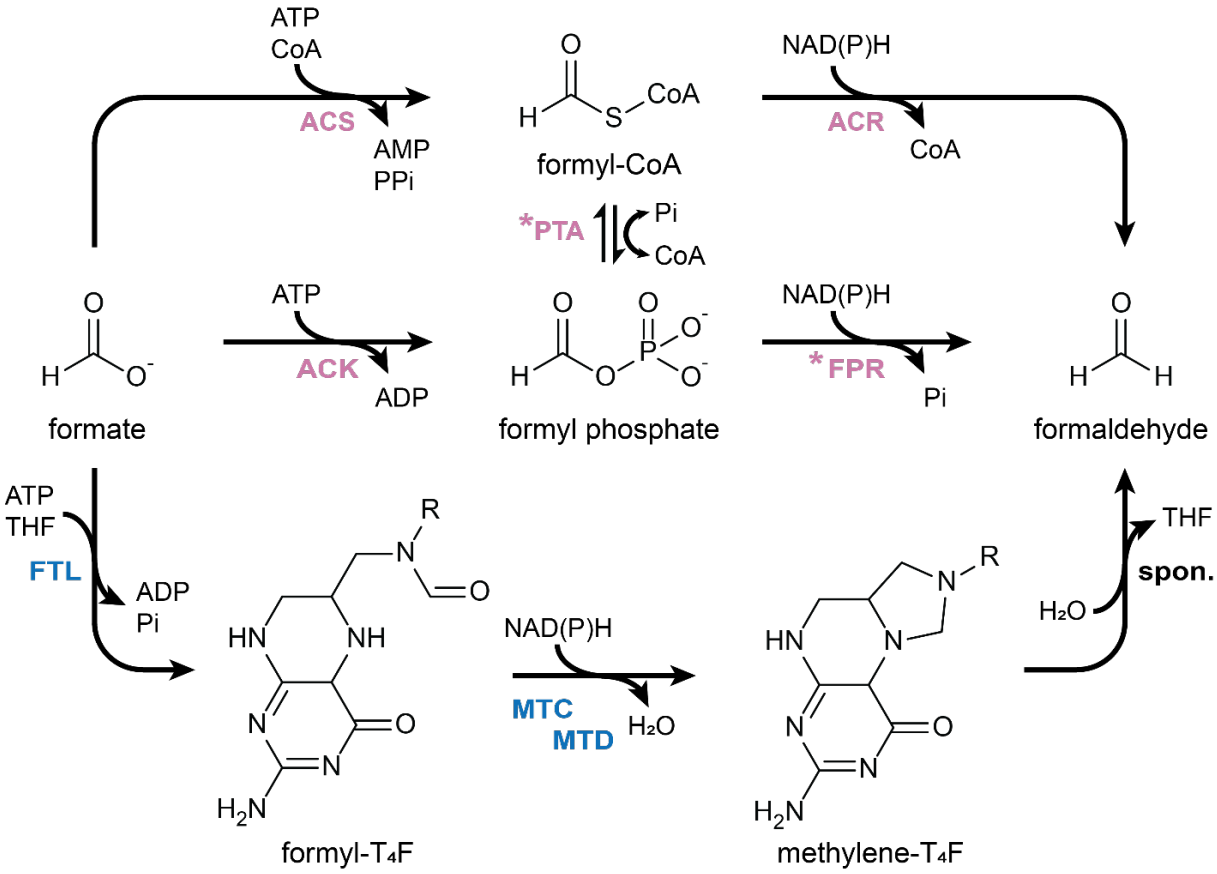


Figure 1 Pathways for the activation of formate to formaldehyde using reactive intermediates. Enzymes are indicated in color under the reaction arrows. Promiscuous activities are indicated in magenta, native reactions are given in blue. * indicates putative reactions. ACS = acetyl-CoA synthase, ACR = acyl-CoA reductase, PTA = phosphate acetyl transferase, ACK = acetate kinase, FPR = formyl phosphate reductase, FTL = formate tetrahydrofolate ligase, MTC = methylene tetrahydrofolate cyclohydrolase, MTD = methylene tetrahydrofolate dehydrogenase, T_4F = tetrahydrofolate, spon. = spontaneous hydrolysis.

Another way to activate carboxylic groups prior to reduction is through phosphorylation. Recently, a three-enzyme route was proposed, in which formate is first activated to formyl phosphate (using acetate kinase, ACK), transferred to CoA via phosphate acetyl transferase (PTA) and further reduced by ACR[24-26]. However, as this route also relies on an ACR, it faces

the same interference with the acetyl-CoA pool. To circumvent these challenges, the direct reduction of formyl phosphate would be a highly attractive alternative; however, such activity has not been described in the literature so far.

Here, we sought to establish the direct reduction of formyl phosphate into formaldehyde through engineering a formyl phosphate reductase (FPR). We described promiscuous FPR activity in four enzyme families and identified a variant of N-acetyl-γ-glutamyl-phosphate reductase from *Denitrovibrio acetiphilus* (DaArgC) that produces relevant levels of formaldehyde from formate *in vivo* when co-produced with ACK. Using a semi-rational enzyme engineering approach, we identified a triple variant (DaArgC3) that shows a complete loss of the native enzyme activity, a 300-fold shift in specificity towards formyl phosphate and decreased side-reactivity with acetyl phosphate. We further showed that in combination with ACK, DaArgC3 operates at the thermodynamically possible limit, outcompeting any other solution for the conversion of formate into formaldehyde. When interfaced with the FORCE pathway[26], our ACK-DaArgC3 route is kinetically and thermodynamically superior to any other cascade, thus representing the most efficient route towards C2-compounds from formate described thus far.

Materials & Methods

Chemicals

Chemicals were obtained from Sigma-Aldrich (Munich, Germany) and Carl Roth GmbH (Karlsruhe, Germany). Biochemicals and commercial enzymes were obtained from Thermo Fisher Scientific (St. Leon-Rot, Germany) and New England Biolabs GmbH (Frankfurt am Main, Germany). Gel and Plasmid purification kits were procured from Machery Nagel (Düren, Germany).

Strains and Plasmids

Strains and plasmids are listed in **Tables 1-2.** Primers (**Table 3**) were obtained from Eurofins MWG GmbH (Ebersberg, Germany). Gene fragments were ordered from TWIST Bioscience

(San Francisco, USA). Sequencing was performed by Microsynth Seqlab (Göttingen, Germany)

Table 1 Bacterial strains

strain	genotype	use	manufacturer
BL21 (DE3)	*E. coli str. B F– ompT gal dcm lon hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)*	protein production	Invitrogen
BL21 (DE3) ΔfrmRAB	*BL21 (DE3) ΔfrmRAB::Kan*	protein production	Ari Satanowski, MPI Marburg
BL21 (DE3) ΔfrmRAB ΔpflAB	*BL21 (DE3) ΔfrmRAB ΔpflAB*	protein production	Hai He, MPI Marburg
DH5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	vector propagation, cloning	Invitrogen

Restriction ligation cloning (RELC)

In vivo construct pCDFDuet-1_EcAckA_DaArgC was constructed using standard RELC. First, pCDFDuet-1 and pET28a_DaArgC were digested at 37 °C, 1 h with NdeI and XhoI, with FastAP supplemented to pCDFDuet-1 (Thermo Scientific), gel purified and ligated at 16 °C, overnight using T4 ligase (NEB). Ligation reactions were heat-inactivated at 65 °C and transformed into chemocompetent *E. coli* DH5α. Individual clones were evaluated for successful integration using Microsynth T7term primer. In the next step, RELC was performed with the resulting pCDFDuet-1_DaArgC and pCA24N_EcAckA using BamHI and HindIII as described above. Successful integration was evaluated by sequencing with DuetDOWN1 (AddGene)

Golden Gate (GG) vector construction

The primer and vector combinations for GG vector construction are given in **Table 4**.

GG parental vectors pCDFDuet-1_EcAckA_GFP and pET28a_GFP were constructed by Golden Gate cloning. Fragments were amplified using NEB Q5 polymerase in a FlexCycler² (Analytic Jena). Annealing temperatures were calculated by NEB Tm calculator v1.13.0. PCR product was purified and eluted in water. GG reactions were set up using NEB T4 ligase and

buffer as well as NEB BsaI-HF V2. Cycles were 15x 1.5 min, 37°C and 3 min, 16 °C. After ligation, the reactions were kept at 16 °C. GG reaction was transformed into chemocompetent *E. coli* DH5α without purification. Successful integration was evaluated by green/white screening of the colonies and sequencing with Microsynth primer T7term.

For GG constructs based on the parental constructs, gene fragments were either produced in the identical fashion or obtained with appropriate GG overhangs and codon-optimized for *E. coli* from TWIST Biosciences. In the GG reaction, after 15 cycles, T4 ligase was inactivated by heat-shock at 50 °C, 5 min, permitting removal of relegated BsaI sites. Finally, BsaI was inactivated at 80°C and the reactions were treated as described above.

ISM library generation of DaArgC

pCDFDuet-1_EcAckA_DaArgC was mutated using degenerated primers. During the course of the ISM, primers were adjusted to include the mutations found during screening. PCR reactions were run in a FlexCycler² (Analytic Jena) using the saturation mutagenesis protocol for NEB Q5 polymerase and annealing temperatures calculated by the NEB Tm calculator v1.13.0. PCR products were gel purified and digested at 37 °C overnight using Thermo Fisher DpnI. After PCR purification, the DNA was transformed into *E. coli* DH5α and evaluated for successful integration by sequencing with Microsynth T7term primer.

Table 2 Plasmids

name	relevant genotype	UniProt-ID insert	source
pCA24N_EcAckA	CamR, *Escherichia coli* ackA, Plac		ASKA collection[27]
pCA24N_ecASD	CamR, *Escherichia coli* asd, Plac		ASKA collection[27]
pCA24N_EcArgC	CamR, *Escherichia coli* argC, Plac		ASKA collection[27]
pCA24N_ecGapA	CamR, *Escherichia coli* gapA, Plac		ASKA collection[27]
pCA24N_ecProA	CamR, *Escherichia coli* proA, Plac		ASKA collection[27]
pCDFDuet-1	StrepR, lacI	-	Novagen
pCDFDuet-1_EcAckA_GFP	StrepR, *Escherichia coli* ackA, sfGFP, Plac	P0A6A3	This study
pCDFDuet-1_EcAckA_AtArgC	StrepR, *Escherichia coli* ackA, *Arabidopsis thaliana* argC, Plac	P0A6A3/ Q93Z70	This study
pCDFDuet-1_EcAckA_BcArgC	StrepR, *Escherichia coli* ackA, *Bacillus clausii* argC, Plac	P0A6A3/ Q5WEW7	This study
pCDFDuet-1_EcAckA_CsArgC	StrepR, *Escherichia coli* ackA, *Caldicellulosiruptor saccharolyticus* argC, Plac	P0A6A3/ A4XJN9	This study
pCDFDuet-1_EcAckA_DaArgC	StrepR, *Escherichia coli* ackA, *Denitrovibrio acetiphilus* argC, Plac	P0A6A3/ D4H3H4	This study
pCDFDuet-1_EcAckA_DaArgC1	StrepR, *Escherichia coli* ackA, *Denitrovibrio acetiphilus* argC G182V, Plac	P0A6A3/ D4H3H4	This study
pCDFDuet-1_EcAckA_DaArgC2	StrepR, *Escherichia coli* ackA, *Denitrovibrio acetiphilus* argC S178V G182V, Plac	P0A6A3/ D4H3H4	This study
pCDFDuet-1_EcAckA_DaArgC3	StrepR, *Escherichia coli* ackA, *Denitrovibrio acetiphilus* argC S178V G182V L233I, Plac	P0A6A3/ D4H3H4	This study
pCDFDuet-1_EcAckA_DaArgC2*	StrepR, *Escherichia coli* ackA, *Denitrovibrio acetiphilus* argC S178V, Plac	P0A6A3/ D4H3H4	This study
pCDFDuet-1_EcAckA_DaArgC3*	StrepR, *Escherichia coli* ackA, *Denitrovibrio acetiphilus* argC S178V L233I, Plac	P0A6A3/ D4H3H4	This study
pCDFDuet-1_EcAckA_EcArgC	StrepR, *Escherichia coli* ackA, *Escherichia coli* argC, Plac	P0A6A3/ P11446	This study
pCDFDuet-1_EcAckA_EcPTArgC	StrepR, *Escherichia coli* ackA, *Pseudidiomarina taiwanensis* argC, Plac	P0A6A3/ A0A432ZG32	This study
pET28a_GFP	KanR, sfGFP	-	This study
pET28a_AtArgC	KanR, *Arabidopsis thaliana* argC, Plac	Q93Z70	This study
pET28a_BcArgC	KanR, *Bacillus clausii* argC, Plac	Q5WEW7	This study
pET28a_CsArgC	KanR, *Caldicellulosiruptor saccharolyticus* argC, Plac	A4XJN9	This study
pET28a_DaArgC	KanR, *Denitrovibrio acetiphilus* argC, Plac	D4H3H4	This study
pET28a_DaArgC1	KanR, *Denitrovibrio acetiphilus* argC G182V, Plac	D4H3H4	This study
pET28a_DaArgC2	KanR, *Denitrovibrio acetiphilus* argC S178V G182V, Plac	D4H3H4	This study
pET28a_DaArgC3	KanR, *Denitrovibrio acetiphilus* argC S178V G182V L233I, Plac	D4H3H4	This study
pET28a_DaArgC2*	KanR, *Denitrovibrio acetiphilus* argC G182V, Plac	D4H3H4	This study
pET28a_DaArgC3*	KanR, *Denitrovibrio acetiphilus* argC S178V L233I, Plac	D4H3H4	This study
pET28a_EcPTArgC	KanR, *Pseudidiomarina taiwanensis* argC, Plac	A0A432ZG32	This study
pET28a_StAckA	KanR, *Salmonella typhimurium* ackA, Plac	P63411	TWIST Bioscience
pMC_V_01_Part Entry sfGFP	AmpR, sfGFP	-	Stukenberg et al.[28]

Table 3 Oligonucleotide primers. Bold font indicates BsaI recognition site, lines indicate oligonucleotide overhangs. Degenerate codons are indicated in italics.

name	Sequence
28a_RDM_fw	GAGCGAGAGAGACCGCTGCTAACAAAGCCCG
28a_RDM_rv	GAGCATAAGAGACCGTGATGATGATGATGATGGC
28a_BsaI_fw	GAGGGTCTCACGAGCACCACCACC
28a_BsaI_rv	GAGGGTCTCACATATGGTATATCTCCTTCTTAAAGTTAAAC
ASKA-argC_BsaI_fw	GTTGGTCTCATATGTTGAATACGCTGATTGTGGG
ASKA_CDFDuet_BsaI_rv	GACGGTCTCGCTCGAGTTAGCGGCCGCATAGG
AtArgC_fw	GTTGGTCTCATATGAGTACCGCAAGC
AtArgC_rv	GACGGTCTCGGCTCGAGCGGGAACAGGGGTTG
BcArgC_fw	GTTGGTCTCATATGGAAGTTGGCATTATTGG
BcArgC_rv	GACGGTCTCGCTCGAGCGGGTACATAGGAACTG
CsArgC_fw	GTTGGTCTCATATGATCAAAGCCAGCATTATTGG
CsArgC_rv	GACGGTCTCGCTCGAGAAAATATTCACCAACAATGACC
DaArgC_RDM_fw	GAGGGTCTCATATGAAGGTTTCGGTTATCGG
DaArgC_RDM_rv	GAGGGTCTCGCTCGAGTTATAAGATCCCG
DaArgC_fw	GTTGGTCTCATATGAAGGTTTCGGTTATCG
DaArgC_rv	GACGGTCTCGCTCGAGTAAGATCCCGAGTGTATCG
DaArgC_S178_NNK	GATCATTGCGGACTCCAAA*NNK*GGGGTGACCGGGGCCGGTC
DaArgC_S178_rv	GGAGTCCGCAATGATCCCCTCTGGTGAGATCAAGCCGGCC
DaArgC_G182_NNK	CTCCAAATCTGGGGTGACC*NNK*GCCGGTCGTAAGGCCGAC
DaArgC_G182_rv	GGTCACCCCAGATTTGGAGTCCGCAATGATCCCCTCTGG
DaArgC_A183_NNK	CAAATCTGGGGTGACCGGG*NNK*GGTCGTAAGGCCGACATTG
DaArgC_A183_rv	CCCGGTCACCCCAGATTTGGAGTCCGCAATGATCCCCTC
DaArgC_Y203_NNK	CAATGAGGACTTTCGTCCT*NNK*GCCATTTTCTCACACCGTC
DaArgC_Y203_rv	GGACGAAAGTCCTCATTGCACTCACAGAAGGAATACGC
DaArgC_L233_WDK	CGTCTTATTCACCCCGCAC*WDK*ATCCCTGCAAGTAAGGGC
DaArgC_L233_rv	GTGCGGGGTGAATAAGACGTTCGTCTCTTTTCCAGTCTC
DaArgC_GV_S178_NNK	GATCATTGCGGACTCCAAA*NNK*GGGGTGACCGTGGCCGGTC
DaArgC_GV_A183_NNK	CAAATCTGGGGTGACCGTG*NNK*GGTCGTAAGGCCGACATTG
DaArgC_GV_A183_rv	CACGGTCACCCCAGATTTGGAGTCCGCAATGATCCCCTC
DaArgC_SV-GV_A183_NNK	CAAAGTGGGGGTGACCGTG*NNK*GGTCGTAAGGCCGACATTG
DaArgC_SV-GV_A183_rv	CACGGTCACCCCCACTTTGGAGTCCGCAATGATCCCCTC
DaArgC_S178V-G182_fw	CAAAGTGGGGGTGACCGGGGCCGGTCGTAAGGCC
DaArgC_S178V-G182_rv	GGCCTTACGACCGGCCCCGGTCACCCCCACTTTG
GFP_BsaI_fw	GAGTATGAGAGACCTCGTGAGACGGAAAGTGAAAC
GFP_BsaI_rv	GAGCTCGAGAGACCCATGAGACGTATAAACGCAG
EcPTArgC_fw	GTTGGTCTCATATGCACCAGGAAAAAAGC
EcPTArgC_rv	GACGGTCTCGCTCGAGCAACACGGCACTACG
pCDF_BsaI_fw	GAGGGTCTCACGAGTCTGGTAAAGAAAC
pCDF_BsaI_rv	GAGGGTCTCACATAATGTATATCTCCTTCTTATAC
Duet_DOWN1	GATTATGCGGCCGTGTACAA
T7term	TGCTAGTTATTGCTCAGCGG
RDM_DaArgC_fw	CATCATCATCATCACGGTCTCTTATG
RDM_DaArgC_rv	GCAGCGGTCTCTCTCGAGTTA

Table 4 Golden Gate construction of plasmids. Gene fragments are indicated by *. Conditions given for DaArgC also apply to DaArgC variants.

construct	insert/template		primer fw	primer rv
DaArgC3_RDM_template	pCDFDuet-1_EcAckA_DaArgC3 pET28a		DaArgC_RDM_fw 28a_RDM_fw	DaArgC_RDM_rv 28a_RDM_rv
pCDFDuet-1_EcAckA_GFP	pCDFDuet-1_EcAckA_DaArgC pMC_V_01_Part sfGFP	Entry	pCDF_BsaI_fw GFP_BsaI_fw	pCDF_BsaI_rv GFP_BsaI_rv
pET28a_GFP	pET28a pMC_V_01_Part sfGFP	Entry	28a_BsaI_fw GFP_BsaI_fw	28a_BsaI_rv GFP_BsaI_rv
pCDFDuet-1_EcAckA_AtArgC	AtArgC* pCDFDuet-1_EcAckA_GFP			
pCDFDuet-1_EcAckA_BcArgC	BcArgC* pCDFDuet-1_EcAckA_GFP			
pCDFDuet-1_EcAckA_CsArgC	CsArgC* pCDFDuet-1_EcAckA_GFP			
pCDFDuet-1_EcAckA_DaArgC	DaArgC* pCDFDuet-1_EcAckA_GFP			
pCDFDuet-1_EcAckA_EcArgC	pCA24N_EcArgC pCDFDuet-1_EcAckA_GFP		ASKA-argC_BsaI_fw	ASKA_CDFDuet_BsaI_rv
pCDFDuet-1_EcAckA_EcPTArgC	EcPTArgC* pCDFDuet-1_EcAckA_GFP			
pET28a_GFP	pCDFDuet-1_EcAckA_GFP pET28a		GFP_BsaI_fw (pET28a) 28a_Cterm_BsaI_fw	GFP_BsaI_rv 28a_Cterm_BsaI_rv
pET28a_AtArgC	pCDFDuet-1_EcAckA_AtArgC pET28a_GFP		AtArgC_fw	AtArgC_rv
pET28a_BcArgC	pCDFDuet-1_EcAckA_BcArgC pET28a_GFP		BcArgC _fw	BcArgC_rv
pET28a_CsArgC	pCDFDuet-1_EcAckA_CsArgC pET28a_GFP		CsArgC_fw	CsArgC_rv
pET28a_DaArgC	pCDFDuet-1_EcAckA_DaArgC pET28a_GFP		DaArgC_fw	DaArgC_rv
pET28a_EcPTArgC	pCDFDuet-1_EcAckA_EcPTArgC pET28a_GFP		EcPTArgC_fw	EcPTArgC_rv

Random mutagenesis of DaArgC3

The template for random mutagenesis, DaArgC3_RDM_template, was constructed using a pET28a backbone by Golden Gate as depicted in **Table 4**. From this template, DaArgC3 was

randomized using primers RDM_DaArgC_fw and RDM_DaArgC_rv in a 50 µl reaction containing 1x Mg-free Taq buffer (NEB), 7 mM MgCl2, 50 µM MnCl2, 400 µM dATP and dGTP, 2 mM dTTP and dCTP, 400 µM of each primer, 5 ng/µl template and 1 µl Taq Polymerase (NEB). The PCR was cycled at 95 °C, 30 sec initial denaturation, 30 cycles of 95°C, 30 sec melting, 66 °C, 1 min annealing, 68 °C, 2 min extension, with a final extension of 68 °C, 10 min. The product was PCR purified and directly used in a Golden Gate reaction with pCDFDuet-1_EcAckA_GFP. After transformation, 10 individual clones were sent for sequencing to determine the error rate. A library of a size of 2500 clones was generated by washing the transformation plates with LB and performing a pooled DNA extraction.

Single site mutagenesis of DaArgC2 and DaArgC3

G182 was reconstructed in vectors pCDFDuet-1_EcAckA_DaArgC2 and pCDFDuet-1_EcAckA_DaArgC3 using primers DaArgC_S178V-G182_fw and DaArgC_S178V-G182_rv. Saturation mutagenesis was performed as described above. The resulting vector were pCDFDuet-1_EcAckA_DaArgC2* and pCDFDuet-1_EcAckA_DaArgC3* were used as template for onboarding of DaArgC2* and DaArgC3* into pET28a using the GG protocol given above.

Protein purification

Protein purification was carried out in *E.coli* BL21 DE3 using the method published before[29].

Phylogenetic analysis of the ArgC family

E.coli ArgC was analyzed using BLASTP [30] and 500 homologs were exported. Using hhfilter [31], all homologs with greater than 60% identity to the *E.coli* variant were excluded, resulting in 118 sequences. The tree was constructed in MEGA X[32]. The evolutionary history was inferred using the Neighbor-Joining method [33]. The bootstrap consensus tree inferred from 50 replicates was taken to represent the evolutionary history of the taxa analyzed[34]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed.

Evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were 1203 positions in the final dataset.

SDS-PAGE analysis of protein production

BL21 DE3 Δ*frmRAB* or BL21 DE3 Δ*frmRAB* Δ*pflAB* containing pCDFDuet-1_EcAckA_xxArgC were grown in 600 µl LB 50 µg/ml streptomycin in 2.0 ml 96 Deep Well Plates with V Bottom (Plate One) and grown overnight. The following day, the cultures were transferred 30 µl into 600 µl M9 20 mM glucose, 50 µg/ml streptomycin, grown for 2 h at 37°C, cooled to 25 °C, induced with 250 µM IPTG and grown for 16 h at 25 °C. After expression, cells were harvested and resuspended in 50 µl H_2O. 5 µl of the suspension were added to 5 µl of water and 10 µl of SDS loading buffer (Thermo Fisher) and heated at 95 °C for 30 min. 10 µl of the resulting sample were separated on a commercial SDS gel (BioRad) and stained with Simply Blue SafeStain (Invitrogen).

Crystallization of DaArgC and DaArgC3

DaArgC and DaArgC3 were purified as described above. Immediately after affinity purification, the eluate was loaded onto a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated in SEC Buffer (75 mM KCl, 25 mM HEPES-KOH pH 7.8). Fractions corresponding to dimeric enzyme were collected, pooled and concentrated to 10 mg/mL on 10,000 MWCO filters (Amicon Ultra). Enzyme purity was evaluated via SDS page. Crystal plates were set up using the sitting drop vapor diffusion method, diluting equal volume of enzyme in reservoir solution. The reservoir solution for DaArgC contained 0.2 M potassium bromide, 0.2 M potassium thiocyanate, 0.1 M sodium cacodylate pH 6.5, 3 % w/v γ-PGA (Na^+ form, LM) and 2 % v/v PEG 500 MME and the reservoir for DaArgC3 contained 0.2 M sodium chloride, 0.1 M HEPES pH 7.5 and 25%(w/v) PEG 4000.

X-ray diffraction data for DaArgC and DaArgC3 were collected at the Beamline ID30B (European Synchrotron Radiation Facility, Grenoble, France). All images were processed

using XDS [35]. The datasets were scaled using Scala of the CCP4 program suite [36]. The Phenix software package [37] was used to perform molecular replacement (PhaserMR) for phasing of both datasets by using the N-acetyl-γ-glutamyl-phosphate reductase from *Shigella flexneri* (PDB 3dr3) as search model [38]. Initial models were built with phenix.autobuild and refined with the phenix.refine and Coot[39]. Data collection and refinement statistics are provided in **Table 5**.

Table 5. Data collection and refinement statistics.

	DaArgC	DaArgC3
Data collection		
Beam line	ESRF ID30B	ESRF ID30B
Wavelength (Å)	0.97625	0.97625
Space group	I 2	P 2_1 2_1 2_1
Cell dimensions		
a, *b*, *c* (Å)	86.7, 77.8, 122.0	91.9, 109.6, 133.2
α, β, γ (°)	90.0, 92.9, 90.0	90.0, 90.0, 90.0
Unique reflections	55250 (5200)	66892 (6707)
Resolution (Å)	42.69 - 1.99 (2.06 - 1.99)	42.31 - 2.194 (2.27 - 2.19)
R_{merge}	0.02854 (0.2347)	0.03863 (0.1318)
$I / \sigma(I)$	13.45 (2.91)	13.50 (5.56)
$CC_{1/2}$	0.999 (0.936)	0.996 (0.952)
Completeness (%)	99.13 (93.33)	96.33 (97.87)
Redundancy	6.5 (6.5)	3.2 (3.2)
Refinement		
No. reflections	55208 (5191)	66884 (6707)
R_{work} / R_{free}	0.1843 / 0.2152	0.1883 / 0.2069
No. atoms (non-hydrogen)	5439	10561
Protein	5217	10023
Ligand/ion	2 Na^+	4 Na^+
Water	220	534
B-factors (average)	45.76	38.56
Protein	45.62	38.44
Ligand/ion	35.60	32.98
Water	49.11	40.84
R.m.s. deviations		
Bond lengths (Å)	0.014	0.004
Bond angles (°)	1.220	0.680
Rotamer outliers (%)	0.35	0.37
Ramachandran favored (%)	97.45	96.26
Ramachandran outliers (%)	0.30	0.31
PDB ID	8AFU	8AFV

Values in parentheses are for highest-resolution shell.

Formyl phosphate synthesis

To an aqueous solution of dipotassium phosphate (1 mL, 1 M, 1 equiv.), acetic formic anhydride (0.106 g, 1.2 equiv.) was added. The solution was mixed shortly and monitored by $^{31}P\{^1H\}$-NMR. The first measurement after 1 minute shows that 59% of dipotassium phosphate was converted to formyl phosphate. (**Supplementary Fig. 2-4**)

The decomposition of formyl phosphate was carried out under similar conditions applied in the enzymatic process. An aliquot of a solution of dipotassium phosphate prepared as described above (0.1 mL) was added to a solution of MOPS buffer 0.5 M and magnesium chloride 0.01 M (0.9 mL) mixed shortly and directly monitored via $^{31}P\{^{1}H\}$-NMR. The concentration of formyl phosphate linearly decreased at a rate of -0.34 $\%min^{-1}$ (**Supplementary Fig. 5**).

NMR analysis

^{1}H- NMR measurements for the formic acetic anhydride synthesis were conducted at room temperature on a Bruker AS 400 (Bruker Corporation, Billerica, MA, USA) spectrometer. The chemical shift δ is given in ppm and was referenced to the solvent residual signal. ^{1}H- and ^{31}P-measurements for the formyl phosphate formation and decomposition were conducted at 30 °C on a Bruker AV 300 spectrometer.

Spectrometric analysis

All photo-spectrometric assays were carried out in 100 µl of reaction volume in a High Precision Cell 10 mm Light Path quartz cuvette (Hellma Analytics) using a CARY 60 UV-Vis Spectrophotometer (Agilent) Kinetic program. Unless otherwise mentioned, components were added to the cuvette in the order that they are listed, and the last indicated component was used to start the reaction. NAD(P)H consumption or production was assessed by tracking the absorbance at 340 nm in 0.5 s intervals over time.

Kinetics of AckA on formate. Reactions were measured at 37 °C and contained 100 mM HEPES-KOH pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 250 µM NADH, 5 mM PEP, 1:40 PK/LDH mix and 0.1 µM ACKA. Formate concentrations were 3, 10, 20, 50, 150 and 300 mM. Initial velocities were determined and a Michaelis Menten kinetic was constructed.

Activity of DaArgC on N-acetyl-γ-glutamyl phosphate. Reactions were measured at 30 °C and contained 100 mM MOPS pH 7.0, 10 mM $MgCl_2$, 5 mM ATP, 250 µM NADPH,1 µM ArgB, 0.1 µM DaArgC and 50 mM N-acetyl-γ-glutamate.

Activity of DaArgC on acetyl phosphate. The Assay was carried out at 30 °C in an Infinite M PLEX plate reader (Tecan) on half-well plates (Greiner) with a liquid column height of 4 mm. Reactions contained 100 mM MOPS pH 7.0, 10 mM $MgCl_2$, 250 μM NADPH, 1 μM DaArgC and 50 mM acetyl phosphate.

Formaldehyde detection via Nash assay

Formaldehyde was detected by reaction with the Nash reagent: 2 M ammonium acetate, 20 mM acetyl acetone and 50 mM acetic acid. Samples were diluted 1:1 into the reagent. Where required, a formaldehyde calibration curve in the sample matrix was added for quantification. After incubation for 1 h at 37 °C, the brightly yellow diacetyldihydrolutidine was formed, which was detected by fluorescence measurements at 412 nm excitation/ 505 nm emission in both 96-well and 384-well formats in an Infinite M PLEX plate reader (Tecan). Side-reactivity with formyl phosphate, formyl-CoA or acetaldehyde was not observed (**Fig. SI 6**).

In vitro reactions

Reactions were carried out at 30°C and started by addition of formate. Samples were taken in regular intervals as indicated below. Samples were quenched 1:1 in 10% formic acid, then one volume equivalent of Nash reagent was added for detection.

Comparison of formyl phosphate reductases. Reactions contained 100 mM MOPS-KOH pH 7.0, 10 mM $MgCl_2$, 5 mM ATP, 1 mM CoA, 250 μM NAD(P)H, 50 mM formate, 2 μM EcAckA and 2 μM of ArgC, ASD, GapA or ProA. Time points were taken at 1, 5, 10 and 15 min.

Comparison of ArgC homologs. Reactions contained 100 mM MOPS-KOH pH 7.0, 10 mM $MgCl_2$, 5 mM ATP, 250 μM NADPH, 50 mM formate, 1 μM EcAckA and 5 μM ArgC. Time points were taken at 1.5, 5, 10 and 20 min.

Characterization of DaArgC. Reactions contained 100 mM MOPS-KOH pH 7.0, 10 mM $MgCl_2$, 5 mM ATP, 250 μM NADPH, 200 mM ammonium formate (and/ or 200 mM ammonium acetate), 10 μM EcAckA and 1 μM DaArgC (if doubled; 20 or 2 μM enzyme). Time points were taken at 0.5, 1, 2 and 3 min.

***In vitro* characterization of variants on NADPH and NADH.** Reactions contained 100 mM MOPS-KOH pH 7.0, 10 mM $MgCl_2$, 5 mM ATP, 250 μM NADPH or 250 μM NADH, 50 mM ammonium formate, 2 μM EcAckA and 2 μM DaArgC. Time points were taken at 1, 5, 10 and 15 min.

Kinetic characterization of DaArgC variants. After measurements, initial slopes were calculated and Michaelis Menten plots were constructed. **Formyl phosphate.** Formyl phosphate was produced as described above. The synthesis was incubated for 1 min before being added 1:10 into the enzymatic reaction containing 500 μM MOPS pH 7.0, 10 mM $MgCl_2$, 1 mM NADPH and 5 μM DaArgC variant. Time points were taken at 1, 2, and 3 min. **NADPH.** Reactions contained 100 mM MOPS-KOH pH 7.0, 10 mM $MgCl_2$, 5 mM ATP, 500 mM ammonium formate, 10 μM EcAckA and 1 μM of DaArgC variants. NADPH was added in the range of 2.5 μM up to 500 μM. Time points were taken at 1, 2 and 3 min for DaArgC, DaArgC2* and DaArgC3* and 1, 3 and 5 min for DaArgC1-3. **NADH.** Reactions contained 100 mM MOPS-KOH pH 7.0, 10 mM $MgCl_2$, 5 mM ATP, 250 mM ammonium formate, 10 μM EcAckA and 1 μM of DaArgC variants NADH was added in the range of 10 μM up to 2000 μM. Time points were taken at 1, 4 and 8 min. Initial slopes were calculated and Michaelis Menten plots were constructed.

Lysate reactions

Lysate activity of pCDFDuet-1_EcAckA_DaArgC. pCDFDuet-1_EcAckA_DaArgC was transformed into BL21 DE3 Δ*frmRAB*. The next day, 30 ml overexpression culture was grown to OD in LB at 37 °C, cooled to 25 °C, induced by addition of 250 μM IPTG and incubated overnight at 25 °C. The next day, the culture was spun down and the resulting pellet resuspended in 500 μl CellLytic™ B Cell Lysis Reagent (Sigma Aldrich). After incubation at room temperature for 10 min, the cell debris was spun down and the lysate stored on ice. Formaldehyde production was assayed in reactions containing 100 mM MOPS-KOH pH 7.0, 10 mM $MgCl_2$, 5 mM ATP, 250 μM NADPH, 50 mM formate and 1:10 lysate. Two controls were additionally supplemented with 5 μM purified EcAckA or DaArgC. Time points were taken at 0, 20, 40 and 60 min.

In vivo reactions

pCDFDuet-1 and pCDFDuet-1_EcAckA_DaArgCx were transformed into BL21 DE3 Δ*frmRAB* or BL21 DE3 Δ*frmRAB* Δ*pflAB*. LB overnight cultures were set up in triplicates. The next day, 600 µl M9 + 20 mM glucose was inoculated with 30 µl overnight culture and grown on 2.0 ml 96 Deep Well Plates with V Bottom (Plate One) for 2 h at 37 °C. Cells were induced by addition of 250 µM IPTG, and incubated 16 h at 25 °C. The next day, formate was added to the reaction, incubated at 30 °C and time points were taken as indicated, stored at -20°C and then quantified.

***In vivo* activity of pCDFDuet-1_EcAckA_DaArgC.** 50 mM ammonium formate was added to the cells. Time points were taken at 0, 1, 2 and 4 h. The identical method was applied to evaluate the *in vivo* activity of the ArgC homologs.

ISM of DaArgC. Single colonies of transformed ISM libraries were directly inoculated into 600 µl M9 + 20 mM glucose and grown for 8 h at 37°C before proceeding as described above. 50, 25 or 5 mM of ammonium formate were added to the supernatant. Samples were taken at 2 or 4 h.

RDM of DaArgC. Libraries and controls were transformed into BL21 DE3 Δ*frmRAB*. Using a PIXL Colony Picking Robot (Singer Instruments), cell were picked into 100 µl LB 50 µg/ml Kanamycin, 50 µg/ml streptomycin in a 384-well format. Cells were grown overnight at 1000 rpm in an Infors Multitron incubator set to 80% humidity. From the master plate, 30 µl culture were transferred into 600 µl M9 50 µg/ml Kanamycin, 50 µg/ml streptomycin + 20 mM glucose and at 37 °C for 2 h on 2.0 ml 96 Deep Well Plates with V Bottom (Plate One). Cultures were induced with 250 µM IPTG and incubated 16 h at 25 °C. 5 mM ammonium formate was added and samples were taken at 4 h.

***In vivo* activity of pCDFDuet-1_EcAckA_DaArgC variants.** 500, 50, 5 or 0 mM ammonium formate were added to the cells. Time points were taken at 0, 1, 2 and 4 h. Slopes were calculated using linear regression.

Statistics

Data was evaluated in GraphPad Prism 9. Data was fit using the linear regression and Michaelis Menten tools using standard settings. Where relevant, mean and standard error were calculated by the program.

LC-MS detection of glycolyl-CoA

Glycolyl-CoA production was detected by LC-MS. Reactions contained 100 mM MOPS-KOH pH 7.0, 10 mM MgCl2, 150 µM ThDP, 0.5 mM ADP, 5 mM ATP, 0.5 mM CoA, 250 µM NADPH, NAD+ and/or NADH and/or 0.5 mM $K_xH_yPO_4$ pH 6.9, 10 µM MeOXC4 and 5 µM of LmACR, EcAckA, DaArgC, BsACS and/or EcPTA as well as 200 mM NH_4 formate.

Quantitative determination of glycolyl-CoA was performed using a LC-MS/MS. The chromatographic separation was performed on an Agilent Infinity II 1290 HPLC system using a Kinetex EVO C18 column (150 × 1.7 mm, 3 µm particle size, 100 Å pore size, Phenomenex) connected to a guard column of similar specificity (20 × 2.1 mm, 5 µm particle size, Phenomoenex) a constant flow rate of 0.25 ml/min with mobile phase A being 0.1 % formic acid in water and phase B being 0.1 % formic acid methanol (Honeywell, Morristown, New Jersey, USA) at 40° C. The injection volume was 2 µl. The mobile phase profile consisted of the following steps and linear gradients: 0 – 1 min constant at 5 % B; 1 – 4.5 min from 5 to 80 % B; 4.5 – 5.5 min constant at 80 % B; 5.5 – 5.6 min from 80 to 5 % B; 5.6 to 8 min constant at 5 % B. An Agilent 6470 mass spectrometer was used in positive mode with an electrospray ionization source and the following conditions: ESI spray voltage 5500 V, nozzle voltage 500 V, sheath gas 400° C at 11 l/min, nebulizer pressure 20 psig and drying gas 100° C at 5 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. Mass transitions, collision energies, Cell accelerator voltages and Dwell times have been optimized using chemically pure standards. Parameter settings are given in **Table 6**.

Table 6 Parameter settings for LC-MS detection of glycolyl-CoA

Compound	Quantifier	Collision Energy	Qualifier	Collision Energy	Dwell	Fragmenter voltage	Cell Accelerater Volatege
Glycolyl-CoA	826.1→319.4	24	826.1→428.4	25	200	140	5

Results

Acetate kinase as formate kinase

To establish the formyl phosphate route, we first focused on identifying a formate kinase (FOK) that could catalyze the phosphorylation of formate. Only very few examples of FOK activity have been reported, however, no corresponding genes have been identified, thus far[24]. Nevertheless, an acetate kinase from *Salmonella typhimurium* (StAckA) was described to catalyze the phosphorylation of formate[25]. We characterized the FOK activity of StAckA and its *E.coli* homolog (EcAckA) in more detail. Based on its slightly more favorable catalytic properties, we chose EcAckA for our formate reduction cascade (**Table 7**, **Supplementary Fig. 7**).

Table 7. Kinetic parameters of acetate kinases on formate.

	$K_{M,\ app}$ (mM)	$k_{cat,\ app}$ (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
EcAckA	106 ± 10	22 ± 0.9	2.1×10^2
StAckA	117 ± 9	16 ± 0.5	1.4×10^2

Phylogenetic screening reveals enzymes with formyl phosphate reductase activity

Next, we focused on identifying a formyl phosphate reductase (FPR). While the reduction of formyl phosphate into formaldehyde has not been described, thus far, the reduction of acyl-phosphates into aldehydes is known to occur naturally. We evaluated four different reductases that catalyze such reactions for activity towards formyl phosphate: N-acetyl-γ-glutamyl phosphate reductase (ArgC), aspartate semialdehyde dehydrogenase (ASD), glyceraldehyde-3-phosphate dehydrogenase (GapA) and γ-glutamyl phosphate reductase (ProA) (for their native reactions, see **Fig. SI 8**). Because formyl phosphate is unstable and

difficult to isolate, we produced it *in situ* using EcAckA. When testing the four candidates from *E. coli* (ecASD, EcArgC, ecGapA, ecProA), we detected FPR activity in all cases, albeit at very low rates (**Fig. 2a**). EcArgC and ecGapA performed best under the chosen conditions, with formaldehyde production rates of 2.2 µM/min and 1.9 µM/min, respectively. GapA is a key enzyme in glycolysis and uses NADH-as cofactor, while ArgC is involved in arginine biosynthesis and NADPH-dependent. We reasoned that an NADPH-dependent enzyme might be better suited for *in vivo* applications, due to the higher NADPH/$NADP^+$ ratio, which would strongly favor the reduction reaction. Hence, we chose ArgC as FPR candidate for further engineering[40].

Next, we constructed a phylogenetic tree of 118 ArgC homologs and selected five representative variants from distinct clades for further biochemical characterization (**Fig. 2b**). While four of the homologs displayed no FPR activity, ArgC from *D. acetiphilus* (DaArgC) showed a formaldehyde production rate of 8 µM/min, a 4-fold higher activity than EcArgC (**Fig. 2c**).

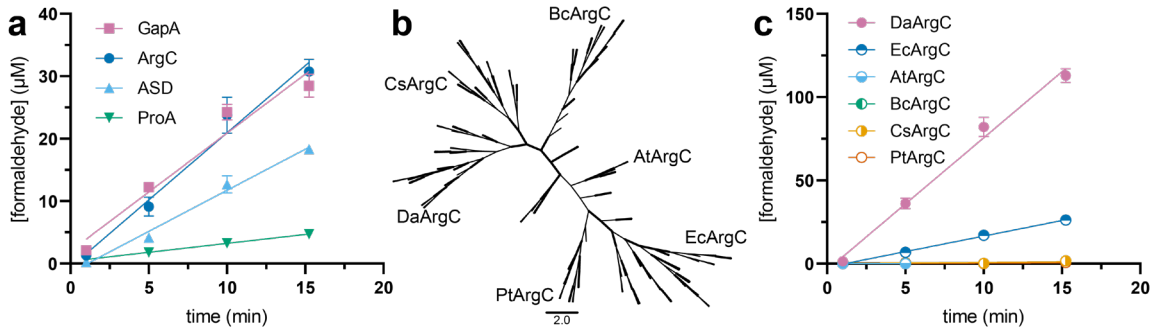


Figure 2 Discovery of promiscuous formyl phosphate reductase activity in diverse metabolic enzymes. a Comparison of multiple E.coli enzymes' formyl phosphate reductase activity. EcAckA was used to produce formyl phosphate in situ. The formaldehyde concentration was determined using the Nash assay. Mean and standard error are shown. Data points were baseline-corrected against a no formate control. Curves were fit using linear regression of the data points in GraphPad Prism 9.0.2. **b** Phylogeny of ArgC homologs. 50x bootstrap Neighbor-Joining tree of homologs with < 60% identity to EcArgC is shown. Chosen representative homologs are indicated on the tree. **c** Comparison of ArgC homologs regarding their activity toward formyl phosphate. Assay was performed as in A. at = Arabidopsis thaliana, bc = Bacillus clausii, cs = Caldicellulosiruptor saccharolyticus, da = Denitrovibrio acetiphilus, ec = Escherichia coli, pt = Pseudidiomarina taiwanensis.

Initial characterization of the formate reduction module

We next reconstituted and tested the complete formyl phosphate route. When combining purified DaArgC with excess concentrations of EcAckA *in vitro*, formaldehyde was produced at a rate of 33 µM/min (**Fig. 3a**). We also observed activity of the cascade in bacterial lysate at a rate of 1.1 µM/min, when using an inducible double expression system (pCDFDuet-1_EcAckA_DaArgC) that produced both enzymes in approximately equal amounts, as judged by SDS page analysis (**Fig. 3b**). Spiking purified enzyme into the lysate confirmed that DaArgC was the rate-limiting enzyme.

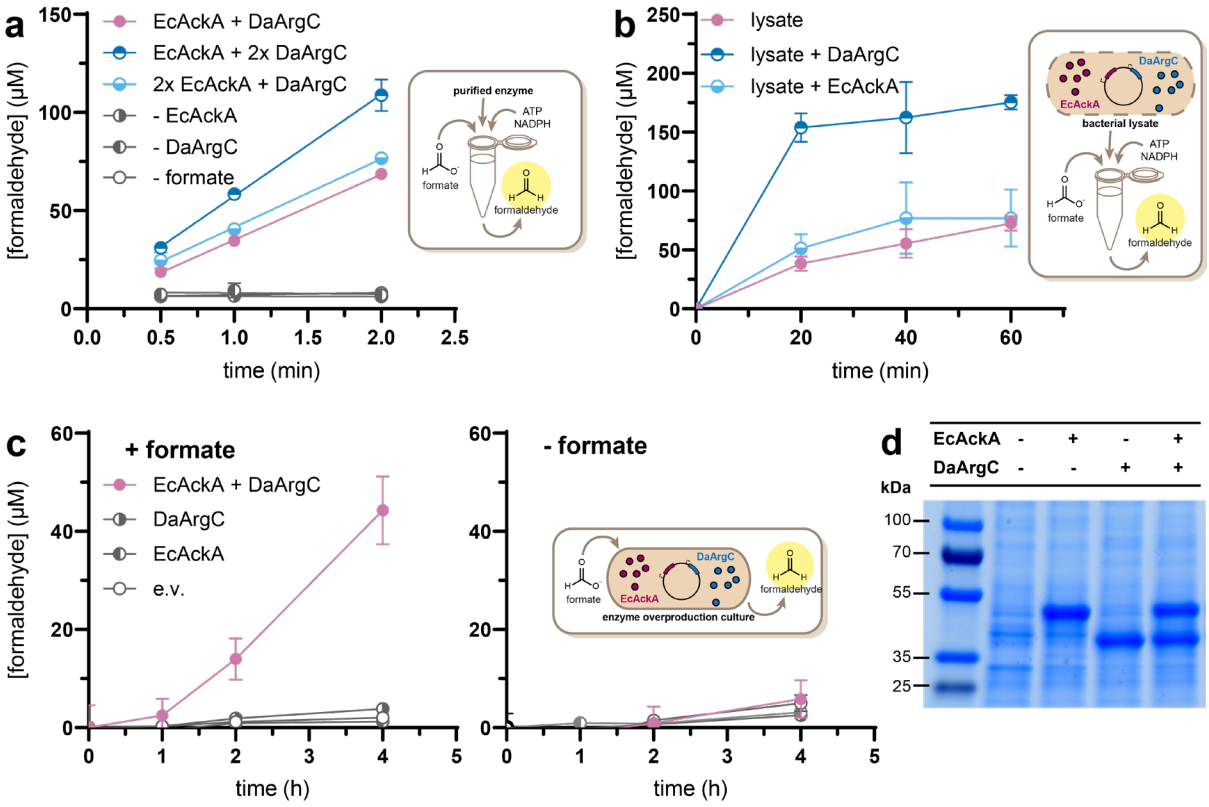


Figure 3 Characterization of the EcAckA + DaArgC cascade. a In vitro formaldehyde production. Formaldehyde was detected by Nash assay. Mean and standard error are shown; n = 3. **b** Activity in bacterial lysate. Purified DaArgC or EcAckA were added to lysate to determine the rate-limiting enzyme. pur. = purified. Formaldehyde was detected as in A. Mean and standard error are shown; n = 3. **c** Formaldehyde production by the cascade in BL21 DE3 ΔfrmRAB in the presence or absence of formate. e.v. = empty vector. Formaldehyde was detected as in A. Mean and standard error are shown; n = 3. d SDS-PAGE of EcAckA and DaArgC production in BL21 (DE3) ΔfrmRAB cells.

In the following, we also demonstrated that the module was active *in vivo*, using *E. coli* BL21 DE3 Δ*frmRAB*, in which the glutathione-dependent formaldehyde detoxification system was knocked out. When adding 50 mM formate to a growing *E. coli* BL21 (DE3) Δ*frmRAB* culture transformed with pCDFDuet-1_EcAckA_DaArgC, we observed formaldehyde production at roughly 10 μM/h. Induction and production of both EcAckA and DaArgC were required for formaldehyde output (**Fig. 3c**). Formaldehyde production in the absence of formate was marginal (likely caused by basal formate levels produced *in cellulo* by pyruvate-formate lyase (PFL), see below) (**Fig. 3c, Supplementary Fig. 9**). Interestingly, DaArgC was the only ArgC homolog that produced significant formaldehyde levels in the *in vivo* system (**Supplementary Fig. 10b**). We also assessed cell survival under the assay conditions, as both formate and formaldehyde are toxic to *E.coli* (**Fig. SI 11**). However, after incubation under the assay conditions, all cells, including the controls (empty vector, DaArgC only, EcAckA only) could be successfully rescued in LB medium (**Supplementary Fig. 12**).

Overall, this data demonstrated that the formyl phosphate route was active and showed *in vitro* rates comparable to published cascades using formyl-CoA as intermediate (5 μM/min in vitro, 10 μM/min in cell lysate, **Supplementary Table 1**), notably without any further optimization[15, 19, 23]. *In vivo* productivity of these routes had been observed, but not quantified.

Engineering DaArgC towards improved FPR activity

Although *in vivo* productivity of the formate-phosphate route was already at par with recently reported systems, our lysate experiments had indicated that DaArgC was rate-limiting, likely limiting formaldehyde production *in vivo*. Thus, we aimed at improving the enzyme's FPR activity through protein engineering. To that end, we solved the enzyme's crystal structure and identified key active site residues (**Fig. 4a, Supplementary Fig. 13**).

The ArgC active site can be divided into three parts, the NADPH-binding domain, the catalytic site, and a substrate-binding channel. The catalytic site consists of catalytically active C149 and the HRH motif (residues 208-210) responsible for binding the phosphate and carbonyl group of the substrate (gray labelling in **Fig. 4a**).

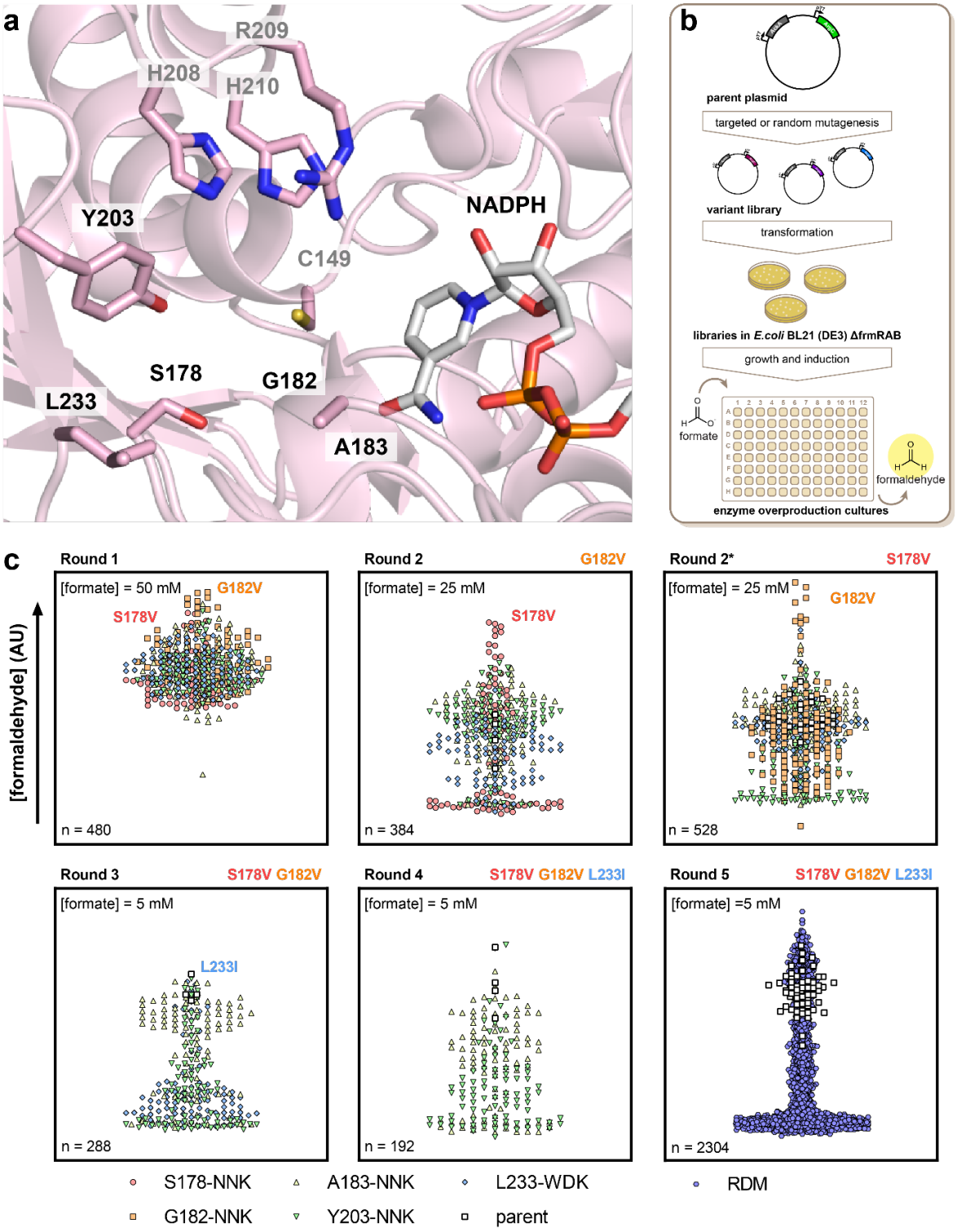


Figure 4 Enzyme engineering of DaArgC. a Cartoon depiction of the active site and substrate-binding domain of DaArgC (PBD-ID 8AFU, resolution 2.0 Å). NADPH (grey) is superimposed based on the structure of Mycobacterium tuberculosis ArgC (PBD-ID 2I3G). The HRH motif responsible for binding of the phosphate and carbonyl group of the native substrate and the cysteine forming a covalent intermediate are labelled in light grey. Residues accommodating the branched side-chain of N-acetyl-γ-glutamyl phosphate are labelled in black. **b** Schematic depiction of the library screening method. **c** Iterative Saturation and Random Mutagenesis of DaArgC. Variants of DaArgC were generated by

Saturation Mutagenesis (Rounds 1-4) or Random Mutagenesis (RDM) and produced in BL21 DE3 *ΔfrmRAB*. Formate was added to the supernatant of the cultures as indicated on the top left of the graphs. After incubation, samples were taken and formaldehyde was detected by Nash assay to evaluate relative performance. Mutants displaying increased activity were used as parent in the next round as indicated above the graphs. The NNK codon encode all 20 possible amino acids and one stop codon. The WDK codon encodes one stop codon and 11 amino acids: C, F, K, I, L, M, N, R, S, W, Y. Symbols represent individual measurements.

The substrate channel includes a GAG motif (residues 182-184) and surrounding residues that accommodate the branched side-chain of N-acetyl-γ-glutamyl phosphate (black labelling in **Fig. 4a**). To increase ArgC activity with formyl phosphate, we aimed at improving accommodation of formyl phosphate in the active site by introducing larger amino acid residues in key positions to close up the active center of the enzyme, aiming to exclude native substrate and improve binding of formyl phosphate. Thus, we targeted five residues of DaArgC (S178, G182, A183, Y203 and L233) (**Fig. 4a**) and used iterative saturation mutagenesis (ISM)[41] to screen for favorable amino acid exchanges. ISM was performed in BL21 DE3 Δ*frmRAB* transformed with variants of pCDFDuet-1_EcAckA_DaArgC. Individual colonies were picked and subjected to *in vivo* screens for formaldehyde production using the Nash assay (**Fig. 4b**). We screened approximately 4000 variants and incrementally lowered the formate concentration from 50 to 5 mM to screen for DaArgC variants with improved K_M. Round 1 of mutagenesis resulted in variant DaArgC G182V (DaArgC1), which was followed by a S178V in round 2 of mutagenesis (DaArgC2). Using double variant DaArgC2 as a template, we discovered a L233I substitution that showed an additional beneficial effect, resulting in triple variant DaArgC3.Although the improvements were rather small, they were consistently discovered in replicates (**Fig. 4c**). When using DaArgC3 as template for iterative or random mutagenesis, no further improvements were discovered: round 4 yielded no further variants and round 5 only silent mutations.

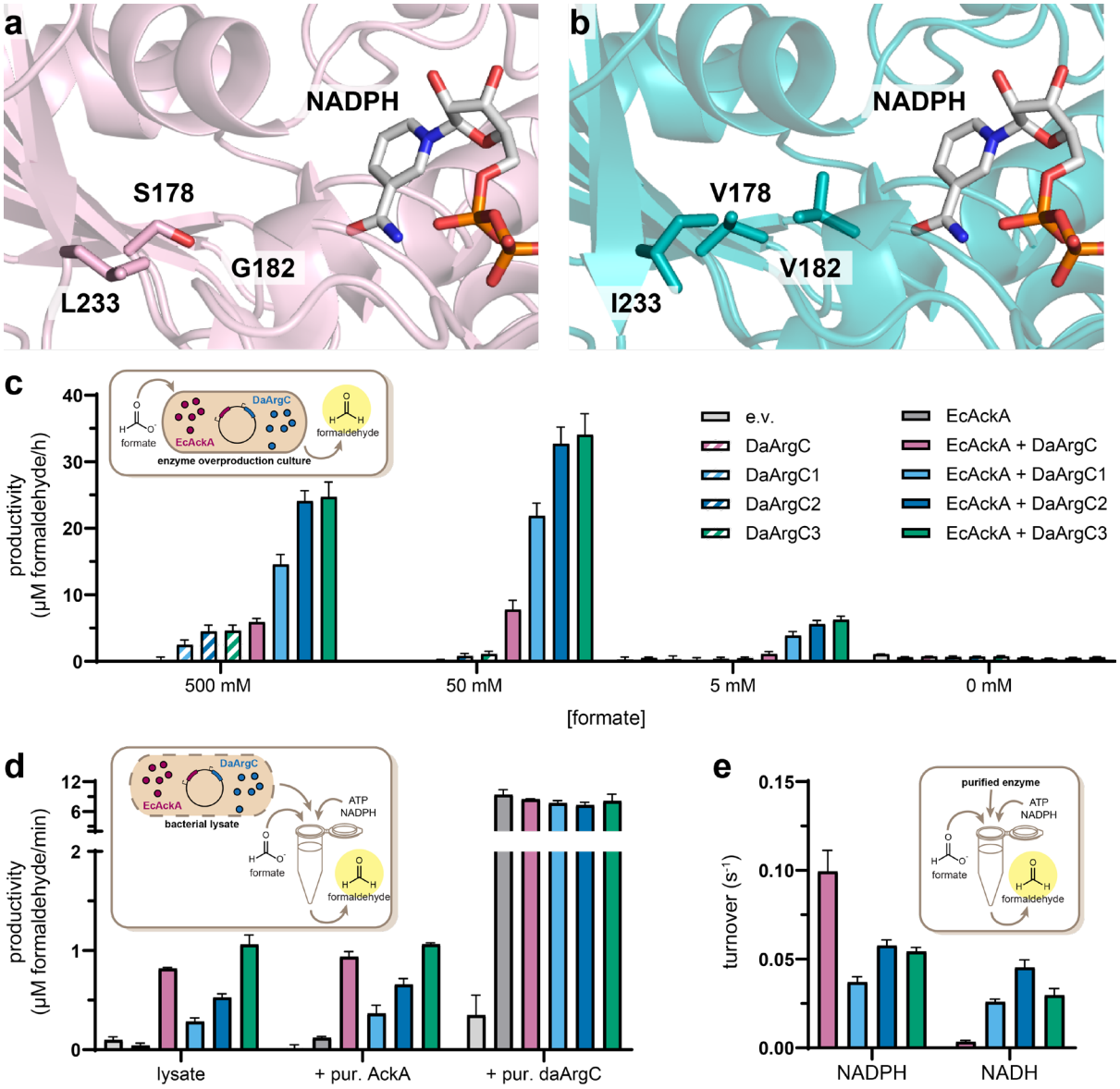


Figure 5 Validation of DaArgC variants. a Crystal structure of DaArgC (PBD-ID 8AFU, 2.0 Å) and **b** DaArgC3 (PBD-ID 8AFV, 2.2 Å). NADPH (grey) is modelled based on the structure of Mycobacterium tuberculosis ArgC (PBD-ID 2I3G). Active site residues affected by mutagenesis are shown as sticks. **c** In vivo productivity of DaArgC variants. EcAckA and DaArgC variants were produced in E. coli BL21 DE3 *ΔfrmRAB ΔpflAB* and formate was added to the supernatant. The rate of formaldehyde production was evaluated by linear regression of four time points. Mean and standard error are shown; n = 3. **d** Formaldehyde production by DaArgC variants in bacterial lysate. Enzymes were produced in in BL21 DE3 *ΔfrmRAB* and chemically lysed. Reactions were supplemented with purified EcAckA or DaArgC to determine the rate-limiting enzyme. Data was evaluated as in d. Mean and standard error are shown; n = 3. **e** Cofactor utilization of DaArgC variants. Purified DaArgC variants were coupled to EcAckA and given NADH or NADPH as reduction equivalent. Data was evaluated as in d. Mean and standard error are shown; n = 3.

Confirmation of improved FPR activity in DaArgC variants

To validate the screened candidates, we expressed the three variants in *E.coli* BL21 DE3 Δ*frmRAB* Δ*pflAB*. Deletion of *pflAB* was used to suppress PFL-mediated formate production in BL21 DE3 Δ*frmRAB*, which posed a problem, when testing the engineered DaArgC variants at low formate feeding conditions (i.e., 5 mM, see **Supplementary Fig. SI**). When testing the different enzyme variants, *in vivo* formaldehyde production increased from DaArgC to DaArgC3, while enzyme production levels remained unchanged, as judged by SDS-PAGE analysis (**Fig. 5c, Supplementary Fig. 14**). Across all substrate concentrations, DaArgC3 showed a four-fold improved formaldehyde productivity over DaArgC. Interestingly, on 500 mM formate, control vectors carrying only DaArgC variants showed formaldehyde formation up to 5 μM/h, suggesting that endogenous EcAckA was sufficient to produce formyl phosphate at high external formate concentrations.

In lysate-based assays, supplementation with either EcAckA or DaArgC showed the EcAckA+DaArgC3 system was still limited by DaArgC3 (**Fig. 5d**). However, much to our surprise, the engineered DaArgC variants did not show improvement over the wild type enzyme in bacterial lysate assays (50 mM formate, **Fig. 5d**). In fact, DaArgC1 and DaArgC2 were 3- and 1.5-fold slower than wild type. Final variant DaArgC3 showed activities comparable to the wild type in lysate. Using purified enzyme, an even more pronounced decrease in activity between wild type and variants was observed (**Fig. 5e**). These findings raised the question how mutagenesis had affected the kinetic parameter of the enzyme.

Changes in kinetic traits of DaArgC variants do not explain *in vivo* improvements.

To understand why the engineered enzymes showed improved performance in *in vivo* assays, but not in lysate assays, we decided to characterize the different variants in more detail *in vitro*. To obtain substrate for assays, we chemically synthesized formyl phosphate. As formyl phosphate rapidly hydrolyses in water (**Supplementary Fig. 2 & 5**), it is not possible to prepare standard aqueous solutions. Therefore, we formed formyl phosphate *in situ* from acetic formic anhydride and dipotassium phosphate and immediately added the crude product solution to the enzyme assays. This limited the maximum substrate concentration in our reaction mixtures to 50 mM. It is worth noting, however, that the hydrolytic

decomposition rate of formyl phosphate of -0.34%min^{-1} is negligible as compared to enzymatic reaction, thus not invalidating the assay (see above).

At 50 mM formyl phosphate, the wild type DaArgC showed an apparent k_{cat} of 0.1 s^{-1}, but was not fully saturated. Variants DaArgC1, DaArgC2 and DaArgC3 showed similar turnover numbers (0.1-0.2 s^{-1}), but were already saturated at apparent K_M's (formyl phosphate) between 2-5 mM (**Table 8, Supplementary Fig. 15**). We next coupled DaArgC variants to EcAckA to determine the kinetic parameters for the NADPH cofactor. Wild type DaArgC showed a k_{cat} of ~0.7 s^{-1} at an apparent K_M of ~0.2 mM. The different variants all showed specific activities of 0.1-0.3 s^{-1}, and a decreased apparent K_M (NADPH) of ~0.05 μM. Surprisingly, when testing NADH as alternative cofactor, which is not accepted by wild type DaArgC, we found that DaArgC1-3 used it at about 10% catalytic efficiency compared to NADPH (**Table 8, Fig. 5e**). The K_M for NADH was approximately 6-fold higher compared to NADPH (~0.3 mM), while the specific activities were at about 50% of the activity with NADPH (~0.1 s^{-1}), indicating that variants DaArgC1-C3 had evolved to additionally accept NADH as redox cofactor.

Table 8. Kinetic parameters of DaArgC variants. Michaelis Menten fit and standard error of triplicates. (1) indicates measurements coupled to EcAckA. (2) maximal activity observed, [formyl phosphate] = 46 mM, (3) activity below detection limit, n.d. = not detected.

enzyme	formyl phosphate			NADPH(1)			NADH(1)			NADPH/ NADH
	$K_{M,\ app}$ (mM)	$k_{cat,\ app}$ (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	$K_{M,\ app}$ (mM)	$k_{cat,\ app}$ (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	$K_{M,\ app}$ (mM)	$k_{cat,\ app}$ (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	$k_{cat}/K_{M\ (NADPH)}$/ $k_{cat}/K_{M\ (NADH)}$
DaArgC	no fit	0.12 ± 0.04(2)	-	0.16 ± 0.02	0.66 ± 0.03	$4.1x10^3$	>500(3).	<0.005(3).	<1(3)	$>4.1x10^3$
DaArgC1	31 ± 7	0.11 ± 0.01	2.4	0.04 ± 0.01	0.14 ± 0.02	$3.5x10^3$	0.28 ± 0.05	0.08 ± 0.00	$2.8x10^2$	$1.3x10^1$
DaArgC2	38 ± 9	0.17 ± 0.02	3.1	0.06 ± 0.02	0.26 ± 0.03	$4.3x10^3$	0.30 ± 0.04	0.12 ± 0.00	$4.0x10^2$	$1.1x10^1$
DaArgC3	18 ± 7	0.12 ± 0.02	4.6	0.05 ± 0.01	0.23 ± 0.02	$4.6x10^3$	0.34 ± 0.05	0.09 ± 0.00	$2.6x10^2$	$1.8x10^1$

G182V relaxes redox cofactor specificity of DaArgC

During the course of experimental evolution, DaArgC1-C3 had gained the ability to accept NADH in addition to NADPH, allowing the variants to access NADH as an alternative

cofactor pool. We thus aimed to understand the molecular basis for this cofactor selectivity. When we re-established the glycine residue at position 182 in DaArgC2 and DaArgC3 (creating variants DaArgC2* and DaArgC3*), NADPH selectivity of the variants was fully restored, indicating that the glycine to valine substitution was responsible for the observed cofactor flexibility (**Supplementary Table 2, Supplementary Fig. 15**).

The ability of G182V to relax cofactor specificity of the enzyme towards NADH is surprising, since changes in cofactor specificity usually require more intense engineering efforts[42, 43]. NADPH specificity in ArgC is achieved in the wild type by a tight hydrogen network of the 2' phosphate with the side- and main chains of residues T9, S33, E34 and T35 (derived from comparison with mtArgC[44]). The position of these residues are not significantly affected in DaArgC3 and remain within hydrogen bond distance of the 2'-phosphate (**Supplementary Fig. 16**). G182 is not proximal to the phosphate group of NADPH, but lies within the conserved GAG motif of the substrate-binding channel (residues 182-184). Notably, two amino acids flanking the GAG motif, T181 and R185, form hydrogen bonds to the nicotinamide group and the diphosphate of NADPH, respectively (**Supplementary Fig. 16**). While we could not observed major changes between the TGAGR loop of DaArgC and the TVAGR loop of DaArgC3 in the $NADP^+$/NADPH-unbound state (RMSD 0.119), it is known that in mtArgC[44] the GAG motif moves closer to $NADP^+$ in the bound state. It is therefore possible that V182 affects loop movement, resulting in a tighter binding of the cofactor and hence a decreased the $K_{M,\,app}$ for NADPH (**Table 8**). A tighter binding of the nicotinamide moiety may also positively affect NADH binding (compensating the missing 2'-phosphate interaction), and explain the decrease in $k_{cat,\,app}$ in the G182V variant, in case $NADP^+$ release becomes rate limiting.

Exclusion of native substrate by the DaArgC variants contributes to *in vivo* activity

To further understand how the engineered enzymes perform in respect to their native reaction, we evaluated the activity of purified DaArgC1-C3 towards N-acetyl-γ-glutamyl-phosphate (**Fig. 6***)*. We produced N-acetyl-γ-glutamyl-phosphate, *in situ* from N-acetyl glutamate and ATP, using *E.coli* acetyl glutamate kinase (ArgB). Wild type DaArgC showed a turnover of about 1 s^{-1} with N-acetyl γ glutamyl phosphate which is in line with activities of

approximately 0.6 s^{-1} for other ArgC homologs[45, 46]. Activity with the native substrate was almost three orders of magnitude reduced in all variants (**Fig. 6a**). When testing for reaction with acetyl phosphate that might act as competing substrate *in vivo*, we observed that activity of DaArgC1-3 towards acetyl phosphate was two orders of magnitudes decreased compared to that of the wild type, and one order of magnitude lower in respect to FPR activity, (**Fig. 6a**). Notably, activity with acetyl phosphate decreased stepwise in the engineered variants (**Fig. 6a**). Our results showed that during the course of engineering, DaArgC1-3 lost their native activities in favor of the engineered FPR activity with an almost 300-fold switch in selectivity (**Fig. 6b**). In summary, the variants no longer interfere with amino acid biosynthesis and show reduced activity with acetyl phosphate, thus limiting interference with the acetyl-CoA pool, which allows DaArgC1-3 to operate completely orthogonal to native metabolism *in vivo*.

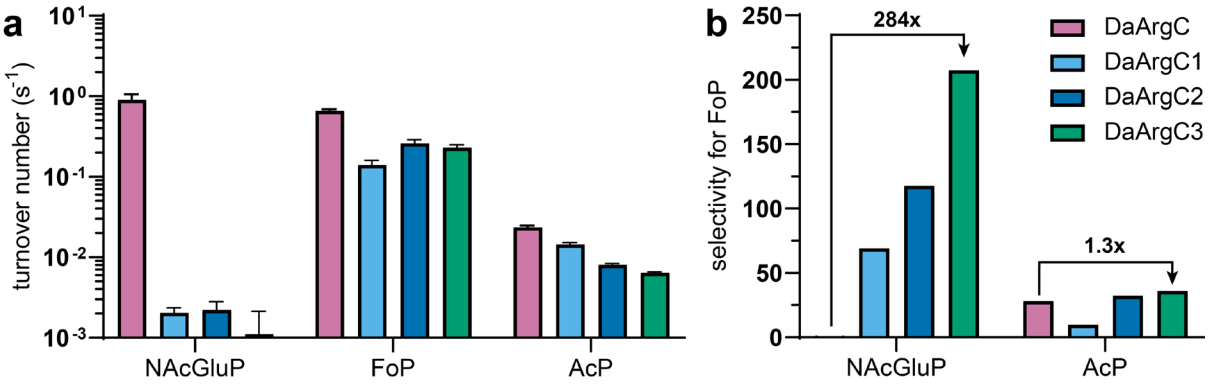


Figure 6 DaArgC variants show an improved selectivity for formyl phosphate over native substrate. a Activity of DaArgC variants on native and promiscuous substrates. Shown are mean and error; n = 3. **b** Selectivity of variants on FoP over the other substrates. Bars represent calculated selectivities (v0(FoP)/v0(NAcGluP) and v0(FoP)/v0(NAcGluP) without error propagation. NAcGluP = N-acetyl-γ-glutamyl-phosphate; FoP = formyl phosphate, AcP = acetyl phosphate

FPR can be integrated into an *in vitro* cascade converting formate to glycolyl-CoA

Recently, the FORCE pathway was developed. In this pathway, formaldehyde is converted to formyl-CoA. One formaldehyde and one formyl-CoA are then condensed into the C2 compound glycolyl-CoA through either members of the 2-hydroxy-acyl-CoA synthase (HACL) or the oxalyl-CoA (OXC) family[22, 26, 29]. To start this pathway directly from formate instead of formaldehyde, we aimed at coupling our formyl phosphate route to acylating CoA-

reductase LmACR (to convert formaldehyde into formyl-CoA) and engineered OXC-family member MeOXC4[29] for glycolyl-CoA formation from formaldehyde and formyl-CoA (**Fig. 7a**).

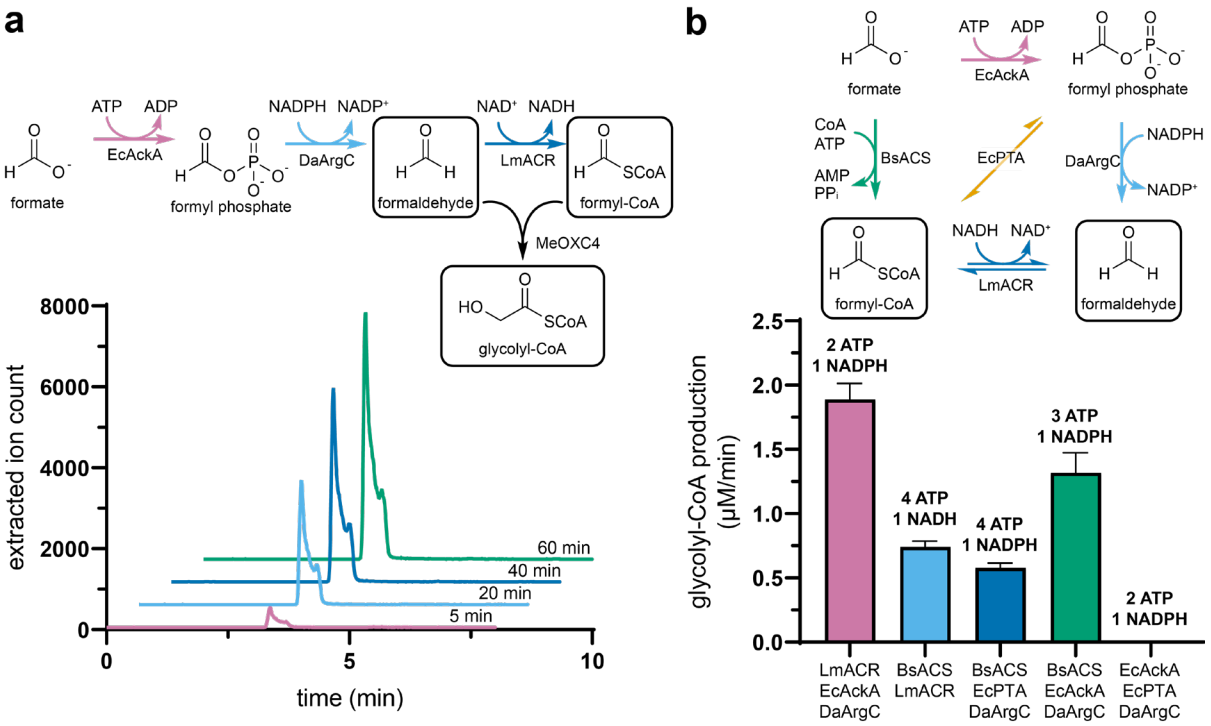


Figure 7 MeOXC4 can be used in an enzymatic cascade to produce glycolyl-CoA from formate. **a** Scheme and LC-MS extracted ion counts for the production of glycolyl-CoA by EcAckA+DaArgC+LmACR+MeOXC4. **b** Scheme and LC-MS data of glycolyl-CoA production by different activation modules. ATP and NAD(P)H consumption per C2 unit formed are indicated above the bars. Shown are mean and error; n = 3.

When testing EcAckA+DaArgC with LmACR[19, 22, 23] and MeOXC4, we observed that our formyl phosphate/CoA route indeed produced glycolyl-CoA (**Fig. 7a**). Notably, this formyl phosphate/CoA route outcompeted a completely CoA-based route (BsACS[19] + LmACR) by 150%. We also established and tested all other potential combinations involving formyl-CoA and formyl phosphate, using EcPTA to link CoA- and phosphate-pools (**Fig. 7b**). However, the EcAckA+DaArgC+LmACR route was the fastest combination *in vitro*. Moreover, the EcAckA+DaArgC+LmACR route operates at the thermodynamic minimum, requiring the hydrolysis of two phosphoanhydride bonds and one reducing equivalent per C2 unit formed. Overall, these results showed that the EcAckA+DaArgC+LmACR route is kinetically and

thermodynamically superior to all other potential glycolyl-CoA forming pathways and at the theoretically possible energetic minimum for the formation of C2- compounds from formate.

Discussion

Here, we established a formyl phosphate route from formate to formaldehyde that operates at the thermodynamically possible minimum and involves two new-to-nature enzyme activities: FOK and FPR. To that end, we harnessed the latent formate kinase activity in EcAckA and detected FPR activity in four *E. coli* reductases; ArgC, GapA, ASD and ProA. Based on the favorable activities of ArgC with formyl phosphate, we focused on further improving this FPR activity. Phylogenetic screening identified a superior ArgC homolog from *D. acetiphilus* that we further improved by directed evolution. We employed directed and (limited) saturation mutagenesis to convert DaArgC into a bona fide FPR. During three rounds of mutagenesis, *in vivo* activity of variant DaArgC3 was improved 4-fold over wild type *in vivo* and a 300-fold selectivity switch for formyl phosphate over its native substrate was observed.

Although we observed increased activity of DaArgC1-3 *in vivo*, this was not reflected in our cell lysates and *in vitro* measurements. Usually, isolated enzymatic bio-catalysis is not directly transferable to complex metabolic networks, for a multiplicity of unpredictable reasons. We speculate, that a combination of increased cofactor availability (NADH and NADPH) and cofactor affinity (decreased K_M for NADPH) allows DaArgC1-3 to operate as an orthogonal FPR *in vivo*. The complete loss of activity with its native substrate, N-acetyl-γ-glutamyl phosphate, and a reduced side activity with acetyl phosphate additionally contributes to that effect.

Through the formyl phosphate route, formaldehyde is readily produced from formate *in vitro* and *in vivo* at the cost of only one energy-rich phosphoanhydride bond and one reducing equivalent (i.e., NAD(P)H). Compared to previously published formate to formaldehyde conversion via formyl-CoA[15, 17, 19, 22, 23, 47-49], the formyl phosphate sequence requires one fewer energy-rich phosphoanhydride bond. It operates entirely orthogonal to the acetyl-CoA pool. Compared to the natural formyl-T_4F route[18], the cascade does not need the complex T_4F

cofactor, requires fewer enzymes and does not depend on spontaneous hydrolysis, which makes the FOK-FPR cascade superior to both routes.

Recently, the landscape of synthetic formate assimilation has expanded upon by the HWLS[18] and SMGF[19] pathways that assimilate formate via formaldehyde. Integration of FOK-FPR into these systems may have beneficial effects due to a lowered protein burden and reduced thermodynamic cost (phosphoanhydride hydrolysis) and/or kinetic constraints (spontaneous methylene-T_4F hydrolysis) compared to CoA- and T_4F-based routes. In a proof-of-principle, we showed that our cascade has the potential to feed the FORCE pathway[26], which assimilates formaldehyde and formyl-CoA into glycolyl-CoA. FOK-FPR, when combined with LmACR and MeOXC4, produced glycolyl-CoA, the first intermediate of the FORCE pathway, at rates outcompeting any other possible cascade and at the theoretically possible energetic minimum *in vitro*. Implementing FOK-FPR *in vivo* could expand FORCE for the assimilation of formate, in addition to methanol and formaldehyde.

The most limiting factor of the FOK-FPR cascade currently is its inhibition by acetate. While the FPR reaction discriminates well against acetyl phosphate, EcAckA is still a native enzyme and readily accepts acetate. A *bona fide* formate kinase would confer complete orthogonality, preventing metabolic crosstalk of the cascade. A specific formate kinase was described in the 1960s [24] and might offer a solution, however, the corresponding gene has never been identified. Alternatively, ACK templates could be engineered towards more specific formate kinase activity.

Here, we designed, established and prototyped a new-to-nature formate reduction module *in vitro* and *in vivo*. Our study provides the basis for a subsequent *in vivo* implementation of this formyl phosphate-based route that extends the existing formate to formaldehyde conversion pathways and enriches the metabolic landscape of formate assimilation by another, highly efficient alternative.

References

[1] Bar-Even, A., Noor, E., Lewis, N. E., and Milo, R. (2010) Design and analysis of synthetic carbon fixation pathways, *Proc. Natl. Acad. Sci. U.S.A. 107*, 8889-8894.

[2] 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.

[3] 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. (2021) A new-to-nature carboxylation module to improve natural and synthetic CO2 fixation, *Nat. Catal. 4*, 105-115.

[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, *Chem. Rev. 115*, 12936-12973.

[5] Schaub, T. (2018) CO2-based hydrogen storage: CO2 hydrogenation to formic acid, formaldehyde and methanol, *Phys. Sci. Rev. 3*.

[6] Artz, J., Müller, T. E., Thenert, K., Kleinekorte, J., Meys, R., Sternberg, A., Bardow, A., and Leitner, W. (2018) Sustainable Conversion of Carbon Dioxide: An Integrated Review of Catalysis and Life Cycle Assessment, *Chem. Rev. 118*, 434-504.

[7] 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.

[8] 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.

[9] 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, *Nat. Microbiol. 5*, 1459-1463.

[10] Large, P. J., Peel, D., and Quayle, J. R. (1962) Microbial growth on C(1) compounds. 3. Distribution of radioactivity in metabolites of methanol-grown Pseudomonas AM1 after incubation with [C]methanol and [C]bicarbonate, *Biochem. J. 82*, 483-488.

[11] 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.

[12] Kemp, M. B., and Quayle, J. R. (1966) Microbial growth on C1 compounds. Incorporation of C1 units into allulose phosphate by extracts of Pseudomonas methanica, *Biochem. J. 99*, 41-48.

[13] Van Dijken, J. P., Harder, W., Beardsmore, A. J., and Quayle, J. R. (1978) Dihydroxyacetone: An intermediate in the assimilation of methanol by yeasts?, *FEMS Microbiol. Lett. 4*, 97-102.

[14] 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. 60*, 1-13.

[15] 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.

[16] Poust, S., Piety, J., Bar-Even, A., Louw, C., Baker, D., Keasling, J. D., and Siegel, J. B. (2015) Mechanistic Analysis of an Engineered Enzyme that Catalyzes the Formose Reaction, *Chembiochem 16*, 1950-1954.

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

[18] Hu, G., Li, Z., Ma, D., Ye, C., Zhang, L., Gao, C., Liu, L., and Chen, X. (2021) Light-driven CO2 sequestration in Escherichia coli to achieve theoretical yield of chemicals, *Nat. Catal. 4*, 395-406.

[19] Hu, G., Guo, L., Gao, C., Song, W., Liu, L., and Chen, X. (2022) Synergistic Metabolism of Glucose and Formate Increases the Yield of Short-Chain Organic Acids in Escherichia coli, *ACS Synth. Biol. 11*, 135-143.

[20] Franz, A. W., Kronemayer, H., Pfeiffer, D., Pilz, R. D., Reuss, G., Disteldorf, W., Gamer, A. O., and Hilt, A. Formaldehyde, In *Ullmann's encycl. ind. chem.*, 1-34.

[21] Cramer, H. H., Chatterjee, B., Weyhermüller, T., Werlé, C., and Leitner, W. (2020) Controlling the Product Platform of Carbon Dioxide Reduction: Adaptive Catalytic Hydrosilylation of CO2 Using a Molecular Cobalt(II) Triazine Complex, *Angew. Chem. Int. Ed. 59*, 15674-15681.

[22] 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.

[23] Wang, J., Anderson, K., Yang, E., He, L., and Lidstrom, M. E. (2021) Enzyme engineering and in vivo testing of a formate reduction pathway, *Synth. Biol. 6*, ysab020.

[24] Sly, W. S., and Stadtman, E. R. (1963) Formate Metabolism. Ii. Enzymatic Synthesis of Formyl Phosphate and Formyl Coenzyme a in Clostridium Cylindrosporum, *J. Biol. Chem. 238*, 2639-2647.

[25] Chittori, S., Savithri, H. S., and Murthy, M. R. (2012) Structural and mechanistic investigations on Salmonella typhimurium acetate kinase (AckA): identification of a putative ligand binding pocket at the dimeric interface, *BMC Struct. Biol. 12*, 24.

[26] Chou, A., Lee, S. H., Zhu, F., Clomburg, J. M., and Gonzalez, R. (2021) An orthogonal metabolic framework for one-carbon utilization, *Nat. Metab. 3*, 1385-1399.

[27] 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.

[28] Stukenberg, D., Hensel, T., Hoff, J., Daniel, B., Inckemann, R., Tedeschi, J. N., Nousch, F., and Fritz, G. (2021) The Marburg Collection: A Golden Gate DNA Assembly Framework for Synthetic Biology Applications in Vibrio natriegens, *ACS Synth. Biol. 10*, 1904-1919.

[29] Nattermann, M., Burgener, S., Pfister, P., Chou, A., Schulz, L., Lee, S. H., Paczia, N., Zarzycki, J., Gonzalez, R., and Erb, T. J. (2021) Engineering a Highly Efficient Carboligase for Synthetic One-Carbon Metabolism, *ACS Catal. 11*, 5396-5404.

[30] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *J. Mol. Biol. 215*, 403-410.

[31] Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A. N., and Alva, V. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core, *J. Mol. Biol. 430*, 2237-2243.

[32] Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, *Mol. Biol. Evol. 35*, 1547-1549.

[33] Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol. 4*, 406-425.

[34] Felsenstein, J. (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap, *Evolution 39*, 783-791.

[35] Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement, *Acta Crystallogr. D 66*, 133-144.

[36] 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., 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 Crystallogr. D 67*, 235-242.

[37] 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 Crystallogr. D 66*, 213-221.

[38] Levin, E. J., Kondrashov, D. A., Wesenberg, G. E., and Phillips, G. N., Jr. (2007) Ensemble refinement of protein crystal structures: validation and application, *Structure 15*, 1040-1052.

[39] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr. D 66*, 486-501.

[40] Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli, *Nat. Chem. Biol. 5*, 593-599.

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

[42] 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, *Proc. Natl. Acad. Sci. U.S.A. 115*, E11455-E11464.

[43] Calzadiaz-Ramirez, L., Calvo-Tusell, C., Stoffel, G. M. M., Lindner, S. N., Osuna, S., Erb, T. J., Garcia-Borras, M., Bar-Even, A., and Acevedo-Rocha, C. G. (2020) In Vivo Selection for Formate Dehydrogenases with High Efficiency and Specificity toward NADP(), *ACS Catal. 10*, 7512-7525.

[44] Cherney, L. T., Cherney, M. M., Garen, C. R., Niu, C., Moradian, F., and James, M. N. (2007) Crystal structure of N-acetyl-gamma-glutamyl-phosphate reductase from Mycobacterium tuberculosis in complex with NADP(+), *J. Mol. Biol. 367*, 1357-1369.

[45] Wandinger-Ness, A. U., Ness, S. A., and Weiss, R. L. (1986) Simultaneous purification of three mitochondrial enzymes. Acetylglutamate kinase, acetylglutamyl-phosphate reductase and carbamoyl-phosphate synthetase from Neurospora crassa, *J. Biol. Chem. 261*, 4820-4827.

[46] Goto, M., Agari, Y., Omi, R., Miyahara, I., and Hirotsu, K. (2003) Expression, purification and preliminary X-ray characterization of N-acetyl-gamma-glutamyl-phosphate reductase from Thermus thermophilus HB8, *Acta Crystallogr. D 59*, 356-358.

[47] Sanchez, L. B. (1998) Aldehyde dehydrogenase (CoA-acetylating) and the mechanism of ethanol formation in the amitochondriate protist, Giardia lamblia, *Arch. Biochem. Biophys. 354*, 57-64.

[48] Brasen, C., Urbanke, C., and Schonheit, P. (2005) A novel octameric AMP-forming acetyl-CoA synthetase from the hyperthermophilic crenarchaeon Pyrobaculum aerophilum, *FEBS Lett. 579*, 477-482.

[49] Zhao, Y., Lei, M., Wu, Y., Wang, C., Zhang, Z., Deng, F., and Wang, H. (2009) Molecular cloning and expression of the complete DNA sequence encoding NAD+-dependent acetaldehyde dehydrogenase fromAcinetobacter sp. strain HBS-2, *Ann. Microbiol. 59*, 97.

[50] Krimen, L. I. (1970) Acetic Formic Anhydride, *Org. Synth. 50*, 1-3.

Acknowledgements

This work was supported by the German Ministry of Education and Research Grant 031B0850B (Met*A*For) and the Max Planck Society. Research on CO2 hydrogenation at RWTH Aachen and MPI CEC is part of the activities in the Exzellenzcluster 2186 "The Fuel Science Center" ID: 390919832 funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy

We would like to thank Irene Martinez for experimental assistance, Ari Satanowski and Dr. Hai He for strain generation, Luca Schulz and Dr. Simon Burgener and the members of the Met*A*For project for continuous theoretical discussions about the trajectory of the project. We also thank the European Synchrotron Radiation Facility for provision of synchrotron radiation

facilities and our special thanks go to David Flot and Andrew McCarthy for assistance in using beamline ID30B.

Competing Interests

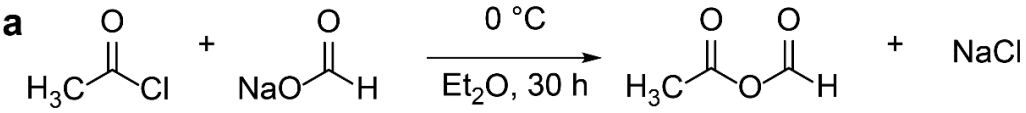
The authors declare no competing interests.

Supplementary Figures

Synthesis of formic acetic anhydride

The synthetic procedure described hereafter has been carried out under an argon atmosphere using Schlenk-technique according to a literature procedure[50]. Sodium formate (1.1 eq., 220 mmol, 14.96 g) was dispersed in dry diethyl ether (50 mL) and cooled to 0 °C. To this slurry, acetyl chloride (1 eq., 200 mmol, 15.698 g) was added dropwise over 25 minutes. The mixture was vigorously stirred at 0 °C for 32 h until acetyl chloride was completely consumed as confirmed by monitoring the reaction via 1H-NMR. The mixture was filtered and the solid residue washed once with diethyl ether (10 mL). Filtrate and washing solution were combined and the volatiles removed at 0 °C under reduced pressure of 50 mbar. Acetic formic anhydride was obtained as colorless liquid (12.96 g) with a purity of 90.6 % as determined by NMR (impurities: acetic anhydride (3.9 %), acetic acid (1.2 %) and diethyl ether (4.7 %)) (**Supplementary Figure 1**).

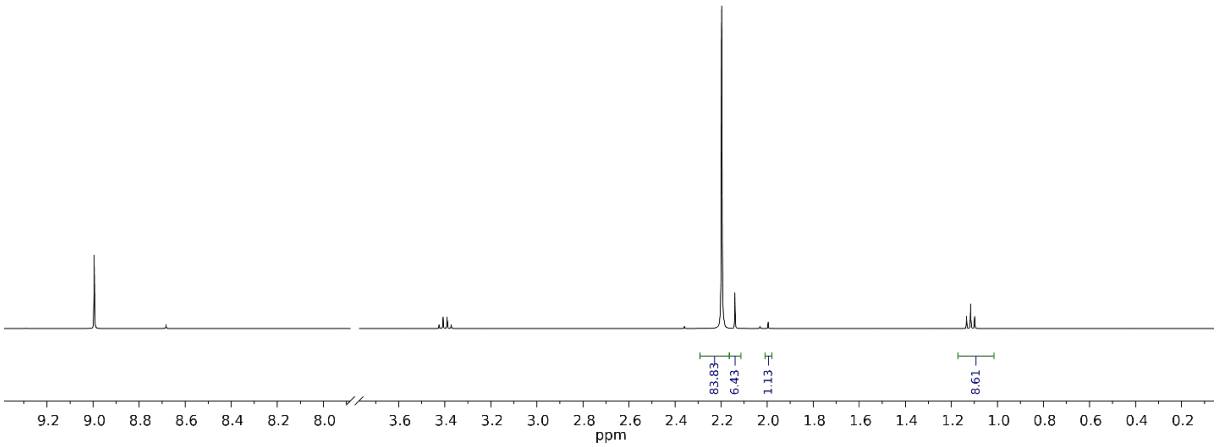
a



acetyl chloride 1.0 eq | sodium formate 1.1 eq | formic acetic anhydride

b

Compound	Integral	Molar ratio [%]
Formic acetic anh.	83.83	90.6
Acetic anh.	6.43 (/2)	3.5
Acetic acid	1.13	1.2
Diethyl ether	8.61(/2)	4.7



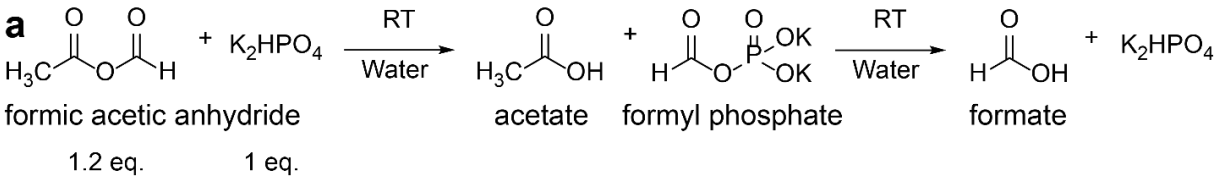
Supplementary Figure 1 Synthesis of formic acetic anhydride a Reaction conditions for the synthesis of formic acetic anhydride. **b** ^{1}H-NMR of reaction mixture. Formic acetic anhydride shows a signal at 2.2 ppm for its CH3 group and at 8.99 ppm for its CH group. Diethyl ether shows a triplet at 1.12 ppm and a quartet at 3.4 ppm. Side products are acetic acid at 2.00 ppm (CH3), acetic anhydride at 2.14 ppm (CH3) and residual signals exist for acetyl chloride (2.36 ppm) and sodium formate (8.68 ppm).

Formyl phosphate synthesis

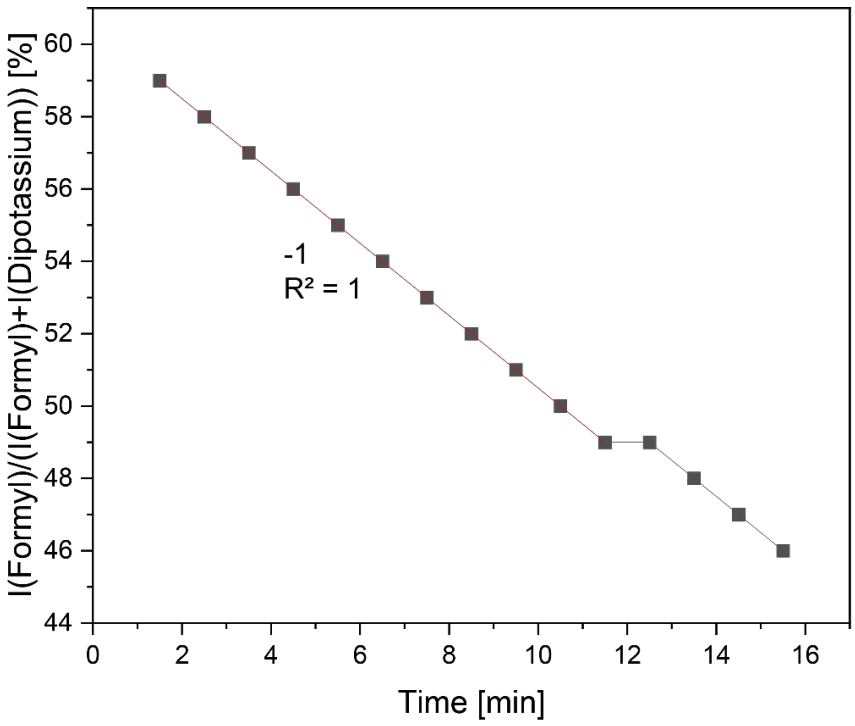
To an aqueous solution of dipotassium phosphate (1 mL, 1 M, 1 equiv.), acetic formic anhydride (0.106 g, 1.2 equiv.) was added. The solution was mixed shortly and monitored by 31P{1H}-NMR. The first measurement after 1 minute shows that 59% of dipotassium phosphate was converted to formyl phosphate The concentration of formyl phosphate linearly decreases at a rate of -1%min-1 (Figure SI2). At the end of the reaction, the pH of the solution was around 4.

The integral ratio of the 31P{1H}-NMR signals was used for quantification. At 0.1 ppm the singlet from dipotassium phosphate is detected, while formyl phosphate shows a singlet at -2.1 ppm (Figure SI 3). This assignment was confirmed by 31P-NMR where formyl phosphate shows a doublet at -2.1 ppm (JPH= 5.8 Hz) and in 1H-NMR a doublet at 8.3 ppm (JPH= 5.8 Hz) for the formyl group (Figure SI 3). 1H-31P-HMBC-NMR confirmed this assignment (Figure SI 4). Both doublets also become singlets in the respective 1H{31P} and 31P{1H}-NMR decoupled spectra (Figure SI 3).

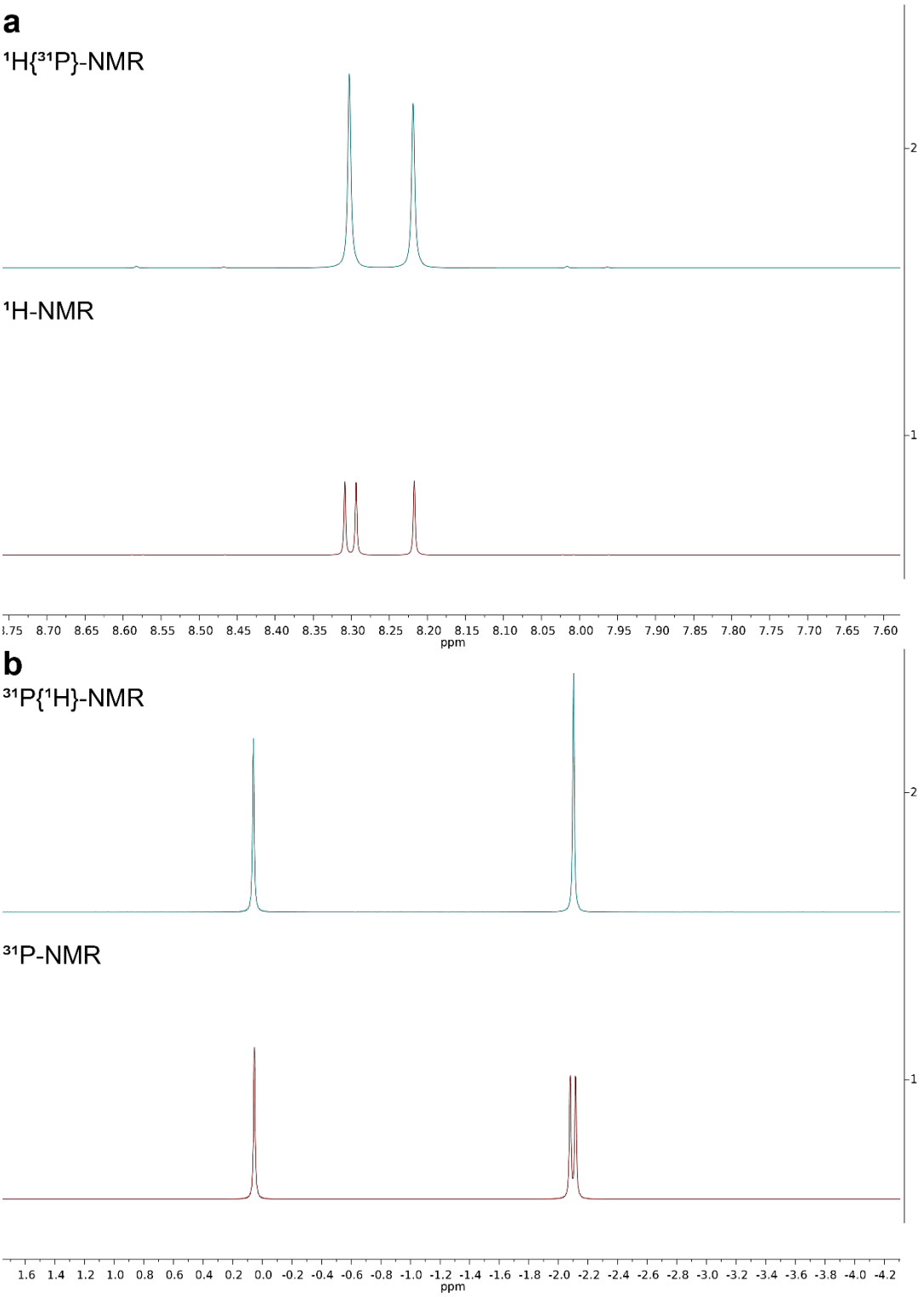
a



b

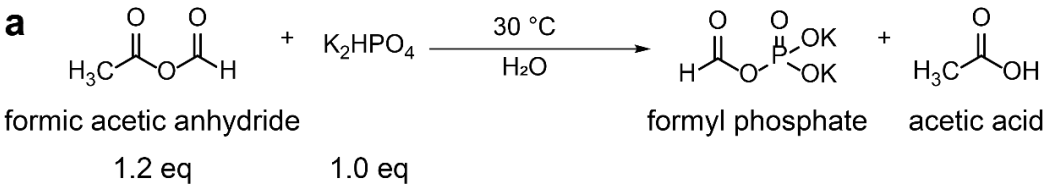


Supplementary Figure 2 Hydrolysis of formyl phosphate a Scheme for the synthesis of formyl phosphate followed by its hydrolysis to formate and phosphate. **b** Hydrolysis rate of formyl phosphate as determined by quantitative $^{31}P\{^{1}H\}$-NMR integrals of formyl phosphate (formyl) and dipotassium phosphate (dipotassium). Only points on red line were included in linear fit.



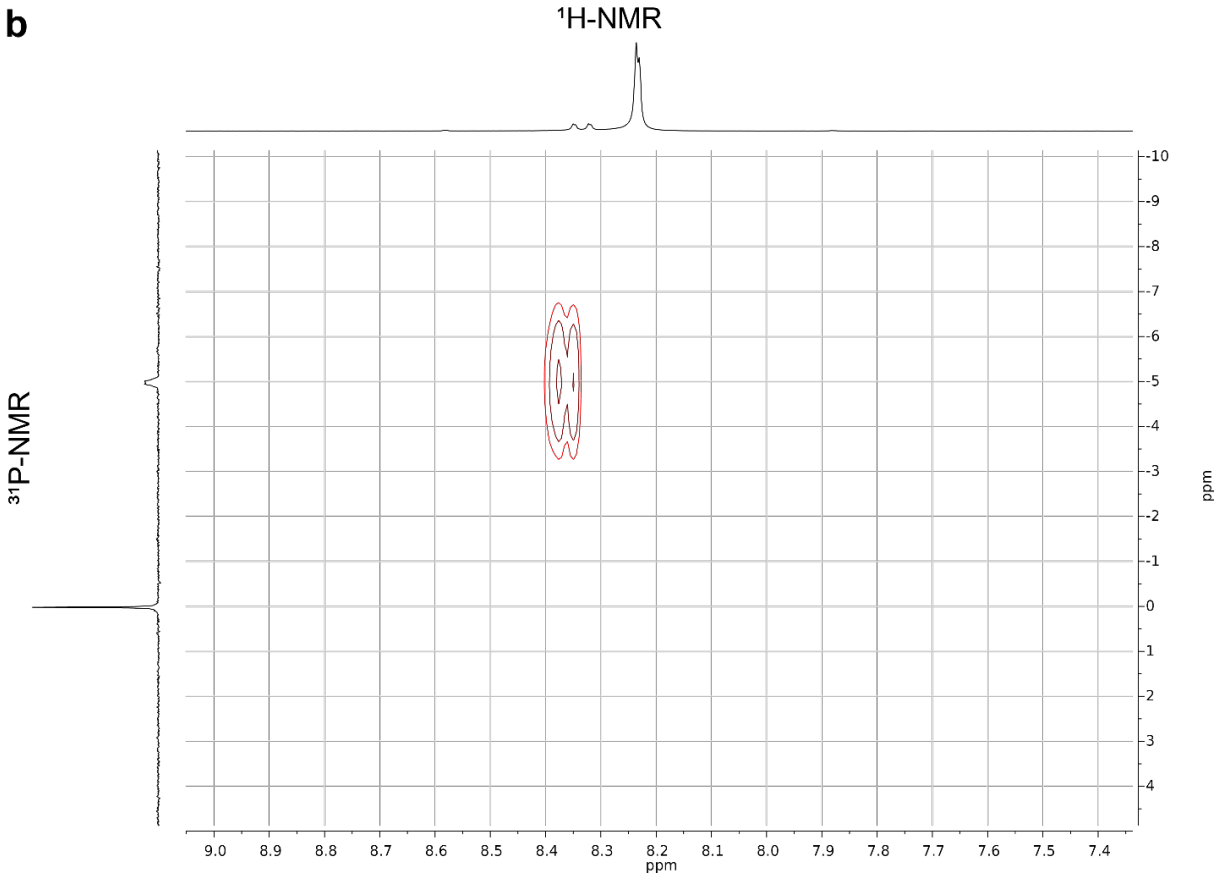
Supplementary Figure 3 Coupled and decoupled NMR of formyl phosphate A doublet is created by coupling between CH group and phosphorous. Measuring with suppressed coupling removes the doublet, confirming the proximity and formation of formyl phosphate. **a** $^{1}H\{^{31}P\}$-NMR and ^{1}H-NMR of formyl phosphate. Doublet indicates formyl phosphate, singlet at 8.22 ppm indicates formic acid. **b** $^{31}P\{^{1}H\}$-NMR and ^{31}P-NMR of formyl phosphate. Doublet indicates formyl phosphate, singlet at 0.1 ppm indicates dipotassium phosphate.

a



formic acetic anhydride 1.2 eq | K_2HPO_4 1.0 eq | formyl phosphate | acetic acid

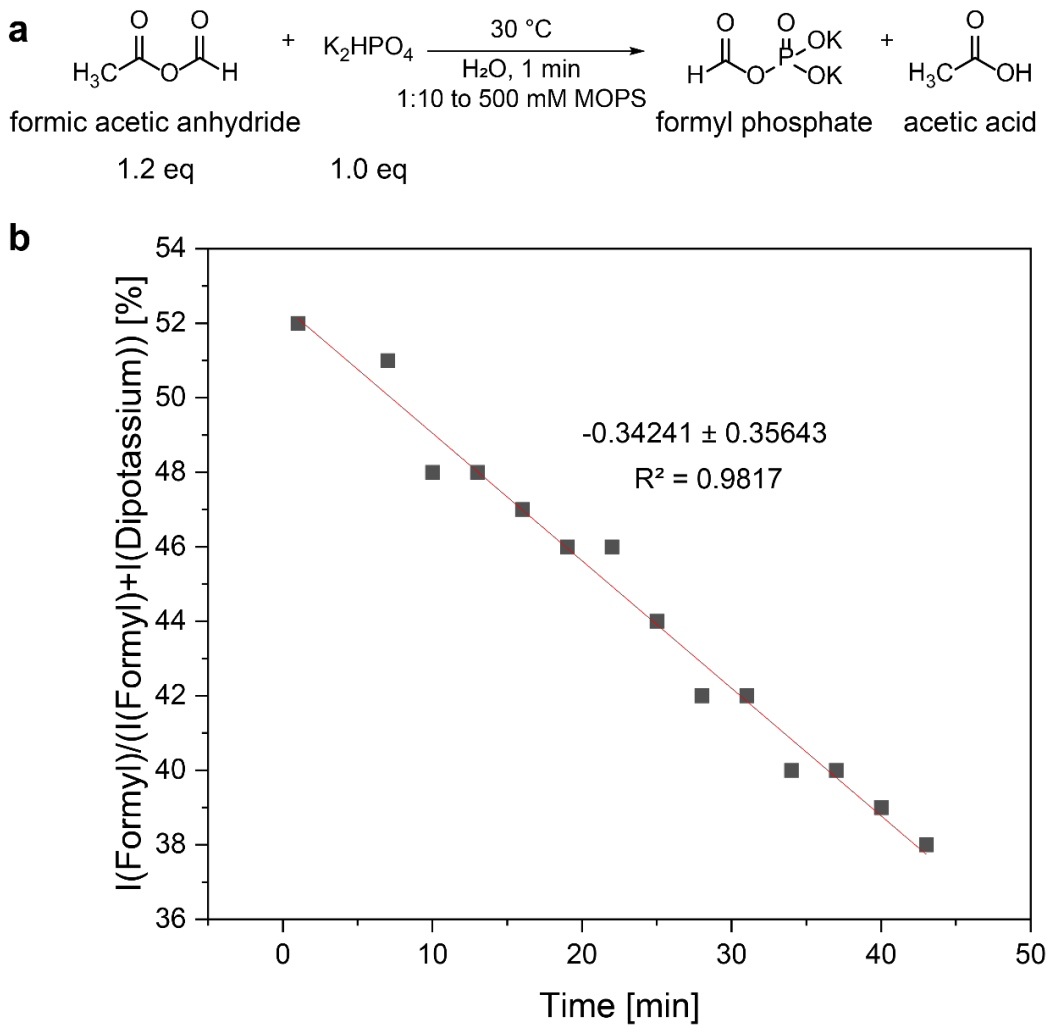
b



Supplementary Figure 4 Synthesis of formyl phosphate. a Reaction conditions for the synthesis of formyl phosphate. **b** ^{1}H-^{31}P-HMBC-NMR for formyl phosphate.

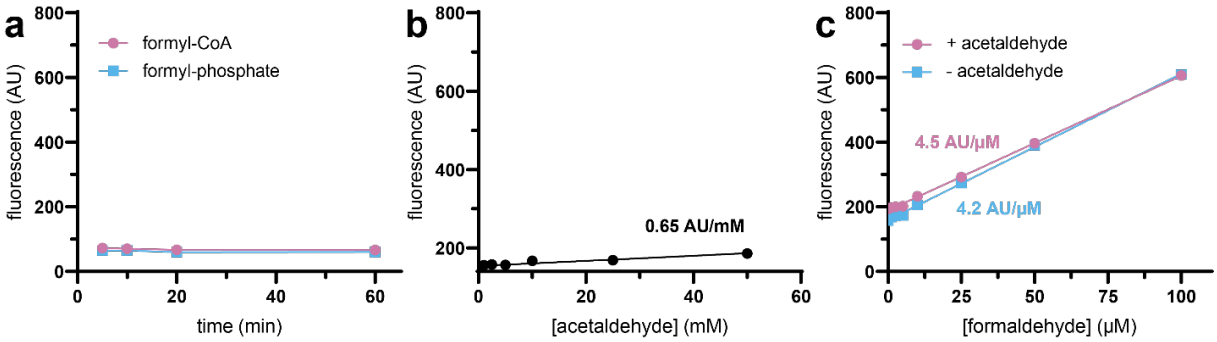
Hydrolysis rate of formyl phosphate under reaction conditions

The decomposition of formyl phosphate under similar conditions applied in the enzymatic process was investigated. An aliquot of a solution of formyl phosphate prepared as described above (0.1 mL) was added to a solution of MOPS buffer 0.5 M and magnesium chloride 0.01 M (0.9 mL) mixed shortly and directly monitored via $^{31}P\{^{1}H\}$-NMR.

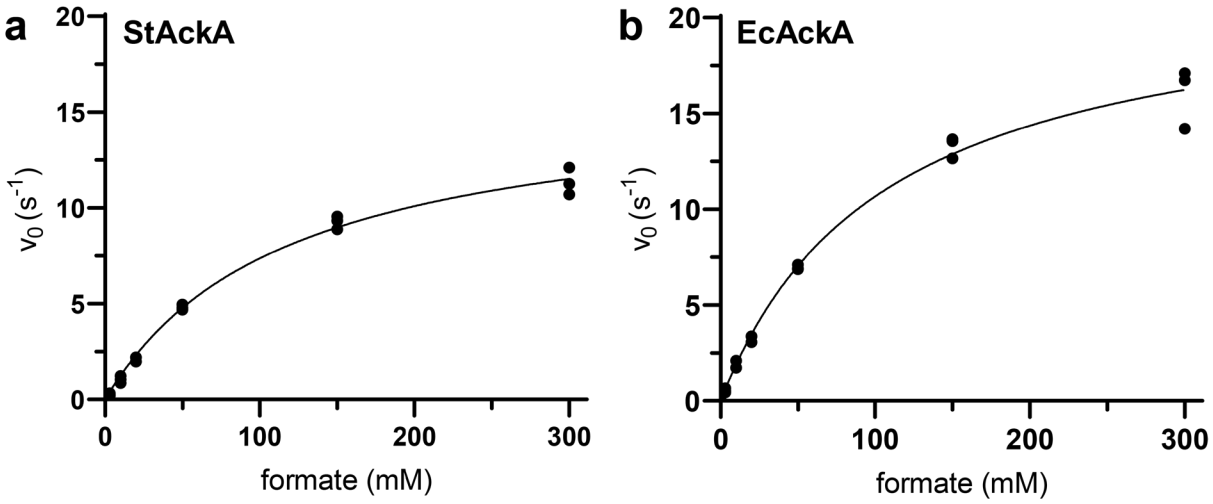


Supplementary Figure 5 Hydrolysis of formyl phosphate in reaction conditions. a Reaction conditions of formyl phosphate preparation for ArgC activity assays. **b** Hydrolysis rate of formyl phosphate as determined by quantitative 31P{1H}-NMR integrals of formyl phosphate (formyl) and dipotassium phosphate (dipotassium).

The concentration of formyl phosphate linearly decreased at a rate of -0.34 %min^{-1}. This experiment confirms that the hydrolytic decomposition is negligible as compared to enzymatic reaction.

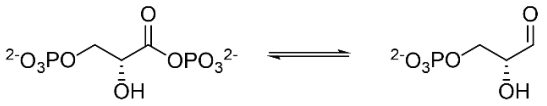


Supplementary Figure 6 Controls for the Nash reaction. **a** reactivity of Nash reagent with formyl-CoA and formyl phosphate. Formyl-CoA and formyl phosphate were produced *in situ* by FCS and FOK, respectively. **b** reactivity of Nash reagent with acetaldehyde. **c** Reactivity of Nash reagent with formaldehyde in the presence and absence of 50 mM acetaldehyde. AU = arbitrary units. AcALD = acetaldehyde. FALD = formaldehyde.



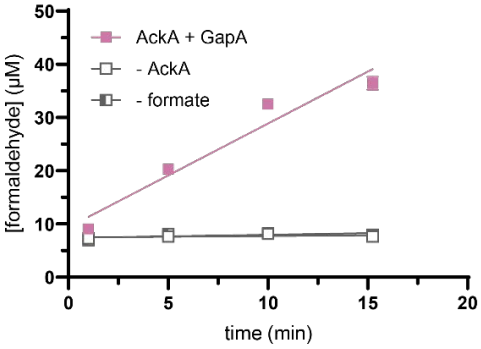
Supplementary Figure 7 Kinetics for formyl-phosphate synthesis from formate by acetate kinases. a Kinetic determination of *S. typhimurium* AckA. **b** Kinetic determination of *E. coli* AckA. Initial velocity was determined in triplicates and a Michaelis Menten kinetic was constructed. Individual replicates are shown.

GapA glyceraldehyde-3-phosphate dehydrogenase

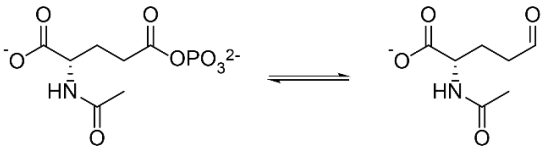


1,3-bisphosphoglycerate

glyceraldehyde-3-phosphate

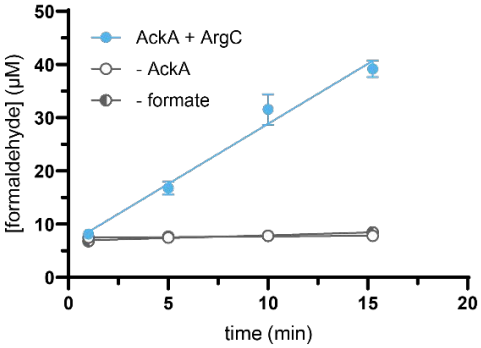


ArgC N-acetyl-gamma-glutamyl phosphate reducase

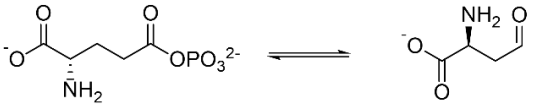


N-acetyl-gamma-glutamyl phosphate

N-acetyl-gamma-L-glutamate semialdehyde

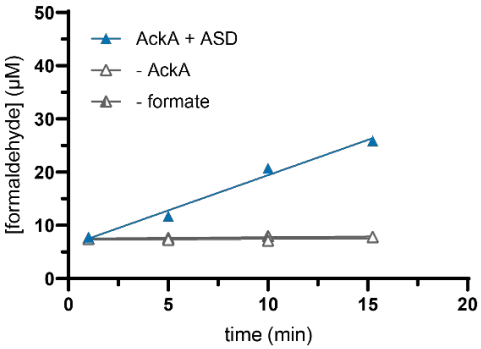


ASD aspartate-semialdehyde dehydrogenase

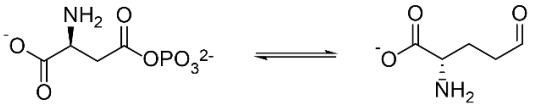


beta-aspartyl-phosphate

L-aspartate-semialdehyde

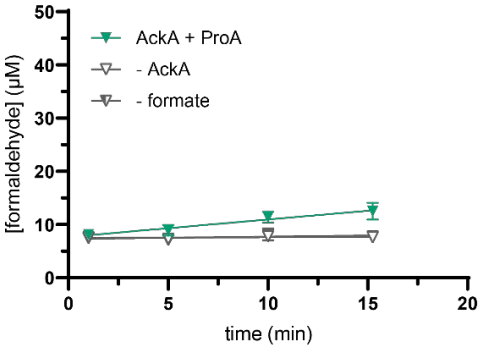


ProA gamma-glutamyl phosphate reductase

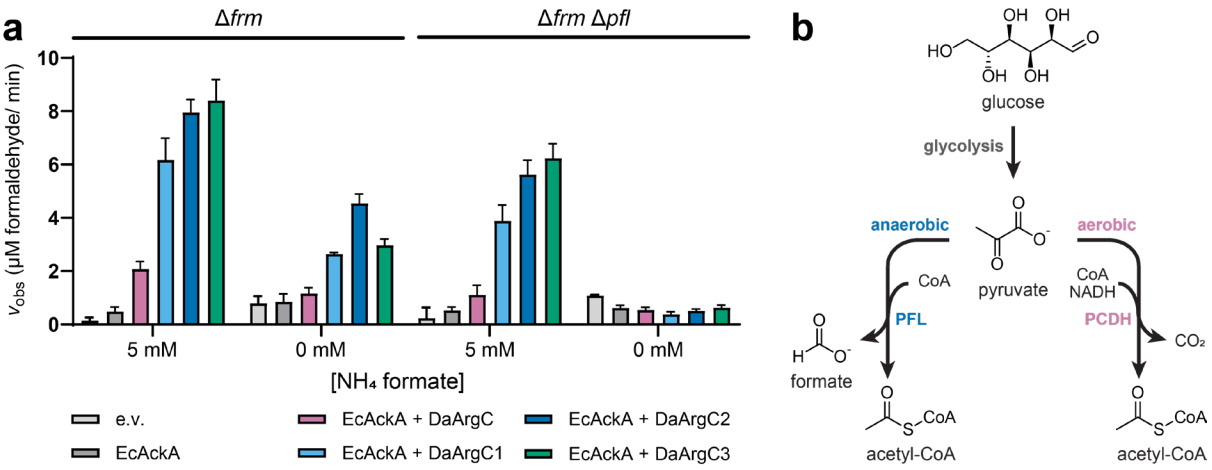


gamma-glutamyl-phosphate

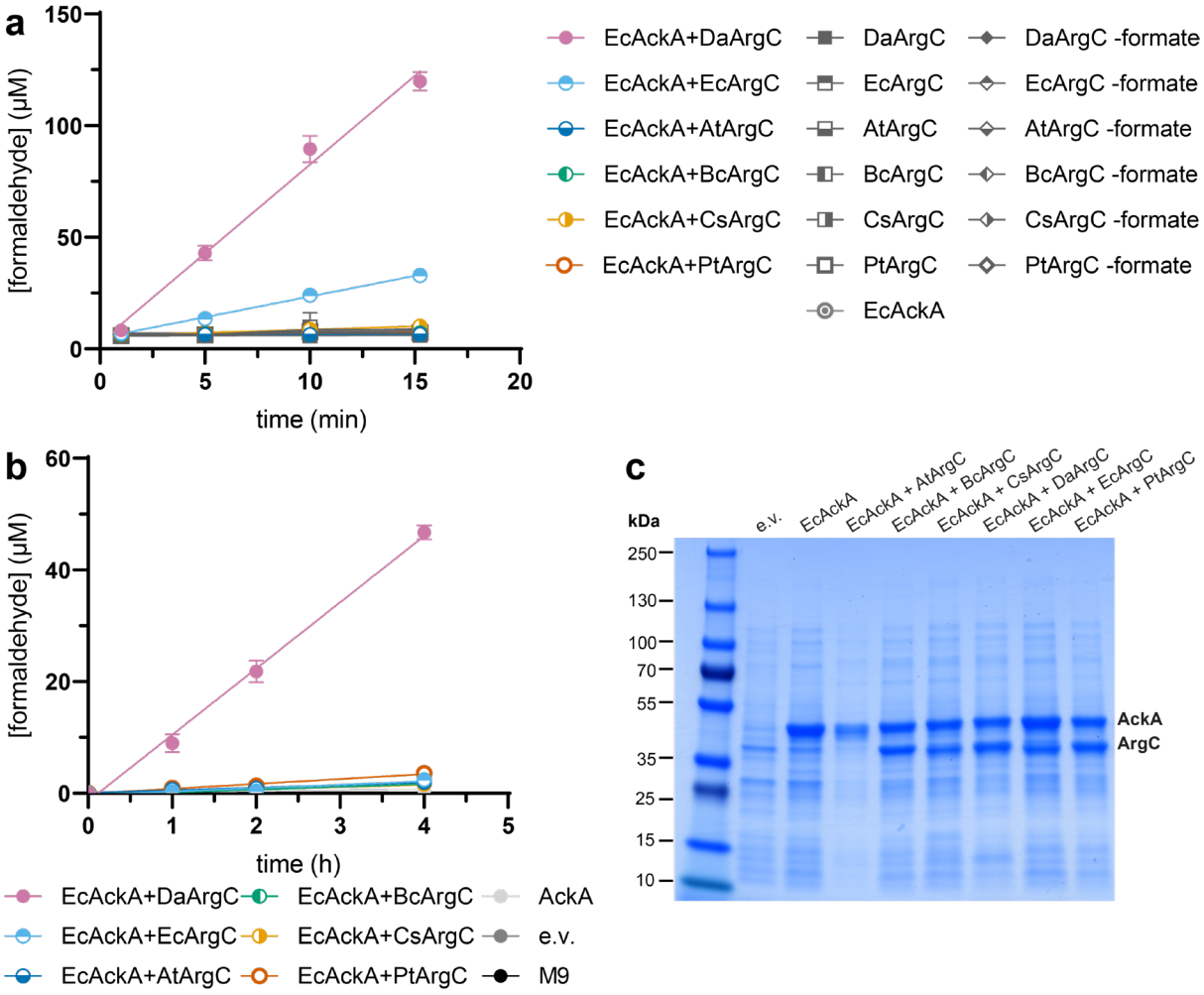
gamma-L-glutamate-semialdehyde



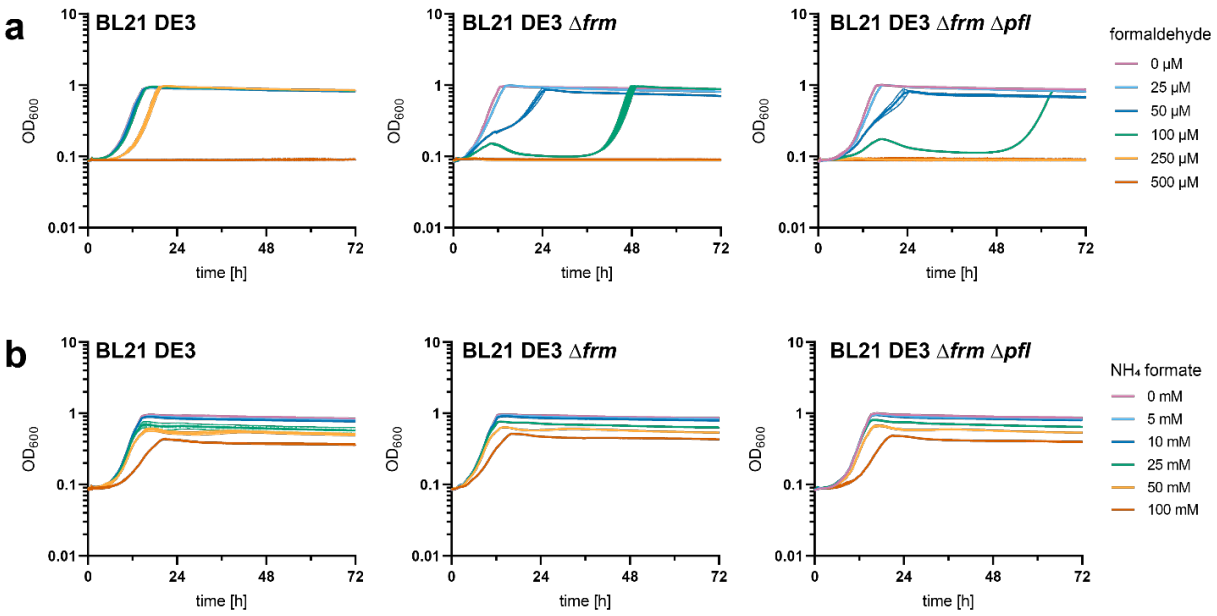
Supplementary Figure 8 Native function and promiscuous activities of formyl phosphate reductase candidates. Chemical structure of native substrate is given. Graphs depict FPR activity and relevant controls. Shown are mean and standard error of triplicates with linear fit.



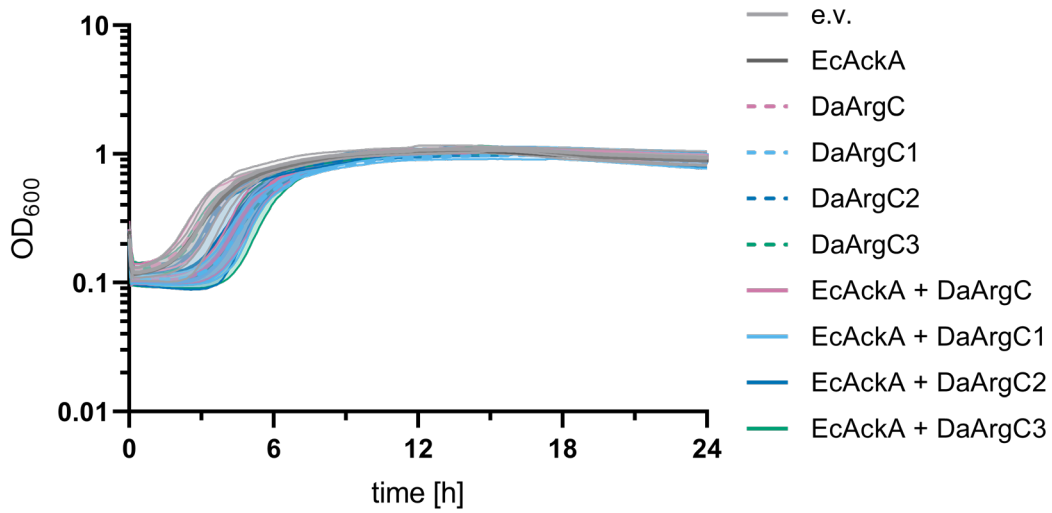
Supplementary Figure 9 Validation of DaArgC variants. a Background formaldehyde production in BL21 DE3 Δ*frmRAB* and BL21 DE3 Δ*frmRAB* Δ*pflAB* on 0 mM and 5 mM formate. Shown are mean and standard error; n = 3. **b** Metabolic network of *E. coli* from glucose to acetyl-CoA. In anerobic conditions, PFL produces formate and acetyl-CoA from pyruvate. PFL = pyruvate formate lyase. PCDH = pyruvate dehydrogenase.



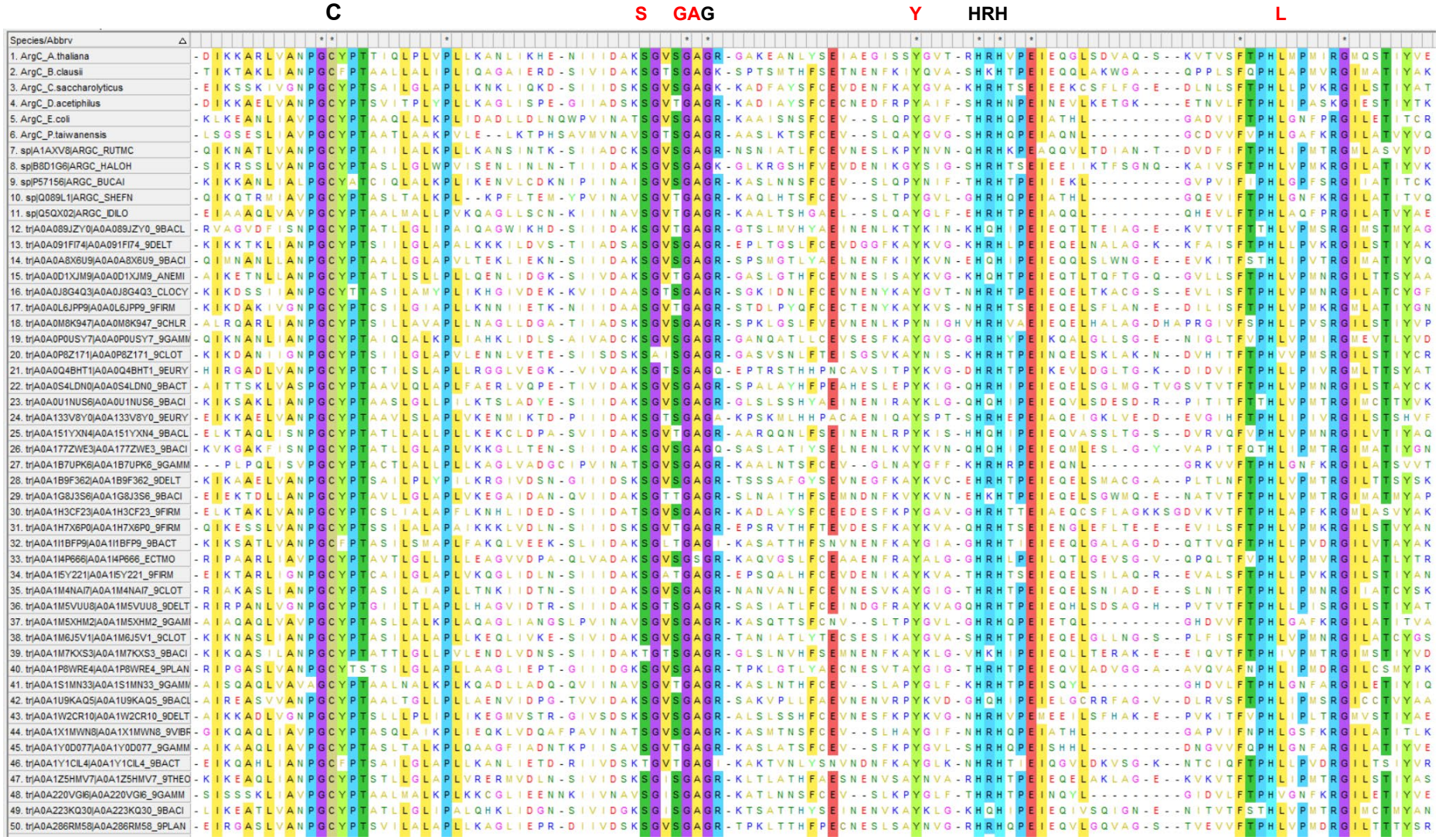
Supplementary Figure 10 ***In vivo*** **formaldehyde production by ArgC homologues**. **a** Formaldehyde production by EcAckA+ArgC *in vitro* including relevant controls. Formaldehyde was detected by Nash assay. Shown are mean and standard error; n = 3. **b** Formaldehyde production by the EcAckA+ArgC module in BL21DE3 Δ*frmRAB*. Formaldehyde was detected by Nash assay. Shown are mean and standard error; n = 3. **c** SDS-Page of protein production. Bands for AckA and ArgC are indicated.



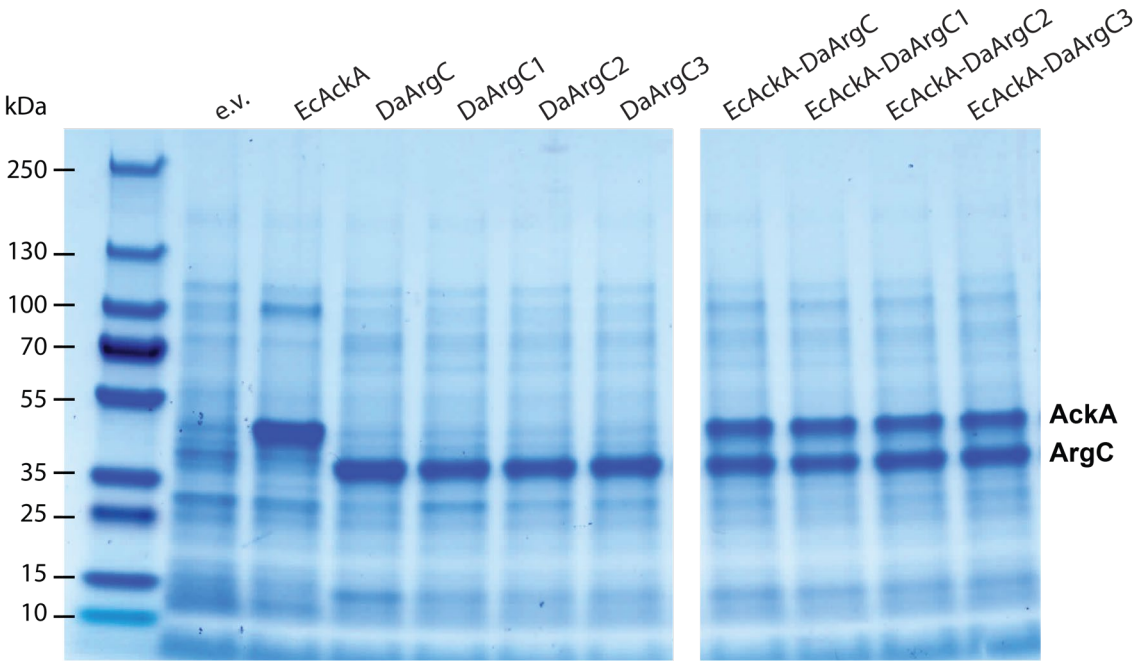
Supplementary Figure 11. Toxicity of formaldehyde (**a**) and ammonium formate (**b**) on the growth of E.coli BL21 DE3 deficient in formaldehyde detoxification system *frmRAB* and/or pyruvate formate lyase *pflAB*. Shown are mean and standard error; n = 3.



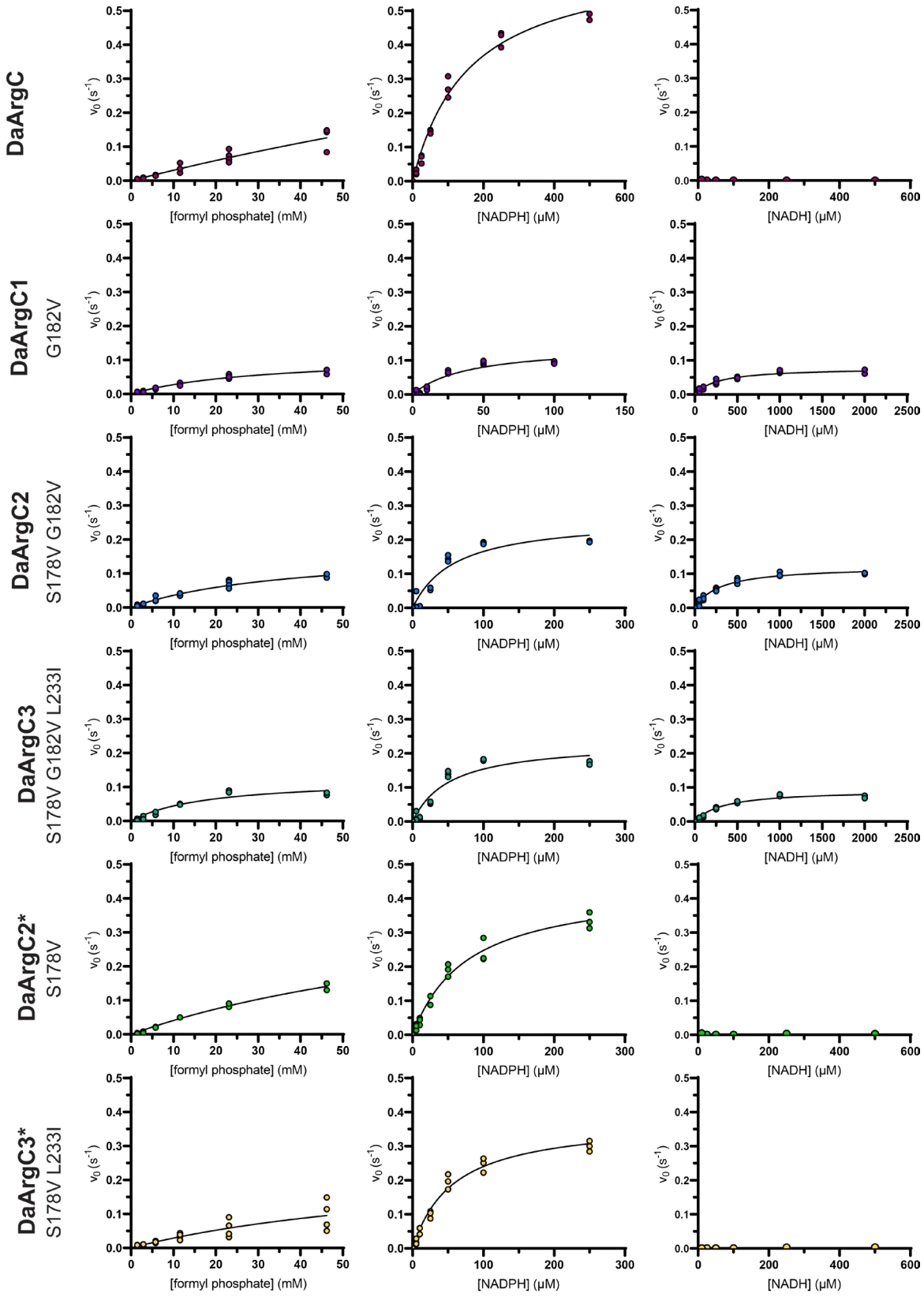
Supplementary Figure 12. Rescue of BL21DE3 Δ*frmRAB* cells carrying EcAckA and/or DaArgC variants after assay conditions (overexpression followed by 4 h incubation with 50 mM formate). Shown are mean an standard error; n = 3.



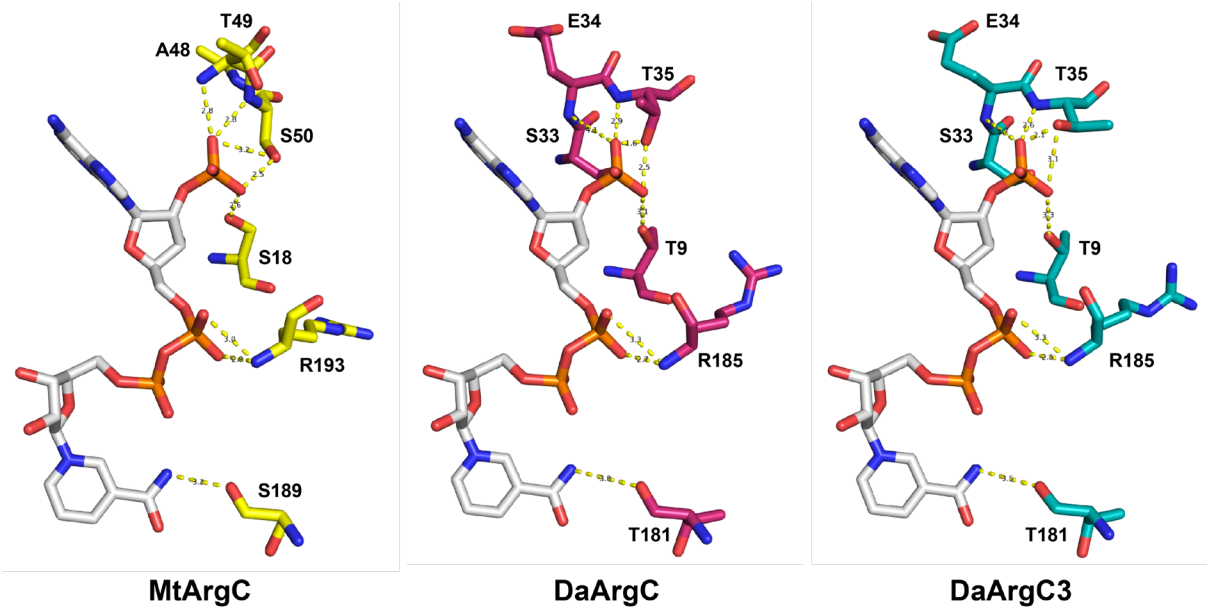
Supplementary Figure 13 Alignment of ArgC homologues. Conserved residues (70%) are highlighted. Conserved active site cysteine, GAG motif and HRH motif are indicated. Target residues for saturation mutagenesis are labelled in red.



Supplementary Figure 14 *In vivo* production of DaArgC variants SDS-Page of protein production in BL21DE3 Δ*frmRAB* Δ*pflAB*. Bands for AckA and ArgC are indicated**.** Both sectors belong to the same gel. Irrelevant bands were cut out.



Supplementary Figure 15 Kinetic evaluation of DaArgC variants. Row 1 kinetics on formyl phosphate. Row 2 kinetics on NADPH, coupled to saturating concentrations of formate and non-limiting amounts of EcAckA. Row 3 kinetics on NADPH, coupled to saturating concentrations of formate and non-limiting amounts of EcAckA. Shown are individual replicates and a Michaelis Menten fit.



Supplementary Figure 16 Enzyme engineering of DaArgC. Cartoon depiction of the NADPH-binding domains of mtArgC (PBD-ID 2I3G) and the putative NADPH-binding domains of DaArgC (PBD-ID 8AFU) and DaArgC3 (PBD-ID 8AFV). Relevant residues are labelled.

Supplementary Tables

Supplementary Table 1 Literature efficiencies of ACS and ACR variants.

enzyme	K_M (mM)	k_{cat} (s^{-1})	productivity (µM formaldehyde/s)	publication
stACS	~ 120	~ 12		Wang et al., 2021[23]
mhACS	~ 90	~ 6		Wang et al., 2021
ecACS	~ 50	~ 8		Wang et al., 2021
arACS	~ 40	~ 6		Wang et al., 2021
ttACS	~ 150	~ 5		Wang et al., 2021
sdACS	~ 150	~ 9		Wang et al., 2021
LmACR***	10.4 ± 2.2	0.99 ± 0.07		Chou et al., 2019[22]
ACS-bmACR*			~ 0.14 (lysate)	Wang et al., 2021
ACS-LmACR*			~ 0.2 (lysate)	Wang et al., 2021
BsACS-LmACR**			~ 0.08 (in vitro)	Hu et al, 2022[19]
BsACS-LmACR**			~ 0.08 (in vitro)	Hu et al, 2022

Supplementary Table 2. Kinetic parameters of DaArgC variants DaArgC2* and DaArgC3*. Parameters were determined by Michaelis Menten fit (see **Figure SI 15**). Standard error of triplicates is indicated. (1) indicates measurements coupled to EcAckA. (2) maximal activity observed, [formyl-phosphate] = 46 mM. n.d. = not detected

enzyme	**mutations**	**formyl phosphate**			**NADPH** (1)			**NADH**(1)		
		$K_{M,\ app}$ (mM)	$k_{cat,\ app}$ (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	$K_{M,\ app}$ (mM)	$k_{cat,\ app}$ (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	$K_{M,\ app}$ (mM)	$k_{cat,\ app}$ (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
DaArgC2*	S178V	no fit	0.14 ± 0.01(2)	-	0.08 ± 0.01	0.44 ± 0.02	$5.5*10^3$	n.d.	n.d.	-
DaArgC3*	S178V L233I	no fit	0.10 ± 0.05(2)	-	0.06 ± 0.01	0.38 ± 0.02	$6.3*10^3$	n.d.	n.d.	-

Supplementary References

[1] Krimen, L. I. (1970) Acetic Formic Anhydride, *Org. Synth. 50*.

[2] Wang, J., Anderson, K., Yang, E., He, L., and Lidstrom, M. E. (2021) Enzyme engineering and in vivo testing of a formate reduction pathway, *Synth. Biol. 6*, ysab020.

[3] 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.

[4] Hu, G., Guo, L., Gao, C., Song, W., Liu, L., and Chen, X. (2022) Synergistic Metabolism of Glucose and Formate Increases the Yield of Short-Chain Organic Acids in Escherichia coli, *ACS Synth. Biol. 11*, 135-143.

Chapter 2

Engineering a highly efficient carboligase for synthetic one-carbon metabolism

Maren Nattermann[1,*], Simon Burgener[1,*], Pascal Pfister[1], Alexander Chou[2], Luca Schulz[1], Seung Hwan Lee[2], Nicole Paczia[1], Jan Zarzycki[1], Ramon Gonzalez[2], Tobias J. Erb[1,3]

[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 Biomedical Engineering, University of South Florida, Tampa, FL, USA

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

ACS Catalysis, ***2021***, *11: 5396-5404*

Author contributions

The manuscript was written through contributions of all authors. S.B. and T.J.E. conceived of the project. M.N. and S.B. performed most of the experiments. P.P. and J.Z. solved the crystal structures. A.C. performed whole-cell bioconversion experiments. L.S. characterized CoA transferases. S.H.L. performed and analyzed protein expression in bioconversion strains. N.P. developed and performed LC-MS analyses. R.G. and T.J.E conceptualized and supervised the work and acquired funding. All authors analyzed the results and wrote the manuscript. All authors have given approval to the final version of the manuscript. * These authors contributed equally.

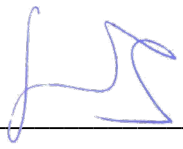
Ein Entwurf der umseitig genannten Veröffentlichung ist Bestandteil einer weiteren Qualifikationsarbeit:

Dr. Simon Burgener

Kumulative Dissertation

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

Zum Erlangen des wissenschaftlichen Grades Dr. rer. nat.



Unterschrift des Mitverfassers

Abstract

One of the biggest challenges to realize a circular carbon economy is the synthesis of complex carbon compounds from one-carbon (C1) building blocks. Since the natural solution space of C1-C1 condensations is limited to highly complex enzymes, the development of more simple and robust biocatalysts may facilitate the engineering of C1 assimilation routes. Thiamine diphosphate-dependent enzymes harbor great potential for this task, due to their ability to create C-C bonds. Here, we employed structure-guided iterative saturation mutagenesis to convert oxalyl-CoA decarboxylase (OXC) from *Methylobacterium extorquens* into a glycolyl-CoA synthase (GCS) that allows for the direct condensation of the two C1 units formyl-CoA and formaldehyde. Quadruple variant MeOXC4 showed a 100,000-fold switch between OXC and GCS activity, a 200-fold increased GCS activity compared to the wild type and a formaldehyde affinity that is comparable to natural formaldehyde-converting enzymes. Notably, MeOCX4 outcompetes all other natural and engineered enzymes for C1-C1 condensations by more than 40-fold in catalytic efficiency and is highly soluble in *Escherichia coli*. In addition to increased GCS activity, MeOXC4 showed up to 300-fold higher activity than the wild type towards a broad range of carbonyl acceptor substrates. When applied *in vivo*, MeOXC4 enables the production of glycolate from formaldehyde, overcoming the current bottleneck of C1-C1 condensation in whole-cell bioconversions and paving the way towards synthetic C1 assimilation routes *in vivo*.

Introduction

The synthesis of complex molecules from one-carbon (C1) compounds is key to a circular economy. C1 compounds, in particular formate and methanol, can be derived directly from CO_2 through several processes, including hydrogenation, photochemistry, electrochemistry or biocatalysis,[1-5] and further serve as feedstock for the formation of value-added products through microbial fermentation.[6-9] Because natural methylo- and formatotrophic microorganisms are not well-suited for large-scale biotechnological processes, current efforts aim at engineering well-established microbial platform organisms by implementing natural and new-to-nature C1 converting pathways into these microbes.[10-13]

As highlighted by several recent studies, linear C1 converting pathways are particularly interesting,[10, 12, 14] since they minimally interfere with the host metabolism, and, unlike cyclic pathways, are robust towards draining of intermediates. However, linear pathways require the direct condensation of two C1 units, which is chemically challenging, due to the difficulty to generate C1 nucleophiles (in contrast to C1 electrophiles, which are readily available, notably in the form of CO_2 and formaldehyde). Only two enzymes are known that catalyze direct C1-C1 condensation in nature: acetyl-CoA synthase (ACS) and glycine synthase (GS). They are the key enzymes of the reductive acetyl-CoA pathway and the reductive glycine pathway, respectively.[15, 16] GS and ACS are structurally and mechanistically highly complex biocatalysts (the latter being limited to strictly anaerobic conditions),[17, 18] which illustrates the challenge of direct C1-C1 condensations and their implementation in microbial platform organisms. Thus, the discovery or development of simpler C1-C1 carboligases could enable new-to-nature linear carbon fixation pathways and ultimately facilitate the engineering of whole-cell catalysts for efficient C1 conversions.[7, 19]

One way to generate nucleophilic (one)-carbon centers is to invert the reactivity of a carbonyl species through Umpolung, which enzymes can achieve through the cofactor thiamine diphosphate (ThDP).[20] In recent years, several ThDP-dependent enzymes have been engineered to catalyze C1-extension reactions, which underscores the general potential of these enzymes for synthetic carbon fixation pathways. The most prominent example is the new-to-nature enzyme formolase (FLS), which was derived from benzaldehyde lyase.[21] Similarly, glycolaldehyde synthase (GLS) was engineered from benzoylformate decarboxylase.[22] Although in both cases a very low initial activity was improved ~100-fold by

directed evolution, the final enzyme variants still exhibited low catalytic efficiencies (k_{cat}/K_M < 10 M^{-1} s^{-1}), as well as poor affinity for the highly toxic formaldehyde (K_M ≥ 170 mM). While these engineered enzymes indeed were able to support new-to-nature linear C1 assimilation pathways *in vitro*, the low substrate affinity and activity limited their implementation *in vivo* so far.[21, 22]

We recently identified another class of ThDP-dependent enzymes as potential formaldehyde carboligases. Members of the 2-hydroxyacyl-CoA lyase (HACL) / oxalyl-CoA decarboxylase (OXC) enzyme family catalyze the condensation of formyl-CoA with formaldehyde to produce glycolyl-CoA. This activity will be referred to as glycolyl-CoA synthase (GCS) (**Figure 1A**).[23, 24] Notably, HACL from *Rhodospirillales bacterium URHD0017* (RuHACL) displayed more than ten-fold higher catalytic efficiency (k_{cat}/K_M = 110 M^{-1} s^{-1}) compared to the engineered FLS and GLS. However, efforts of further engineering HACL towards higher GCS-activity have not been successful[23], mainly because of the lack of structural information. So far, no HACL structure is available and homology models fail to accurately predict the structure of the C-terminal active site loop, which exhibits low sequence homology throughout the HACL/OXC family (**Figure S1**). Moreover, another limitation of HACLs is their poor production in *E. coli*.[23, 24]

To overcome these challenges with HACLs, we recently focused on repurposing OXCs, which naturally catalyze the decarboxylation of oxalyl-CoA, as formyl-CoA condensing enzymes.[24] Here, we aimed at improving the GCS-activity of OXC to enable the production of glycolyl-CoA at high rates under physiologically relevant formaldehyde concentrations (<0.5 mM). Using structure-guided enzyme engineering, we converted OXC into a bona fide GCS through several rounds of iterative site mutagenesis (ISM)[25] and demonstrated its function in an *E. coli* whole-cell bioconversion system to improve current production limitations and pave the way towards synthetic C1 fixation pathways.

Results

MeOXC structure identifies targets for mutagenesis

Previously, we showed that besides the decarboxylation of oxalyl-CoA into formyl-CoA, OXC from *Methylorubrum extorquens* (MeOXC) is also capable of catalyzing the acyloin condensation

of formyl-CoA with various aldehydes, including formaldehyde.[24] However, the catalytic efficiency of MeOXC with formaldehyde (k_{cat}/K_M = 2 $M^{-1}s^{-1}$) is far below that of RuHACL, which is the best performing HACL known to date (k_{cat}/K_M = 110 $M^{-1}s^{-1}$). Additionally, the K_M for both substrates is extremely high (formaldehyde: 100 mM and formyl-CoA: 3 mM; **Table 1**). We therefore sought to engineer MeOXC towards a more efficient GCS.

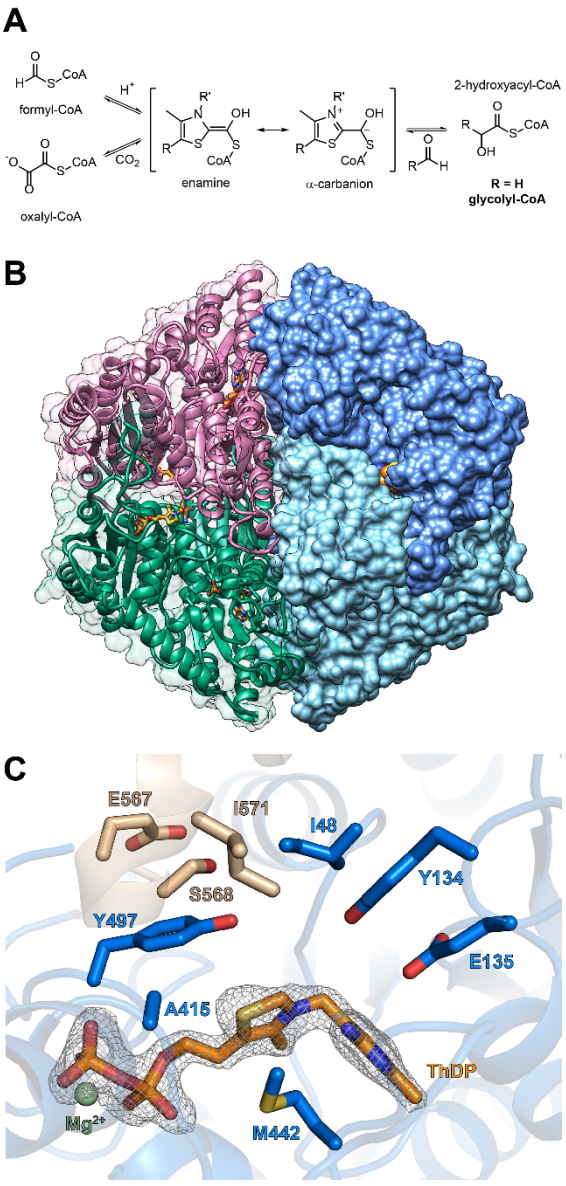


Figure 1. Crystal structure of MeOXC reveals amino acids surrounding the ThDP cofactor as potential targets for engineering GCS-activity. A) Reaction scheme of OXC and HACL. The reaction proceeds via the α-hydroxyl-CoA-ThDP covalent intermediate on the ThDP cofactor, which has an enamine and an α-carbanion resonance structure. The intermediate can perform a nucleophilic attack on an aldehyde, giving rise to 2-hydroxyacyl-CoA (glycolyl-CoA in the case of the GCS reaction). B) Overall structure of MeOXC with bound ThDP (orange). Shown is a homotetramer. C) Amino acids of MeOXC (blue) in the environment of ThDP (orange). The C-terminal loop of the OfOXC structure is shown in light brown. The simulated annealing omit map around the refined ThDP is shown at σ = 4. The amino acids I48, E134, Y135, A415, Y497, S568, E567 and I571 were selected for saturation mutagenesis, based on their probable proximity to the α-carbanion/enamine intermediate (see A and Figure S2).

MeOXC is closely related to OXC from *Oxalobacter formigenes* (OfOXC, 63% identity) and *E. coli* (EcOXC, 61%), for both of which structures are available.[26-28]

To gain insights into the specific active site topology of MeOXC, we solved the crystal structure of the enzyme at a resolution of 1.9 Å (**Figure 1B**). Overall, the structure is very similar to that

of OfOXC (rmsd = 0.425 Å for 7230 aligned atoms). Because we could not observe electron density after residue E567, we modeled the C-terminal part (16 residues) based on the OfOXC structure (PDB ID 2ji7). In OfOXC this part is flexible and electron density for the closed conformation was obtained only after soaking the enzyme with substrate or product.[28]

During catalysis, OXC and HACL both form the same α-carbanion/enamine intermediate (**Figure 1A and Figure S2**). In HACL this intermediate is generated through proton abstraction from formyl-CoA or by cleaving off the aldehyde moiety from a 2-hydroxyacyl-CoA thioester,[29] whereas in OXC this intermediate is formed by decarboxylation of oxalyl-CoA.[28] To increase the GCS-activity of MeOXC, we sought to systematically alter the active site around the ThDP cofactor with ISM.[25] Based on the crystal structure, we identified eight residues in proximity of the ThDP cofactor as targets: I48, Y134, E135, A415, Y497, E567, S568, and I571 (**Figure 1C**).

Establishing a high-throughput screen for GCS-activity

ISM requires screening of thousands of variants.[30] We thus conceived a high-throughput screen based on the conversion of glycolyl-CoA to glycolate, which is subsequently oxidized to glyoxylate by glycolate oxidase (GOX) under stoichiometric production of H_2O_2 (**Figure 2A**). H_2O_2 is quantified by horseradish peroxidase (HRP) that catalyzes the oxidation of Ampliflu Red® to the fluorophore resorufin.

To convert glycolyl-CoA into glycolate, we sought to establish a glycolyl-CoA:formate CoA-transferase (GFT). Such an enzyme would not only turn glycolyl-CoA into glycolate, but at the same time also regenerate formyl-CoA from formate, thereby closing a catalytic cycle for the continuous conversion of formate and formaldehyde into glycolate. We screened formyl-CoA:oxalate CoA-transferase from *Oxalobacter formigenes*[31] (FRC), propionate CoA-transferase from *Cupriavidus necactor*[32] (CnPCT) and *Clostridium propionicum*[33] (CpPCT), as well as 4-hydroxybutyrate CoA-transferase from *Clostridium aminobutyricum*[34] (AbfT). The latter showed best GFT activity ($k_{cat,app.}$ = 0.40 ± 0.01 s^{-1}, $K_{M,app}$(glycolyl-CoA) = 12 ± 1 μM; **Figure 2C&D and S3**) and was chosen for the assay.

We validated our screen with purified MeOXC, AbfT, human GOX[35] and HRP (**Figure 2B & S4**) and further demonstrated that this assay could be used to quantify MeOXC activity from *E. coli* lysates in 384-well plates.

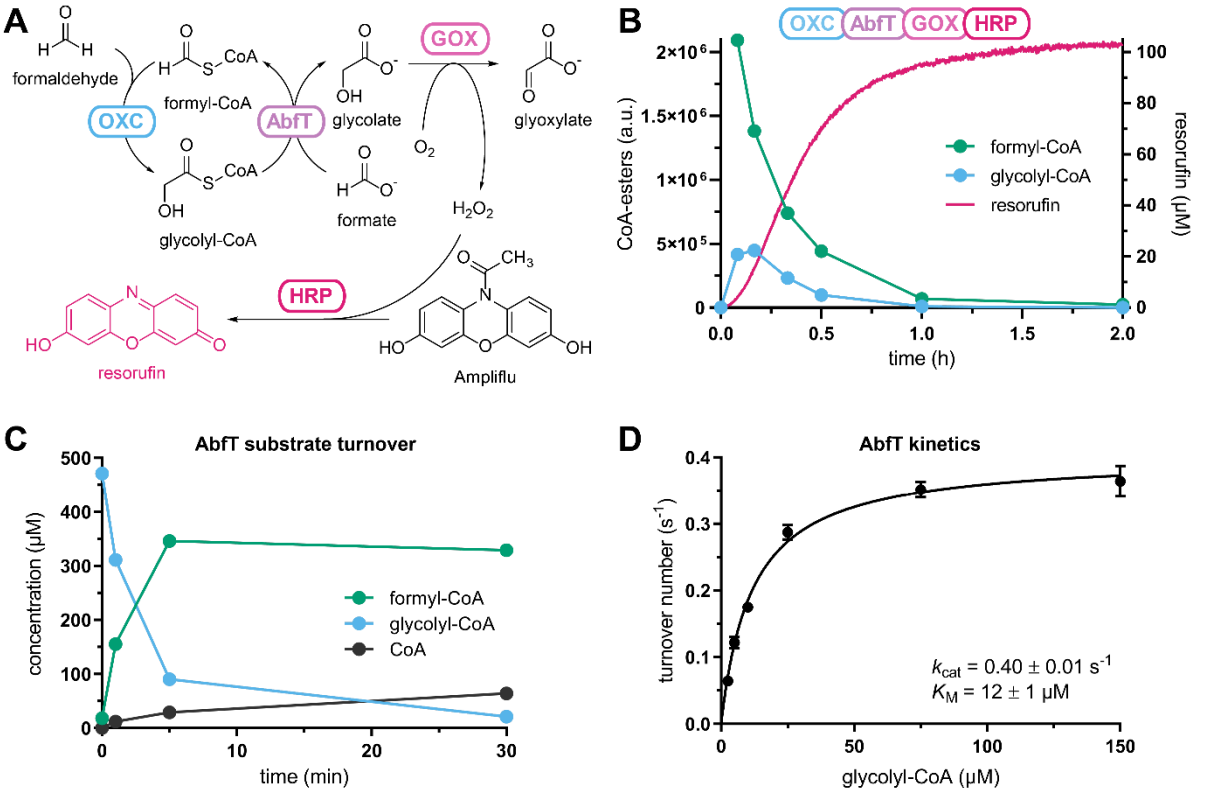


Figure 2. Establishing a high-throughput read-out for GCS-activity. A) Reaction scheme indicating enzymes for each step. Overall, glycolyl-CoA formation is detected via the fluorophore resorufin. **B)** Reaction progress of the OXC-AbfT-GOX-HRP cascade. Resorufin was detected via fluorescence in a plate-reader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

Iterative Saturation Mutagenesis of MeOXC

Having established a high-throughput screen, we created saturation mutagenesis libraries of the eight identified active site residues using the 22c trick (**Figure S5**).[30] We employed sub-saturating concentrations of formaldehyde and formyl-CoA (50 mM and 0.5 mM, respectively) to screen for MeOXC variants with improved GCS-activity and higher affinity for both substrates. In the first round of ISM (R1), libraries of position 48 and 571 contained only

variants with reduced activity (**Figure 3**), suggesting that the isoleucines in both of these positions are critical for catalysis. Therefore, these residues were not screened in the following rounds R2 to R6. The libraries of the remaining six residues mostly contained variants with WT-like activity but also some variants with significantly increased GCS-activity, validating that these positions are potential targets to improve enzyme activity (**Figure 3**). The best performing variant in R1 was an alanine to cysteine substitution in position 415.

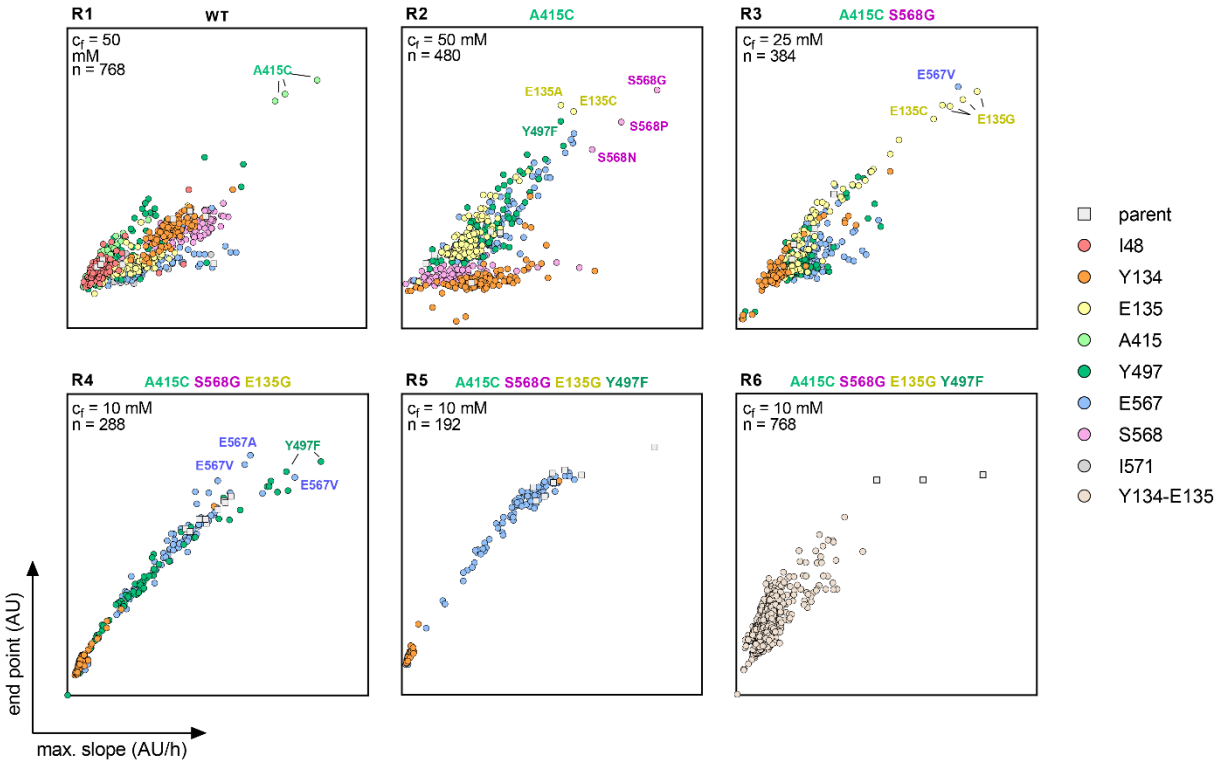


Figure 3. ISM of MeOXC for improved GCS-activity. MeOXC saturation mutagenesis libraries were screened for GCS-activity using a fluorescent-based readout in cell-free extracts. Plotted is the maximal slope (AU/h) versus end point (AU) of product formation. Each panel contains all libraries of the corresponding round. The best performing variants of each round were sequenced and their substitution is shown in the graphs with colored labels. The parent variant of each round is indicated above the corresponding panel and was included in the screen as a reference. c_f = formaldehyde concentration in the screen, n = total number of screened clones per round, AU = arbitrary units.

Steady-state kinetics with purified enzyme showed that in the A415C variant the K_M for formaldehyde and formyl-CoA, decreased 3-fold and 19-fold, respectively (**Table 1**),

confirming that sub-saturating formaldehyde concentrations could be used in our screens to identify enzyme variants with improved kinetics.

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 substitution S568G conferred a ~3-fold improvement in formaldehyde affinity, prompting us to decrease the formaldehyde concentration to 25 mM in R3, in which we identified E135G. For R4, we further lowered the concentration of formaldehyde to 10 mM to identify the substitution 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 possible variants), but this library also contained no improved variants (**Figure 3**). Thus, after screening ~3,600 clones in six rounds, we obtained final variant MeOXC4 carrying four substitutions E135G, A415C, Y497F, and S568G.

Table 1. Apparent steady-state parameters of MeOXC variants.

	Formaldehyde			**Formyl-CoA**		
Enzyme	$k_{cat,app}$ (s^{-1})	$K_{M,app}$ (mM)	k_{cat}/K_M (s^{-1} M^{-1})	$k_{cat,app}$ (s^{-1})	$K_{M,app}$ (mM)	k_{cat}/K_M (s^{-1} M^{-1})
RuHACL[a]	3.3 ± 0.3	29 ± 8	110	1.5 ± 0.1	0.20 ± 0.05	7,200
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	1.6 ± 0.1	12 ± 3	130	1.6 ± 0.1	0.17 ± 0.04	9,400
MeOXC4	2.0 ± 0.2	5 ± 1	400	2.4 ± 0.2	0.22 ± 0.03	11,000

Errors reflect the standard deviation of three independent measurements. Michaelis-Menten plots are shown in **Figure S7**. a) data taken from Chou et al.[23]

Carboligation activity is improved at the expense of decarboxylation activity in MeOXC1-4

Next, we characterized MeOXC4 and the intermediate variants from R1-3 (MeOXC1-3) in more detail. All variants were produced in *E. coli* at levels comparable to the native EcOXC (**Figure 4A and S6**), indicating that their improvement was based on increased catalytic properties, and not on improved solubility and/or stability. The catalytic efficiency of the GCS activity improved over each round of the ISM, ultimately resulting in a 200- and 110-fold improvement in k_{cat}/K_M for formaldehyde and formyl-CoA, respectively, due to a lowered K_M for both substrates, in combination with a tenfold increased k_{cat} (**Table 1**).

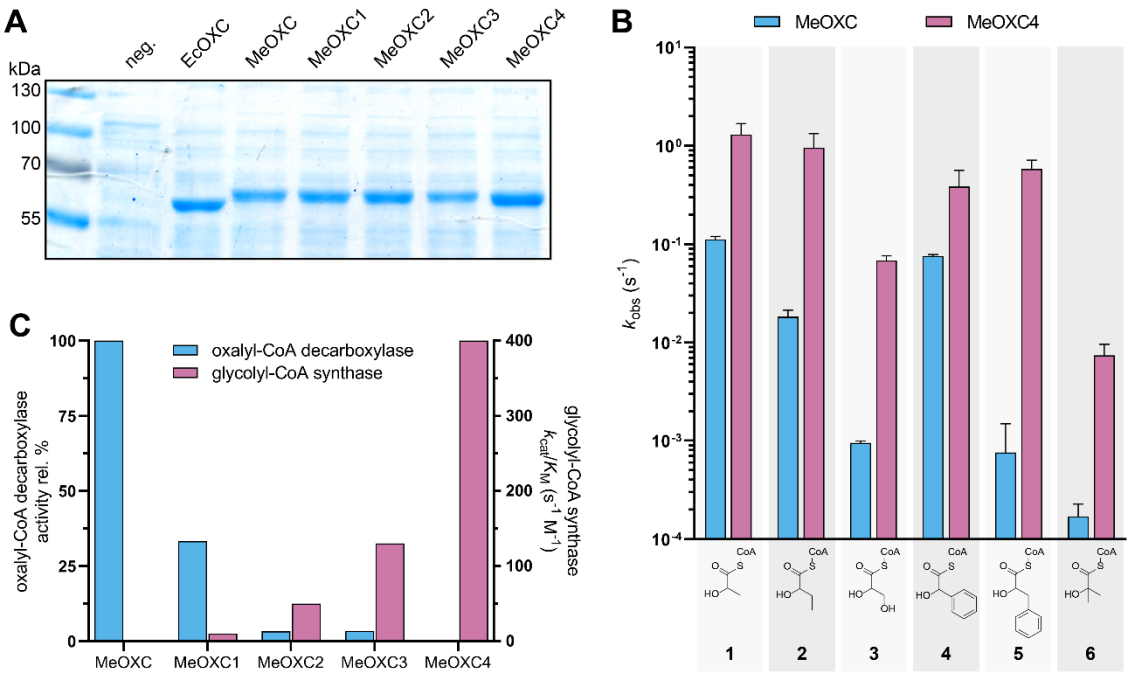


Figure 4. Characterization of MeOXC mutants from the ISM. A) SDS-PAGE analysis of the variants expressed in *E. coli* BL21(DE3). A total of 5 µg protein was loaded in each lane and empty pET-16b was used as negative control. EcOXC is OXC from *E. coli*. **B)** Comparison of the aldehyde substrate scope of MeOXC and MeOXC4. Reactions contained 1 mM formyl-CoA and the corresponding acceptor substrate (formaldehyde, acetaldehyde, propionaldehyde, glycolaldehyde 100 mM; benzaldehyde, phenylacetaldehyde 10 mM; acetone 1 M). Initial velocity was determined by LC-MS. Error bars show the standard deviation of three replicates. Chemical structures of products are shown below the graph: 1 lactyl-CoA, **2** 2-hydroxybutyryl-CoA, **3** glyceryl-CoA, **4** mandelyl-CoA, **5** 3-phenyllactyl-CoA, **6** 2-hydroxyisobutyryl-CoA. Full datasets are shown in **Figure S8**.**C)** Comparison of the OXC- and GCS-activities. OXC-activity is the velocity of oxalyl-CoA decarboxylation at 1 mM oxalyl-CoA, relative to the WT. For GCS-activity, the absolute catalytic efficiency for formaldehyde is shown.

While GCS-activity strongly increased over the course of the ISM, our MeOXC variants gradually lost their native OXC-activity (**Figure 4C**). Final variant 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), which was exclusively due to a reduced k_{cat}, as the K_M for oxalyl-CoA remained virtually identical to WT MeOXC (105 μM).[24] This finding is in line with previous studies on OfOXC and MeOXC, where residues Y134, E135, Y497 and S568 were replaced by alanine without significant changes to the K_M for oxalyl-CoA.[24, 28]

Interestingly, when added to the reaction, oxalyl-CoA did not affect the GCS activity of MeOXC4 (**Figure S9**), even at 1 mM, indicating that the enzyme's original substrate does not act as inhibitor, despite still showing a very favorable apparent K_M. This is probably due to a decreased on-rate (k_{on}) and/or an increased off-rate (k_{off}) for oxalyl-CoA, and an opposite trend for formyl-CoA. The latter was supported by further experiments, which showed that in the presence of formaldehyde, MeOXC4 preferred carboligation and converted oxalyl-CoA after decarboxylation directly into glycolyl-CoA (k_{obs} = 0.15 s^{-1}), releasing formyl-CoA only at a very slow rate of <0.01 s^{-1}. This is in stark contrast to MeOXC WT, which shows a high formyl-CoA release rate of 98 s^{-1} after oxalyl-CoA decarboxylation, followed by slow carboligation (0.31 s^{-1}) (**Figure S9**).

Combined, the four substitutions in MeOXC4 directly affected catalytic activity, as well as k_{on} and k_{off} rates of the different substrates, causing a specificity switch between native OXC (oxalyl-CoA decarboxylation) and GCS-activity (formyl-CoA condensation with formaldehyde) of greater than 100,000-fold.

The reversible (de)carboxylation activity of MeOXC is lost in MeOXC4

To confirm that the switch in activity was achieved by suppressing the native OXC reaction, we sought to also test the reverse reaction of OXC (i.e., the carboxylation of formyl-CoA). We envisioned an enzyme cascade in which the product of the reverse reaction, oxalyl-CoA, is constantly removed from the equilibrium by further reduction into glycolate (**Figure S10A**). This setup would render the overall reaction thermodynamically favorable (ΔG ≈ -19 kJ mol^{-1})[36] and allow to measure the reverse reaction, in contrast to earlier efforts, which had failed.[37]

To convert oxalyl-CoA into glyoxylate, we characterized putative oxalyl-CoA reductase PanE2 from *M. extorquens*[38] (**Figure S10B**). For the reduction of glyoxylate into glycolate, we employed GhrB from *E. coli*.[39] Combined with PanE2 and GhrB, MeOXC catalyzed the carboxylation of formyl-CoA at a rate of ~0.5 min^{-1} (**Figure S10C-E**), while formyl-CoA carboxylation was not detectable in MeOXC4 (**Figure S10D**), confirming that this variant had indeed lost its native OXC-activity.

Structural basis for improved GCS-activity in MeOXC4

To rationalize the effect of the different substitutions on catalysis, we solved the crystal structure of MeOXC4 to a resolution of 2.4 Å, modelled its flexible C-terminus beyond E567 as described before, and compared it to MeOXC WT and OfOXC.

Substitution A415C, which replaced fully conserved A415 in the HACL/OXC family by a cysteine, caused a decrease in the apparent K_M for formyl-CoA by more than an order of magnitude (**Table 1**).

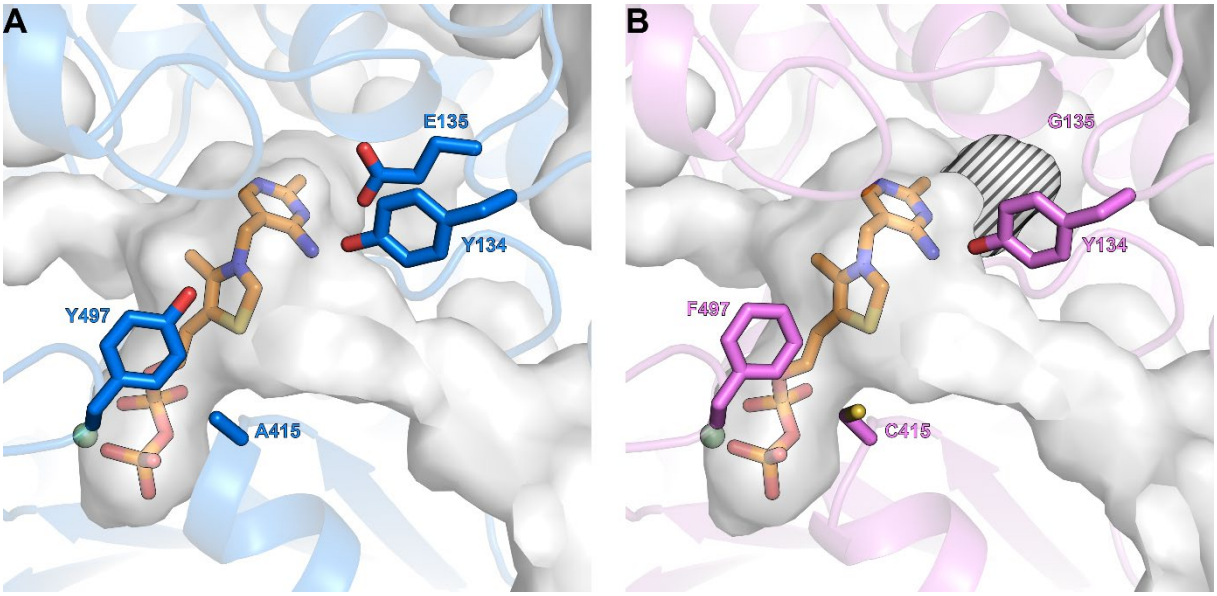


Figure 5. Comparison of ThDP binding sites. ThDP is shown in orange, Mg^{2+} in pale green. **A)** Active site of MeOXC WT. **B)** Active site of MeOXC4. The substitution E135G gave rise to a cavity in the ThDP-binding pocket, highlighted with the striped pattern. Consequently, Y134 may be optimally positioned for proton transfer reactions of the glycolyl-CoA-ThDP adduct (see **Figure S2**).

Structural analysis showed that C415 is in close proximity to S568 of the flexible C-terminal loop reaching the active site, suggesting that this substitution is important to (re-) organize the active site for the new substrate. Notably, no other amino acid substitution at this position has a beneficial effect on catalytic efficiency (**Figure 3**), indicating that the sulfide moiety of the cysteine seems to be important for formyl-CoA accommodation. However, when testing corresponding substitution A389C in RuHACL, GCS-activity was decreased by almost 50% at sub-saturating formyl-CoA concentrations (100 μM) (**Figure S11**), suggesting that formyl-CoA accommodation in MeOXC1-4 differs from RuHACL.

Substitution S568G lies within the flexible C-terminal loop, adjacent to A415C and likely provides space to accommodate the bulkier side chain of C415. This is in line with the observation that S568G was only beneficial in combination with A415C and was not detected in the first ISM round (**Figure 3**).

The third substitution E135G caused the greatest structural change by creating an extra cavity at the active site (**Figure 5**). Notably, none of the ISM libraries at residue 134 contained a variant of improved activity (**Figure 3**), suggesting that Y134 is important for GCS catalysis, likely by facilitating protonation of the glycolyl-CoA-ThDP intermediate (**Figure S2**). Thus, the E135G substitution does not serve in substrate accommodation, but likely allows optimal positioning of Y134 during catalysis, which is supported by the fact that k_{cat} was increased approximately fourfold by the E135G substitution, while the apparent K_M for formyl-CoA was slightly increased.

In OfOXC active site water molecules play a role during catalysis. Three ordered water molecules (W1-3) are observed in a structure with a trapped α-carbanion/enamine intermediate, with W2 (in close contact to W1 and W3) proposed to protonate Cα of the intermediate (**Figure S2**).[28] W1 is hydrogen-bonded to residues corresponding to Y134 and E135 , while Y497 and S568 form hydrogen bonds to W3.[28] In MeOXC4 this hydrogen bonding network to W1 and W3 is lost due to the E135G, Y497F and S568G substitutions. This change in the water network likely helps to promote GCS-activity at the expense of the OXC reaction. In summary, our structural analysis showed that the active site of MeOXC4 is re-organized for improved substrate accommodation and increased catalytic activity.

MeOXC4 shows a broad substrate scope for different C1 extensions

Having improved the catalytic efficiency of C1 extensions of formaldehyde by MeOXC4 by more than two orders of magnitude, we wondered whether the engineered enzyme would also promote the acyloin condensation of formyl-CoA with acceptor substrates other than formaldehyde. Indeed, MeOXC4 accepted a broad range of aldehyde substrates, including small hydrophilic and aliphatic aldehydes, bulky hydrophobic aldehydes and acetone (**Figure 4B**). The activity for all tested substrates was significantly higher than WT MeOXC (up to a factor of 340 in the case of phenylacetaldehyde). This broad substrate scope makes MeOXC4 a versatile catalyst for C1 extension of carbonyl acceptors that can be exploited for the biocatalytic production of valuable 2-hydroxy acids, such as lactic acid, mandelic acid and 3-phenyllactic acid.

Application of MeOXC4 to one-carbon bioconversion in *E. coli* whole-cells

Finally, we aimed at testing the performance of our engineered enzymes also *in vivo*. We previously established an *Escherichia coli* whole-cell bioconversion system based on RuHACL.[23] When combined into one pathway with an acyl-CoA reductase from *Lysteria monocytogenes* (LmACR) and *E. coli* aldehyde dehydrogenase AldA (EcAldA), whole cells converted formaldehyde into glycolate (**Figure 6A**). However, despite several attempts to optimize glycolate production, carbon flux was insufficient to support biotechnological applications, mainly due to the low abundance and the high apparent formaldehyde K_M of RuHACL, which posed a major bottleneck.[23]

To test the effects of engineered MeOXC *in vivo*, we replaced RuHACL with variants MeOXC1-4 (**Figure 6B**). In line with the increasing GCS-activities of the different MeOXC variants, glycolate production was successively increased. The best mutant, MeOXC4, outperformed RuHACL in glycolate productivity twofold (**Figure 6B**), likely because of the enzyme's more favorable kinetics (k_{cat}/K_M = 400 M^{-1} s^{-1} versus 110 M^{-1} s^{-1} for RuHACL) as well as improved expression of MeOXC4 compared to RuHACL, even when the latter was codon-optimized, as confirmed for the MG1655-derived host strain (**Figure S12**).

These result suggested that the other enzymes of the pathway, in particular LmACR (k_{cat}/K_M = 95 M^{-1} s^{-1})[23], became rate-limiting. To optimize concentration of all enzymes in the pathway,

we tested different expression levels, using a two plasmid system with independent inducible promoters. MeOXC4 was expressed from the IPTG-inducible T7 promoter, while LmACR and EcAldA were co-expressed from a cumate-inducible T5 promoter (**Figure 6C**). Combinatorically screening of several IPTG and cumate concentrations revealed that highest glycolate production levels were reached at low induction levels of MeOXC4 ([IPTG] = 40 µM) and high induction levels of LmACR/EcAldA ([cumate] = 400 µM) (**Figure 6C**).

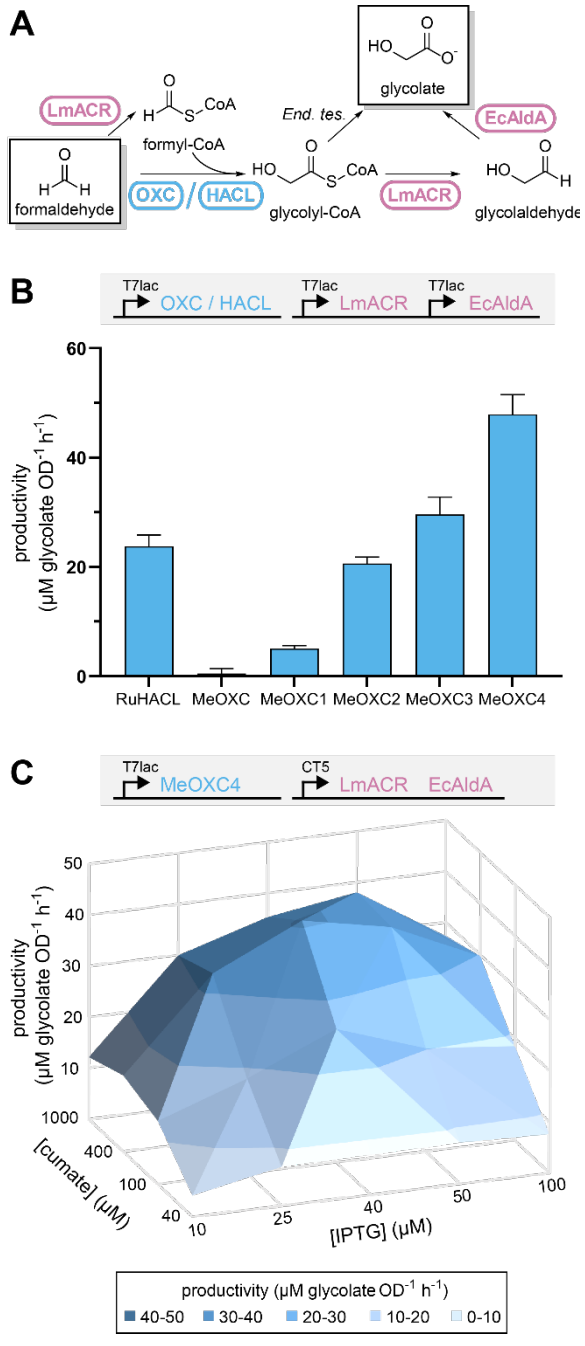


Figure 6. Assessing the performance of MeOXC in a synthetic one-carbon bioconversion pathway using *E. coli* whole-cells. A) The synthetic pathway converting formaldehyde to glycolate used for screening OXC/HACL variants. *End. tes.*: endogenous thioesterases. **B)** Productivity of *E. coli* whole-cells expressing the indicated OXC/HACL variant based on glycolate formation in 3-hour bioconversions of 5 mM formaldehyde. Expression of all pathway enzymes was controlled by the IPTG-inducible T7lac promoter, with 100 µM IPTG used for induction. Bars are drawn to the mean values, and error bars represent standard deviation of triplicate experiments. **C)** Inducer concentration dependency of the productivity of *E. coli* whole-cells expressing MeOXC4. Productivity is the same as defined for panel B. Expression of MeOXC4 was controlled by the IPTG-inducible T7lac promoter, while expression of LmACR and EcAldA was controlled by a cumate-inducible T5 promoter (CT5). Mean of n = 4 replicates is shown as a surface plotted against IPTG and cumate concentration.

This is in contrast to our previous results, where productivity was highest at high RuHACL and low LmACR/EcAldA induction levels.[23] Additionally, glycolate productivity was decreased by more than two-fold, when we replaced LmACR by a different ACR, *Rhodopseudomonas palustris* PduP[40], which shows a higher k_{cat} but also higher K_M for formyl-CoA (**Figure S13**). Taken together, these results support the hypothesis that the glycolate productivity of our system is currently not limited by C1-C1 condensation, but ACR activity *in vivo*. Thus, enhancing ACR activity will be key towards improving and further developing this C1 fixation pathway for biotechnological applications.

Discussion

Through ISM we successfully evolved OXC from *M. extorquens* into a *bona fide* GCS by switching decarboxylation and carboligation activity of the enzyme by 100,000-fold (**Figure 4C**). Notably, none of the newly introduced amino acids of MeOXC4 is found in any HACL homolog (**Figure S1**), indicating that through directed evolution an alternative (presumably local) maximum in the GCS activity landscape was found. This raises the question how the reactivity of OXC and HACL is determined.

It has been proposed that in OXC after decarboxylation the α-carbanion/enamine intermediate is nonplanar, rendering the Cα more basic and facilitating the rate-limiting protonation step, yielding formyl-CoA.[28] In contrast, an enamine-like planar structure was observed in ThDP–dependent carboligases that require a carbonyl acceptor substrate.[41-43] It is tempting to speculate that HACLs and MeOXC4 also stabilize the enamine state of the intermediate and thereby favor nucleophilic attack on the carbonyl acceptor substrate.

Notably, engineered MeOXC4 shows kinetic parameters that are comparable with natural formaldehyde converting biocatalysts. Only two naturally occurring enzymes are known to be involved in formaldehyde fixation, 3-hexulose-6-phosphate synthase (HPS) in the ribulose monophosphate pathway and formaldehyde transketolase (FTK) in the dihydroxyacetone pathway. The reported values of the Michaelis constant for formaldehyde range between 0.15 to 3 mM and 0.4 mM to 1.9 mM for HPS and FTK, respectively.[44-47] With an apparent K_M of 5 mM for formaldehyde, engineered MeOXC4 is fully compatible with an *in vivo* application, in contrast to other (engineered) C1-C1 carboligases that show apparent K_M values of 170 mM

(GLS) and 29 mM (RuHACL) respectively, formaldehyde concentrations that are toxic to *E. coli*. Compared to ACS and GS, MeOXC4 is a low-complexity C1-C1 carboligase that is homotetrameric (i.e. requires only one gene), oxygen-insensitive and requires only ThDP and Mg^{2+} as cofactors. Additionally, MeOXC4 can be produced at high levels in *E. coli*, which makes it a versatile tool for C1 extensions. Taken together, our results highlight the potential of enzyme engineering to create new-to-nature C1 enzymes and pathways for a sustainable (bio)catalysis and biotechnology.

Acknowledgements

We thank N. S. Cortina and P. Claus for LC-MS measurements. We acknowledge G. Santoni at the European Synchrotron Radiation Facility Grenoble, France (ESRF, beamlines ID23-1) as well as D. von Stetten at the Deutsches Elektronen-Synchrotron Hamburg, Germany (DESY, beamline P13). We thank P. Wichmann for help with CoA-ester synthesis. This work was supported by the German Ministry of Education and Research Grant FormatPlant (BMBF grant 031B0194 (FormatPlant) and 031B0850B (Met*A*For)) and by the U.S. Department of Energy (Award Number: DE-EE0008499).

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. 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, *Nat. Energy 5*, 703-710.

[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, *Chem. Rev. 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. Biotechnol. 62*, 168-180.

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

[9] Satanowski, A., and Bar-Even, A. (2020) A one-carbon path for fixing CO2 : C1 compounds, produced by chemical catalysis and upgraded via microbial fermentation, could become key intermediates in the valorization of CO2 into commodity chemicals, *EMBO Rep. 21*, e50273.

[10] 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.

[11] 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*, 933-946.

[12] 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.

[13] Keller, P., Noor, E., Meyer, F., Reiter, M. A., Anastassov, S., Kiefer, P., and Vorholt, J. A. (2020) Methanol-dependent Escherichia coli strains with a complete ribulose monophosphate cycle, *Nat. Commun. 11*, 5403.

[14] 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, *Nat. Microbiol. 5*, 1459-1463.

[15] Ljungdahl, L. G. (1986) The autotrophic pathway of acetate synthesis in acetogenic bacteria, *Annu. Rev. Microbiol. 40*, 415-450.

[16] Sanchez-Andrea, I., Guedes, I. A., Hornung, B., Boeren, S., Lawson, C. E., Sousa, D. Z., Bar-Even, A., Claassens, N. J., and Stams, A. J. M. (2020) The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans, *Nat. Commun. 11*, 5090.

[17] 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.

[18] Kikuchi, G., Motokawa, Y., Yoshida, T., and Hiraga, K. (2008) Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia, *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci. 84*, 246-263.

[19] 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.

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

[21] 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.

[22] 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.

[23] 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.

[24] 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.

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

[26] Werther, T., Zimmer, A., Wille, G., Golbik, R., Weiss, M. S., and Konig, S. (2010) New insights into structure-function relationships of oxalyl CoA decarboxylase from Escherichia coli, *FEBS J. 277*, 2628-2640.

[27] Berthold, C. L., Moussatche, P., Richards, N. G. J., 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.

[28] 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.

[29] 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, *J. Biol. Chem. 280*, 9802-9812.

[30] 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 Synth. Biol. 2*, 83-92.

[31] Baetz, A. L., and Allison, M. J. (1990) Purification and characterization of formyl-coenzyme A transferase from Oxalobacter formigenes, *J. Bacteriol. 172*, 3537-3540.

[32] Volodina, E., Schurmann, M., Lindenkamp, N., and Steinbuchel, A. (2014) Characterization of propionate CoA-transferase from Ralstonia eutropha H16, *Appl. Microbiol. Biotechnol. 98*, 3579-3589.

[33] 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, *Eur. J. Biochem. 269*, 372-380.

[34] Scherf, U., and Buckel, W. (1991) Purification and properties of 4-hydroxybutyrate coenzyme A transferase from Clostridium aminobutyricum, *Appl. Environ. Microbiol. 57*, 2699-2702.

[35] 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.

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

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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, *Proc. Natl. Acad. Sci. U.S.A. 99*, 591-595.

[42] 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.

[43] 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, *Nat. Chem. Biol. 2*, 324-328.

[44] 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.

[45] 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, *Appl. Microbiol. Biotechnol. 76*, 439-445.

[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, *J. Bacteriol. 179*, 6041-6047.

[47] 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.

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).

Synthesis of CoA-esters

Formyl-CoA and oxalyl-CoA were synthesized and purified as described previously.[1] Glycolyl-CoA, lactyl-CoA, mandelyl-CoA, 3-phenyllactyl-CoA and 2-hydroxyisobutyryl-CoA were synthesized via carbonyldiimidazole coupling, as described previously[2, 3]. Glyceryl-CoA and 2-hydroxybutyryl-CoA were synthesized enzymatically. For glyceryl-CoA, 100 mM sodium phosphate pH 6.5, 5 mM $MgCl_2$, 0.15 mM ThDP, 100 mM propionaldehyde, 10 mM formyl-CoA were mixed and the reaction was initiated by adding MeOXC Y497A[1] to a final concentration of 2 mg/mL. After 23 minutes at 30 °C, the reaction was quenched with formic acid (final concentration 5% (v/v)). For 2-hydroxybutyryl-CoA, 100 mM sodium phosphate pH 6.5, 5 mM $MgCl_2$, 0.15 mM ThDP, 100 mM propionaldehyde, 10 mM formyl-CoA were mixed and the reaction was initiated by adding MeOXC4 to a final concentration of 2 mg/mL. After 23 minutes at 30 °C, the reaction was quenched with formic acid (final concentration 5% (v/v)).

All CoA-esters were purified by preparative HPLC-MS with a methanol gradient in 25 mM ammonium formate pH 4.2 (formyl-, glycolyl- and oxalyl-CoA) or 8.1 (glyceryl-, 2-hydroxybutyryl-, 2-hydroxyisobutyryl-, lactyl-, mandelyl-, 3-phenyllactyl-CoA). 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[4], following NADH consumption at 340 nm.

Cloning and Mutagenesis

Strains and plasmids are listed in **Table S1**. Human *gox* was codon optimized and synthesized by Twist Bioscience, PCR amplified using the primers listed in **Table S2** and cloned into pET28b using NdeI and BamHI. *panE2* was PCR-amplified from *M. extorquens* chromosomal DNA using the corresponding primers (**Table S2**). The purified PCR products were digested with NdeI and BamHI and ligated into pET-16b.

Point mutations were introduced by PCR using mismatch primers (**Table S2**). 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. Correct cloning was confirmed by sequencing (Eurofins Genomics, Ebersbach, Germany and Microsynth AG, Balgach, Switzerland)

For *E. coli* whole-cell bioconversion experiments, MeOXC and mutant variants 1-4 were PCR-amplified using primers (**Table S2**) having 15-bp homology to the destination vector, pCDFDuet-1 (Novagen). pCDFDuet-1 was linearized by restriction digest using EcoRI-HF (NEB Inc.). PCR and digestion products were gel purified and ligated using the In-Fusion HD EcoDry cloning kit (Takara Bio USA) following manufacturer directions. Resulting plasmids isolated from single colonies were verified by sequencing (Eurofins Genomics LLC). Sequence verified plasmids were further used to create bioconversion strains (**Table S1**) by electroporation using standard protocols.

For *E. coli* whole-cell bioconversion experiments, MeOXC and mutant variants 1-4 were PCR-amplified using primers (**Table S2**) having 15-bp homology to the destination vector, pCDFDuet-1 (Novagen). pCDFDuet-1 was linearized by restriction digest using EcoRI-HF (NEB Inc.). PCR and digestion products were gel purified and ligated using the In-Fusion HD EcoDry cloning kit (Takara Bio USA) following manufacturer directions. Resulting plasmids isolated from single colonies were verified by sequencing (Eurofins Genomics LLC). Sequence

verified plasmids were further used to create bioconversion strains (**Table S1**) by electroporation using standard protocols.

Protein Production and Purification

All proteins were heterologously produced in *E. coli* BL21 (DE3). 500 mL TB containing 100 μg/mL ampicillin (AbfT, FRC, GOX, MeOXC(1-4), PanE2, PCTs, PduP, RuHACL G390N, RuHACL A389C G390N) 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. 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 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 7.8, 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. MeOXC WT and mutants were flash-frozen in liquid nitrogen and stored at −80 °C.

SDS-PAGE analysis of OXC expression

Protein was produced as described above. Protein concentration in the clarified cell lysate was determined with the Bradford assay, using bovine serum albumin for calibration.[5] 5 µg of total protein was loaded in each lane. As negative control, empty pET-16b plasmid was used and MeOXC expression was compared to EcOXC (overexpressed from plasmid pCA24N-EcOXC).

Crystallization & Structure Determination

MeOXC and MeOXC4 were purified as described above. Immediately after affinity purification, the eluate was loaded onto a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) 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 ThDP and 1 mM CoA and crystal plates were set up using the sitting drop vapor diffusion method, diluting equal volume of enzyme in reservoir solution. For MeOXC the reservoir solution contained 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 glycolyl-CoA improved the final resolution of the crystals without resulting in clear 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.[6] The dataset was scaled using the program suite ccp4.[7] The Phenix software package[8] 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.[9] Final B-facor refinement and water positioning was also performed via phenix.refine.

LC-MS/MS analysis of CoA-esters

Samples were prepared for LC-MS/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. CoA-esters were separated on an Agilent Infinity II 1290 HPLC system equipped with an EVO C18 column (50 × 2.1 mm, 1.7 μm particle size, 100 Å pore size) and a suitable guard column (20 × 2.1 mm, 5 μm particle size) (Phenomoenx, Torrance, CA, USA). Flow rate was constant, at 0.25 ml/min and 25 °C. Mobile phase A consisted of 50 mM ammonium formate and phase B of 99.9% methanol. 2 μl of sample were injected per run. The following steps were employed: 0 – 4.5 min 3% B, 4.5 – 5.5 min linear gradient 3 from 80% B, 5.5 – 6.5 min 80 % B, 6.5 – 7.5 min from 80 to 3 % B, 7.5 – 8.5 min 3% B. Masses were detected as follows on an Agilent 6495 ion funnel mass spectrometer in positive mode using an electrospray ionization source: ESI spray voltage 1000 V, sheath gas 400° C at 12 l/min, nebulizer pressure 20 psig, drying gas 100° C at 11 l/min. Components were identified by retention time and mass transition. Integration of chromatograms was performed using the MassHunter software (Agilent, Santa Clara, CA, USA). From the integrals, absolute concentrations were calculated from an external calibration curve prepared in sample matrix.

LC-MS/MS analysis of glycolate and glyoxylate

Samples were prepared for analysis as described for the CoA-esters. An Agilent Infinity II 1290 HPLC system equipped with a Kinetex Evo C18 column (150 x .12mm, 100 A, 1.7 μm, Phenomenex) and a 20 X 2.1 mm guard column was used for separation. Flow rate was constant, at 0.1 ml/min and 25 °C. Mobile phase A was 0.1% formic acid in water and mobile phase B was 0.1% formic acid in methanol. 0.5 μl were injected. The mobile phase profile was as follows: 0 –5 min at 0 % B, 5 – 7 min from 0 to 100 % B; 7 – 10 min at 100 % B; 10 – 10.1 min from 100 to 0 % B; 10.1 – 20 min at 0 % B. Detection occurred in an Agilent 6470A mass spectrometer in negative mode with an electrospray ionization source. Settings were ESI spray voltage 4500 V, Nozzle Voltage at 500 V sheath gas 400° C at 11 l/min, nebulizer pressure 45 psig and drying gas 170° C at 5 l/min. Evaluation of the LC-MS/MS data was performed as described for the CoA-esters.

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.

ISM of MeOXC

ISM was performed using the 22c trick[10] and the primers listed in **Table S2**. 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 S4). 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. Activity assays were performed on Nunc 384 shallow well plates (Thermo Fisher) 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.

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 ThDP 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 S4**. 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 ThDP. 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 S4**. 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 S5**. Michealis-Menten graphs are shown in **Figure S6**.

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 ThDP, 1 mM formyl-CoA, 10 µM MeOXC or 5 µM MeOXC4 and either 100 mM formaldehyde, acetaldehyde, propionaldehyde or glycolaldehyde, or 10 mM benzaldehyde or phenylacetaldehyde, or 1 M acetone. Reactions were incubated at 30 °C and samples were taken after 0, 0.5, 1, 2, 3 and 5 minutes and analysed via LC-MS.

Inhibition of MeOXC4 by oxalyl-CoA

The impact of oxalyl-CoA on the GCS reaction was determined via LC-MS. Reactions contained 100 mM potassium phosphate pH 6.9, 10 mM $MgCl_2$, 150 µM ThDP, 1 mM formyl-CoA and/ or 1 mM oxalyl-CoA, 50 mM formaldehyde and 1 µM MeOXC4. Reactions were incubated at 30 °C and samples were taken after 0, 0.5, 1, 2, 3 and 5 minutes and analysed via LC-MS.

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.

MeOXC forward. The oxalyl-CoA decarboxylase activity was measured as described previously.[1] 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 (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.

MeOXC 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 ThDP, 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}).

PduP. Kinetic parameters for the activation of formaldehyde to formyl-CoA by PduP were determined. Reactions contained 50 mM MOPS-KOH pH 7.8, 2.5 mM NAD^+, 1 mM CoA and 180 nM PduP. Reactions were pre-incubated for 2 minutes and started by addition of formaldehyde. Assays were performed at 30 °C. Reaction velocity was monitored as reduction of NAD^+ at 340 nm.

E. coli whole-cell bioconversions

E. coli strains were pre-grown in a media containing M9 basal salts (6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl, 2 mM $MgSO_4$, and 15 μM thiamine-HCl) additionally supplemented with 20 g/L glycerol, 10 g/L tryptone, 5 g/L yeast extract, 50 μg/mL carbenicillin, 50 μg/mL spectinomycin, and the micronutrient solution of Neidhardt.[11] To test RuHACL and MeOXC variants, colonies of each strain were picked from LB-agar plates (with appropriate antibiotics) and used to inoculate 200 μL of the above media in a 2 mL deep, square 96-well plate (PlateOne). To test varying inducer concentrations, the strain expressing MeOXC4 was pre-cultured overnight in LB media containing appropriate antibiotics and used to inoculate an appropriate volume of the above media (1% inoculation). 200 μL of the inoculated media was then distributed into the wells of a 2 mL deep, square 96-well plate. Plates containing

inoculated media were covered with a rayon microporous film (USA Scientific) and incubated in an incubating microplate shaker (VWR International) at 30°C and 1000 rpm. After 2.5 hours, chemical inducers (IPTG and cumate) were added in the indicated concentrations.24 hours after inoculation, cells were pelleted by centrifugation (3220×g, 8 min) and the pellet washed with 1 mL of the M9 basal salt media with the micronutrient solution of Neidhardt and without other additional supplements. Cells were pelleted again by centrifugation (3220×g, 8 min) and the cell pellets were resuspended in 1 mL of the M9 basal salt media supplemented with the micronutrient solution of Neidhardt and 5 mM formaldehyde. The plates were covered and returned to the incubating microplate shaker (30°C, 1000 rpm). After 3 hours, final OD600 of the cell suspension was measured using a Synergy H1 microplate reader (Biotek) by transferring 50 μL of cells to a 96 well plate with 50 μL media in each well for dilution. Wells containing 100 μL media were used as blanks. The deep-well plates were centrifuged (3220×g, 8 min) and glycolate in the supernatant was quantified by HPLC analysis using previously reported conditions.[12]

Cell pellets harvested after bioconversion were resuspended to 20OD in B-PER® Bacterial Protein Extraction Reagent (Thermo Fisher) supplemented with 0.1 mg/mL chicken egg white lysozyme (Fisher) and 5 U/mL Benzonase® nuclease (Sigma) for cell lysis. After incubation in room temperature for 15 minutes, 100 μL of each cell lysate was transferred to 1.5 mL microcentrifuge tubes for centrifugation at 15,000×g for 5 minutes. The soluble cell lysates obtained from the supernatant were analyzed using Protein 80 Kit (Agilent) by 2100 Bioanalyzer system (Agilent), following the manufacturer instructions.

Supporting Tables

Table S1. Strains and plasmids used in this study.

Strain/Plasmid	Description/Genotype	Source
AC440	MG1655 λ(DE3) ΔfrmA ΔfdhF ΔfdnG ΔfdoG ΔglcD::FRT	Chou et al.[12]
BL21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS	Invitrogen
DH5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	Invitrogen
pCDFDuet-1-P1-ntH6-RuHACL^{G390N}	pCDFDuet-1 with codon optimized 6xHis-tagged *Rhodospirillales bacterium URHD0017 hacl* (Uniprot: A0A1H8YFL8) with a G390N mutation under control of the *T7lac* promoter and *lacI*	Chou et al.[12]
pCDFDuet-1-P1-ntH6-MeOXC(1-4)	pCDFDuet-1 with 6xHis-tagged *Methylorubrum extorquens oxc* (or indicated mutant) under control of the *T7lac* promoter and *lacI*	This study
pET-16b-s-PduP	pET-16b with strep-tagged *Rhodopseudomonas palustris pduP* under control of *T7* promoter and *lacI*	Zarzicky et al.[4]
pCA24N-GhrB	pCA24N with *Escherichia coli ghrB* under control of T5 promoter and *lacI*	ASKA collection[13]
pCA24N-EcOXC	pCA24N with *Escherichia coli oxc* under control of T5 promoter and *lacI*	ASKA collection[13]
pET-16b-AbfT	pET-16b with *Clostridium aminobutyricum abfT* under control of *T7* promoter and *lacI*	Scheffen et al. [14]
pSEVA581-CnPCT	pSEVA581 with *Cupriavidus necator pct* under control of *T7* promoter	Gifted by I. Bernhardsgrütter
pET-16b-CpPCT	pET-16b with *Clostridium pripionicum pct* under control of *T7* promoter and *lacI*	Gifted by I. Berg
pET-16b-FRC	pET-16b with *Oxalobacter formigenes frc* under control of *T7* promoter and *lacI*	Gifted by A. Bar-Even
pET-16b-MeOXC(1-4)	pET-16b with *Methylorubrum extorquens oxc* (or indicated mutant) under control of *T7* promoter and *lacI*	Burgener et al.[1], this study
pET-16b-PanE2	pET-16b with *Methylorubrum extorquens panE2* under control of *T7* promoter and *lacI*	This study
pET-16b-RuHACL G390N	pET-16b with *Rhodospirillales bacterium URHD0017 hacl* with a G390N mutation under control of *T7* promoter and *lacI*	Chou et al.[12]
pET-16b-RuHACL A389C G390N	pET-16b with *Rhodospirillales bacterium URHD0017 hacl* with the mutations A389C and G390N under control of *T7* promoter and *lacI*	This study
pET-28b-GOX	pET-28b with *Homo sapiens gox* under control of *T7* promoter and *lacI*	This study
pETDuet-1-P1-LmACR-P2-EcAldA	pETDuet-1 with codon optimized *Lysteria monocytogenes acr* in the P1 cloning site and *E. coli aldA* in the P2 cloning site both under control of the *T7lac* promoter and *lacI*	Chou et al.[12]
pETDuet-1-P^{CT5}-LmACR-EcAldA	pETDuet-1 with codon optimized *Lysteria monocytogenes acr* and *Escherichia coli aldA* in a synthetic operon under control of the cumate inducible CT5 promoter and *cymR*	Chou et al.[12]
pETDuet-1-P^{CT5}-PduP-EcAldA	pETDuet-1 with codon optimized *Rhodopseudomonas palustris pduP* and *Escherichia coli aldA* in a synthetic operon under control of the cumate inducible CT5 promoter and *cymR*	This study
pTWIST-GOX	pTWIST with codon optimized *Homo sapiens gox*	This study

Table S2. 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
gox_fw	GATACATATGCTGCCGCGTCTGATCTG
gox_rv	GTATGGATCCGCCTGCTAGCTCTAGATTAG
oxc_I48_NDT	CGATCTACAACGTGCCCGGCNDTCCGATCACCGATCTCGG
oxc_I48_VHG	CGATCTACAACGTGCCCGGCVHGCCGATCACCGATCTCGG
oxc_I48_TGG	CGATCTACAACGTGCCCGGCTGGCCGATCACCGATCTCGG
oxc_I48_rv	GCCGGGCACGTTGTAGATCGTCTCGATGCCGTTGAGCTTG
oxc_A415_NDT	CGATCCTCGTCAACGAGGGTNDTAACACCCTCGATCTGGC
oxc_A415_VHG	CGATCCTCGTCAACGAGGGTVHGAACACCCTCGATCTGGC
oxc_A415_TGG	CGATCCTCGTCAACGAGGGTTGGAACACCCTCGATCTGGC
oxc_A415_rv	CCCTCGTTGACGAGGATCGCGTCGGGCCGCTCCTTGATG
oxc_Y497_NDT	CTTCAACAACAACGGCATCNDTCGCGGCACCGACACCGATC
oxc_Y497_VHG	CTTCAACAACAACGGCATCVHGCGCGGCACCGACACCGATC
oxc_Y497_TGG	CTTCAACAACAACGGCATCTGGCGCGGCACCGACACCGATC
oxc_Y497_rv	GATGCCGTTGTTGTTGAAGATGACGATGCAGACCGGCAGC
oxc_I571_NDT	CCGGCAGCGAGAGCGGCAATNDTGGCAGCCTCAACCCGCAG
oxc_I571_VHG	CCGGCAGCGAGAGCGGCAATVHGGGCAGCCTCAACCCGCAG
oxc_I571_TGG	CCGGCAGCGAGAGCGGCAATTGGGGCAGCCTCAACCCGCAG
oxc_I571_rv	GCCGCTCTCGCTGCCGGCGGCCGGGTCGATCTCCGCATTG
oxc_Y134_NDT	CGATCTGCAGCAGGGCGACNDTGAGGAGATGGACCAGC
oxc_Y134_VHG	CGATCTGCAGCAGGGCGACVHGGAGGAGATGGACCAGC
oxc_Y134_TGG	CGATCTGCAGCAGGGCGACTGGGAGGAGATGGACCAGC
oxc_Y134_rv	GTCGCCCTGCTGCAGATCGACGATCTCGCGCTCGGAGGAG
oxc_E135_NDT	GATCTGCAGCAGGGCGACTACNDTGAGATGGACCAGCTCGC
oxc_E135_VHG	GATCTGCAGCAGGGCGACTACVHGGAGATGGACCAGCTCGC
oxc_E135_TGG	GATCTGCAGCAGGGCGACTACTGGGAGATGGACCAGCTCGC
oxc_E135_rv	GTAGTCGCCCTGCTGCAGATCGACGATCTCGCGCTCGGAG
oxc_E567_NDT	CGACCCGGCCGCCGGCAGCNDTAGCGGCAATATCGGCAGC
oxc_E567_VHG	CGACCCGGCCGCCGGCAGCVHGAGCGGCAATATCGGCAGC
oxc_E567_TGG	CGACCCGGCCGCCGGCAGCTGGAGCGGCAATATCGGCAGC
oxc_E567_rv	GCTGCCGGCGGCCGGGTCGATCTCCGCATTGATCAGAGTC
oxc_S568_NDT	CCCGGCCGCCGGCAGCGAGNDTGGCAATATCGGCAGCCTC
oxc_S568_VHG	CCCGGCCGCCGGCAGCGAGVHGGGCAATATCGGCAGCCTC
oxc_S568_TGG	CCCGGCCGCCGGCAGCGAGTGGGGCAATATCGGCAGCCTC
oxc_S568_rv	CTCGCTGCCGGCGGCCGGGTCGATCTCCGCATTGATCAG
oxc_S568G_E567_NDT	CGACCCGGCCGCCGGCAGCNDTGGTGGCAATATCGGCAGC
oxc_S568G_E567_VHG	CGACCCGGCCGCCGGCAGCVHGGGTGGCAATATCGGCAGC
oxc_S568G_E567_TGG	CGACCCGGCCGCCGGCAGCTGGGGTGGCAATATCGGCAGC
oxc_E135G_Y134_NDT	CGATCTGCAGCAGGGCGACNDTGGTGAGATGGACCAGC
oxc_E135G_Y134_VHG	CGATCTGCAGCAGGGCGACVHGGGTGAGATGGACCAGC
oxc_E135G_Y134_TGG	CGATCTGCAGCAGGGCGACTGGGGTGAGATGGACCAGC
oxc_Y134E135_NDTNDT	GATCTGCAGCAGGGCGACNDTNDTGAGATGGACCAGCTCGC
oxc_Y134E135_NDTVHG	GATCTGCAGCAGGGCGACNDTVHGGAGATGGACCAGCTCGC
oxc_Y134E135_NDTTGG	GATCTGCAGCAGGGCGACNDTTGGGAGATGGACCAGCTCGC
oxc_Y134E135_VHGNDT	GATCTGCAGCAGGGCGACVHGNDTGAGATGGACCAGCTCGC
oxc_Y134E135_VHGVHG	GATCTGCAGCAGGGCGACVHGVHGGAGATGGACCAGCTCGC
oxc_Y134E135_VHGTGG	GATCTGCAGCAGGGCGACVHGTGGGAGATGGACCAGCTCGC
oxc_Y134E135_TGGNDT	GATCTGCAGCAGGGCGACTGGNDTGAGATGGACCAGCTCGC
oxc_Y134E135_TGGVHG	GATCTGCAGCAGGGCGACTGGVHGGAGATGGACCAGCTCGC
oxc_E70A_fw	CATCTCCTTCCGCCACGCGCAGAATGCGGGCAAC
oxc_E70A_rv	GTTGCCCGCATTCTGCGCGTGGCGGAAGGAGATG
panE2_fw	GCGCACATATGAGCATCGCGATCGTCG
panE2_rv	CAGAGGATCCTCATGCTCCCTGGATCGC
pNov-P1-ntH6-MeOxc-L	GCCAGGATCCGAATTCTATGACCGTCCAGGCCCA
pNov-P1-ntH6-MeOxc-R	CGCCGAGCTCGAATTTCACTTCTTCTTCAAGGTGCTC
ruHACL_A389C_fw	CCTGTCTGCTGAAGGTTGCAACACCATGGACATCG
ruHACL_A389C_rv	CGATGTCCATGGTGTTGCAACCTTCAGCAGACAGG

Table S3. Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

	MeOXC	MeOXC4
Beam line	ESRF_ID23-1, Grenoble, France	DESY P13, Hamburg, Germany
PDB ID	7AYG	7B2E
Ligands	ThDP, ADP, Mg^{2+}	ThDP, ADP, Mg^{2+}
Wavelength (Å)	0.972	0.976
Resolution range (Å)	39.11 - 1.9 (1.968 - 1.9)	29.97 – 2.8 (2.9 - 2.2.8)
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	161.094, 180.337, 202.008
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)	99.37 (99.97)
Mean (I)/σ (I)	8.82 (1.99)	10.11 (3.85)
Wilson B-factor ($Å^2$)	23.99	28.91
R_{merge}	0.04624 (0.367)	0.04856 (0.1611)
R_{meas}	0.06539 (0.519)	0.06867 (0.2279)
CC1/2	0.997 (0.883)	0.998 (0.978)
Reflections used in refinement	454225 (44467)	142866 (14161)
Reflections used for R_{free}	1987 (195)	1996 (197)
R_{work}	0.1765 (0.3061)	0.2012 (0.2597)
R_{free}	0.2078 (0.3227)	0.2302 (0.3175)
Number of non-hydrogen atoms	36890	34264
macromolecules	32548	32560
ligands	432	432
solvent	3910	1272
Protein residues	4392	4405
RMS(bonds) (Å)	0.008	0.006
RMS(angles) (°)	1.04	0.86
Ramachandran favored (%)	97.54	97.30
Ramachandran allowed (%)	2.27	2.52
Ramachandran outliers (%)	0.18	0.18
Rotamer outliers (%)	0.94	0.97
Average B-factor	27.86	29.30
macromolecules	27.02	29.39
ligands	22.52	26.73
solvent	35.41	28.00

Table S4. 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	1	50, 20, 10, 5, 2	1	1	35	1, 0.6, 0.3, 0.2, 0.1, 0.05
MeOXC4	1	25, 10, 5, 2, 1	1	1	30	1, 0.6, 0.3, 0.2, 0.1, 0.05

Supporting Figures

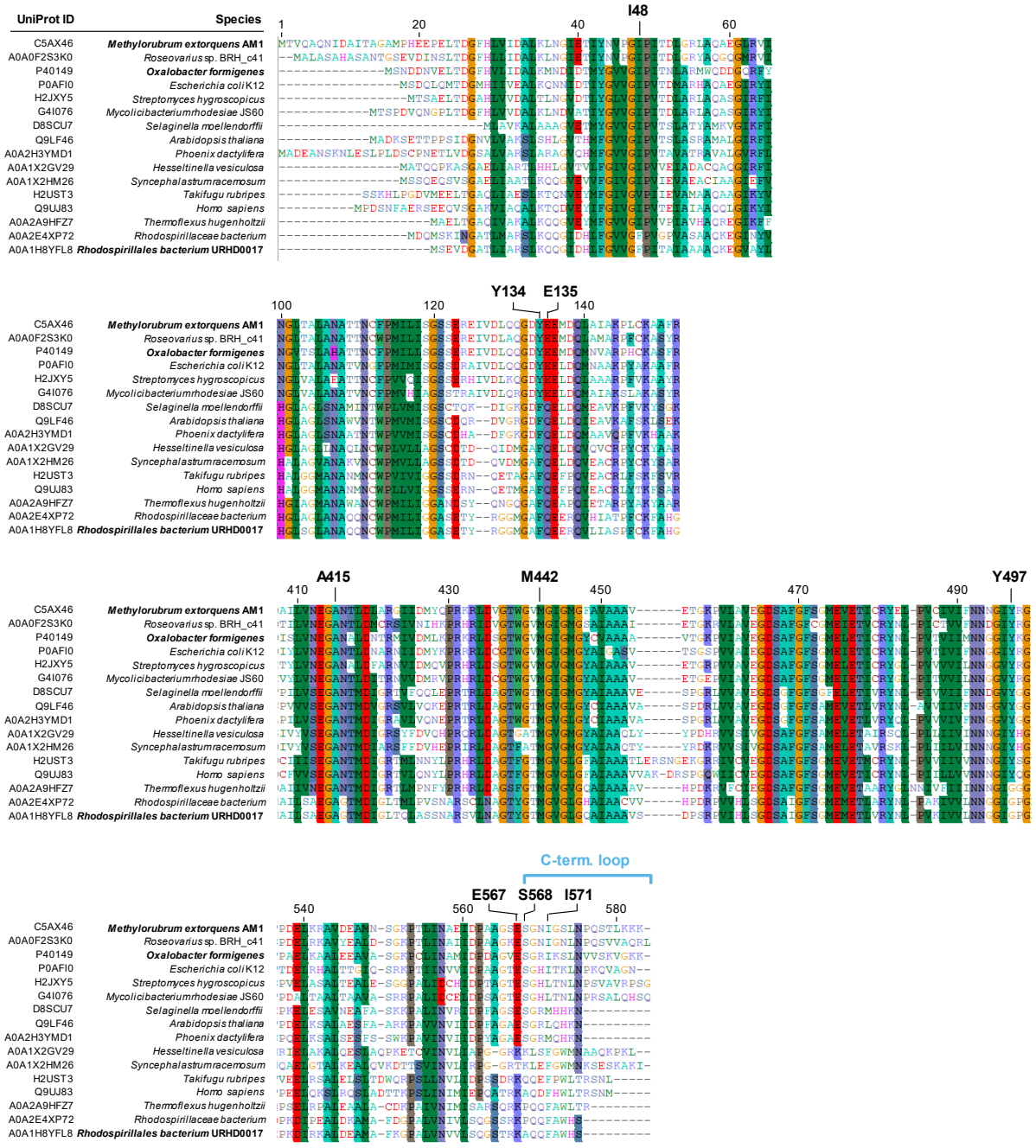


Figure S1. Representative MSA of the HACL/OXC enzyme family. Shown are the sequence regions around the amino acids that are in proximity of the ThDP in MeOXC (see **Figure 1C**). MeOXC, OfOXC and RuHACL (first, third and last from the top) are in bold. The C-terminal loop (marked in blue) is highly flexible in OXC and forms part of the active site upon substrate binding. Note that this region shows low sequence conservation.

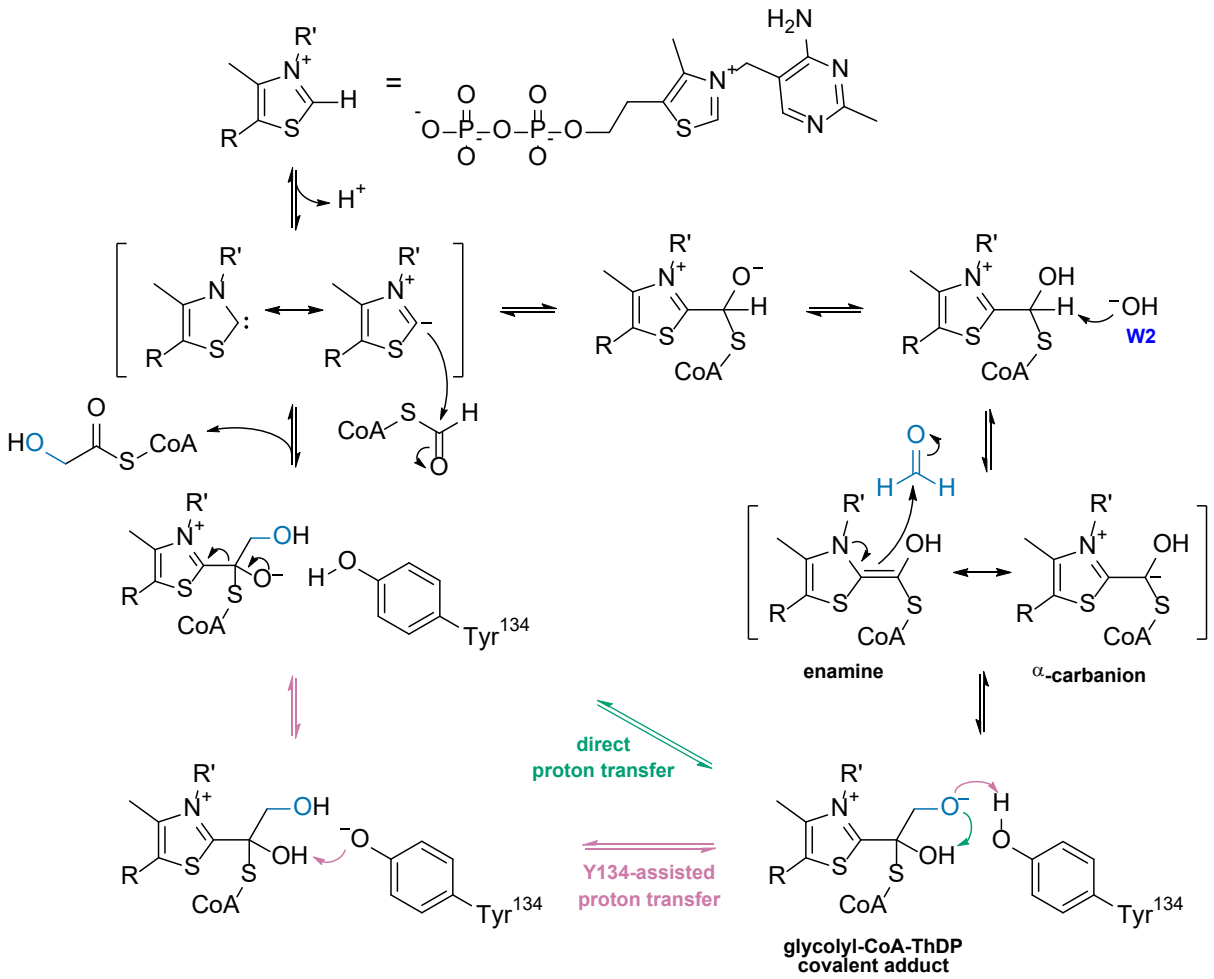


Figure S2. Proposed catalytic cycle of GCS. Catalysis is initiated by proton abstraction from the ThDP cofactor, mediated by E70A.[15] This generates the carbene/ylid state of ThDP, which covalently binds formyl-CoA. Deprotonation of the α-carbon by hydroxide (forming water molecule W2) then gives rise to the α-carbanion/enamine intermediate[15] that subsequently performs a nucleophilic attack onto formaldehyde. The resulting intermediate (glycolyl-CoA-ThDP covalent adduct) undergoes proton transfer and is then released as glycolyl-CoA, regenerating the ThDP-carbene. The proton transfer can either proceed directly between the hydroxyl groups (green arrows) or via Y134 (pink arrows).

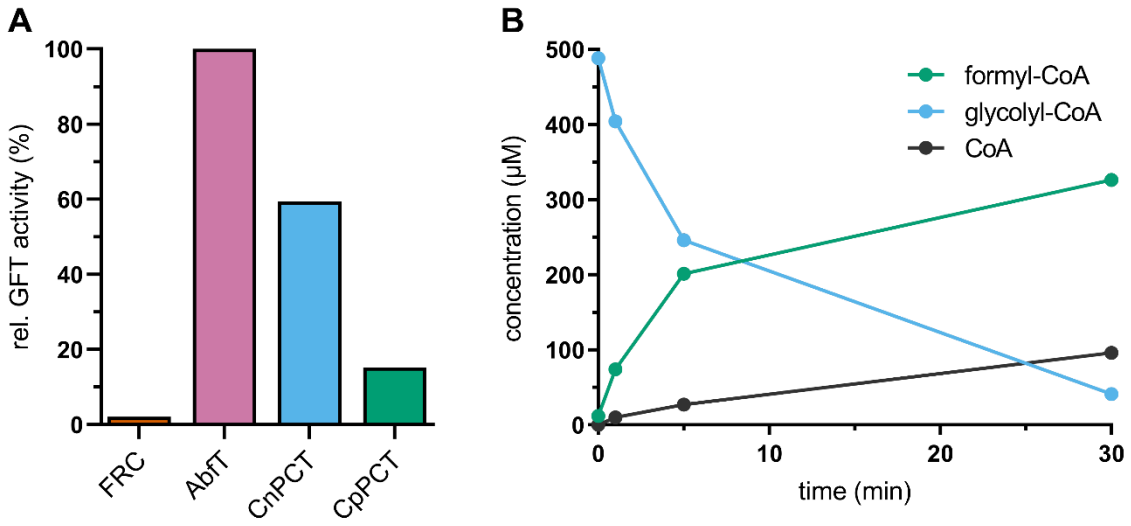


Figure S3. Screening CoA-transferases for GFT activity. A) GFT activity was screened by detecting the product formyl-CoA via LC-MS. Activity was normalized to the highest activity (AbfT). FRC is formyl-CoA:oxalate CoA-transferase from *Oxalobacter formigenes*, AbfT is 4-hydroxybutyrate CoA-transferase from *Clostridium aminobutyricum*, CnPCT is propionyl-CoA transferase from *Cupriavidus necator*, CpPCT is propionyl-CoA transferase from *Clostridium propionicum*. **B)** The reaction progress of CnPCT-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.

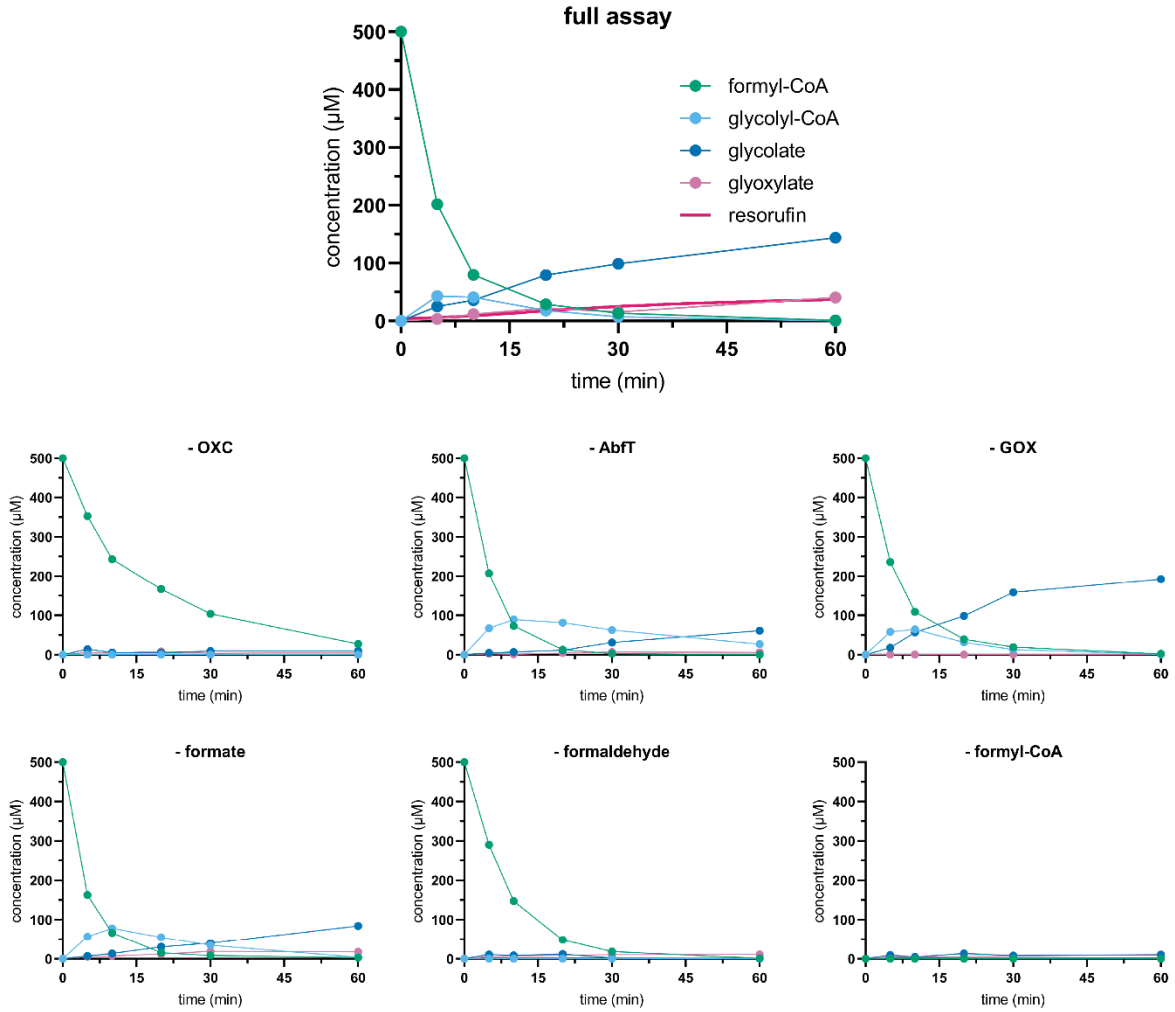


Figure S4. Validation of the GCS screen. The production of key intermediates was assessed by LC-MS and appropriate control reactions were performed to confirm the proper function of the screen. For the full assay, the complementary resorufin production was assessed by fluorescence readout in a plate reader. It was confirmed that resorufin is produced in equimolar amounts to glyoxylate.

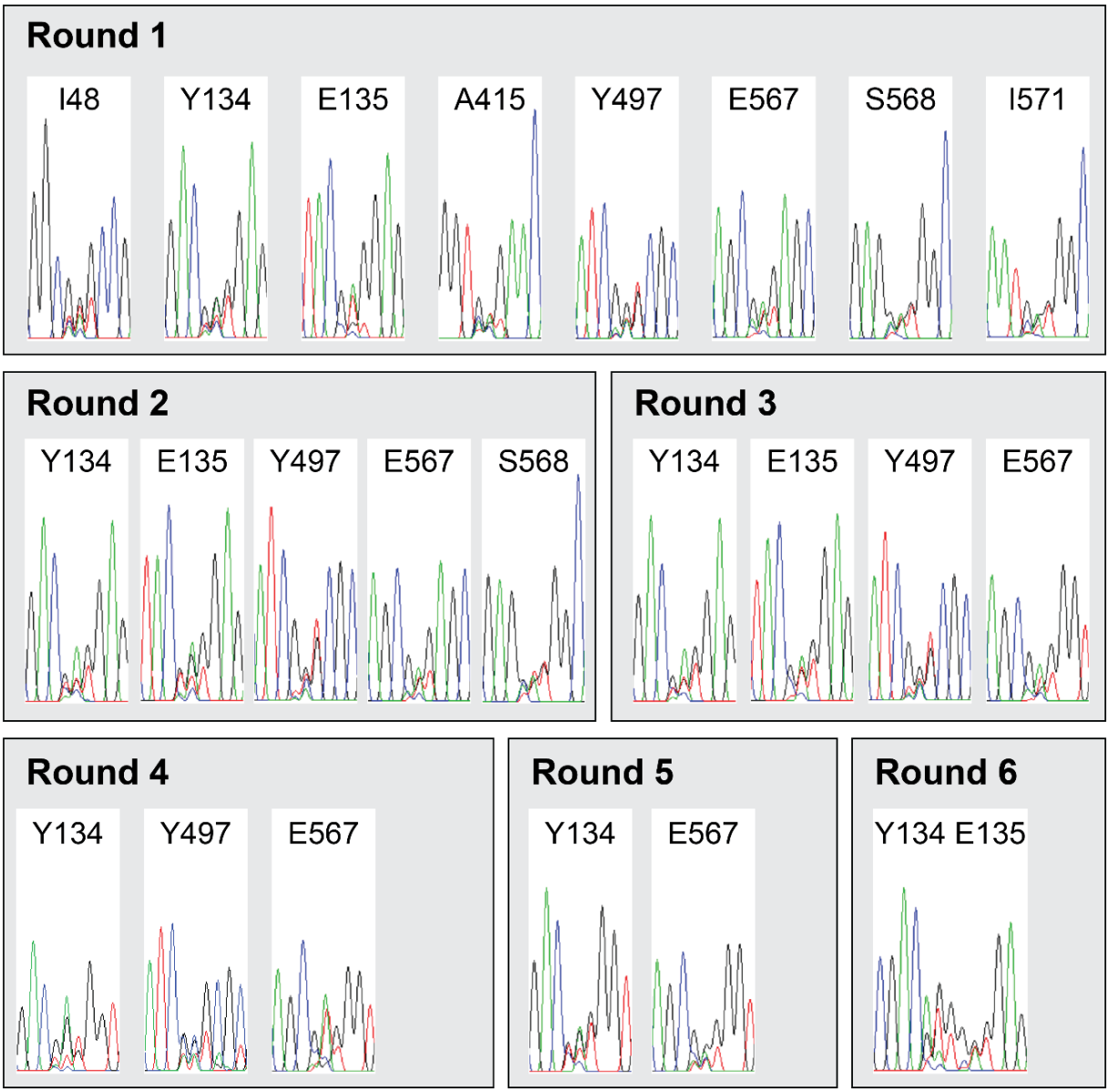


Figure S5. Sequencing chromatograms of the 22c-trick[10] libraries. Green, A. Blue, C. Black, G. Red, T.

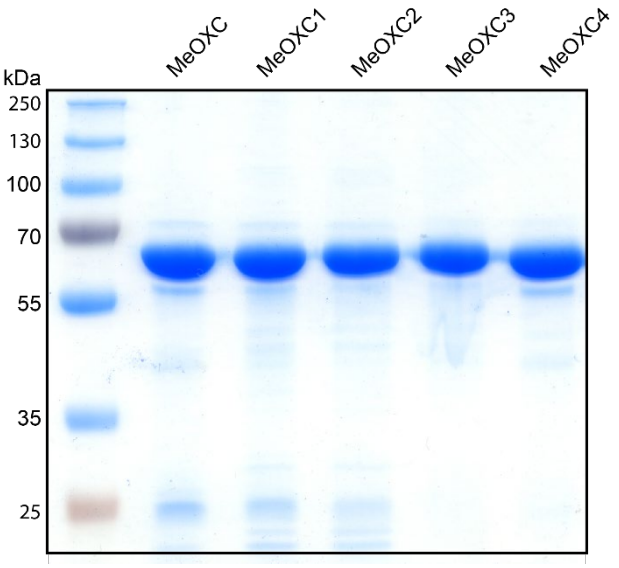


Figure S6. SDS-PAGE of purified MeOXC variants. 5 µg protein was loaded in each lane.

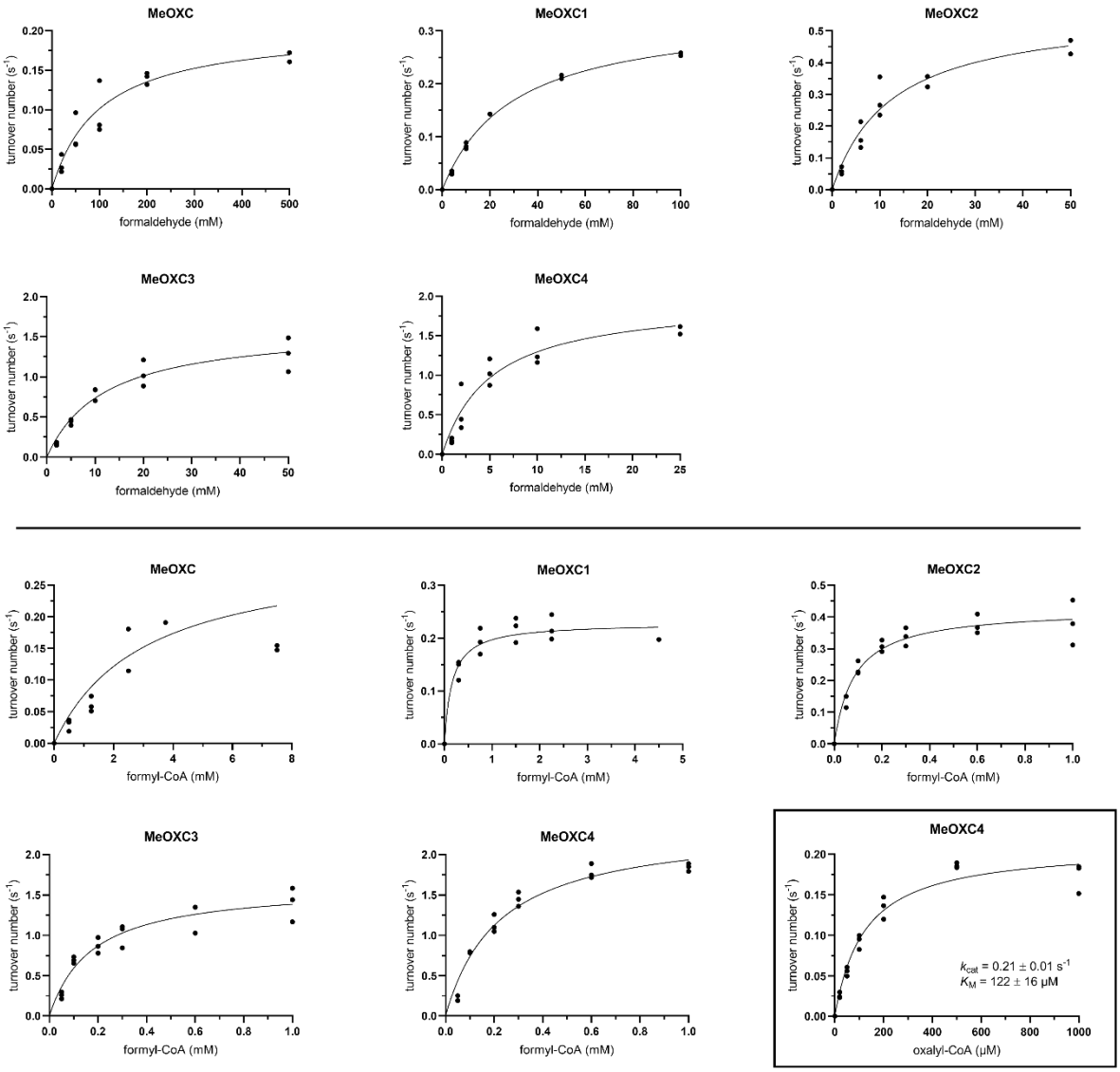


Figure S7. Michaelis-Menten graphs of MeOXC variants. Substrates are formaldehyde (top) and formyl-CoA (bottom). In the box on the bottom left is MeOXC4 with oxalyl-CoA as substrate.

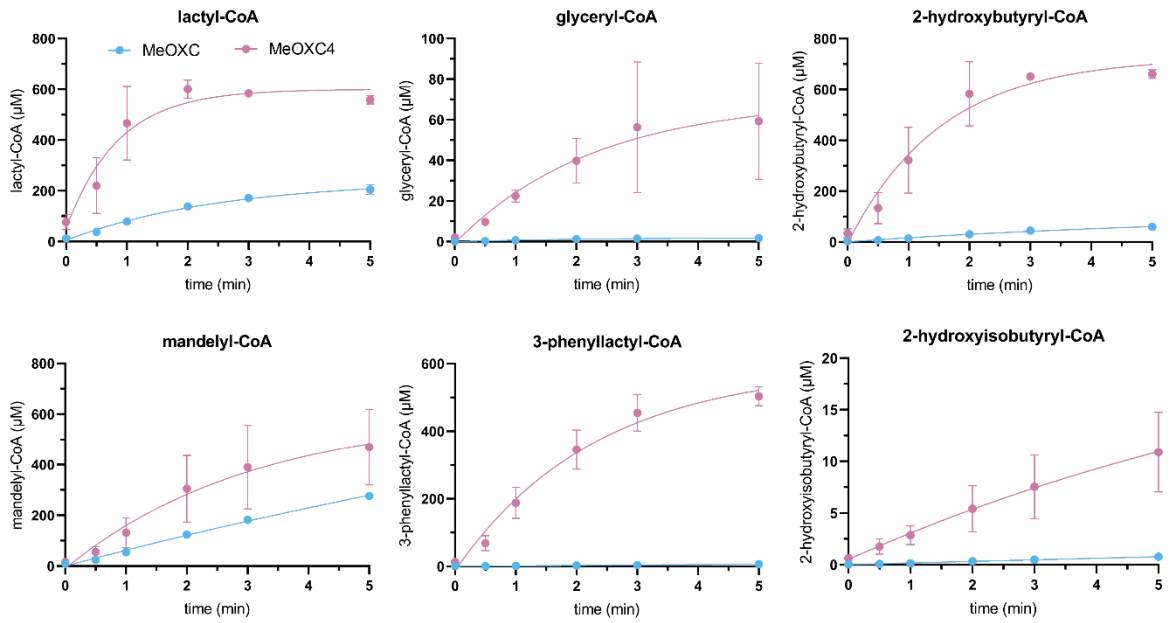


Figure S8. Aldehyde condensation reactions of MeOXC and MeOXC4. From top left to bottom right: condensation of 1 mM formyl-CoA with 100 mM acetaldehyde, 100 mM glycolaldehyde, 100 mM proprionaldehyde, 10 mM benzaldehyde, 10 mM phenylacetaldehyde and 1 M acetone. 10 µM MeOXC and 5 µM MeOXC4 were used.

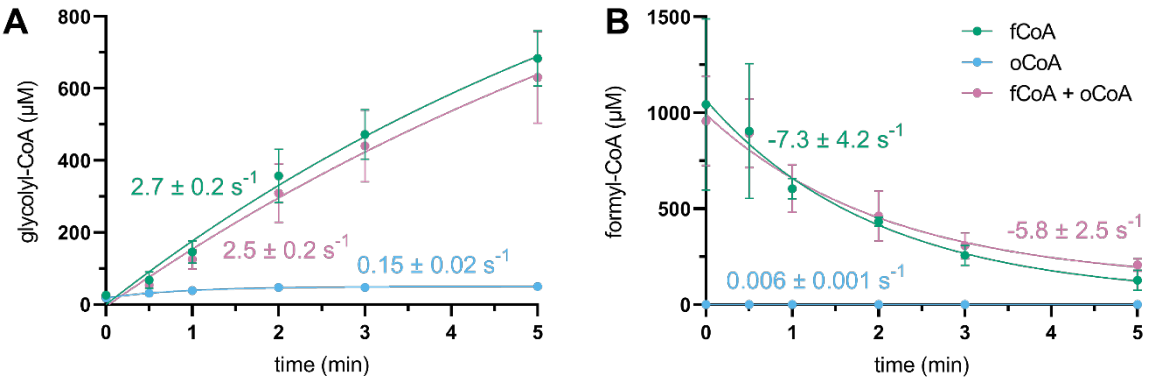


Figure S9. Effect of oxalyl-CoA on GCS activity of MeOXC4. Reactions contained 50 mM formaldehyde and 1 mM of formyl-CoA and/ or oxalyl-CoA. Error bars reflect standard deviation of three replicates. Initial velocity of each reaction is indicated (mean ± s.d.). Data is fit using a one-phase decay. fCoA, formyl-CoA; oCoA, oxalyl-CoA. **A)** glycolyl-CoA production as calculated using a standard curve. **B)** formyl-CoA consumption as estimated using the initial concentration as a calibration point.

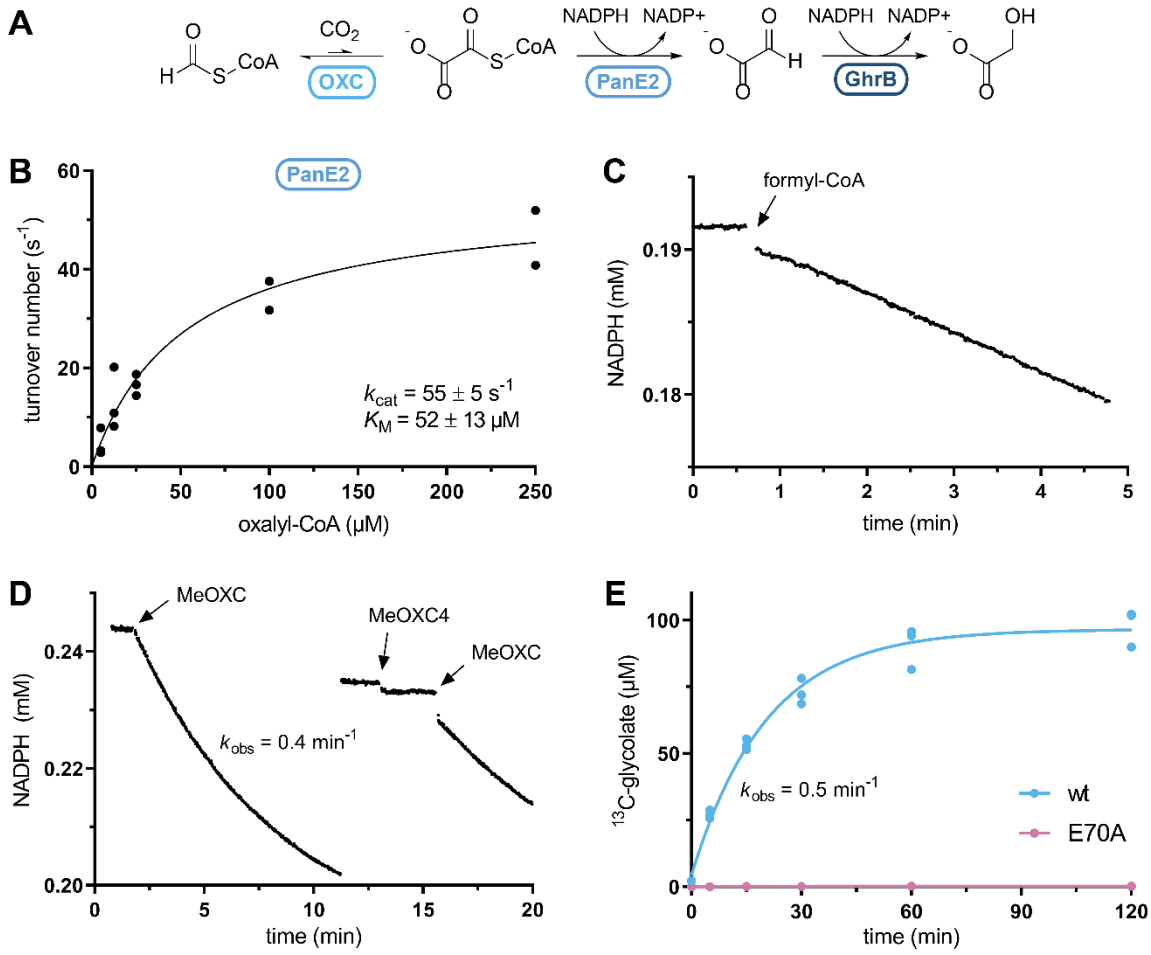


Figure S10. Reverse reaction of OXC. A) Reaction scheme of the OXC-PanE2-GhrB cascade, converting formyl-CoA and CO_2 into glycolate under consumption of 2 NADPH. **B)** Michaelis-Menten plot of PanE2 catalyzing the NADPH-dependent reduction of oxalyl-CoA to glyoxylate. The NADPH concentration was 0.25 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 WT 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.

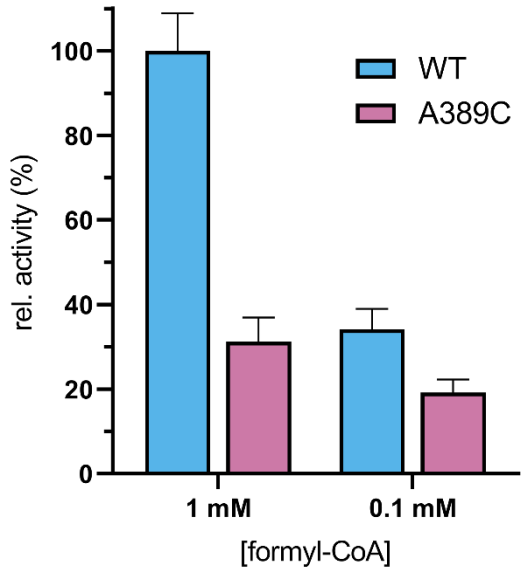


Figure S11. Comparing the GCS-activity of RuHACL G390N (WT) and RuHACL A389C G390N. The mutation A389C corresponds to MeOXC A415C. Formaldehyde concentration was 100 mM and the activity was determined by measuring glycolyl-CoA formation via LC-MS. Activity is shown relative to WT at 1 mM formyl-CoA. Error bars reflect standard deviation of three replicates.

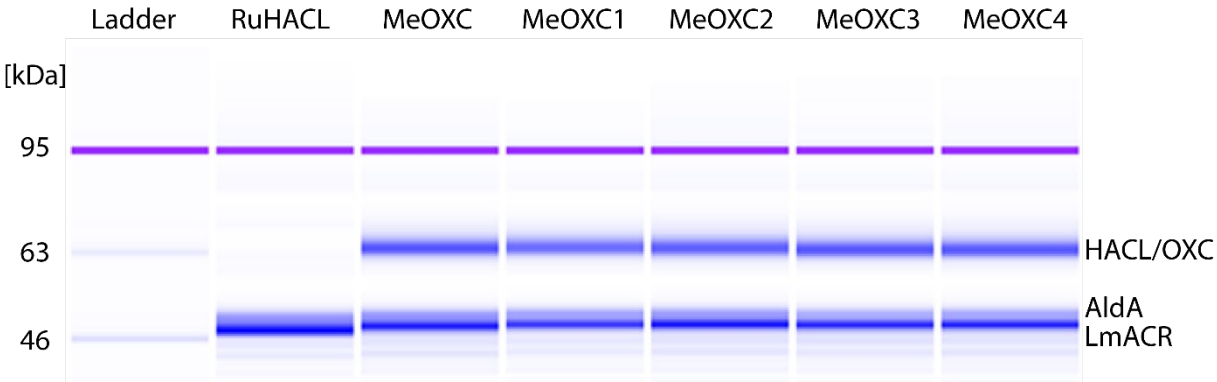


Figure S12. Expression analysis of formaldehyde to glycolate conversion pathway enzymes. RuHACL (~61 kDa) expresses at a substantially lower level than MeOXC variants (~63.5 kDa) in strain AC440, derived from *E. coli* MG1655. The gel representation was generated by 2100 Bioanalyzer Expert software (Agilent) with intensity normalized on a per lane basis.

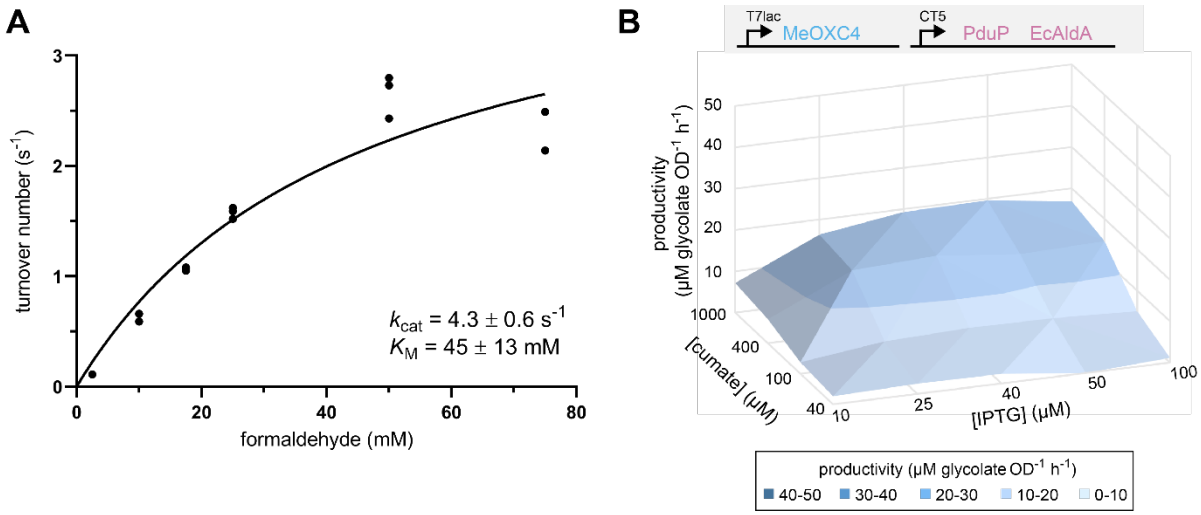


Figure S13. Effect of PduP as an ACR on the whole-cell conversion of formaldehyde to glycolate. **A)** Kinetic parameters for the conversion of formaldehyde to formyl-CoA. **B)** Glycolate production by *E. coli* whole cells expressing MeOXC4, PduP and LmACR under different inducer concentrations. Productivity and promoters are defined as in Figure 6C. Mean of n = 2 replicates is shown as a surface plotted against IPTG and cumate concentration.

Supporting References

[1] 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.

[2] 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, *Proc. Natl. Acad. Sci. U.S.A. 115*, E11455-E11464.

[3] 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*, 517.

[4] 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.

[5] Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem. 72*, 248-254.

[6] Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement, *Acta Crystallogr. D. 66*, 133-144.

[7] 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 Crystallogr. D. 67*, 235-242.

[8] 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 Crystallogr. D. 66*, 213-221.

[9] Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D. 60*, 2126-2132.

[10] 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 Synth. Biol. 2*, 83-92.

[11] Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974) Culture Medium for Enterobacteria, *J. Bacteriol. 119*, 736-747.

[12] 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.

[13] 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.

[14] 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. (2021) A new-to-nature carboxylation module to improve natural and synthetic CO2 fixation, *Nat. Catal. 4*, 105-115.

[15] 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.

General Discussion

General Discussion

The overall aim of this thesis was an expansion of the formate bioeconomy proposed by Yishai *et al*. in 2016[1] by engineering novel enzyme activities for the conversion of formate into metabolic intermediates. Among the natural C1 compounds, formate is the least harmful to the environment – it is not a greenhouse gas such as CO_2 or methane and it is not as toxic as methanol or formaldehyde. It is also assimilated at the highest energetic efficiency of any C1 compound[2, 3]. Formate can be generated from CO_2 by via electrochemical[4-6] or enzymatic[7] processes, before being valorized by natural or synthetic formatotrophs into biomass or value-added chemicals. This thesis adds to this vision by creating the novel FOK-FPR cascade for the activation of formate to formaldehyde (*CHAPTER 1*) and by designing GCS, an enzyme capable of performing said conversion of formaldehyde and formyl-CoA into the C2 compound glycolyl-CoA via efficient C1-C1 condensation (*CHAPTER 2*)[8]. While these distinct modules can be integrated into existing pathways individually, they are also easily combined to produce glycolyl-CoA from formate, expanding the scope of synthetic formate assimilation. However, even with these advances, a full realization of the formate bio-economy demands *in vivo* implementation, posing a future challenge.

FOK-FPR: an energy-efficient path from formate t formaldehyde

The FOK-FPR cascade established in *CHAPTER 1* was developed with modular implementation in mind. Using formyl phosphate as intermediate, it is a direct alternative to existing routes of aero-tolerant formate to formaldehyde reduction via formyl-THF and formyl-CoA, operating at the lowest possible energetic cost. While FOK activity has been described in literature[9, 10] and was confirmed in this thesis in EcAckA, FPR activity was first discovered in this thesis, making a valuable addition to the metabolic toolbox of synthetic formate reduction. In fact, FPR activity is present in multiple families of phosphate reductases, with N-acetyl-γ-glutamyl phosphate reductase (ArgC) identified as the most promising one. Phylogenetic homologue screening of the ArgC family and subsequent targeted enzyme engineering yielded DaArgC3, an enzyme with a 300-fold switch in selectivity from its native substrate to formyl phosphate – thus acting as a *bona fide* FPR. With metabolic integration already in mind, DaArgC3 was optimized for *in vivo* applications, offering a 4-fold

improvement in activity over wild type, counterbalanced by a 2-fold decrease of *in vitro* enzymatic activity. This illustrates an often-underestimated facet of enzyme engineering for metabolic integration: living systems are highly complex and can affect enzymatic activity in multitudinous ways – from the genetic all the way to the metabolic level. *CHAPTER 1* demonstrates such a setting, in which DaArgC3 escapes competition of native substrate with formyl phosphate, reduces side-reactivity with acetyl phosphate, and opens a secondary cofactor pool by gaining promiscuous activity on NADH on top of its native NADPH-dependence. Those three factors, in conjunction, are sufficient to not only counterbalance its decreased turnover number, but also vastly increase its overall formaldehyde production *in vivo*.

This thesis reached a local maximum in FPR activity that could no longer be increased by rational design. The limitations in throughput posed by activity screening of individual variants can be overcome by growth selection. Biosensor systems for formaldehyde detection are developed in collaboration with Dr. Sebastian Wenk (MPI for plant physiology, Golm, Germany). Preliminary results indicate that FOK-FPR can indeed support growth of the selection strain which requires the module to generate ~4% of its biomass (cysteine and threonine), but validation is still ongoing. The use of growth biosensors enables evolution of both FOK and FPR at much higher throughput, e.g. by adaptive laboratory evolution (ALE)[11].

While FPR remains the rate limiting step of FOK-FPR, FOK is another suitable target for improvement, as in the current system, the phosphorylation of formate is performed by acetate kinase. Full exclusion of acetate as a substrate would be desirable as it minimizes interactions with natural metabolism and competing reactions, permitting an orthogonal channeling of formate to desired product. While an obscure FOK has been described in *Clostridium cylidrosporum*[10], this enzyme has never been identified or purified. Discovery of this enzyme outstanding, engineering of acetate kinase toward a bona fide FOK is the best way forward.

In principle, FOK-FPR enables the production of value-added compounds from formate. To this end, a downstream conversion module is required. Natural formaldehyde assimilation via the RuMP[12] and XuMP[13] pathways could be re-routed to formate using the module, with product formation occurring after metabolic integration. The synthetic homoserine[14], HWLS [15] and SMGF[16] pathways would be similarly suited for FOK-FPR implementation. In the latter two, FOK-FPR implementation would be exceptionally straightforward, as they already utilize

the formyl-T_4F and formyl-CoA dependent routes of formate reduction, which could be replaced by FOK-FPR to reduce energy burden on the cell. Finally, FOK-FPR could be implemented into orthogonal systems that intentionally separate production from central metabolism, such as FORCE[17] (**Figure 1**).

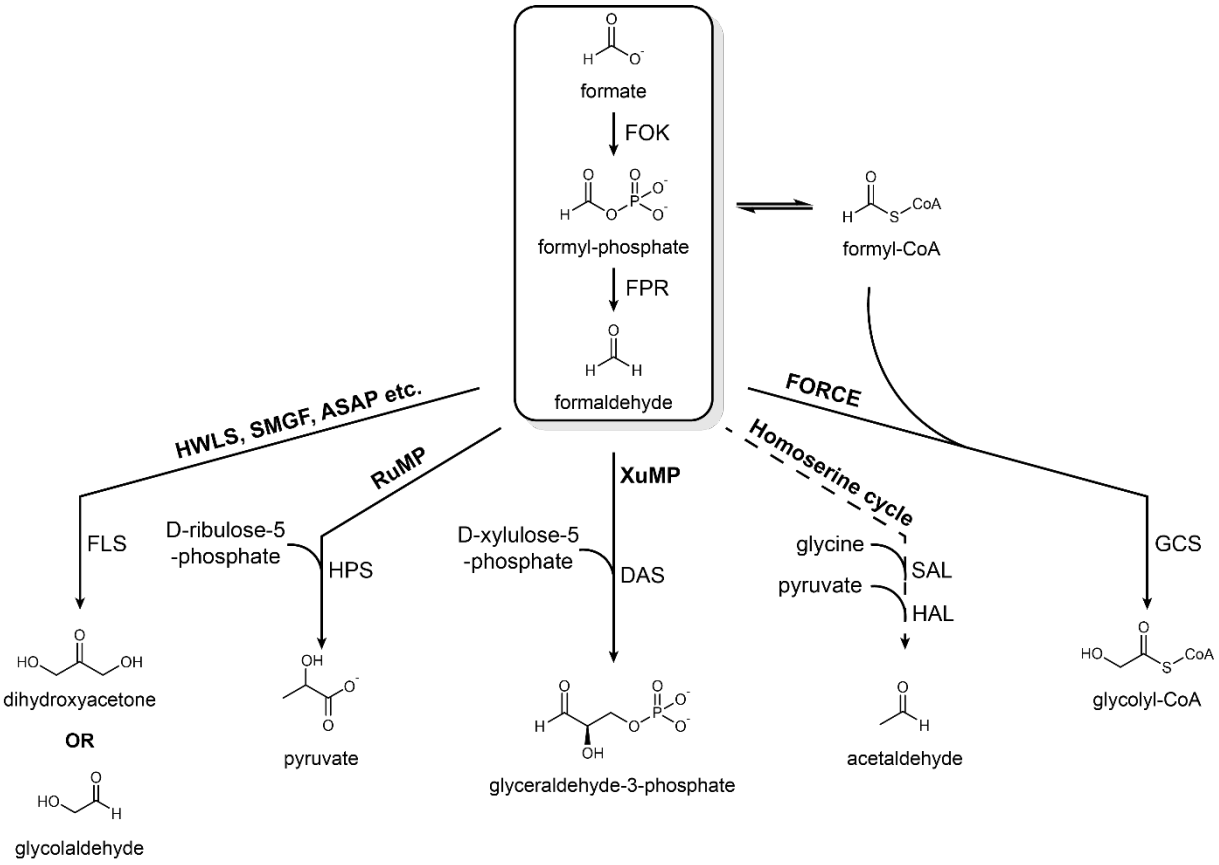


Figure 1 Downstream valorization of the FOK-FPR cascade. Pathways are indicated in bold letters. Enzymes carrying out core steps of C1 assimilation are given next to the arrows. For cycles, the product released from the cycle as well as the required acceptor molecules are shown. For linear pathways, only the product of condensation is shown. FOK = formate kinase, FPR = formyl-phosphate reductase, FLS = formolase, HPS = 3-hexulose-6-phosphate synthase, DAS = dihydroxyacetone synthase, SAL = serine aldolase, HAL = 4-hydroxy-oxo-butanoate aldolase, GCS = glycolyl-CoA synthase.

GCS: an efficient C1-C1 carboligase

Linear pathways for C1 assimilation are highly desirable as they reduce the need for costly acceptor molecule regeneration. Direct C1-C1 condensation enables such pathways; however, the solutions found in nature are highly complex enzymatic machineries. The development of

low-complexity alternatives has therefore come into focus. The ThDP-dependent enzymatic formose reaction catalyzed by multiple ligase families has been found as an alternative, enabling straight-forward metabolic integration[15, 16, 18, 19] as well as generation of value-added product[20-22]. However, this reaction is notoriously inefficient (*General* ***Introduction*****, Table 3**). This shortfall was overcome by utilization of the much more efficient GCS reaction carried out by RuHACL[17, 23]. Around this GCS reaction, the FORCE pathway was established, producing glycolyl-CoA from formyl-CoA and formaldehyde and converting it in multiple downstream steps to value-added compounds like 2-hydroxy-acids, carboxylic acids, 2-hydroxyaldehdyes, diols, acyl-CoAs and alcohols by reduction and dehydration of glycolyl-CoA as well as the iterative C1 elongation of the aldehyde intermediates by RuHACL. However, RuHACL is limited by its low production titers in *E.coli* and a high K_M on formaldehyde[23].

CHAPTER 2 investigated the development of a GCS that avoids these shortfalls. A candidate for such an enzyme was not sought in the HACL family, but in the related OXC family, whose members are generally more soluble and, importantly, structurally characterized. As the active site was well-described, iterative saturation mutagenesis of substrate-proximal residues was performed, yielding incremental improvements in enzymatic activity to final variant MeOXC4 that had comparable kinetic characteristics to natural formaldehyde assimilating enzymes and had almost fully lost its native activity. However, MeOXC4 had not turned into a *bona fide* GCS – formaldehyde aside, it readily and efficiently accepted longer-chained aldehyde substrates, all at improved efficiencies compared to wild type. Arguably, this is a feature, not a bug, as FORCE proposes application of the carboligase in a multi-functional role – first for the condensation of formyl-CoA with formaldehyde, then with acetaldehyde, then proprionaldehyde and so on, creating iterative C1 elongation[17]. Moving toward implementation, it could be shown that MeOXC4 indeed outperforms HACL in the resting cell system used for FORCE.

Combining the enzymes developed in this study with a few existing modules, production of lactic acid, a precursor for bioplastic, from formate can be envisioned (**Figure 2**). EcAckA+DaArgC3+LmACR produces formyl-CoA and formaldehyde, MeOXC4 condenses them to glycolyl-CoA, the FORCE pathway's reductive cascade yields acetaldehyde, which can be C1-elongated by MeOXC4 to lactyl-CoA and hydrolyzed to lactic acid.

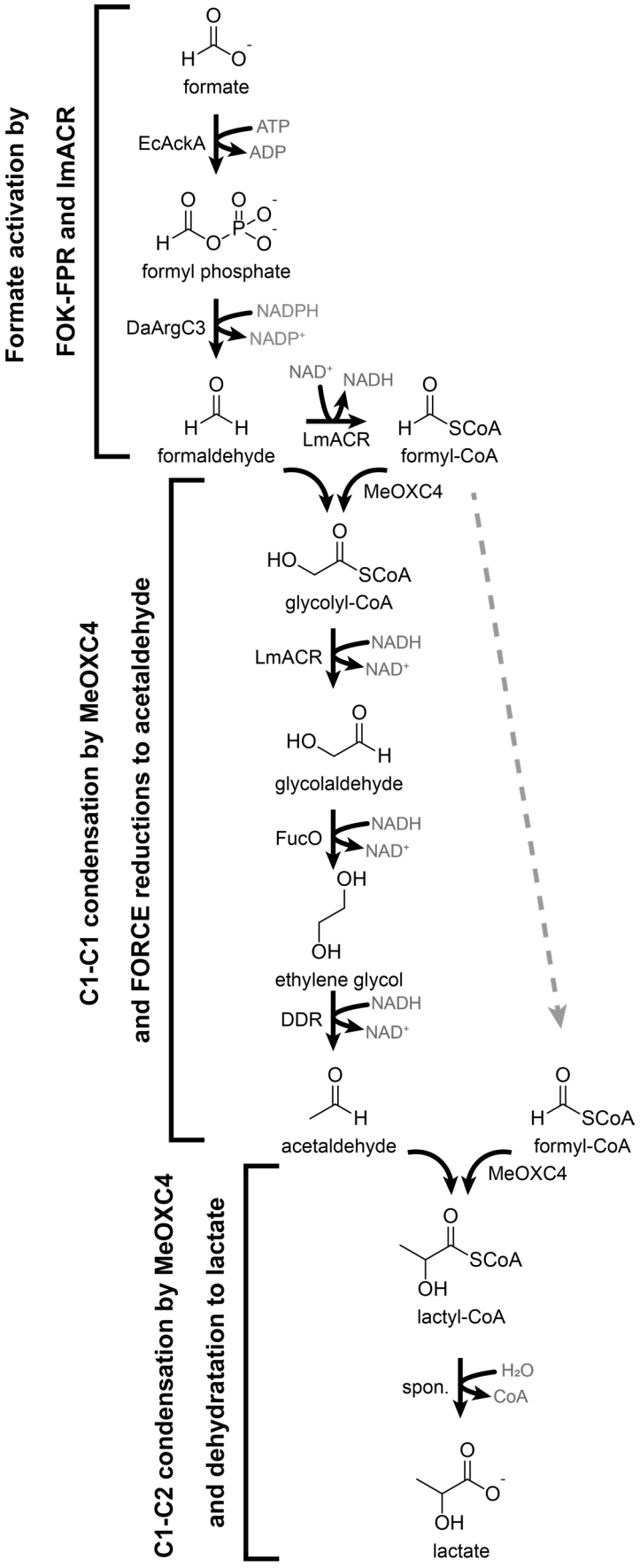


Figure 2 Production of lactate from 3 molecules of formate using the FOK-FPR module developed in *CHAPTER 1*, reactions of the FORCE pathway[17] and the C1-C1 and C1-C2 condensation reactions catalyzed by MeOXC4, as demonstrated in *CHAPTER 2*. EcAckA: *E. coli* acetate kinase, DaArgC: *D. acetiphilus* N-acetyl-gamma-glutamyl phosphate reductase, MeOXC4: *M. extorquens* oxalyl-CoA carboxylase variant, LmACR: *L. monocytogenes* acyl-CoA reductase, FucO: 1,2-diol oxidoreductase, DDR: diol dehydratase, spon. = spontaneous reaction

This pathway requires 7 enzymes and, accounting for oxidative phosphorylation, consumes 13 ATP equivalents, compared to more conventional approaches to produce lactic acid from C1 compounds, which require a full metabolic integration of C1 compounds into pyruvate. As an example, using the XuMP pathway combined with FOK-FPR for conversion of formate to lactic acid requires 15 enzymes and 12 ATP equivalents – making the FORCE module much shorter while operating at a similar energetic cost.

Removed from C1 compounds as a sole feedstock, more targeted biochemical synthesis can be envisioned around MeOXC4. A previous publication by Burgener *et al.* as well as *CHAPTER 3*, show that MeOXC and its variants enable C1 elongation of multiple (aromatic) aldehyde substrates and acetone[8, 24]. This permits creation of a simple three enzyme module for the production of 2-hydroxy-acids (i.e. value-added compounds like phenyllactic and mandelic acid) from the corresponding aldehyde and formate – activating formate to formyl-CoA using ACS[25, 26], condensing formyl-CoA and aldehyde to yield the 2-hydroxyacyl-CoA using MeOXC4, and releasing the 2-hydroxy-acids by hydrolysis or CoA transfer.

Outlook in the context of the formate bio-economy

Some key steps can be defined for a successful realization of the formate bio-economy. Starting from the bottom up, designing and engineering novel enzyme functions enables shorter, more energy efficient formate assimilation pathways. Their integration permits establishment of synthetic formatotrophy in common platform organisms[27-29]. Another option is the creation of production platforms out of natural formatotrophs such as *Moorella thermoacetica*[30] or *Acetobacterium woodii*[31]. These organisms then require a reliable supply of green formate. This could potentially take the form of bio-hybrid systems cultivating bacteria in a CO_2 hydrogenation chamber, as has been shown for other C1 assimilating systems[32, 33] and relies heavily on efficient renewable energy production. Alternatively, anaerobic systems may directly reduce CO_2 to formate using formate dehydrogenase[34]. Downstream production cascades, as exist for isoprenes, acetone, butanol, ethanol and many more[35-37], need to be implemented and optimized. Finally, once all these considerations are accumulated into a functional strain, production needs to be scaled up to industrial relevance, as has been

achieved for the production of acetone and isopropanol from CO_2/H_2 by the acetogen *Clostridium autoethanogenum*[38].

The results obtained in this thesis contribute in major ways to the diversity, energy-efficiency and simplification of formate assimilation. Compared to existing formate reduction cascades[15, 16, 26], the FOK-FPR cascade of *CHAPTER 1* conserves energy for the organism and avoids costly interactions with intracellular THF or CoA pools. Meanwhile, the engineered GCS of *CHAPTER 2* offers high versatility as a promiscuous aldehyde acceptor while providing high efficiency for most desired reactions. Overlain with appropriate genetic regulations, production routes containing either or both can be tailored for high selectivity and efficiency. Overall, this thesis demonstrates the power of enzyme engineering in the realization of a sustainable bio-economy.

References

[1] 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.

[2] Claassens, N. J., Cotton, C. A. R., Kopljar, D., and Bar-Even, A. (2019) Making quantitative sense of electromicrobial production, *Nat. Catal. 2*, 437-447.

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

[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, *Chem. Rev. 115*, 12936-12973.

[5] Artz, J., Muller, T. E., Thenert, K., Kleinekorte, J., Meys, R., Sternberg, A., Bardow, A., and Leitner, W. (2018) Sustainable Conversion of Carbon Dioxide: An Integrated Review of Catalysis and Life Cycle Assessment, *Chem. Rev. 118*, 434-504.

[6] Schaub, T. (2018) CO2-based hydrogen storage: CO2 hydrogenation to formic acid, formaldehyde and methanol, *Phys. Sci. Rev. 3*, 20170015.

[7] Ljungdahl, L. G., and Andreesen, J. R. (1978) Formate dehydrogenase, a selenium-tungsten enzyme from Clostridium thermoaceticum, In *Meth. Enzymol.* (Fleischer, S., and Packer, L., Eds.), 360-372, Academic Press.

[8] Nattermann, M., Burgener, S., Pfister, P., Chou, A., Schulz, L., Lee, S. H., Paczia, N., Zarzycki, J., Gonzalez, R., and Erb, T. J. (2021) Engineering a Highly Efficient Carboligase for Synthetic One-Carbon Metabolism, *ACS Catal. 11*, 5396-5404.

[9] Chittori, S., Savithri, H. S., and Murthy, M. R. (2012) Structural and mechanistic investigations on Salmonella typhimurium acetate kinase (AckA): identification of a putative ligand binding pocket at the dimeric interface, *BMC Struct. Biol. 12*, 24.

[10] Sly, W. S., and Stadtman, E. R. (1963) Formate Metabolism. Ii. Enzymatic Synthesis of Formyl Phosphate and Formyl Coenzyme a in Clostridium Cylindrosporum, *J. Biol. Chem. 238*, 2639-2647.

[11] Dragosits, M., and Mattanovich, D. (2013) Adaptive laboratory evolution -- principles and applications for biotechnology, *Microb. Cell Fact. 12*, 64.

[12] Kemp, M. B., and Quayle, J. R. (1966) Microbial growth on C1 compounds. Incorporation of C1 units into allulose phosphate by extracts of Pseudomonas methanica, *Biochem. J. 99*, 41-48.

[13] Van Dijken, J. P., Harder, W., Beardsmore, A. J., and Quayle, J. R. (1978) Dihydroxyacetone: An intermediate in the assimilation of methanol by yeasts?, *FEMS Microbiol. Lett. 4*, 97-102.

[14] 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. 60*, 1-13.

[15] Hu, G., Li, Z., Ma, D., Ye, C., Zhang, L., Gao, C., Liu, L., and Chen, X. (2021) Light-driven CO2 sequestration in Escherichia coli to achieve theoretical yield of chemicals, *Nat. Catal. 4*, 395-406.

[16] Hu, G., Guo, L., Gao, C., Song, W., Liu, L., and Chen, X. (2022) Synergistic Metabolism of Glucose and Formate Increases the Yield of Short-Chain Organic Acids in Escherichia coli, *ACS Synth. Biol. 11*, 135-143.

[17] Chou, A., Lee, S. H., Zhu, F., Clomburg, J. M., and Gonzalez, R. (2021) An orthogonal metabolic framework for one-carbon utilization, *Nat. Metab. 3*, 1385-1399.

[18] 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.

[19] Wang, X., Wang, Y., Liu, J., Li, Q., Zhang, Z., Zheng, P., Lu, F., and Sun, J. (2017) Biological conversion of methanol by evolved Escherichia coli carrying a linear methanol assimilation pathway, *Bioresour. Bioprocess. 4*.

[20] Cai, T., Sun, H., Qiao, J., Zhu, L., Zhang, F., Zhang, J., Tang, Z., Wei, X., Yang, J., Yuan, Q., Wang, W., Yang, X., Chu, H., Wang, Q., You, C., Ma, H., Sun, Y., Li, Y., Li, C., Jiang, H., Wang, Q., and Ma, Y. (2021) Cell-free chemoenzymatic starch synthesis from carbon dioxide, *Science 373*, 1523-1527.

[21] Güner, S., Wegat, V., Pick, A., and Sieber, V. (2021) Design of a synthetic enzyme cascade for the in vitro fixation of a C1 carbon source to a functional C4 sugar, *Green Chem. 23*, 6583-6590.

[22] Jo, H.-J., Kim, J.-H., Kim, Y.-N., Seo, P.-W., Kim, C.-Y., Kim, J.-W., Yu, H.-n., Cheon, H., Lee, E. Y., Kim, J.-S., and Park, J.-B. (2022) Glyoxylate carboligase-based whole-cell

biotransformation of formaldehyde into ethylene glycol via glycolaldehyde, *Green Chem. 24*, 218-226.

[23] 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.

[24] 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. 59*, 5526-5530.

[25] 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.

[26] Wang, J., Anderson, K., Yang, E., He, L., and Lidstrom, M. E. (2021) Enzyme engineering and in vivo testing of a formate reduction pathway, *Synth. Biol. 6*, ysab020.

[27] 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.

[28] 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.

[29] 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.

[30] Kato, J., Takemura, K., Kato, S., Fujii, T., Wada, K., Iwasaki, Y., Aoi, Y., Matsushika, A., Murakami, K., and Nakashimada, Y. (2021) Metabolic engineering of Moorella thermoacetica for thermophilic bioconversion of gaseous substrates to a volatile chemical, *AMB Express 11*, 59.

[31] Arslan, K., Schoch, T., Hofele, F., Herrschaft, S., Oberlies, C., Bengelsdorf, F., Veiga, M. C., Durre, P., and Kennes, C. (2022) Engineering Acetobacterium woodii for the production of isopropanol and acetone from carbon dioxide and hydrogen, *Biotechnol. J 17*, e2100515.

[32] Cestellos-Blanco, S., Chan, R. R., Shen, Y.-x., Kim, J. M., Tacken, T. A., Ledbetter, R., Yu, S., Seefeldt, L. C., and Yang, P. (2022) Photosynthetic biohybrid coculture for tandem and tunable CO₂ and N₂ fixation, *Proc. Natl. Acad. Sci. U.S.A. 119*, e2122364119.

[33] Amrut Pawar, A., Karthic, A., Lee, S., Pandit, S., and Jung, S. P. (2022) Microbial electrolysis cells for electromethanogenesis: Materials, configurations and operations, *Environ. Eng. Res. 27*, 200484-200480.

[34] Calzadiaz-Ramirez, L., and Meyer, A. S. (2022) Formate dehydrogenases for CO2 utilization, *Curr. Opin. Biotechnol. 73*, 95-100.

[35] Lee, S. Y., Kim, H. U., Chae, T. U., Cho, J. S., Kim, J. W., Shin, J. H., Kim, D. I., Ko, Y.-S., Jang, W. D., and Jang, Y.-S. (2019) A comprehensive metabolic map for production of bio-based chemicals, *Nat. Catal. 2*, 18-33.

[36] David, Y., Baylon, M. G., Lee, S. Y., and Park, S. J. (2017) Bioproduction of Chemicals: An Introduction, In *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals* (Lee, S. Y., Ed.), 1-16, Springer International Publishing, Cham.

[37] Scully, S. M., and Orlygsson, J. (2019) Chapter 5 - Biological Production of Alcohols, In *Advanced Bioprocessing for Alternative Fuels, Biobased Chemicals, and Bioproducts* (Hosseini, M., Ed.), 83-108, Woodhead Publishing.

[38] Liew, F. E., Nogle, R., Abdalla, T., Rasor, B. J., Canter, C., Jensen, R. O., Wang, L., Strutz, J., Chirania, P., De Tissera, S., Mueller, A. P., Ruan, Z., Gao, A., Tran, L., Engle, N. L., Bromley, J. C., Daniell, J., Conrado, R., Tschaplinski, T. J., Giannone, R. J., Hettich, R. L., Karim, A. S., Simpson, S. D., Brown, S. D., Leang, C., Jewett, M. C., and Köpke, M. (2022) Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale, *Nat. Biotechnol. 40*, 335-344.

Acknowledgements

All my time at the Erb lab, part of me felt that I was standing on the shoulders of giants, that helped lift me up to where I am today. This thesis would not have been possible without the support of the people around me.

Thank you to Tobi Erb for giving me the opportunity to work in his lab for both my Master and PhD projects, for giving support during the low phases of my project and his trust in me and my work. To my TAC committee, Prof. Dr. Johann Heider and Prof. Dr. Lars-Oliver Essen for their continued feedback on my work. To my thesis committee for agreeing to evaluate my work.

Thank you to Simon Burgener for being an amazing supervisor both in my Master thesis and then, on a voluntary basis, throughout the beginning of my PhD. To Seb Wenk for being an amazing grant partner, for all the hours he put into making our shared projects work, for staying on top of the things I forgot. To both of you, for proofreading this thesis.

Thank you to all my co-authors and labmates for all of the experimental assistance and discussion of my work. Shout-outs to my long-suffering X-Ray team, Jan Zarzycki and Pascal Pfister, for all the times they didn't give up on me and my proteins, regardless how long they took to diffract. To Nicole Paczia and Peter Claus (and of course Emmma and Mmmolly) for the thousands of LC-MS measurements. To my current and former students, Lennart Nickel, Philipp Wichmann, Tiago Lopes, Irene Martinez and Charlotte Wallner, for giving me the chance to learn how to supervise, for doing excellent experimental work and taking a lot of load off my shoulders. To Daniel Marchal for being an amazing bench neighbor in the last years. To Luca Schulz for jumping in anytime I needed a second opinion on my silly ideas. To everyone that ever took one of my agar plates out of the incubator or turned on my PC when I wasn't at the institute. It takes a village to finish a PhD project.

Finally, thank you to my family and friends. To Jasper for being my rock during the past four years, for picking me up whenever I was tempted to drop everything, for loving me. To Sabine, for that one drive you made halfway through Germany when I told you I wasn't in a good place. To Peter, for the countless times you just dropped by on your way home, those brief calls just to check up on me from time to time. For carrying all that furniture when I moved.

To Torben for always lending an open ear and the TFT and League games you suffered through. To Rosi for our Sunday calls, for all your advice and belief in me. To Anne and Edna for the little group chat that we use to pick each other up, and for one of the best birthday gifts ever. To Jürgen and Frauke for welcoming me into your family these past years, and for those two weeks you had to feed me while I was stuck in your house with COVID (sorry about that). Another shout-out to Jasper, Jürgen and Peter for proofreading and comments.

I said finally already, but the final space will always belong to the people that will never know I finished my PhD. To Karl, I wish we had more time together, and could have celebrated all those milestones together. To Waldemar, I wish I could see you smile when I tell you I've graduated. I know you'd be proud, because you've always been unwaveringly proud of Torben and I, no matter what. To Ongellaus, I never even got to tell you I got my Master's degree, and I'm now taking another step I'll never get to celebrate with you. I know you'd be over the moon, and I know we'd be planning a trip to a musical right now.