



Exploring and engineering formaldehyde assimilation

Hai He

Univ.-Diss. (Kumulative Dissertation)

**zur Erlangung des akademischen Grades
"doctor rerum naturalium"**

(Dr. rer. nat.)

in der Wissenschaftsdisziplin Systeme und Synthetische Biologie

**eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
Institut für Biochemie
der Universität Potsdam**

**Potsdam-Golm, Deutschland
Juli 2020**

Supervisor

Dr. habil Arren Bar-Even

Reviewers

Prof. Dr. Volker F. Wendisch

Prof. Dr. Yu Wang

Dr. Arren Bar-Even

Published online on the
Publication Server of the University of Potsdam:
<https://doi.org/10.25932/publishup-47386>
<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-473867>

Declaration of Work

The work described in this dissertation, submitted to the Institute of Biochemistry and Biology at the University of Potsdam, has been carried out between October 2016 and March 2020 in the research group of Dr. Arren Bar-Even at the Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany. I declare that this thesis has not been submitted to another examining body, either partially or wholly, as part of a doctoral degree. The thesis as a whole has not been published or submitted for publication.

This cumulative thesis is bound from four publications which are marked with “*” in the list below. My contributions to these publications are declared hereby:

* **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(6): 1601-1611.

Personal contribution: Conducted the entire experimental work;

Wrote the paper together with supervisor and coworkers.

* **He, H.**, Höper, R., Dodenhöft, M., Marlière, P. and Bar-Even, A., **2020**. Promiscuous aldolases facilitate an optimized methanol assimilation pathway in *E. coli*. *Metabolic engineering*, 60: 1-13.

Personal contribution: The entire work was carried out by me and students under my supervision; the students however only did preliminary tests;

Wrote the paper together with supervisor.

* **He, H.**, Noor E., Ramos-Parra, P.A., García-Valencia, L.E., Patterson, A.J., Hanson, A.D., de la Garza, R.I.D. and Bar-Even, A., **2020**. *In vivo* rate of formaldehyde condensation with tetrahydrofolate. *Metabolites*, 10(2): 65.

Personal contribution: Designed and conducted the major part of experiments;

Assisted computational analysis with E. Noor.

Wrote the paper together with supervisor and coworkers.

* Patterson, A.J.#, **He, H.**#, Folz, J.S.#, Li, Q., Wilson, M.A., Fiehn, O., Bruner, S.D., Bar-Even, A., Hanson, A.D., Thioproline formation as a driver of formaldehyde toxicity in *Escherichia coli*. *Biochemical Journal*. 477(9): 1745-1757.

Personal contribution: Designed and conducted the major part of experiments together with supervisor and coworkers.

Reviewed the paper together with supervisor and coworkers.

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.

Personal contribution: Wrote the review together with supervisor and colleagues.

Lindner, S.N., Ramirez, L.C., Krüsemann, J.L., Yishai, O., Belkhef, S., **He, H.**, Bouzon, M., Döring, V. and Bar-Even, A., **2018**. NADPH-auxotrophic *E. coli*: a sensor strain for testing *in vivo* regeneration of NADPH. *ACS Synth. Biol.* 7(12): 2742-2749.

Personal contribution: Conducted computational analysis.

Krüsemann, J.L., Lindner, S.N., Dempfle, M., Widmer, J., Arrivault, S., Debacker, M., **He, H.**, Kubis, A., Chayot, R., Anissimova, M., Marlière, P., Cotton, C.A.R. and Bar-Even, A., **2018**. Artificial pathway emergence in central metabolism from three recursive phosphoketolase reactions. *FEBS J.*, 285(23): 4367-4377.

Personal contribution: Constructed some of the selection strains in the study.

Cotton, C.A.R., Bernhardsgrütter, I., **He, H.**, Erban A., Toman, S., Dempfle, M., Kopka, J., Lindner, S.N., Erb, T.J., and Bar-Even, A. Underground isoleucine biosynthesis pathways in *E. coli*. *Resubmitted*.

Personal contribution: Conducted metabolomics experiments with coworkers.

Analyzed NGS sequence results.

Reviewed the paper together with coworkers.

Wenk, S.[#], Schann, K.[#], **He, H.**, Rainaldi, V., Kim, S.Y., Lindner, S.N., and Bar-Even, A., An “energy-auxotroph” *E. coli* provides an *in vivo* platform for assessing synthetic NADH regeneration systems. *Biotechnology and Bioengineering*, DOI: 10.1002/bit.27490.

Personal contribution: Conducted computational analysis.

Constructed MDH overexpress plasmids.

Reviewed the paper together with coworkers.

([#] Theses authors contributed equally to the work.)

Conformation of Supervisor:

I hereby confirm the contributions mentioned by Hai He to mentioned studies.

Dr. Arren Bar-Even (PI)

Table of Contents

Declaration of Work.....	1
Summary	iii
Zusammenfassung	v
Overarching Introduction	1
C1 compounds as feedstocks.....	1
Formaldehyde as a versatile intermediate	5
Formaldehyde assimilation pathways	6
Modularity and selection metabolic engineering strategies	10
Outline of the thesis studies	12
Chapter 1: Ribulose monophosphate shunt provides nearly all biomass and energy required for growth of <i>E. coli</i>	15
Chapter 2: An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in <i>E. coli</i>	33
Chapter 3: <i>In vivo</i> rate of formaldehyde condensation with tetrahydrofolate	55
Chapter 4: Thioproline formation as a driver of formaldehyde toxicity in <i>Escherichia coli</i> ...	71
Overarching Discussion	93
<i>In vivo</i> formaldehyde production	93
Formaldehyde assimilation	95
Metabolic network.....	96
Closing remarks.....	97
References (overarching introduction and discussion)	99
Acknowledgements	105

Summary

Increasing concerns regarding the environmental impact of our chemical production have shifted attention towards possibilities for sustainable biotechnology. One-carbon (C1) compounds, including methane, methanol, formate and CO, are promising feedstocks for future bioindustry. CO₂ is another interesting feedstock, as it can also be transformed using renewable energy to other C1 feedstocks for use. While formaldehyde is not suitable as a feedstock due to its high toxicity, it is a central intermediate in the process of C1 assimilation. This thesis explores formaldehyde metabolism and aims to engineer formaldehyde assimilation in the model organism *Escherichia coli* for the future C1-based bioindustry.

The first chapter of the thesis aims to establish growth of *E. coli* on formaldehyde via the most efficient naturally occurring route, the ribulose monophosphate pathway. Linear variants of the pathway were constructed in multiple-gene knockouts strains, coupling *E. coli* growth to the activities of the key enzymes of the pathway. Formaldehyde-dependent growth was achieved in rationally designed strains. In the final strain, the synthetic pathway provides the cell with almost all biomass and energy requirements.

In the second chapter, taking advantage of the unique feature of its reactivity, formaldehyde assimilation via condensation with glycine and pyruvate by two promiscuous aldolases was explored. Facilitated by these two reactions, the newly designed homoserine cycle is expected to support higher yields of a wide array of products than its counterparts. By dividing the pathway into segments and coupling them to the growth of dedicated strains, all pathway reactions were demonstrated to be sufficiently active. The work paves a way for future implementation of a highly efficient route for C1 feedstocks into commodity chemicals.

In the third chapter, the *in vivo* rate of the spontaneous formaldehyde tetrahydrofolate condensation to methylene-tetrahydrofolate was assessed in order to evaluate its applicability as a biotechnological process. Tested within an *E. coli* strain deleted in essential genes for native methylene-tetrahydrofolate biosynthesis, the reaction was shown to support the production of this essential intermediate. However, only low growth rates were observed and only at high formaldehyde concentrations. Computational analysis dependent on *in vivo* evidence from this strain deduced the slow rate of this spontaneous reaction, thus ruling out its substantial contribution to growth on C1 feedstocks.

The reactivity of formaldehyde makes it highly toxic. In the last chapter, the formation of thioproline, the condensation product of cysteine and formaldehyde, was confirmed to contribute this toxicity effect. Xaa-Pro aminopeptidase (PepP), which genetically links with folate metabolism, was shown to hydrolyze thioproline-containing peptides. Deleting *pepP* increased strain sensitivity to formaldehyde, pointing towards the toxicity of thioproline-containing peptides and the importance of their removal. The characterization in this study could be useful in handling this toxic intermediate.

Overall, this thesis identified challenges related to formaldehyde metabolism and provided novel solutions towards a future bioindustry based on sustainable C1 feedstocks in which formaldehyde serves as a key intermediate.

Zusammenfassung

Die zunehmende Besorgnis über die Auswirkungen der chemischen Produktion auf die Umwelt hat den Fokus auf die Vorzüge der nachhaltigen Biotechnologie gelenkt. Ein-Kohlenstoff-Verbindungen (C1), wie Methan, Methanol, Formiat und CO, sind vielversprechende Ausgangsstoffe für eine künftige biobasierte Industrie. Darüber hinaus ist CO₂ ein weiterer interessanter Rohstoff, da es ebenfalls mit erneuerbarer Energie in andere C1-Verbindungen umgewandelt werden kann. Formaldehyd ist zwar aufgrund seiner Toxizität und Reaktivität nicht als Kohlestoffquelle geeignet, stellt aber ein zentrales Intermediat der C1-Assimilation dar. In dieser Arbeit wurde der Formaldehyd-Stoffwechsel untersucht, um Formaldehyd-Assimilations-Wege im Modellorganismus *Escherichia coli* für die zukünftige C1-basierte Bioindustrie zu entwickeln.

Das erste Kapitel dieser Arbeit zielt darauf ab, das Wachstum von *E. coli* auf Formaldehyd über den energetisch effizientesten natürlich vorkommenden Weg, den Ribulose-Monophosphat-Weg (RuMP-Weg), zu etablieren. Durch Gendelektionen wurden lineare Varianten des RuMP-Weges in *E. coli* Stämmen konstruiert, um das Wachstum mit der Aktivität der Schlüsselenzyme der Formaldehyd-Assimilation zu verbinden. So konnte auf Formaldehyd-abhängiges Wachstum in diesen Stämmen getestet werden. Im finalen Stamm wurde nahezu die gesamte Biomasse und der Energiebedarf der *E. coli* Zellen über den RuMP-Weg bereitgestellt.

Im zweiten Kapitel wurde unter Ausnutzung der einzigartigen Reaktivität des Formaldehyds die Aktivitäten zweier promiskuitiver Aldolasen für die Kondensation von Formaldehyd mit Glycin oder Pyruvat *in vivo* untersucht. Diese beiden Reaktionen dienen als Schlüsselreaktionen des neu gestalteten Homoserin-Zyklus, der verglichen mit anderen bekannten Assimilierungswegen höhere Ausbeuten einer breiten Palette von Produkten ermöglichen kann. Durch die Unterteilung des Stoffwechselweges in Segmente und deren Selektion in auxotrophen *E. coli* Stämmen wurde gezeigt, dass alle Reaktionen des Pfades ausreichend aktiv sind. Diese Arbeit ebnet den Weg für die zukünftige Nutzung des Homoserin-Zyklus als hocheffiziente Route für die Umwandlung von C1-Rohstoffen in grundlegende Chemikalien.

Im dritten Kapitel wurde die *in vivo*-Rate der spontanen Kondensation von Formaldehyd mit Tetrahydrofolat zu Methylen-Tetrahydrofolat untersucht, um ihr Potential als Formaldehyd Assimilierungs-Reaktion in einem biotechnologischen Prozess zu bewerten. Die Reaktion

wurde in einem *E. coli*-Stamm getestet, in welchem essentielle Gene für die native Methylen-Tetrahydrofolat-Biosynthese entfernt wurden. Es konnte gezeigt werden, dass die spontane Reaktion die Produktion dieses wesentlichen Zwischenprodukts ermöglicht. Die erreichten Wachstumsraten waren jedoch gering und nur bei hohen Formaldehydkonzentrationen zu beobachten. Eine Computer-gestützte Analyse basierend auf *in vivo*-Erkenntnissen über diesen Stamm ergab eine langsame Reaktionsgeschwindigkeit dieser spontanen Reaktion und schloss somit ihren wesentlichen Beitrag zum Wachstum auf C1-Einsatzmaterial aus.

Seine Reaktivität macht Formaldehyd hochtoxisch. Im letzten Kapitel wurde bestätigt, dass die Bildung von Thioprolin, dem Kondensationsprodukt von Cystein und Formaldehyd, zu dieser Toxizitätswirkung beiträgt. Es wurde gezeigt, dass die Xaa-Pro-Aminopeptidase (PepP), die genetisch mit dem Folat-Stoffwechsel verbunden ist, thioprolinhaltige Peptide hydrolysiert. Die Deletion von PepP erhöhte die Empfindlichkeit des Stammes gegenüber Formaldehyd, was auf die Toxizität der Thioprolin-haltigen Peptide und die Bedeutung ihrer Entfernung hinweist. Diese Studie kann für den Umgang mit der Toxizität von Formaldehyd nützlich sein.

Insgesamt identifizierte diese Arbeit Herausforderungen im Zusammenhang mit dem Formaldehyd-Stoffwechsel und lieferte neue Lösungen für eine C1-basierte Bioindustrie, in der Formaldehyd als wichtiges Zwischenprodukt dient.

Overarching Introduction

Driven by the concerns on current fossil based industry and facilitated by the rapid development of biotechnology, biomanufacturing is taking more responsibility for sustainable and environmentally friendly fuel and chemical production (Clomburg et al., 2017; Lee et al., 2019). Future industrial biomanufacturing of commodity chemicals requires cheap, widely available, renewable and non-food feedstocks. While sugars and starches are currently commonly used feedstocks, their utilization compete with human consumption. Sugars from lignocellulosic biomass fulfil these requirements, but involve difficult, costly and deleterious decomposition processes (Sanderson, 2011). Its supply chain also faces various challenges (Hassan et al., 2019). One-carbon (C1) compounds, specifically methane, methanol, formate, CO, and CO₂, have been widely proposed to be attractive alternatives as they fit the abovementioned requirements (Fig. 1) (Dürre and Eikmanns, 2015; Haynes and Gonzalez, 2014; Schrader et al., 2009; Venkata Mohan et al., 2016; Yishai et al., 2016). Especially, the former substrates are expected to be sustainably converted from CO₂, making a carbon cycle.

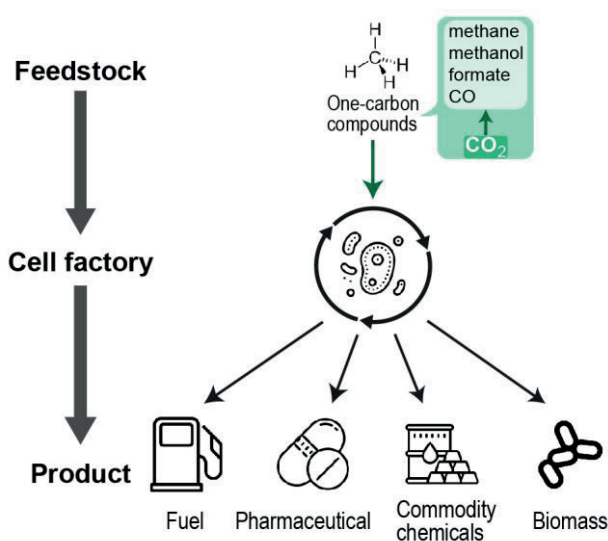


Figure 1. One-carbon compounds for future industrial biomanufacturing feedstocks. The shifting of industry from fossil-based to microbial cell factories on C1 substrates, i.e. methane, methanol, formate, CO and CO₂, aligns with the Sustainable Development Goals announced by the United Nations (D. Yang et al., 2017). Of these, the former feedstocks can be transformed from CO₂.

C1 compounds as feedstocks

C1 compounds include an array of substrates. Carbon dioxide (CO₂), carbon monoxide (CO), formate (HCOOH), methanol (CH₃OH) and methane (CH₄) have gained great attention as feedstocks for microbial production.

CO₂ is an unlimited resource (IEA, 2019). As the terminal product of carbon oxidation, CO₂ fixation for carbon recycling has received great attention. Especially in response to the unceasingly increasing atmospheric concentration of this greenhouse gas, there has been extensive studies on CO₂ assimilation for bio-production. Among the six natural CO₂ fixation routes, the reductive acetyl-CoA pathway, also known as the Wood-Ljungdahl pathway (Ljungdahl and Wood, 1965), is the most energy efficient one (Fig. 2) (Claassens et al., 2019a; Fuchs, 2011). However, this pathway operates only under anaerobic conditions and transfers low amount of electrons. Both features constrain the product spectrum via the pathway (Müller, 2019). Moreover, the pathway needs an electron and energy source. Hydrogen (H₂), which can be renewably produced by biological processing of organic waste and electrochemical water split (Nikolaidis and Poullikkas, 2017), can serve as a clean source of electrons and energy. However, its low solubility (Henry's constants of 7.8×10^{-4} M atm⁻¹, refer to Sander (2015)) in water limits the mass transfer to the microbial culture, constraining the industrial usage in bio-production.

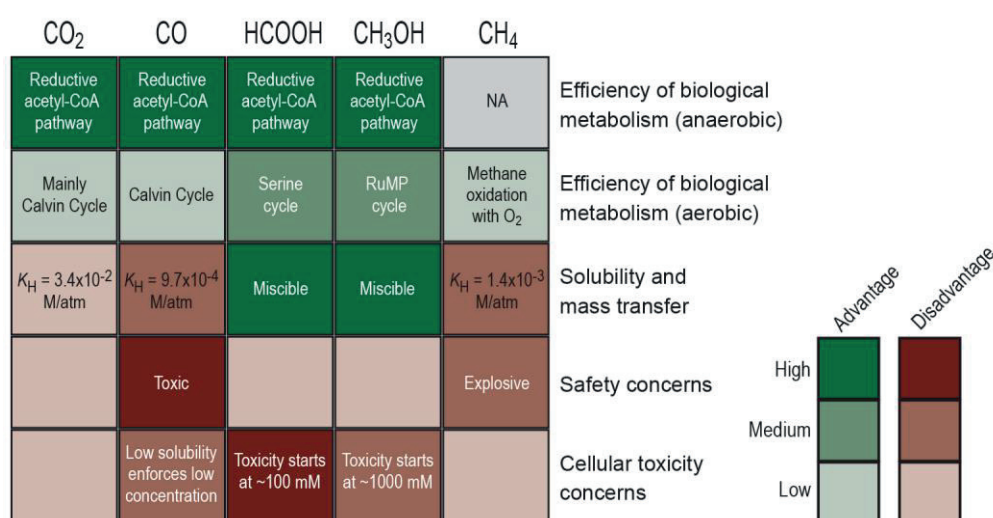


Figure 2. Comparison of C1 feedstocks. K_H , Henry's law constant; NA, not applicable. Adapted from Claassens et al. (2019a).

The most dominating CO₂ assimilation pathway is the Calvin-Benson-Bassham (CBB) cycle (Calvin and Benson, 1948), used by autotrophic photosynthetic microorganisms, such as microalgae and cyanobacteria (plants are problematic to be used for industrial biomanufacturing because of factors like land use, cultivation time, downstream process, etc.). A lot of effort has been put into biofuel production by microalgae and cyanobacteria, harvesting sunlight energy. However the light capturing efficiency and light distribution are the major challenges (Liao et al., 2016). Additionally, the limiting metabolic engineering knowledge and techniques for these photosynthetic organisms motivated studies of engineering the CBB

pathway in other well-characterized organisms. Recently, the CBB cycle has been established heterologously in *Escherichia coli* (Gleizer et al., 2019) and *Pichia pastoris* (Gassler et al., 2020), using formate or methanol as an electron and energy source while fixing CO₂, respectively. However, the CBB cycle requires high amounts of ATP and reducing power. Its carboxylase, RuBisCO (Ribulose-1,5-bisphosphate carboxylase/oxygenase), is slow and with strong side activity with oxygen (Erb and Zarzycki, 2018). Therefore, synthetic carbon fixation architectures, such as the CETCH (crotonyl-CoA/ethylmalonyl-CoA/hydroxybutyryl-CoA) cycle (Schwander et al., 2016) and the MOG (Malonyl-CoA-oxaloacetate-glyoxylate) pathways (Bar-Even et al., 2010), were designed and tested.

Similar to CO₂, CO has also attracted much interest as a feedstock due to the need of reducing industrial gas waste emission (Zeng, 2019). Moreover, syngas, containing CO, H₂, and CO₂, from organic waste gasification provides a globally available resource (Latif et al., 2014). A diverse range of microbes can use CO as carbon and energy source. Some acetogens fix CO anaerobically via the reductive acetyl-CoA pathway at high efficiency (Fig. 2), thus have been exploited for the production of acetate, 2,3-butanediol, butanol or butyrate (Schiel-Bengelsdorf and Dürre, 2012). Chemolithoautotrophic bacteria assimilate CO aerobically through the CBB cycle by first oxidized it to CO₂. However, their biotechnological potentials await to be characterized and their utilization is still in its infancy (Dürre and Eikmanns, 2015). Moreover, the challenges inherited from traits of CO, i.e. low solubility, high toxicity and explosive under aerobic conditions (Fig. 2), need to be addressed for its industrial utilization (Takors et al., 2018).

Another gaseous C1 substrate, methane, suffers similar problems as a feedstock (Fig. 2). Methane is the main component of natural gas and biogas, widely accessible from landfills and syngas. Extensive analyses on its biological conversion to liquid fuels (Bio-GTL) and chemicals have been made in recent years (Averesch and Kracke, 2018; Clomburg et al., 2017; Fei et al., 2014; Haynes and Gonzalez, 2014; Hwang et al., 2018; Lee et al., 2016; Liao et al., 2016). Especially, Gonzalez and colleagues as well as other studies estimated that methane bioconversion offers an economically feasible potential because of its low cost (Clomburg et al., 2017; Comer et al., 2017; Fei et al., 2014; Haynes and Gonzalez, 2014). Natural methanotrophic microorganisms or methanotrophs can use methane as sole carbon and energy source. Some aerobic methanotrophs have been investigated for the production of lipids, poly- β -hydroxybutyrate (PHB), or biomass. However, engineering of these bacteria is limited by metabolic knowledge and available toolboxes (Kalyuzhnaya et al., 2015). Although the

structure of the copper-containing methane monooxygenase was recently characterized (Ross et al., 2019), its heterologous expression has largely failed, making synthetic methanotrophs infeasible. Moreover, aerobic methane oxidation is energetically wasteful and requires special designed bioreactors to solve the problem of mass transfer and safety concerns regarding explosive.

The biological utilization of methane is closely related to biological assimilation of methanol, as the oxidation product of the former is methanol. Methanol itself is a promising feedstock for industrial applications and microbial growth (Cotton et al., 2020; Olah, 2005; Schrader et al., 2009). Proposed early in 1985 as a potential feedstock for bioprocesses (Dijkhuizen, 1985), methanol is now in the spotlight and studies on biological methanol assimilation were conducted extensively in the last decade. Unlike methane, CO₂ and CO, methanol is liquid and, importantly, miscible with water, making it easy to handle. Currently, methanol is mainly produced via CO hydrogenation from syngas, which is mainly from methane. Direct conversion of methane, biologically and chemically, has been intensively studied (Kim et al., 2019; Sushkevich et al., 2017; Zakaria and Kamarudin, 2016). Production from renewable resources, e.g. CO₂ (Olah, 2013; Patterson et al., 2019) and biomass (Devi et al., 2019), are on their way towards industrialization. These together will increase the supply and decrease the price. Importantly, methanol is more reduced than most sugars (as well as CO₂, CO and formate) (Wang et al., 2020), equipping it with the potential of higher product yields as it can provide more reducing power. One problem is that methanol is toxic (Fig. 2), this on the other hand could avoid contaminations by bacteria. There are three natural pathways for methanol assimilation, including the ribulose monophosphate (RuMP) cycle, the xylulose monophosphate (XuMP) cycle or dihydroxyacetone (DHA) cycle, and the serine cycle. The RuMP cycle is the most efficient pathway in aerobic conditions, the reductive acetyl-CoA pathway is the most efficient under anaerobic conditions (Fig. 2) (Claassens et al., 2019a).

Formate is another potential microbial feedstock. It can be renewably produced from CO₂ by hydrogenation (Wang et al., 2015), electrochemical or photochemical processes (H. Yang et al., 2017; Zhou et al., 2016), as well as biomass oxidation (Albert et al., 2014). Yishai et al. (2016) proposed a “formate bioeconomy” and discussed formate production and biological assimilation. The natural serine cycle and reductive acetyl-CoA pathway are the most efficient pathways of formate assimilation under aerobic and anaerobic conditions, respectively (Fig. 2) (Claassens et al., 2019a), however synthetic pathways, such as the reductive glycine pathway, might be more interesting for biotechnological applications (Bar-Even, 2016; Bar-Even et al.,

2013). Apart from being a carbon source, formate is also a promising electron donor in bioindustry (Claassens et al., 2018). A challenge in the use of formate as microbial feedstock is its cellular toxicity, special cultivation mode and bioreactor design that might be necessary to address this problem (Cotton et al., 2020).

Overall, methanol and formate are the superior microbial feedstocks compared to the C1 gases, since they are completely miscible and easier to handle. The gaseous C1 substrates can be transformed to methanol or formate abiotically, CO₂ especially holds the interest. The biomanufacturing aims at the overall process feasible, economical, efficient, green, and sustainable.

Formaldehyde as a versatile intermediate

Formaldehyde, another C1 compound, is a metabolic hub in C1 metabolism (Yurimoto et al., 2005). Formaldehyde itself is not suitable as a feedstock, mainly because of its high toxicity. However, it can be a central intermediate between the C1 feedstocks and microbes, because (1) the C1 feedstocks can be converted to formaldehyde, as shown by the aerobic methane and methanol assimilation pathways which are all run via formaldehyde. (2) Formaldehyde is an important and ubiquitous cellular metabolite participating in different metabolic processes. Additionally, (3) formaldehyde has a similar reduction state as the average cellular carbon, indicated by the similar molecular formulas of formaldehyde (CH₂O) and the average cell mass (CH_{1.8}O_{0.5}N_{0.2}) (Villadsen et al., 2011). This suggests lower resource requirements for biomass incorporation of formaldehyde. Most importantly, (4) formaldehyde is highly reactive and ready to participate in biochemical reactions.

Formaldehyde can react with a diverse set of molecules, such as thiols, amines, pterins, and sugars (Fig. 3). The rapid reactions with free thiol and amine groups (Fig. 3a and b) on proteins or nucleic acids lead to cross-links resulting in cytotoxicity of formaldehyde (Chen et al., 2016). Hopkinson and coworkers (Kamps et al., 2019; Shishodia et al., 2018) systematically investigated reactions between formaldehyde and nucleobases and amino acids, revealing that the condensations occur at different rates and the products vary in stabilities. Among amino acids, cysteine which has both thiol and amine groups reacts most efficiently with formaldehyde forming stable thioproline (thiazolidine-4-carboxylate) (Kamps et al., 2019). Through the similar mechanism, thiol-containing glutathione, mycothiol and bacillithiol, or amine-containing tetrahydrofolate (THF) and tetrahydromethanopterin (H₄MPT) are used to capture formaldehyde for detoxification or assimilation (Marx et al., 2003; Schug, 2018). More

importantly, by combining with phosphosugars, formaldehyde can be fixed into central metabolism without losing electrons. For example, in the RuMP cycle formaldehyde is added to ribulose 5-phosphate by 3-hexulose-6-phosphate synthase (EC 4.1.2.43). Similarly in the XuMP cycle formaldehyde is transferred to xylulose 5-phosphate by dihydroxyacetone synthase forming dihydroxyacetone and glyceraldehyde 3-phosphate. They are represented by the reaction of Fig. 3c and d, respectively. The key enzymes activities depend on metal, i.e. Mg^{2+} or Mn^{2+} , or thiamine-diphosphate, respectively. The versatile reactivity of formaldehyde makes it interesting for biological applications (Desmons et al., 2019). Based on this, novel enzymes and pathways have been designed, for example the formolase (Siegel et al., 2015), the glycolaldehyde synthase (Lu et al., 2019), and an array of other pathways (Yang et al., 2019).

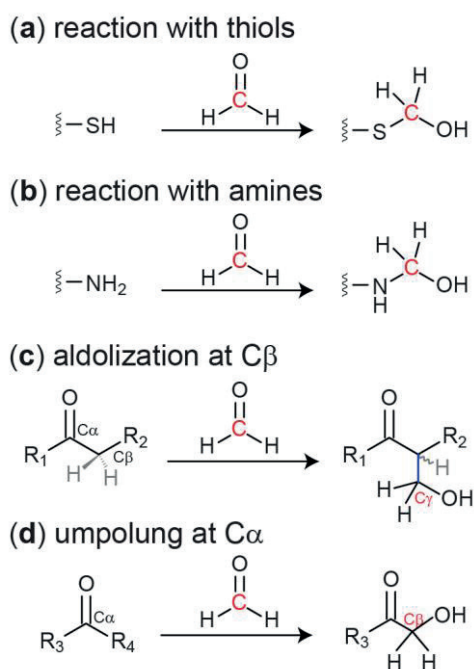


Figure 3. Biochemical reactions of formaldehyde. Formaldehyde can react with thiol groups (a) and amine groups (b), or work in aldol reactions (c) as well as thiamine diphosphate (ThDP) dependent umpolung (polarity inversion) reactions (d). Panel (c) and (d) are adapted from Desmons et al. (2019).

Formaldehyde assimilation pathways

As mentioned above, methylotrophs have evolved three different metabolic pathways to fix formaldehyde while using environmental C1 substrates as carbon and energy sources, i.e. the RuMP pathway and the serine cycle in prokaryotes, the XuMP pathway in yeast (Fig. 4).

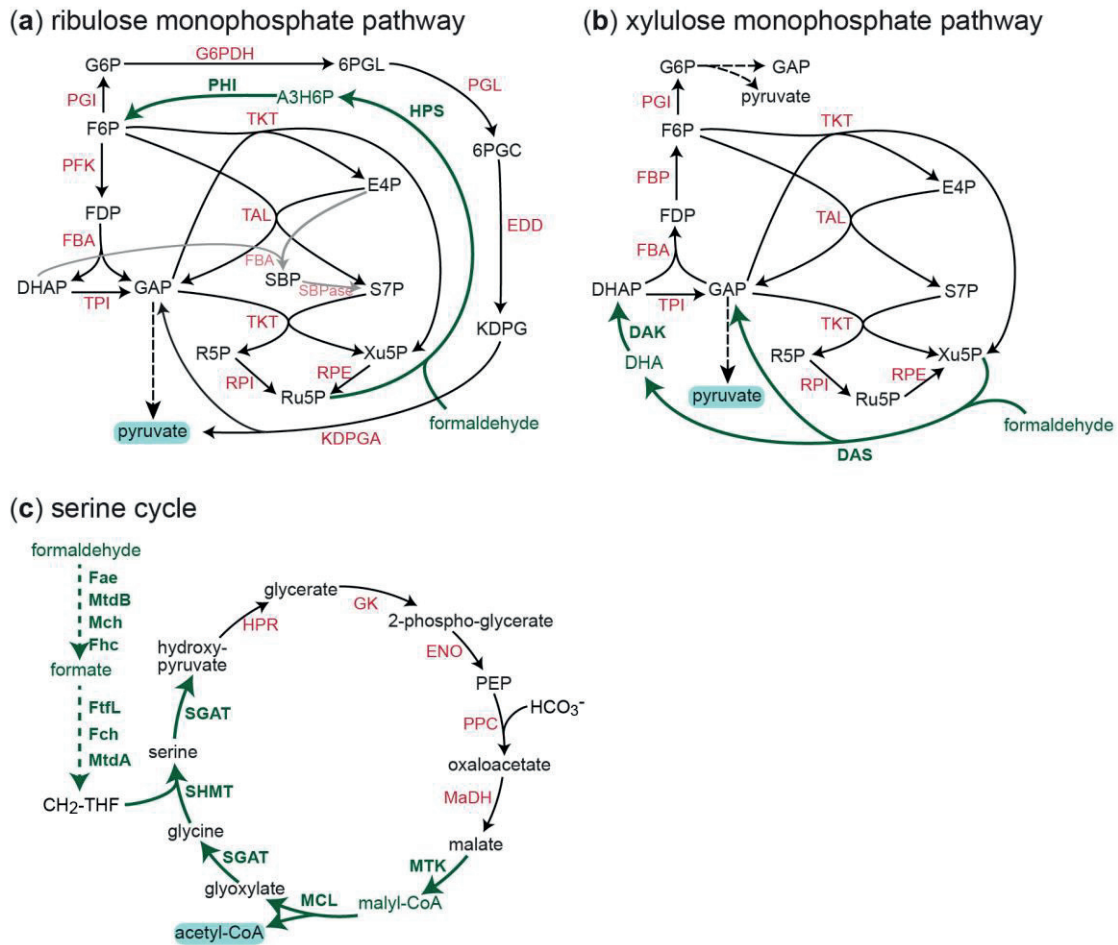


Figure 4. Natural formaldehyde assimilation pathways. The RuMP pathway (a) and the serine cycle (b) are used by bacterial and archaea, while the XuMP pathway (c) was found in yeast. Abbreviations of enzymes: DAK, dihydroxyacetone kinase; DAS, dihydroxyacetone synthase; ENO, enolase; Fae, formaldehyde-activating enzyme; FBA, fructose-bisphosphate aldolase; FBP, fructose bisphosphatase; Fch, methenyl-tetrahydrofolate (THF) cyclohydrolase; Fhc, formyltransferase/hydrolase; FtfL, formate-THF ligase; G6PDH, glucose 6-phosphate dehydrogenase; GK, glycerate kinase; HPR, hydroxypyruvate reductase; HPS, 3-hexulose-6-phosphate synthase; KDPGA, 2-keto-3-deoxygluconate 6-phosphate aldolase; MaDH, malate dehydrogenase; Mch, methyl-tetrahydromethanopterin cyclohydrolase; MCL, malyl-CoA lyase; MtdA, methylene-THF dehydrogenase; MtdB, methylene-tetrahydromethanopterin dehydrogenase; MTK, malate thiokinase; PFK, phosphofructokinase; PGI, glucose-6-phosphate isomerase; PGL, 6-phosphogluconolactonase; PHI, 6-phospho-3-hexuloisomerase; PPC, phosphoenolpyruvate carboxylase; RPE, ribulose 5-phosphate 3-epimerase; RPI, ribose 5-phosphate isomerase; SBPase, sedoheptulose bisphosphatase; SGAT, serine-glyoxylate aminotransferase; SHMT, serine hydroxymethyltransferase; TAL, transaldolase; TKT, transketolase; TPI, triose-phosphate isomerase. Abbreviations of metabolites: 6PGC, D-gluconate 6-phosphate; 6PGL, 6-phospho D-glucono-1,5-lactone; A3H6P, arabino-3-hexulose 6-phosphate; DHA, dihydroxyacetone; DHAP, glyceraldehyde 3-phosphate; E4P, D-erythrose 4-phosphate; F6P, D-fructose 6-phosphate; FDP, D-fructose 1,6-bisphosphate; G6P, D-glucose 6-phosphate; GAP, D-glyceraldehyde 3-phosphate; KDPG, 2-keto-3-deoxygluconate 6-phosphate; PEP, phosphoenolpyruvate; R5P, D-ribose 5-phosphate; Ru5P, D-ribulose 5-phosphate; S7P, D-sedoheptulose 7-phosphate; SBP, D-sedoheptulose 1,7-bisphosphate; Xu5P, D-xylulose 5-phosphate.

The RuMP pathway (Fig. 4a) and XuMP pathway (Fig. 4b) have a similar metabolic architecture: formaldehyde reacts with a pentose sugar, ribulose 5-phosphate (Ru5P) or xylulose 5-phosphate (Xu5P) respectively, following the reaction mechanisms of aldolization and umpolung activation respectively (Fig. 3c and d). The key enzymes responsible for these reactions are 3-hexulose-6-phosphate synthase (HPS) and 3-hexulose-6-phosphate isomerase (PHI) for the RuMP pathway, dihydroxyacetone synthase (DAS) and dihydroxyacetone kinase (DAK) for the XuMP pathway. The products of these formaldehyde fixation stages enter glycolysis and the pentose phosphate pathway, forming cellular growth precursors and regenerating the pentose substrates, Ru5P and Xu5P. In the serine cycle (Fig. 4c), formaldehyde first is combined with H₄MPT and oxidized to formate, which is then attached to THF forming methylene-THF. The carbon is subsequently transferred to glycine generating serine by serine hydroxymethyltransferase (SHMT). Several enzymatic steps of gluconeogenesis and the tricarboxylic acid (TCA) cycle convert serine to malate, involving a CO₂-fixing carboxylation step. Malate is then cleaved to glyoxylate for glycine regeneration and acetyl-CoA as biomass building block.

Instead of the natural formaldehyde assimilation pathways, which do not necessarily represent optimal solutions but just frozen accidents in the evolution, better pathways could be designed and implemented in order to better fit to the host organism and improve C1 utilization efficiency. Some reported examples are illustrated in Fig. 5. By combining non-oxidative glycolysis (Bogorad et al., 2013) with the RuMP pathway, Bogorad et al (2014) built a carbon-conserved and ATP-independent acetyl-CoA producing methanol condensation cycle (MCC, Fig. 5a), which runs in two different modes when using xylulose 5-phosphate phosphoketolase (XPK) or fructose 6-phosphate phosphoketolase (PFK). To heterologously implement the serine cycle in *E. coli*, Yu and Liao (2018) replaced the long H₄MPT-dependent formaldehyde oxidation with formaldehyde dehydrogenase (Faldh) and bypassed the formation of the highly reactive intermediate hydroxypyruvate via serine deaminase (SdaA) and phosphoenolpyruvate synthase (PPS, Fig. 5b). The third example, the reductive glycine pathway (Fig. 5c), integrated from natural occurring enzymes, converts two formaldehyde molecules and a CO₂ molecule to a pyruvate molecule with little energy and resources consumptions (Kim et al., 2020).

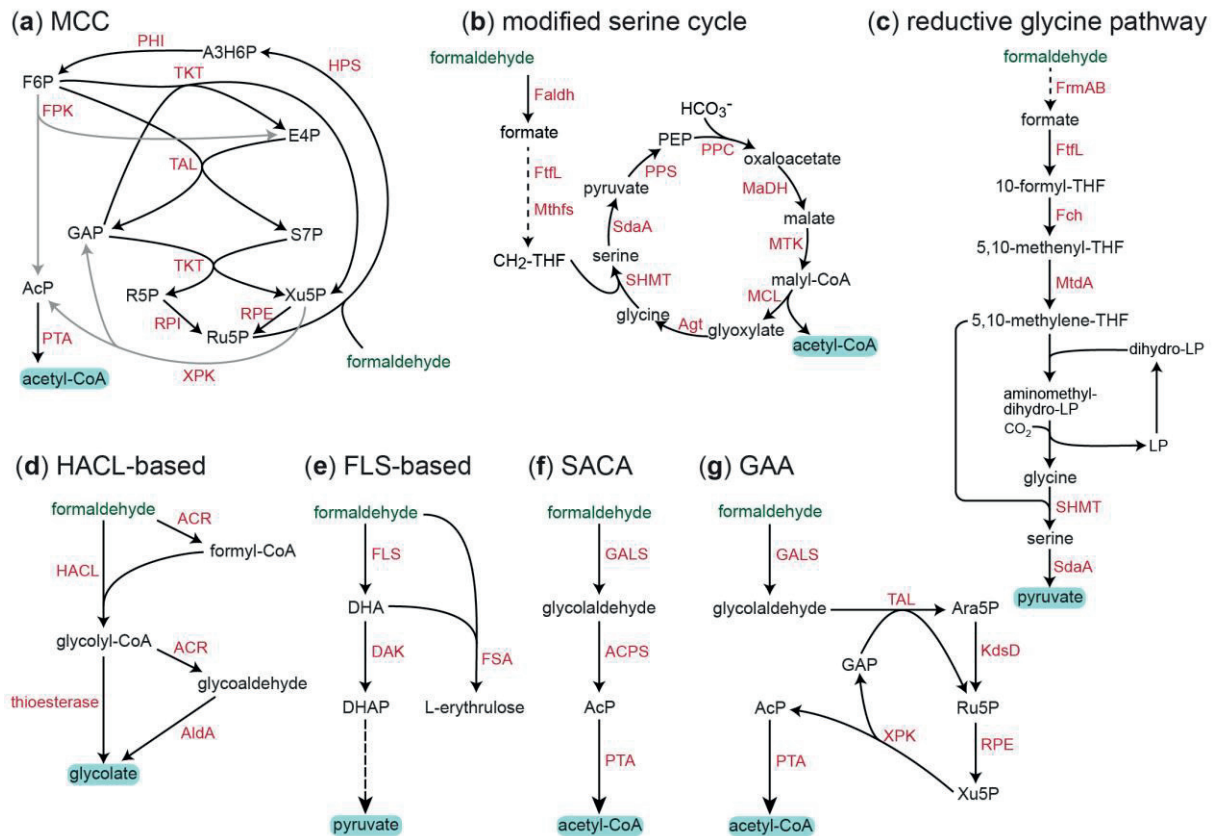


Figure 5. Reported synthetic formaldehyde pathways. Abbreviations of enzymes (red): ACPS, acetyl-phosphate synthase; ACR, acyl-CoA reductase; AldA, aldehyde dehydrogenase A; Agt, alanine-glyoxylate transaminase; DAK, dihydroxyacetone kinase; Faldh, formaldehyde dehydrogenase; Fch, methenyl-THF cyclohydrolase; FLS, formolase; FPK, fructose 6-phosphate phosphoketolase; FrmAB, glutathione-dependent formaldehyde dehydrogenase system; FSA, fructose 6-phosphate aldolase; FtlL, formate-THF ligase; GASL, glycolaldehyde synthase; HAACL, 2-hydroxyacyl-CoA lyase; HPS, 3-hexulose-6-phosphate synthase; KdsD, arabinose 5-phosphate isomerase; MaDH, malate dehydrogenase; MCL, malyl-CoA lyase; MtdA, methylene-THF dehydrogenase; Mthfs, methylene-THF synthase; MTK, malate thiokinase; PHI, 6-phospho-3-hexuloisomerase; PPC, phosphoenolpyruvate carboxylase; PPS, phosphoenolpyruvate synthase; PTA, phosphate acetyltransferase; RPE, ribulose 5-phosphate 3-epimerase; RPI, ribose 5-phosphate isomerase; SdaA, serine deaminase; SHMT, serine hydroxymethyltransferase; TAL, transaldolase; TKT, transketolase; XPK, xylulose 5-phosphate phosphoketolase. Abbreviations of metabolites: A3H6P, arabino-3-hexulose 6-phosphate; AcP, acetylphosphate; Ara5P, arabinose 5-phosphate; DHA, dihydroxyacetone; DHAP, glyceraldehyde 3-phosphate; E4P, D-erythrose 4-phosphate; F6P, D-fructose 6-phosphate; GAP, D-glyceraldehyde 3-phosphate; PEP, phosphoenolpyruvate; R5P, D-ribose 5-phosphate; Ru5P, D-ribulose 5-phosphate; S7P, D-sedoheptulose 7-phosphate; Xu5P, D-xylulose 5-phosphate.

Taking advantage of the versatile reactivity of formaldehyde, pathways based on new-to-nature enzymatic reactions were designed and tested. The 2-hydroxyacyl-CoA lyase (HAACL) was used to catalyze a previously unknown condensation reaction between formaldehyde and formyl-CoA (Fig. 5d) (Chou et al., 2019), supported by the ThDP dependent umpolung reaction (Fig.

3d). Its product glycolyl-CoA can be converted to a common metabolite glycolate. Employing both the umpolung and aldolization activities of formaldehyde, Siegel et al (2015) used the computationally designed formolase (FLS) enzyme, which condenses three formaldehyde molecules into one DHA (C3) molecule (Fig. 5e). Yang J.G. et al (2017) further cascaded the FLS reaction with fructose 6-phosphate aldolase and its mutant (FSA), using its side activity, converting four formaldehyde molecules to one L-erythrulose (C4) molecule (Fig. 5e). Similar to FLS, Lu et al (2019) engineered a glycolaldehyde synthase (GALS) enzyme, which produces glycolaldehyde from formaldehyde. Glycolaldehyde can be transformed into acetyl-CoA through different routes, i.e. the synthetic acetyl-CoA (SACA) pathway (Fig. 5f) or the glycolaldehyde assimilation (GAA) pathway (Fig. 5g) (Yang et al., 2019). Additionally, Yang et al (2019) predicted more possible aldolase reactions and pathways for formaldehyde assimilation using available databases.

Employing methylotrophs, which are naturally able to utilize C1 compounds as sole sources of carbon and energy via formaldehyde, seems the most straightforward way for bio-industry to use C1 feedstocks. Methylotrophs were used to produce biomass and single-cell proteins from C1 compounds in industrial scale (Ritala et al., 2017). Recently, the two model methylotrophic bacteria, *Methylobacterium extorquens* and *Bacillus methanolicus* MGA3, were engineered for methanol-based production of amino acids, PHB, butanol etc. (Brautaset et al., 2007; Ochsner et al., 2015). Although they grow fast on methanol and produce various chemicals, their industrial utilization is limited by multiple factors, including a narrow product spectrum, low yields, titers, and productivities, a restricted genetic toolbox for engineering, and gaps in our understanding of their cellular physiology and metabolism (Kalyuzhnaya et al., 2015; Whitaker et al., 2015). To overcome these difficulties, recent metabolic engineering efforts are aiming to introduce formaldehyde assimilation pathways, especially synthetic pathways that could theoretically support increased yields, into biotechnological relevant model microorganisms, such as *E. coli*, *Saccharomyces cerevisiae*, and *Corynebacterium glutamicum*, which are easier to engineer and which can be better optimized for industrially relevant conditions.

Modularity and selection metabolic engineering strategies

Engineering synthetic C1 metabolism requires the overexpression of pathway enzymes, especially those that are missing in the host or that are natively expressed at insufficient levels. However, simple overexpression is unlikely to be sufficient for realizing the activity of the entire pathway. This is mainly due to the overlap between the introduced pathway and the host's central metabolism, resulting in disrupted fluxes through both systems. To better identify and

resolve problematic metabolic interactions, it is helpful to divide the synthetic pathway into smaller metabolic modules, i.e. sub-pathways consisting of several reactions (Fig. 6). The *in vivo* implementation of these modules can be considerably easier than the full pathway and provide vital information on the metabolic context that enables or constrains the newly introduced activities. To probe the implementation of metabolic modules, it is useful to couple their activity with the growth of the host.

Coupling module activity with growth usually requires modifying the metabolic network of the host by performing strategic gene deletions. These are made to generate a strain auxotrophic for certain essential metabolic intermediates – for example, an amino acid – which can be exclusively synthesized via the synthetic module. As a result, cellular growth becomes dependent on the activity of the module. A range of selection strains can be designed with increasing selection pressure for pathway activity: a “minimal” selection is sustained if the module provides low required metabolites, higher selection pressure is obtained when module activity is responsible for the biosynthesis of multiple building blocks, and very high selection pressure is imposed when the biosynthesis of all or most biomass is dependent on the module.

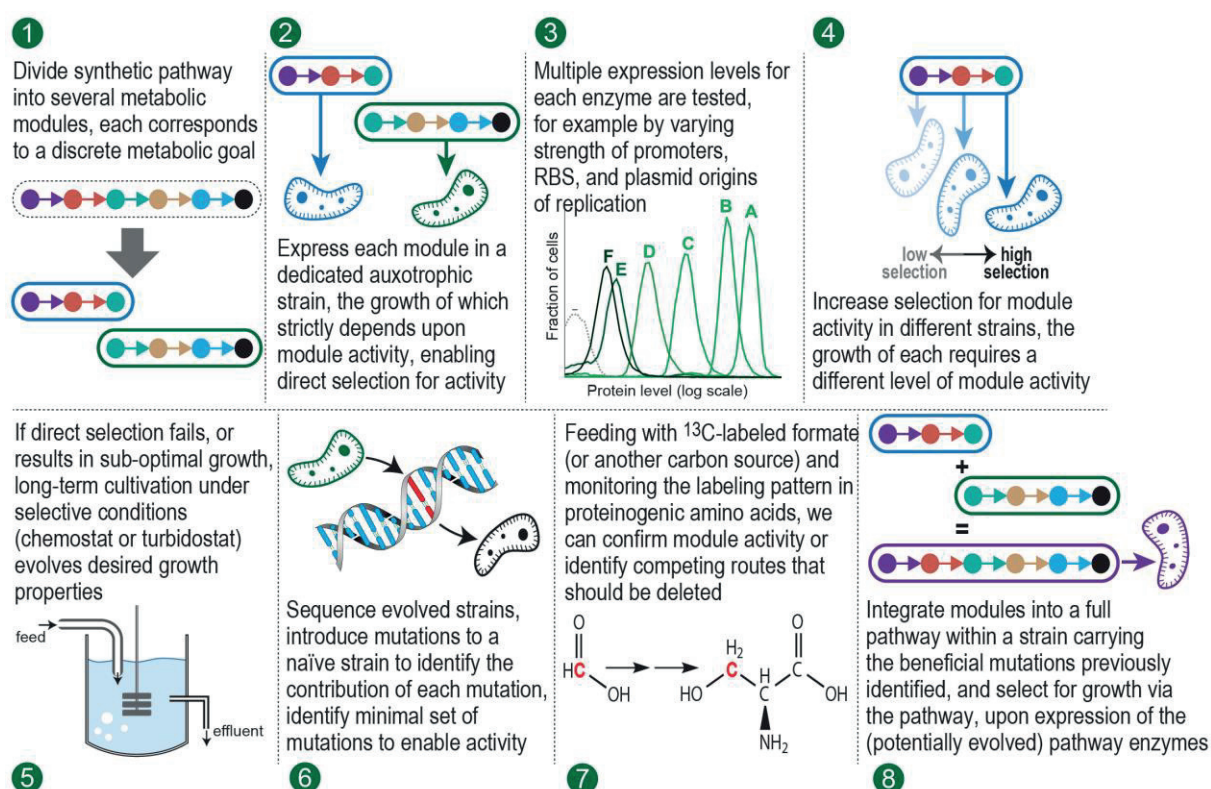


Figure 6. A schematic representation of the modular and selection approach. (Claassens et al., 2019b)

The design of modules and selection strains can be assisted by computational tools based on Flux Balance Analysis, for example OptKnock or FlexFlux (Burgard et al., 2003; Marmiesse et al., 2015; Meyer et al., 2018). Yet, in most cases, manual design based on biochemical and metabolic knowledge suffices. Specifically, when dividing a pathway into metabolic modules, several factors should be taken into consideration (Wenk et al., 2018): (1) the module should be linked to a clear growth phenotype within an appropriate selection strain, resulting in a growth versus no growth readout; (2) the number of enzymes in a module should be limited, to enable easy expression and optimization, and to allow straightforward interpretation of growth phenotypes; (3) modules should together cover the whole pathway and could overlap with one another, such that enzymes occurring in multiple modules can be tested in different metabolic contexts; (4) ideally, modules should be easy to combine, that is, dedicated selection strains – whose growth is dependent on the activity of several consecutive modules – should be easy to construct.

When direct selection for module activity fails or results in poor growth, further optimization is required (Fig. 6). This can be achieved by modulating the expression levels of the pathway enzymes (Wenk et al., 2018; Zelcbuch et al., 2013), and potentially also of related host enzymes, e.g. deletion or down-regulation of enzymes that divert metabolic intermediates from the pathway. In addition, different enzyme variants or codon optimization of the relevant genes can support increased expression and activity. Another method, which does not rely on genetic tools, is the addition of small molecules that specifically inhibit interfering enzymes, as demonstrated for the glycolytic glyceraldehyde 3-phosphate dehydrogenase in the engineering of methanol assimilation in *E. coli* (Woolston et al., 2018a). If these approaches fail to establish module activity, adaptive laboratory evolution (ALE) can be performed to increase module functionality and establish module-dependent growth (Nielsen and Keasling, 2016; Portnoy et al., 2011).

Outline of the thesis studies

Aiming at the goal of a future C1 bioeconomy, this thesis mainly focused on engineering formaldehyde pathways, both natural and newly designed, in the model organism *E. coli*. To establish synthetic methylotrophic growth of *E. coli* via formaldehyde, modularity and selection strategies were applied. The RuMP pathway, the most efficient naturally occurring pathway, was introduced in *E. coli*. Formaldehyde-dependent growth was achieved through the RuMP shunt, a linear variant of the pathway, without applying laboratory evolution. The engineered pathway provides the cell with almost all biomass and energy needs (Chapter 1). As

formaldehyde is highly reactive, novel reactions could be explored for novel pathway designs. Threonine aldolase and pyruvate aldolases, which are able to promiscuously catalyze aldol condensations of formaldehyde with glycine and pyruvate respectively, were used in Chapter 2. Facilitated by these two reactions, the newly designed homoserine cycle displays clear advantages when compared *in silico* to all of its natural counterparts. Employing *E. coli* native enzymes, this work demonstrated that the homoserine cycle can work in a physiological relevant manner.

Similarly using the selection approach, investigations were carried out to assess the *in vivo* spontaneous condensation rate of formaldehyde with THF. The results from this “build-to-learn” manner indicated that although the spontaneous condensation of formaldehyde and THF is feasible under physiological conditions in *E. coli*, it is unlikely to serve as an effective route for formaldehyde assimilation because of its low rate (Chapter 3). Nevertheless, identification of formaldehyde damage-control mechanisms would help to handle this toxic substrate. In Chapter 4, biochemical and genetic evidence was provided showing that cysteine thiazolidination, forming thioproline, contributes substantially to formaldehyde toxicity and Xaa-Pro aminopeptidase (encoded by *pepP*) is involved in hydrolysis of damaged thioproline-containing proteins. Overall, this thesis identified challenges and provided novel solutions to the community towards a future green and sustainable C1 bioeconomy.



Ribulose Monophosphate Shunt Provides Nearly All Biomass and Energy Required for Growth of *E. coli*

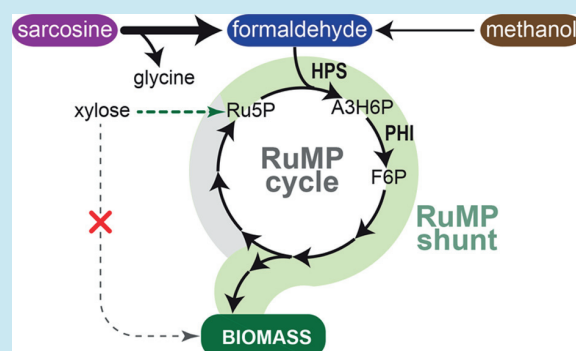
Hai He,¹ Christian Edlich-Muth, Steffen N. Lindner, and Arren Bar-Even*¹

Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany

Supporting Information

ABSTRACT: The ribulose monophosphate (RuMP) cycle is a highly efficient route for the assimilation of reduced one-carbon compounds. Despite considerable research, the RuMP cycle has not been fully implemented in model biotechnological organisms such as *Escherichia coli*, mainly since the heterologous establishment of the pathway requires addressing multiple challenges: sufficient formaldehyde production, efficient formaldehyde assimilation, and sufficient regeneration of the formaldehyde acceptor, ribulose 5-phosphate. Here, by efficiently producing formaldehyde from sarcosine oxidation and ribulose 5-phosphate from exogenous xylose, we set aside two of these concerns, allowing us to focus on the particular challenge of establishing efficient formaldehyde assimilation via the RuMP shunt, the linear variant of the RuMP cycle. We have generated deletion strains whose growth depends, to different extents, on the activity of the RuMP shunt, thus incrementally increasing the selection pressure for the activity of the synthetic pathway. Our final strain depends on the activity of the RuMP shunt for providing the cell with almost all biomass and energy needs, presenting an absolute coupling between growth and activity of key RuMP cycle components. This study shows the value of a stepwise problem solving approach when establishing a difficult but promising pathway, and is a strong basis for future engineering, selection, and evolution of model organisms for growth via the RuMP cycle.

KEYWORDS: ribulose monophosphate cycle, formaldehyde assimilation, methylotrophy, metabolic engineering, growth selection, carbon labeling, flux modeling



Formaldehyde is a metabolic hub in one-carbon metabolism. Methane, methanol, and formate—all promising feedstocks for sustainable bioproduction^{1–3}—can be converted to this key compound and subsequently assimilated into central metabolism.^{4,5} The most efficient route for formaldehyde assimilation, in terms of ATP consumption and biomass yield, is the ribulose monophosphate (RuMP) cycle.^{4,6,7} In this pathway (Figure 1a), ribulose 5-phosphate (Ru5P) reacts with formaldehyde to give arabino-3-hexulose-6-phosphate, catalyzed by 3-hexulose-6-phosphate synthase (HPS), which is isomerized to fructose 6-phosphate (F6P), catalyzed by 6-phospho-3-hexuloisomerase (PHI). F6P is then metabolized by the pentose phosphate cycle to regenerate ribulose 5-phosphate and provide biomass precursors (e.g., triose phosphate) for cellular growth.

The biotechnological utilization of microorganisms that naturally assimilate reduced C1 compounds via the RuMP cycle, for example, *Bacillus methanolicus*, presents numerous challenges, associated with their growth requirements, genetic manipulation and metabolic complexity.^{8–11} Hence, in the past decade, considerable research efforts have been directed toward implementing the RuMP cycle in biotechnological model organisms, for example, *Escherichia coli*, *Corynebacterium glutamicum*, and *Saccharomyces cerevisiae*, with the aim of

enabling these heterotrophs to grow on methanol as sole carbon and energy source.^{8,9,11–19} Previous studies have demonstrated the activity of all key enzymes, that is, methanol dehydrogenase (MDH), HPS, and PHI, in the desired hosts. Methanol assimilation has been further confirmed by tracing the ¹³C-labeling pattern of central metabolism intermediates upon feeding with labeled methanol. However, these previous studies did not couple the activity of the RuMP cycle to the growth of the host microorganism. Without this coupling, further adaptation of the host toward the use of the synthetic route is very challenging. Arguably, the main reason these studies fail to establish such coupling is that they tried to address significant challenges simultaneously: (i) thermodynamically and kinetically constrained NAD-dependent methanol oxidation to formaldehyde ($\Delta_r G^{m'}$ \sim +30 kJ/mol at 7 < pH < 8 and ionic strength of 0.25 mM,²⁰ V_{MAX} < 1 μ mol/min/mg^{21–23}); (ii) sustaining balanced activity of an autocatalytic cycle (a cycle whose product is also an intermediate) which necessitates a delicate tuning of the kinetic parameters of the relevant enzymes to avoid metabolite depletion and ensure sufficient regeneration of Ru5P.²⁴

Received: March 1, 2018

Published: May 14, 2018

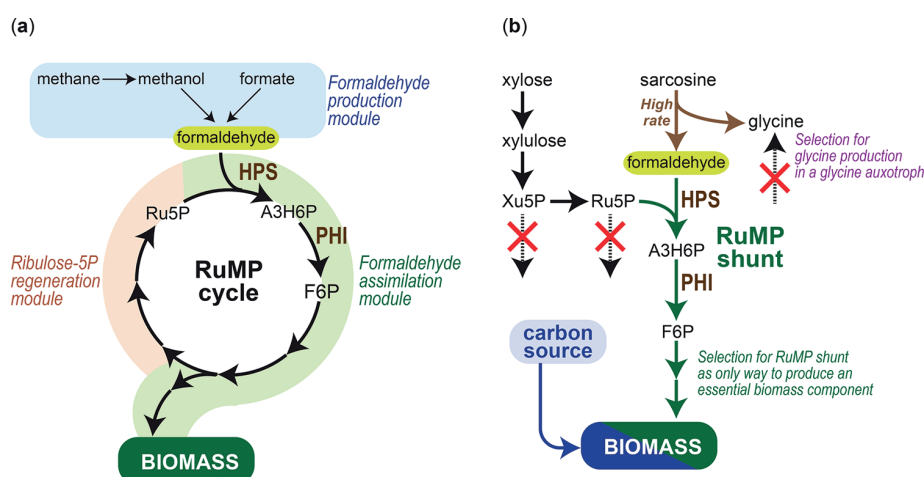


Figure 1. Schematics of the RuMP cycle and RuMP shunt. (a) RuMP cycle can be divided into three general segments: formaldehyde production, formaldehyde assimilation, and ribulose 5-phosphate (Ru5P) regeneration. (b) In this study, we bypass the challenges associated with formaldehyde production by using sarcosine oxidation as an effective formaldehyde source. We further bypass Ru5P regeneration by feeding with xylose. This enables us to focus on formaldehyde assimilation *via* the RuMP shunt. We generate multiple deletion strains to select for the activity of the RuMP shunt as a source of essential biomass precursors. Abbreviations: A3H6P, arabino-3-hexulose-6-phosphate; F6P, fructose 6-phosphate; Xu5P, xylulose 5-phosphate; HPS, 3-hexulose-6-phosphate synthase; and PHI, 6-phospho-3-hexuloisomerase.

Here, to set aside these challenges, we break the activity of RuMP cycle into three modules: formaldehyde production, formaldehyde assimilation, and Ru5P regeneration (Figure 1a). By using an alternative, effective source of formaldehyde, and providing xylose in the medium, we avoid the metabolic challenges associated with the first and the last module. This enables us to focus on the formaldehyde assimilation module and to demonstrate, for the first time, coupling between the growth of *E. coli* and the *in vivo* activity of HPS and PHI. Specifically, instead of slow and limiting methanol oxidation, or direct addition of highly reactive formaldehyde, we use sarcosine (*N*-methyl-glycine) oxidation as a thermodynamically and kinetically favorable source of formaldehyde. Instead of cyclic, autocatalytic activity, we pursue a RuMP shunt architecture: microbial growth dependent on the co-metabolism of xylose and formaldehyde *via* the activity of HPS and PHI (Figure 1b). We design and implement several pathway selection strategies, with increasing dependency of microbial growth on the activity of the RuMP shunt. Finally, we present an *E. coli* strain in which the biosynthesis of ~85% of biomass precursors and almost all energy requirements are dependent on the activity of the RuMP shunt.

RESULTS

Sarcosine As an Effective Source of Glycine and Formaldehyde. Sarcosine oxidase (SOX) from *Bacillus* sp. (strain B-0618) is a fast, irreversible enzyme, catalyzing the oxidative cleavage of sarcosine to glycine and formaldehyde with $k_{\text{cat}} > 100 \text{ s}^{-1}$,²⁵ more than 2 orders of magnitude higher than that of MDH. First, we tested whether SOX is active in *E. coli*. For this, we rationally designed and engineered a glycine auxotroph strain, deleted in serine hydroxymethyltransferase (ΔglyA), 2-amino-3-ketobutyrate CoA-ligase (Δkbl), threonine aldolase (ΔltaE), and isocitrate lyase (ΔaceA , to avoid production of glyoxylate which might be transaminated to glycine). This strain could grow only when glycine was added to the minimal medium. As shown in Figure 2, only when SOX is expressed, can sarcosine replace glycine in a concentration dependent manner, where addition of 5 mM sarcosine was

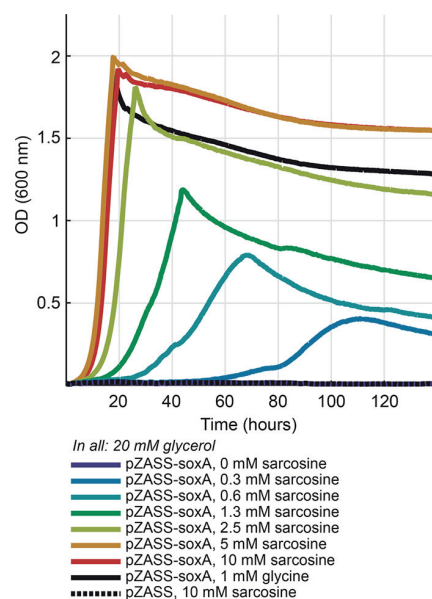


Figure 2. Sarcosine oxidation can efficiently generate glycine in a glycine auxotroph strain. A $\Delta\text{glyA} \Delta\text{kbl} \Delta\text{ltaE} \Delta\text{aceA}$ strain cannot grow without the addition of glycine (dashed line). Supplementation with glycine restores growth (black line). Only when sarcosine oxidase (SOX) is expressed, can sarcosine replace glycine in a concentration-dependent manner. Each growth curve represents the average of three replicates, which differ from each other by less than 5%.

almost indistinguishable from glycine supplementation. As the production of glycine is coupled to that of formaldehyde, this experiment indicates that sarcosine oxidation can serve to produce formaldehyde. We emphasize that within a $\Delta\text{glyA} \Delta\text{kbl} \Delta\text{ltaE} \Delta\text{aceA}$ background, glycine cannot be further assimilated into central metabolism. In fact, in such a strain, glycine can only be cleaved by the glycine cleavage system to provide the cell with C1-moieties.

We reasoned that if formaldehyde is to be used as a key metabolic intermediate, we should eliminate its major cellular

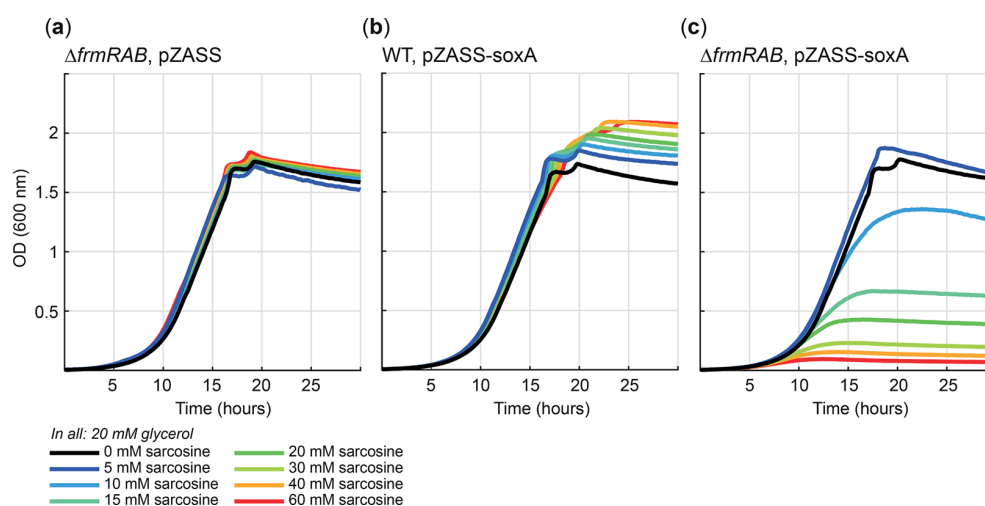


Figure 3. Sarcosine oxidation is toxic only when the glutathione-dependent formaldehyde detoxification system (encoded by *frmRAB*) is deleted. (a) $\Delta frmRAB$ strain with an empty plasmid. (b) WT strain expressing SOX. (c) $\Delta frmRAB$ strain expressing SOX. Each growth curve represents the average of three replicates, which differ from each other by less than 5%.

sink, *i.e.*, the glutathione-dependent formaldehyde detoxification system (*frmRAB*).^{11,26} However, the deletion of this system could result in high sensitivity towards formaldehyde. Indeed, we found that while a WT strain could grow in the presence of up to 500 μM of formaldehyde in the medium, the $\Delta frmRAB$ strain could only grow with up to 200 μM . These concentration boundaries are rather strict, making the direct addition of formaldehyde to the medium challenging and confirming that a formaldehyde-source, like sarcosine, is advantageous for a controlled *in situ* production of this reactive intermediate.

As shown in Figure 3a, a $\Delta frmRAB$ strain was insensitive to increased levels of sarcosine if SOX was not expressed. Increased sarcosine concentrations also did not inhibit the growth of a WT strain expressing SOX (Figure 3b), as the cell could efficiently detoxify formaldehyde *via* the glutathione-dependent formaldehyde detoxification system. However, as shown in Figure 3c, sarcosine was found to inhibit the growth of a $\Delta frmRAB$ strain expressing SOX in a concentration-dependent manner. This further confirms the activity of SOX as an efficient source of formaldehyde, which cannot be easily detoxified in this strain. We note that while SOX produces hydrogen peroxide, this byproduct does not seem to inhibit growth: sarcosine is not toxic, even at very high concentrations, to a WT strain overexpressing SOX (Figure 3b).

To better quantify formaldehyde production we used a NASH assay²⁷ (see methods). Figure 4 shows the time-dependent accumulation of formaldehyde within a WT and a $\Delta frmRAB$ strain upon expression of SOX (feeding with 10 mM sarcosine). Further shown is the accumulation of formaldehyde upon addition of methanol (1 M) and expression of three previously researched MDH variants from *Corynebacterium glutamicum* R (CgAdhA),²¹ *Bacillus stearothermophilus* (BsAdh2),¹⁶ and *Bacillus methanolicus* MGA3 (BmMdh3).²⁸ As expected, formaldehyde did not accumulate in the WT strain, regardless of the expressed enzyme and the formaldehyde source. The MDH-expressing $\Delta frmRAB$ strains accumulated formaldehyde to a level of 40–80 μM , similar to previously reported concentrations from methanol oxidation.^{11,16,29} Notably, formaldehyde accumulation upon feeding with sarcosine (and expressing SOX) was found to be an order

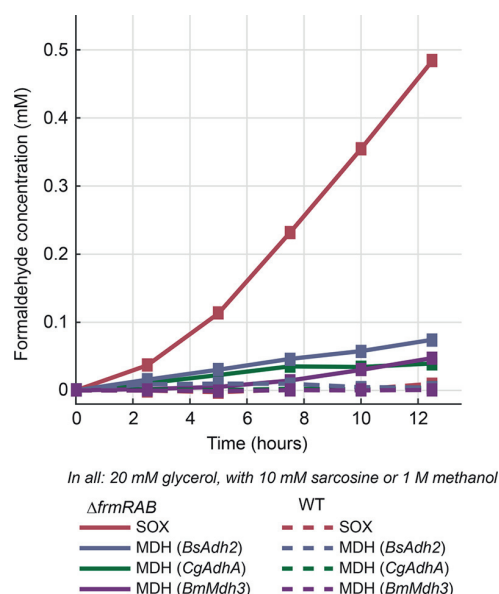


Figure 4. Sarcosine oxidase produces formaldehyde faster than methanol dehydrogenases. We used a NASH assay (Methods) to quantify the time-dependent concentration of formaldehyde following the addition of sarcosine or methanol to the cultivation medium of a WT or $\Delta frmRAB$ strain expressing SOX or a MDH variant. Each experiment was performed in triplicates, which were <10% different in values; shown are the average values at each time point.

of magnitude higher than with methanol, reaching a final concentration of $\sim 500 \mu\text{M}$. These results clearly show that sarcosine oxidation provides a highly efficient way for the *in vivo* production of formaldehyde.

Minimal Selection for the Activity of the RuMP Shunt.

After confirming efficient production of formaldehyde, we aimed to couple it to microbial growth. As an initial step, we constructed an *E. coli* deletion strain which is dependent on low activity of the RuMP shunt, where the biosynthesis of only a few biomass precursors rely on the synthetic route. As shown in Figure 5a, in the RPGF strain ribose 5-phosphate (R5P) isomerase ($\Delta rpiAB$) was deleted, forcing the assimilation of

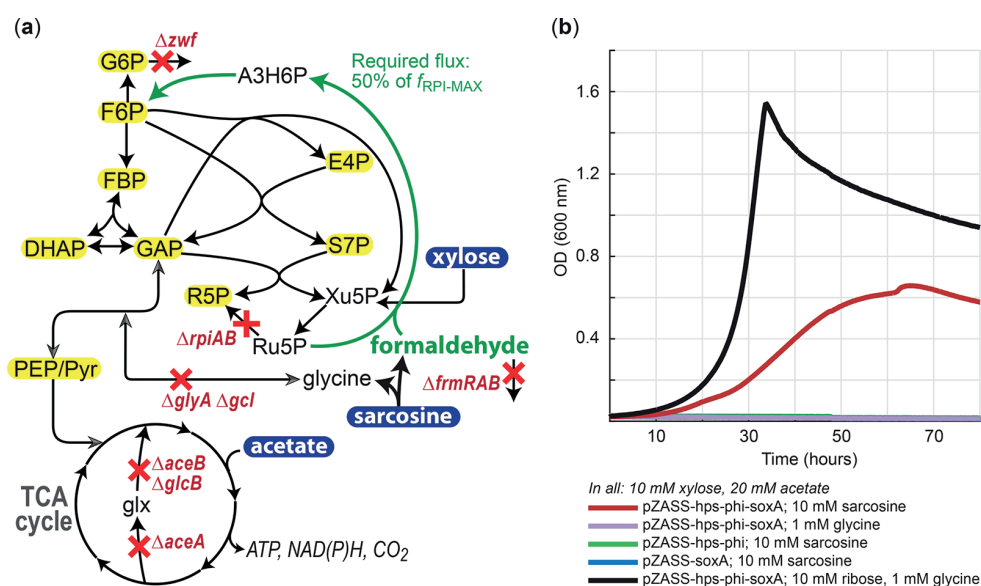


Figure 6. Medium selection pressure for the activity of the RuMP shunt. (a) A schematic representation of the RAGF strain. Exogenous carbon sources are marked with a blue background. Enzyme deletions are shown in red. Cellular metabolites whose biosynthesis depends on the activity of the RuMP shunt are shown in yellow. The flux *via* the shunt needs to be half of the flux required to provide the entire cellular biomass and energy requirement. (b) Sarcosine can replace ribose and glycine only when HPS, PHI, and SOX are expressed in the RAGF strain. Doubling times: black curve, 3.6 h; red curve, 6 h. Each growth curve represents the average of three replicates, which differ from each other by less than 5%. Abbreviations: G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; FBP, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; E4P, erythrose 4-phosphate; S7P, sedoheptulose 7-phosphate; Xu5P, xylulose 5-phosphate; Ru5P, ribulose 5-phosphate; R5P, ribose 5-phosphate; PEP, phosphoenolpyruvate; Pyr, pyruvate; and glx, glyoxylate.

the activity of the RuMP shunt. As the theoretical maximal flux required from the RuMP shunt in a $\Delta rpiAB$ background, $f_{RPI-MAX}$ serves as a useful reference point by which the different selection schemes can be compared. We find that in this low selection scheme, $f_{RPF}/f_{RPI-MAX} = 19\%$, that is, the flux *via* the shunt needs to be only 19% of the flux required if the entire cellular biomass and energy requirements would be dependent on the activity of the RuMP shunt (see [Methods](#)).

Medium Selection for the Activity of the RuMP Shunt.

Next, we aimed to increase the selection pressure for the activity of the RuMP shunt by establishing a strain whose growth is dependent to a higher extent on the activity of the synthetic route. This RAGF strain was deleted in R5P isomerases as well as in isocitrate lyase ($\Delta aceA$) and malate synthase ($\Delta aceB \Delta gclB$), thus deactivating the glyoxylate shunt ([Figure 6a](#)). This strain can oxidize acetate *via* the TCA cycle for the production of reducing power and energy, but cannot use acetate for biomass production (with the exception of the few biomass precursors that are derived from acetyl-CoA³¹). Other deletions within the RAGF strain are shown and explained in [Table 1](#). This strain is expected to be dependent on the activity of the RuMP shunt for the conversion of Ru5P to F6P from which almost all biomass precursors are derived ([Figure 6a](#)).

As shown in [Figure 6b](#), the RAGF strain could not grow on a minimal medium supplemented with xylose, acetate, and glycine. Further supplementation with ribose restored growth as R5P and Xu5P can be metabolized, *via* the pentose phosphate pathway, to provide all biomass precursors. Only when SOX, HPS, and PHI were expressed together, could sarcosine replace both ribose and glycine. In this RuMP shunt-dependent growth, Ru5P is combined with formaldehyde to generate F6P that supports the biosynthesis of almost all

cellular building blocks, whereas energy and reducing power are provided by acetate oxidation ([Figure 6a](#)). Using flux balance analysis we find that $f_{RAGF}/f_{RPI-MAX} = 50\%$, that is, the flux *via* the shunt needs to be half of the flux required to provide the entire cellular biomass and energy requirement.

Strong Selection for the Activity of the RuMP Shunt.

We then aimed to further increase the selection pressure on the activity of RuMP shunt by making it responsible for providing the cell with almost all biomass and energy. Toward this aim, we used the strain RGF in which R5P isomerases are deleted (as well as other genes, [Table 1](#)), as shown in [Figure 7a](#). For this strain, $f_{RGF}/f_{RPI-MAX} = 96\%$, whereas the remaining 4% originates from the glycine produced from sarcosine oxidation.

As shown in [Figure 7b](#), while the positive control—that is, RGF strain expressing SOX, HPS, and PHI and supplemented with xylose, ribose, and glycine—grew well, replacing ribose and glycine with 10 mM or 20 mM sarcosine resulted in rather poor growth. To test whether this growth is indeed fully dependent on the RuMP shunt, we fed the bacteria with ¹³C-labeled sarcosine, i.e., sarcosine-(methyl-¹³C), and monitored the labeling within four representative proteinogenic amino acids: (i) histidine, derived from R5P and thus representing “upper” metabolism; (ii) serine, derived from 3-phosphoglycerate (3PG) and thus representing “middle” metabolism; (iii) valine, derived from pyruvate and thus representing “middle-lower” metabolism (alanine was not used as it has the same retention time and molecular weight as sarcosine); and (iv) proline, derived from 2-ketoglutarate and thus representing “lower” metabolism.

As shown in [Figure 7c](#), the amino acids within a control strain ($\Delta zwf \Delta frmRAB$) fed with sarcosine-(methyl-¹³C) were not labeled. However, upon feeding with sarcosine-(methyl-¹³C), the amino acids of the RGF strain expressing SOX,

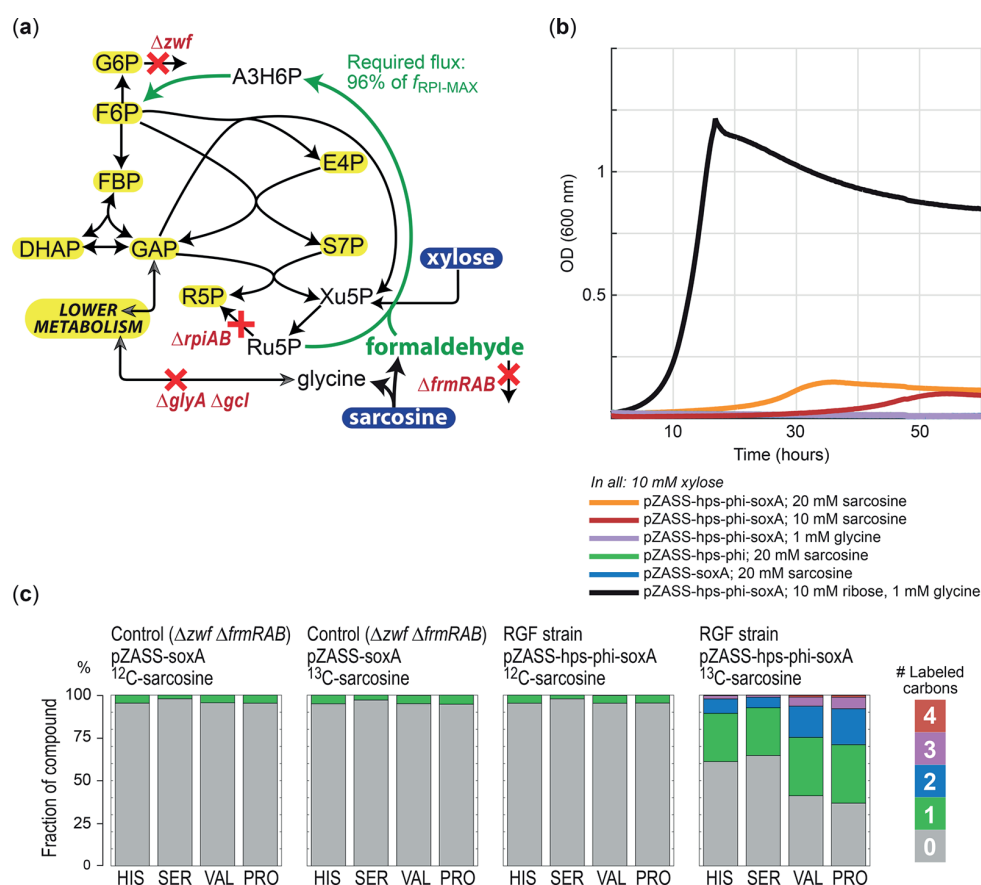
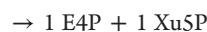
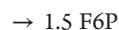
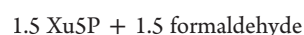


Figure 7. Strong selection pressure for the activity of the RuMP shunt. (a) A schematic representation of the RGF strain. Exogenous carbon sources are marked with a blue background. Enzyme deletions are shown in red. Cellular metabolites whose biosynthesis depends on the activity of the RuMP shunt are shown in yellow. The flux via the shunt needs to be 96% of the flux required to provide the entire cellular biomass and energy requirement. (b) Sarcosine can replace ribose and glycine only when HPS, PHI, and SOX are expressed in the RGF strain. Doubling times: black curve, 2.3 h; red curve, 8.7 h; orange curve, 6.1 h. Each growth curve represents the average of three replicates, which differ from each other by less than 5%. (c) Labeling pattern with proteinogenic histidine (HIS), serine (SER), valine (VAL), and proline (PRO), within different strains fed with labeled and unlabeled sarcosine. Abbreviations: G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; FBP, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; E4P, erythrose 4-phosphate; S7P, sedoheptulose 7-phosphate; Xu5P, xylulose 5-phosphate; Ru5P, ribulose 5-phosphate; and R5P, ribose 5-phosphate.

HPS, and PHI were labeled, confirming the activity of the RuMP shunt. We note that the labeling of serine and valine was slightly lower than expected: as the cleavage of $^{13}\text{C}_1$ -F6P results in half of the glyceraldehyde 3-phosphate (GAP) molecules being labeled, half of the 3PG and pyruvate molecules should be labeled. Consequently, half of the serine molecules (derived from 3PG) should be once labeled and the labeling of valine is expected to result in a mixture of double, single and unlabeled in a 1:2:1 ratio (as valine is synthesized from two pyruvate molecules). The observed, lower than expected labeling for these amino acids can be attributed to the activity of the pentose phosphate pathway: transaldolase produces unlabeled glyceraldehyde 3-phosphate, lowering the fraction of labeled 3PG and pyruvate.

Why is the RuMP shunt-dependent growth of the RGF strain so poor? One possibility is that the activity of SOX, HPS and PHI is not high enough to provide the cell with almost all biomass and energy needs. Alternatively, it could be that a strain carrying the *rpiAB* deletion suffers from inherent metabolic deficiencies that limit its RuMP shunt-dependent growth. Two observations strengthen the latter hypothesis. First, in the $\Delta rpiAB$ background, the metabolites E4P and

R5P—important biomass precursors providing 3% and 10% of the carbon atoms in biomass, respectively³¹—can be produced only via the activity of transketolase. Xu5P serves as a co-product in the biosynthesis of both. While the biosynthesis of E4P releases only one Xu5P molecule, the biosynthesis of R5P releases two molecules. Hence, the biosynthesis of E4P and R5P result in a cycle, where Xu5P is converted to F6P just to be metabolized to Xu5P again:



Since the flux toward E4P and R5P is quite high, the futile production of Xu5P is similarly substantial, considerably burdening the activity of the RuMP shunt. The flux balance analysis indicates that the futile production of Xu5P increases

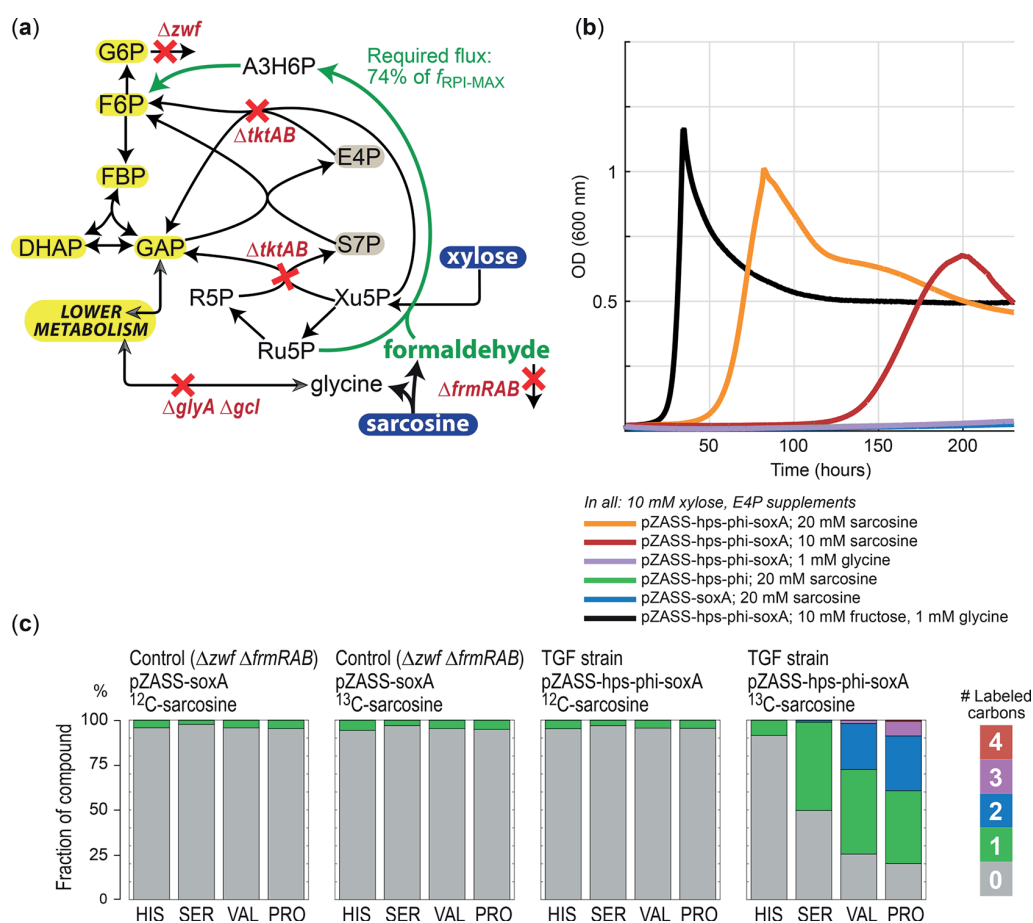


Figure 8. Strong selection pressure for the activity of the RuMP shunt within a transketolase-deleted background. (a) A schematic representation of the TGF strain. Exogenous carbon sources are marked with a blue background. Enzyme deletions are shown in red. Cellular metabolites whose biosynthesis depends on the activity of the RuMP shunt are shown in yellow. The flux *via* the shunt needs to be 74% the flux required to provide the entire cellular biomass and energy requirement with a $\Delta rpiAB$ background. (b) Sarcosine can replace ribose and glycine only when HPS, PHI, and SOX are expressed in the TGF strain. Doubling times: black curve, 2.4 h; red curve, 10.6 h; orange curve, 6.4 h. Each growth curve represents the average of three replicates, which differ from each other by less than 5%. (c) Labeling pattern with proteinogenic histidine (HIS), serine (SER), valine (VAL), and proline (PRO), within different strains fed with labeled and unlabeled sarcosine. Abbreviations: G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; FBP, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; E4P, erythrose 4-phosphate; S7P, sedoheptulose 7-phosphate; Xu5P, xylulose 5-phosphate; Ru5P, ribulose 5-phosphate; and R5P, ribose 5-phosphate.

the flux required *via* the RuMP shunt by more than 20%. This could result in a kinetic bottleneck that limits central metabolism flux.

Second, the expected high concentration of Xu5P, due to the assimilation of xylose, might thermodynamically and kinetically restrict the activity of the transketolase reactions in the direction of E4P and R5P biosynthesis. This could lead to E4P and R5P depletion and severe growth retardation.

Strong Selection for the Activity of the RuMP Shunt: Changing the Metabolic Context. To test whether the RuMP shunt could operate more favorably within a different metabolic context we constructed the TGF strain, shown in Figure 8a, in which the transketolase genes are deleted ($\Delta tktAB$) (see Table 1 for the full list of deleted genes). Similar to the deletion of R5P isomerases, the deletion of transketolase provides a single path for growth on xylose: the RuMP shunt (Figure 8a). However, unlike the $\Delta rpiAB$ strain, the $\Delta tktAB$ strain does not require the RuMP shunt for the biosynthesis of E4P and R5P. Instead, R5P is produced directly from Xu5P. E4P cannot be produced in the $\Delta tktAB$ strain,

requiring the supplementation of small amounts of essential cellular components whose biosynthesis is E4P-dependent: phenylalanine, tyrosine, tryptophan, shikimate, pyridoxine, 4-aminobenzoate, 4-hydroxybenzoate, and 2,3-dihydroxybenzoate (referred to as E4P-supplements, see Methods).

As shown in Figure 8b, the TGF strain expressing SOX, HPS, and PHI grew to a high density when provided with xylose and sarcosine (as well as E4P-supplements). This confirms that a change of metabolic context indeed supported better RuMP shunt-dependent growth. We note that in this strain, $f_{TGF}/f_{RPI-MAX} = 74%$, which is considerably lower than $f_{RGF}/f_{RPI-MAX} = 96%$. The lower flux required within the $\Delta tktAB$ background is attributed to the fact that the futile cycling of Xu5P is averted and the RuMP shunt is not required for the biosynthesis of R5P and E4P. Still, almost all biomass and energy needs of the TGF strain are provided by the activity of the synthetic shunt. Figure 8c shows the labeling of the representative proteinogenic amino-acids upon feeding with sarcosine-(methyl- ^{13}C). Unlike the reduced labeling observed in the RGF strain, the labeling pattern in the TGF strain is precisely as expected: half of the

serine molecules are labeled, and the labeling of valine follows the 1:2:1 pattern. This exact labeling is attributed to the deactivation of the pentose phosphate pathway such that F6P is converted into the glycolytic intermediates solely *via* glycolysis. Histidine is unlabeled as R5P is produced not by the activity of the RuMP shunt but rather by direct isomerization of Xu5P.

DISCUSSION

Engineering an active RuMP cycle in a heterotrophic host is an immense task, as it requires addressing multiple daunting challenges, including the fast production of formaldehyde, efficient assimilation of formaldehyde to central metabolism, and sufficient regeneration of the formaldehyde acceptor, Ru5P. This might explain why the heterologous establishment of the pathway has yet to be completed despite considerable research. In this study, we set aside two of these metabolic challenges, enabling us to focus our research on the efficiency with which formaldehyde is assimilated. By using multiple gene deletion strains we incrementally increased the selection pressure on the activity of the RuMP shunt, culminating in a strain in which almost all biomass precursors and energy are provided by the activity of this synthetic pathway.

Our study revealed that different metabolic contexts strongly affect RuMP shunt-dependent growth. Within the R5P isomerase deletion background, futile production of Xu5P increases the burden on the RuMP shunt. Furthermore, we speculate that this metabolic background presents an inherent thermodynamic and kinetic difficulty: producing the essential metabolites E4P and R5P in the presence of predicted high concentration of the co-product Xu5P. In another metabolic background, in which transketolase was deleted, these problems are averted, resulting in much improved growth.

This study paves the way for a series of sequential selection and evolution experiments, resulting in the establishment of a fully functional RuMP cycle. Importantly, each experiment will select for the activity of a different module of the RuMP cycle, starting with a strain evolved in the previous experiment. This will enable the dissection of the overall metabolic task into small, manageable steps.

The first step would be the replacement of sarcosine, as a source of formaldehyde, with parallel enzymatic systems to convert sustainable C1 feedstocks, such as methanol² or formate,³ into formaldehyde. We tried to replace SOX with methanol dehydrogenase¹¹ or a formate reduction module consisting of formyl-CoA synthetase and formaldehyde dehydrogenase,⁵ but were unable to select for either in a strain that requires the lowest formaldehyde assimilation flux (RPGF strain). This clearly indicates that formaldehyde production presents a severe thermodynamic and/or kinetic bottleneck, which could be addressed by adaptive laboratory evolution.³² For example, continuous cultivation of the TGF strain overexpressing both SOX and methanol dehydrogenase, on a medium containing limiting amounts of sarcosine but saturating amounts of methanol (and xylose), is expected to select for increased utilization of methanol, until sarcosine becomes dispensable (see for example ref 33).

The last step in establishing the fully functional RuMP cycle would be to select for a high rate of Ru5P regeneration. For this aim, the RGF strain might serve as a better starting point than the TGF strain, as it already requires a substantial cyclic flux that effectively partially regenerates Ru5P. This strain could be further evolved for high growth rate and yield, increasing and optimizing this desired flux. Following this adaptation, R5P

isomerase could be reintroduced and an adaptive laboratory evolution would serve to wean growth from dependence on external xylose.

We believe that the use of sequential selection steps, followed by long-term evolution when necessary, provides a highly promising approach to rewire the central metabolism of a heterotrophic organism to support growth on reduced C1 compounds. While our study solves one discrete problem in RuMP implementation—demonstrating a direct selection for the activity of HPS and PHI in a heterotroph—it remains for future studies to solve the two remaining challenges and evolve the relevant strains until a fully functional RuMP cycle is established.

METHODS

Strains and Genomic Modifications. All strains used in this study are listed in [Supplementary Table S1](#). An *E. coli* SIJ488 strain,³⁴ derived from wild-type strain MG1655, was used as the parental strain for genomic modifications and further experiments. Multiple-gene knockouts were obtained either by iterative rounds of λ -Red recombineering³⁴ or by P1 phage transduction.³⁵ For the recombineering, kanamycin resistance cassettes were generated *via* PCR (knockout primers with 50 bp homologous arms are listed in [Supplementary Table S2](#)) using the FRT-PGK-gb2-neo-FRT (Km) cassette (Gene Bridges, Germany). Fresh cultures of parental strains at OD₆₀₀ ~ 0.3 were induced by addition of 15 mM L-arabinose for 45 min at 37 °C. ~300 ng of cassette fragment was transformed using electroporation. After kanamycin selection, gene knockouts were confirmed by determining the size of a PCR product using “Ver” primers ([Supplementary Table S2](#)). For the *gcl* gene, which has a similar size to the Km cassette, K2 paired with the Reverse primer was used. Flippase was induced using 50 mM L-rhamnose for removing the Km cassette. It was verified that colonies can only grow on LB plates without antibiotic, and the Km cassette removal was further confirmed by PCR. For P1 transduction,³⁵ donor lysates were prepared using “donor” strains ([Supplementary Table S1](#)).

Synthetic-Operon Construction. All cloning procedures were carried out in *E. coli* DH5 α strain. The genes of 3-hexulose-6-phosphate synthase (*hps*, UniProt: I3DZR0), 6-phospho-3-hexuloisomerase (*phi*, UniProt: I3DZQ9) from *Bacillus methanolicus* MGA3, sarcosine oxidase from *Bacillus sp.* B-0618 (*soxA*, UniProt: P40859), and methanol dehydrogenases from *Corynebacterium glutamicum* R (CgAdhA, GenBank: BAF55711), *Bacillus stearothermophilus* (BsAdh2, UniProt: P42327) as well as *Bacillus methanolicus* MGA3 (*BmMdh3*, GenBank: EIJ80770), were *de novo* synthesized by Gen9 (Cambridge, MA, USA), after codon optimization using JCat.³⁶ Gene sequences are provided in [Supplementary Table S3](#). Genes were individually inserted into a pNivC vector downstream of ribosome binding site “C” (AAGTTAAGAG-GCAAGA) and with a HIS-tag directly downstream of the start codon. The genes were assembled into one operon using BioBrick enzymes: *BcuI*, *SallI*, *NheI* and *XhoI* (FastDigest, Thermo Scientific).³⁷ Using the restriction enzymes *EcoRI* and *PstI* (FastDigest, Thermo Scientific), the synthetic operon was cut and inserted into an expression pZASS vector, which has a p15A medium-copy origin of replication, a constitutive strong promoter pgi-20,³⁸ and a streptomycin selection marker.

Media and Growth Conditions. LB medium (1% NaCl, 0.5% yeast extract, 1% tryptone) was used for genomic strain modifications and recombinant gene cloning. Antibiotics were

used at the following concentrations: kanamycin, 50 $\mu\text{g}/\text{mL}$; ampicillin, 100 $\mu\text{g}/\text{mL}$; streptomycin, 100 $\mu\text{g}/\text{mL}$; chloramphenicol, 30 $\mu\text{g}/\text{mL}$. Growth experiments were performed in M9 minimal media (47.8 mM Na_2HPO_4 , 22 mM KH_2PO_4 , 8.6 mM NaCl, 18.7 mM NH_4Cl , 2 mM MgSO_4 and 100 μM CaCl_2), supplemented with trace elements (134 μM EDTA, 31 μM $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, 6.2 μM ZnCl_2 , 0.76 μM $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, 0.42 μM $\text{CoCl}_2 \cdot 2\text{H}_2\text{O}$, 1.62 μM H_3BO_3 , 0.081 μM $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$). Carbon sources were added according to the strain and the specific experiment as detailed in the text.

Strains were precultured in 4 mL M9 medium with the carbon sources that were used as positive controls in experiments and with 100 $\mu\text{g}/\text{mL}$ streptomycin. For the TGF strain, the medium was supplemented with E4P supplements:³⁹ 1 mM shikimic acid, 1 μM pyridoxine, 250 μM tyrosine, 500 μM phenylalanine, 200 μM tryptophan, 6 μM 4-aminobenzoic acid, 6 μM 4-hydroxybenzoic acid and 50 μM 2,3-dihydroxybenzoic acid. The precultures were harvested and washed three times in M9 medium, then inoculated in M9 media with suitable carbon sources, with a starting OD_{600} of 0.02 (selection strains RPGF, RAGF, RGF, and TGF) or 0.005 (all other strains). In 96-well microplates (Nunc Delta Surface, Thermo Scientific), 150 μL of culture were added to each well and covered with 50 μL mineral oil (Sigma-Aldrich) to avoid evaporation. (We note that mineral oil enables gas exchange and thus keeps the medium aerobic). The culture was incubated at 37 $^\circ\text{C}$ in a plate reader (Infinite M200 pro, Tecan). The shaking program cycle (controlled by Tecan I-control v1.11.1.0) has 4 shaking phases, lasting 60 s each: linear shaking followed by orbital shaking, both at an amplitude of 3 mm, then linear shaking followed by orbital shaking both at an amplitude of 1 mm. The absorbance (OD_{600}) in each well was monitored and recorded after every three shaking cycles (~ 12.5 min). Raw data from the plate reader were calibrated to normal cuvette measured OD_{600} values according to $\text{OD}_{\text{cuvette}} = \text{OD}_{\text{plate}} / 0.23$. Growth parameters were calculated using Matlab based on three technical triplicates; the average values were used to generate the growth curves. In all cases variability between triplicate measurements were less than 5%.

Growth curves were calculated using a sliding window methodology. First, we converted the linear values of the OD_{600} into logarithmic values. Then, using linear fitting (polyfit function) we calculated the growth rate for given windows of 5 h: starting from each time point (increments of 12.5 min), we defined a period of 5 h, and fitted the logarithmic OD_{600} values in this time window into linear function, the slope of which corresponds to the growth rate within this window. We choose the maximal slope calculated across all time windows to represent the growth rate of the strain.

Formaldehyde Concentration Assay. Cells at a starting OD_{600} of 0.02 were cultured within test tubes in 5 mL M9 medium with 20 mM glycerol and 10 mM sarcosine or 1 M methanol. At the indicated time points, culture were collected and centrifuged for 10 min at 3300g, 4 $^\circ\text{C}$. Formaldehyde concentration in culture supernatants was determined by Nash assay.²⁷ 125 μL of culture supernatant was mixed with 125 μL of Nash reagent (2 M ammonium acetate, 20 mM acetylacetone, and 50 mM acetic acid) in 96-well plate, and incubated at 37 $^\circ\text{C}$ for 1 h. The absorption at 412 nm was measured for each sample using plate reader (Infinite M200 pro, Tecan).

Isotopic-Labeling Experiments. For stationary isotope tracing of proteinogenic amino acids, strains were grown in M9

with 10 mM xylose and 20 mM either unlabeled sarcosine or methyl-labeled¹³ sarcosine (Sigma-Aldrich), E4P supplements were added for the TGF strains. Experiments were performed in duplicate. Cells were harvested at late exponential phase ($\text{OD}_{600} \sim 0.8$). Upon reaching stationary phase, the equivalent volume of 1 mL of culture at OD_{600} of 1 was harvested and washed by centrifugation. Protein biomass was hydrolyzed with 6 M HCl, at 95 $^\circ\text{C}$ for 24 h.⁴⁰ The samples were completely dried under a stream of air at 95 $^\circ\text{C}$.

Hydrolyzed amino acids were analyzed with UPLC–ESI–MS as previously described.⁴¹ Chromatography was performed with a Waters Acquity UPLC system (Waters), using an HSS T3 C_{18} reversed phase column (100 mm \times 2.1 mm, 1.8 μm ; Waters). 0.1% formic acid in H_2O (A) and 0.1% formic acid in acetonitrile (B) were the mobile phases. The flow rate was 0.4 mL/min and the gradient was: 0–1 min, 99% A; 1–5 min, linear gradient from 99% A to 82%; 5–6 min, linear gradient from 82% A to 1% A; 6–8 min, kept at 1% A; 8–8.5 min, linear gradient to 99% A; 8.5–11 min, re-equilibrate. Mass spectra were acquired using an Exactive mass spectrometer (Thermo Scientific) in positive ionization mode, with a scan range of 50.0 to 300.0 m/z . The spectra were recorded during the first 5 min of the LC gradients. Data analysis was performed using Xcalibur (Thermo Scientific). Determination of retention times was performed by analyzing amino-acid standards (Sigma-Aldrich) under the same conditions.

Flux Balance Analysis. Flux balance analysis was implemented in Matlab with the COBRA toolbox^{42,43} using the “standard” *E. coli* core model.⁴⁴ To allow glycine and C1 (*i.e.*, THF-CH_2)—which can be produced from sarcosine—to contribute to biomass directly we modified the biomass reaction to require glycine, serine and C1 instead of 3-phosphoglycerate. It was assumed that 70% of all 3-phosphoglycerate in the standard biomass reaction is converted to biomass *via* glycine and C1, and the remaining 30% *via* serine.³¹ The new biomass reaction consumes exactly the same number of carbons, hydrogens, oxygens, electrons, *etc.*, as the original one. Reactions for serine synthesis from 3-phosphoglycerate (SerABC), serine hydroxymethyltransferase (GlyA), and the glycine cleavage system (GCV) were added to the model, as well as an export reaction for excess glycine. **Supplementary Table S4** presents the reactions that were added to the model.

The value of the modified biomass reaction was fixed at a value of 1.0 (upper and lower bounds equal to 1) and the coefficient for the objective function was set to 0. Instead, the objective function coefficient for HPS was set to -1 . This has the effect that FBA finds the flux distribution resulting in a biomass flux of 1 at the lowest possible flux through HPS, representing the minimal flux *via* the synthetic shunt necessary to support growth. For comparison, the calculations were also carried out for the theoretical situation where all biomass is made from formaldehyde and xylose (but not from glycine) with ΔRPI background: $f_{\text{RPI-MAX}}$. In all simulations the carbon sources were not limiting (the relevant bounds were set to ± 1000) and the ATP “maintenance” reaction was switched off. All models had sarcosine and xylose as carbon sources and the ZWF and GlyA reactions were deleted from all models. In this background, four combinations of additional deletions and additional carbon sources were simulated: (i) model 1, RPGF strain, ΔRPI ΔFBP with succinate as an additional carbon source; (ii) model 2, RAGF strain, ΔRPI ΔAceA with acetate as an additional carbon source; (iii) model 3, RGF strain, ΔRPI ;

and (iv) model 4, TGF strain, Δ TKT with E4P as an additional carbon source.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00093.

Supplementary Tables S1–S4 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +49 331 567-8910. E-mail: bar-even@mpimp-golm.mpg.de.

ORCID

Hai He: 0000-0003-1223-2813

Arren Bar-Even: 0000-0002-1039-4328

Author Contributions

H.H., C.E.M., S.N.L., and A.B.E. designed the experiments; H.H. conducted the experiments; C.E.M. performed the computational analysis; and H.H., C.E.M., S.N.L., and A.B.E. wrote the paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Anne Michaelis for vital help with the metabolomic analysis. The authors thank Sophia Borowski for experimental assistance, and Charlie Cotton and Nico Claassens for critical review of the manuscript. This work was funded by the Max Planck Society. Hai He is funded by the China Scholarship Council.

■ REFERENCES

- (1) Strong, P. J., Xie, S., and Clarke, W. P. (2015) Methane as a resource: can the methanotrophs add value? *Environ. Sci. Technol.* *49*, 4001–18.
- (2) Schrader, J., Schilling, M., Holtmann, D., Sell, D., Filho, M. V., Marx, A., and Vorholt, J. A. (2009) Methanol-based industrial biotechnology: current status and future perspectives of methylotrophic bacteria. *Trends Biotechnol.* *27*, 107–15.
- (3) Yishai, O., Lindner, S. N., Gonzalez de la Cruz, J., Tenenboim, H., and Bar-Even, A. (2016) The formate bio-economy. *Curr. Opin. Chem. Biol.* *35*, 1–9.
- (4) Anthony, C. (1982) *The Biochemistry of Methylotrophs*, Academic Press, London.
- (5) 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*, 3704–9.
- (6) Goldberg, I., Rock, J. S., Ben-Bassat, A., and Mateles, R. I. (1976) Bacterial yields on methanol, methylamine, formaldehyde, and formate. *Biotechnol. Bioeng.* *18*, 1657–68.
- (7) 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, Bioenerg.* *1827*, 1039–47.
- (8) Whitaker, W. B., Sandoval, N. R., Bennett, R. K., Fast, A. G., and Papoutsakis, E. T. (2015) Synthetic methylotrophy: engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization. *Curr. Opin. Biotechnol.* *33*, 165–75.
- (9) Zhang, W., Zhang, T., Wu, S., Wu, M., Xin, F., Dong, W., Ma, J., Zhang, M., and Jiang, M. (2017) Guidance for engineering of synthetic

methylotrophy based on methanol metabolism in methylotrophy. *RSC Adv.* *7*, 4083–4091.

(10) Bennett, R. K., Steinberg, L. M., Chen, W., and Papoutsakis, E. T. (2017) Engineering the bioconversion of methane and methanol to fuels and chemicals in native and synthetic methylotrophs. *Curr. Opin. Biotechnol.* *50*, 81–93.

(11) Müller, J. E., Meyer, F., Litsanov, B., Kiefer, P., Potthoff, E., Heux, S., Quax, W. J., Wendisch, V. F., Brautaset, T., Portais, J. C., and Vorholt, J. A. (2015) Engineering *Escherichia coli* for methanol conversion. *Metab. Eng.* *28*, 190–201.

(12) Witthoff, S., Schmitz, K., Niedenfuhr, S., Noh, K., Noack, S., Bott, M., and Marienhagen, J. (2015) Metabolic engineering of *Corynebacterium glutamicum* for methanol metabolism. *Appl. Environ. Microbiol.* *81*, 2215–25.

(13) Lessmeier, L., Pfeifenschneider, J., Carnicer, M., Heux, S., Portais, J. C., and Wendisch, V. F. (2015) Production of carbon-13-labeled cadaverine by engineered *Corynebacterium glutamicum* using carbon-13-labeled methanol as co-substrate. *Appl. Microbiol. Biotechnol.* *99*, 10163–76.

(14) Price, J. V., Chen, L., Whitaker, W. B., Papoutsakis, E., and Chen, W. (2016) Scaffoldless engineered enzyme assembly for enhanced methanol utilization. *Proc. Natl. Acad. Sci. U. S. A.* *113*, 12691.

(15) Pfeifenschneider, J., Brautaset, T., and Wendisch, V. F. (2017) Methanol as carbon substrate in the bio-economy: Metabolic engineering of aerobic methylotrophic bacteria for production of value-added chemicals. *Biofuels, Bioprod. Biorefin.* *11*, 719.

(16) Whitaker, W. B., Jones, J. A., Bennett, R. K., Gonzalez, J. E., Vernacchio, V. R., Collins, S. M., Palmer, M. A., Schmidt, S., Antoniewicz, M. R., Koffas, M. A., and Papoutsakis, E. T. (2017) Engineering the biological conversion of methanol to specialty chemicals in *Escherichia coli*. *Metab. Eng.* *39*, 49–59.

(17) Dai, Z., Gu, H., Zhang, S., Xin, F., Zhang, W., Dong, W., Ma, J., Jia, H., and Jiang, M. (2017) Metabolic construction strategies for direct methanol utilization in *Saccharomyces cerevisiae*. *Bioresour. Technol.* *245*, 1407–1412.

(18) Bennett, R. K., Gonzalez, J. E., Whitaker, W. B., Antoniewicz, M. R., and Papoutsakis, E. T. (2017) Expression of heterologous non-oxidative pentose phosphate pathway from *Bacillus methanolicus* and phosphoglucose isomerase deletion improves methanol assimilation and metabolite production by a synthetic *Escherichia coli* methylotroph. *Metab. Eng.* *45*, 75.

(19) Gonzalez, J., Bennett, R. K., Papoutsakis, E. T., and Antoniewicz, M. R. (2017) Methanol assimilation in *Escherichia coli* is improved by co-utilization of threonine and deletion of leucine-responsive regulatory protein. *Metab. Eng.*, DOI: 10.1016/j.ymben.2017.11.015.

(20) Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012) eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Res.* *40*, D770–5.

(21) Kotrbova-Kozak, A., Kotrba, P., Inui, M., Sajdok, J., and Yukawa, H. (2007) Transcriptionally regulated *adhA* gene encodes alcohol dehydrogenase required for ethanol and n-propanol utilization in *Corynebacterium glutamicum* R. *Appl. Microbiol. Biotechnol.* *76*, 1347–56.

(22) Wu, T. Y., Chen, C. T., Liu, J. T., Bogorad, I. W., Damoiseaux, R., and Liao, J. C. (2016) Characterization and evolution of an activator-independent methanol dehydrogenase from *Cupriavidus necator* N-1. *Appl. Microbiol. Biotechnol.* *100*, 4969–83.

(23) Krog, A., Heggeset, T. M., Müller, J. E., Kupper, C. E., Schneider, O., Vorholt, J. A., Ellingsen, T. E., and Brautaset, T. (2013) Methylotrophic *Bacillus methanolicus* encodes two chromosomal and one plasmid born NAD⁺ dependent methanol dehydrogenase paralogs with different catalytic and biochemical properties. *PLoS One* *8*, e59188.

(24) Barenholz, U., Davidi, D., Reznik, E., Bar-On, Y., Antonovsky, N., Noor, E., and Milo, R. (2017) Design principles of autocatalytic cycles constrain enzyme kinetics and force low substrate saturation at flux branch points. *eLife*, DOI: 10.7554/eLife.20667.

- (25) Hassan-Abdallah, A., Zhao, G., Chen, Z. W., Mathews, F. S., and Schuman Jorns, M. (2008) Arginine 49 is a bifunctional residue important in catalysis and biosynthesis of monomeric sarcosine oxidase: a context-sensitive model for the electrostatic impact of arginine to lysine mutations. *Biochemistry* 47, 2913–22.
- (26) Gonzalez, C. F., Proudfoot, M., Brown, G., Korniyenko, Y., Mori, H., Savchenko, A. V., and Yakunin, A. F. (2006) Molecular basis of formaldehyde detoxification. Characterization of two S-formylglutathione hydrolases from *Escherichia coli*, FrmB and YeiG. *J. Biol. Chem.* 281, 14514–22.
- (27) Nash, T. (1953) The colorimetric estimation of formaldehyde by means of the Hantzsch reaction. *Biochem. J.* 55, 416–21.
- (28) Heggeset, T. M., Krog, A., Balzer, S., Wentzel, A., Ellingsen, T. E., and Bratset, T. (2012) Genome sequence of thermotolerant *Bacillus methanolicus*: features and regulation related to methylotrophy and production of L-lysine and L-glutamate from methanol. *Appl. Environ. Microbiol.* 78, 5170–81.
- (29) Woolston, B. M., Roth, T., Kohale, I., Liu, D. R., and Stephanopoulos, G. (2018) Development of a formaldehyde biosensor with application to synthetic methylotrophy. *Biotechnol. Bioeng.* 115, 206–2C.
- (30) Donahue, J. L., Bownas, J. L., Niehaus, W. G., and Larson, T. J. (2000) Purification and characterization of glpX-encoded fructose 1, 6-bisphosphatase, a new enzyme of the glycerol 3-phosphate regulon of *Escherichia coli*. *J. Bacteriol.* 182, 5624–7.
- (31) Neidhardt, F. C., Ingraham, J. L., and Schaechter, M. (1990) Building blocks needed to produce 1 g of *E. coli* protoplasm, In *Physiology of the Bacterial Cell: A Molecular Approach*, pp 134–143, Sinauer Associates Inc.
- (32) Dragosits, M., and Mattanovich, D. (2013) Adaptive laboratory evolution – principles and applications for biotechnology. *Microb. Cell Fact.* 12, 64.
- (33) Antonovsky, N., Gleizer, S., Noor, E., Zohar, Y., Herz, E., Barenholz, U., Zelcbuch, L., Amram, S., Wides, A., Tepper, N., Davidi, D., Bar-On, Y., Bareia, T., Wernick, D. G., Shani, I., Malitsky, S., Jona, G., Bar-Even, A., and Milo, R. (2016) Sugar Synthesis from CO₂ in *Escherichia coli*. *Cell* 166, 115–25.
- (34) Jensen, S. I., Lennen, R. M., Herrgard, M. J., and Nielsen, A. T. (2016) Seven gene deletions in seven days: Fast generation of *Escherichia coli* strains tolerant to acetate and osmotic stress. *Sci. Rep.* 5, 17874.
- (35) Thomason, L. C., Costantino, N., and Court, D. L. (2007) *E. coli* genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.*, 1.17.1.
- (36) Grote, A., Hiller, K., Scheer, M., Munch, R., Nortemann, B., Hempel, D. C., and Jahn, D. (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 33, W526–31.
- (37) Zelcbuch, L., Antonovsky, N., Bar-Even, A., Levin-Karp, A., Barenholz, U., Dayagi, M., Liebermeister, W., Flamholz, A., Noor, E., Amram, S., Brandis, A., Bareia, T., Yofe, I., Jubran, H., and Milo, R. (2013) Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.* 41, e98.
- (38) Braatsch, S., Helmark, S., Kranz, H., Koebmann, B., and Jensen, P. R. (2008) *Escherichia coli* strains with promoter libraries constructed by Red/ET recombination pave the way for transcriptional fine-tuning. *BioTechniques* 45, 335–7.
- (39) Zhao, G., and Winkler, M. E. (1994) An *Escherichia coli* K-12 tktA tktB mutant deficient in transketolase activity requires pyridoxine (vitamin B6) as well as the aromatic amino acids and vitamins for growth. *J. Bacteriol.* 176, 6134–8.
- (40) You, L., Page, L., Feng, X., Berla, B., Pakrasi, H. B., and Tang, Y. J. (2012) Metabolic pathway confirmation and discovery through (13)C-labeling of proteinogenic amino acids, Journal of visualized experiments. *J. Visualized Exp.*, DOI: 10.3791/3583.
- (41) Giavalisco, P., Li, Y., Matthes, A., Eckhardt, A., Hubberten, H. M., Hesse, H., Segu, S., Hummel, J., Kohl, K., and Willmitzer, L. (2011) Elemental formula annotation of polar and lipophilic metabolites using (13) C, (15) N and (34) S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J.* 68, 364–76.
- (42) Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmanian, S., Kang, J., Hyduke, D. R., and Palsson, B. O. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* 6, 1290–307.
- (43) Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. O., and Herrgard, M. J. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protoc.* 2, 727–38.
- (44) Orth, J. D., Fleming, R. M., and Palsson, B. O. (2010) Reconstruction and Use of Microbial Metabolic Networks: the Core *Escherichia coli* Metabolic Model as an Educational Guide. *EcoSal Plus*, DOI: 10.1128/ecosalplus.10.2.1.

Supplementary Table 1

List of *E. coli* strains used and constructed in this study.

Strain	Genotype	Reference or source
MG1655	K-12 F ⁻ λ ⁻ <i>ilvG</i> ⁻ <i>rfb</i> -50 <i>rph</i> -1	Lab collection
SIJ488	MG1655 Tn7::para-exo-beta-gam; prha-FLP; xylSpm-lscel	Jensen <i>et al.</i> , Sci. Rep. 5, 17874 (2015).
DH5α	F ⁻ <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20</i> φ80dlacZΔM15 Δ(<i>lacZYA-argF</i>)U169, <i>hsdR17</i> (r _K ⁺), λ ⁻	Lab collection
glyAux	MG1655 Δ <i>glyA</i> Δ <i>kbl</i> Δ <i>ltaE</i> Δ <i>aceA</i>	Lab collection
	SIJ488 Δ <i>frmRAB</i>	This study
	SIJ488 Δ <i>zwf</i> Δ <i>frmRAB</i>	This study
Donor_1	SIJ488 Δ <i>zwf</i> Δ <i>frmRAB</i> Δ <i>rpiA</i> :: <i>Km</i>	This study
Donor_2	SIJ488 Δ <i>zwf</i> Δ <i>frmRAB</i> Δ <i>rpiB</i> :: <i>Km</i>	This study
Donor_3	SIJ488 Δ <i>zwf</i> Δ <i>frmRAB</i> Δ <i>tktA</i> :: <i>Km</i>	This study
Donor_4	SIJ488 Δ <i>zwf</i> Δ <i>frmRAB</i> Δ <i>tktB</i> :: <i>Km</i>	This study
Donor_5	SIJ488 Δ <i>zwf</i> Δ <i>frmRAB</i> Δ <i>rpiB</i> Δ <i>glpX</i> :: <i>Km</i>	This study
Donor_6	SIJ488 Δ <i>zwf</i> Δ <i>frmRAB</i> Δ <i>rpiB</i> Δ <i>gcl</i> :: <i>Km</i>	This study
Donor_7	SIJ488 Δ <i>glyA</i> :: <i>Km</i>	Lab collection
Donor_8	MG1655 Δ <i>fbp</i> :: <i>Km</i>	Lab collection
Donor_9	SIJ488 Δ <i>aceBA</i> :: <i>Km</i>	Lab collection
Donor_10	MG1655 Δ <i>ltaE</i> Δ <i>kbl</i> Δ <i>aceA</i> Δ <i>glcB</i> :: <i>Km</i>	Lab collection
RGF	SIJ488 Δ <i>zwf</i> Δ <i>frmRAB</i> Δ <i>rpiB</i> Δ <i>gcl</i> Δ <i>glyA</i> Δ <i>rpiA</i> :: <i>Km</i>	This study
RPGF	SIJ488 Δ <i>zwf</i> Δ <i>frmRAB</i> Δ <i>rpiB</i> Δ <i>gcl</i> Δ <i>glyA</i> Δ <i>glpX</i> Δ <i>fbp</i> Δ <i>rpiA</i> :: <i>Km</i>	This study
RAGF	SIJ488 Δ <i>zwf</i> Δ <i>frmRAB</i> Δ <i>rpiB</i> Δ <i>gcl</i> Δ <i>glyA</i> Δ <i>glcB</i> Δ <i>aceBA</i> Δ <i>rpiA</i> :: <i>Km</i>	This study
TGF	SIJ488 Δ <i>zwf</i> Δ <i>frmRAB</i> Δ <i>tktB</i> Δ <i>gcl</i> Δ <i>glyA</i> Δ <i>tktA</i> :: <i>Km</i>	This study

Supplementary Table 2

List of oligo primers used in this study.

Primer	Sequence (5'→3')
frmRAB_KO_F	tcacgccgcatccggcagtcgtgctattatcaacgcatattcagttattaattaaccctcactaaagggcg
frmRAB_KO_R	tagaataccccctatagtatattgcatgcagatgatgaggtgcgaaatgtaatacgactcactatagggctc
frmRAB_Ver_F	ggatgaagacgacccgtagtc
frmRAB_Ver_R	attctgattccttctgccc
zwf_KO_F	aaaataaccataaaggataagcgcagatattactcaaactcattccaggaaattaaccctcactaaagggcg
zwf_KO_R	gtataccctggcttaagtagccgggttagtaactaaggagaatgacatgtaatacgactcactatagggctc
zwf_Ver_F	gcacgaggcctgaaagtga
zwf_Ver_R	aaagcagtagcagtcaccgt
rpiB_KO_F	cacaaatgtttttgattgtgaagtttgcacggacggggaagatgaatgattaaccctcactaaagggcg
rpiB_KO_R	tactcatccatgcaagtagtgatgaatctcaatttctccgctgctctattaatacgactcactatagggctc
rpiB_Ver_F	atgttgctggatggtgagtg
rpiB_Ver_R	atagtctcgcccacggtag
rpiA_KO_F	ttttaacgggggaggttccccgtagatcattcacaatggtttgacattaaccctcactaaagggcg
rpiA_KO_R	ttttgatataatgctgtgaaatttcataccacaggcgaacgatcatgtaatacgactcactatagggctc
rpiA_Ver_F	cgcgtagttgagaacggtag
rpiA_Ver_R	cgccctggtgtgtgaatg
gcl_KO_F	aatttgaaagtggaaaaattttccaataaatagaggttaggaataaaaatgattaaccctcactaaagggcg
gcl_KO_R	gcagagaaacgtaacattttatctcccttattcatagtgcatgaagcataatacgactcactatagggctc
gcl_Ver_F	ccgggcttcatcaagactg
gcl_Ver_R	cgccggtaaatgtgcagcg
glyA_KO_F	agcacattgacagcaaatcaccgttctgcttagcgtaaaccgggtaacgaattaaccctcactaaagggcg
glyA_KO_R	cctataaaggccaaaaattttatgttagctgagtcaggagatgaggatgtaatacgactcactatagggctc
glyA_Ver_F	cgcgataacgtagaaaggct
glyA_Ver_R	gtagaaatggcggttaact
glpX_KO_F	ggactggaaggctcaatcgatcaaatcaatcagaggatgtgcacctgcatattaaccctcactaaagggcg
glpX_KO_R	cttcgcgccattccttactgcttagagttgctatgagacgagaacttgctaatacgactcactatagggctc
glpX_Ver_F	gaaaccactgagcgaatta
glpX_Ver_R	gttcgccgtagattcaagg
fbp_KO_F	aaaaacgcctctccgtgtggagaggcgcaggagattacgctccgggaacattaaccctcactaaagggcg
fbp_KO_R	ctttactccataaacattgcagggaagttttatgaaaacgtaggtgaataatacgactcactatagggctc
fbp_Ver_F	agtggtaattggcgtacgc
fbp_Ver_R	agtgctcatcggggatac
tktB_KO_F	ggaagatcttctgcccactataaaccagccacggagtggttatgaattaaccctcactaaagggcg
tktB_KO_R	atccggcaatcagcatccggcaatcaccatcaggcaccttctactcccagtaatacgactcactatagggctc
tktB_Ver_F	ccacctctcagacgtccc
tktB_Ver_R	gcttgacggtagcgtttt
tktA_KO_F	ccgaagcaccttttaccgaaatgctaattacagcagttctttgctttaattaaccctcactaaagggcg
tktA_KO_R	gtcaagtcgtaagggcgtgccctcatcatccgatctggagtaaaaatgtaatacgactcactatagggctc
tktA_Ver_F	gcaattgagcagggaaaccg

tktA_Ver_R	gctcctgatgagagcagtcc
glcB_KO_F	taaaactgccggggctttttgacgctattaatgactttcttttcgCGaattaaccctcactaaagggcg
glcB_KO_R	cggcagcagcgggtgtggCGaaataagCGaaaacgaggagataaacaatgtaatacgactcactatagggctc
glcB_Ver_F	ctcctccagcaacgcacca
glcB_Ver_R	ggcgtaccggtgtgtgga
aceBA_Ver_F	gatagtcgatcgtaagcgattcagcacc
aceBA_Ver_R	gctctccgcaatctcgaagc
pZ_seqS-F	gcattatcagggttattgtctcatg
pZ_seq-R	ctagggcggcggatttgcctac
pNiv_seq-F	cctgtagtagtacttaagctcg
pNiv_seq-R	gctggccttttgctcacatggt
K1	cagtcatagccgaatagcct
K2	cggtgccctgaatgaactgc

Supplementary Table 3

Codon optimized gene sequences used in this study. 6×His tags are underlined.

Gene	Sequence (5' to 3')
<i>Bacillus methanolicus</i> MGA3 <i>hps</i>	<p>ATGCATCATCACCATCACCACGAACTGCAACTGGCTCTGGACCTGGTTAA CATCGAAGAAGCTAAACAGGTTGTTGCTGAAGTTCAGGAATACGTTGACA TCGTTGAAATCGGTACCCCGGTTATCAAAATCTGGGGTCTGCAAGCTGTT AAAGCTGTTAAAGACGCTTCCCGCATCTGCAAGTCTGGCTGACATGAA AACCATGGACGCTGCTGCTTACGAAGTTGCTAAAGCTGCTGAACACGGTG CTGACATCGTTACCATCCTGGCTGCTGCTGAAGACGTTTCTATCAAAGGTG CTGTTGAAGAAGCTAAAAAACTGGGTAAAAAAATCCTGGTTGACATGATC GCTGTTAAAAACCTGGAAGAACGTGCTAAACAGGTTGACGAAATGGGTG TTGACTACATCTGCGTTCACGCTGGTTACGACCTGCAAGCTGTTGGTAAA AACCCGCTGGACGACCTGAAACGTATCAAAGCTGTTGTTAAAAACGCTAA AACCGCTATCGCTGGTGGTATCAAACCTGGAAACCCTGCCGGAAGTTATCA AAGCTGAACCGGACCTGGTTATCGTTGGTGGTGGTATCGCTAACAGACC GACAAAAAAGCTGCTGCTGAAAAAATCAACAACTGGTTAACAGGGTC TGTA</p>
<i>Bacillus methanolicus</i> MGA3 <i>phi</i>	<p>ATGCATCATCACCATCACCACATCTCTATGCTGACCACCGAATTTCTGGCT GAAATCGTTAAAGAAGCTGAACTCTTCTGTTAACCAGATCGCTGACGAAGA AGCTGAAGCTCTGGTTAACGGTATCCTGCAATCTAAAAAAGTTTTCTGTTG CTGGTGCTGGTCGTTCTGGTTTCATGGCTAAATCTTTCGCTATGCGTATGA TGCACATGGGTATCGACGCTTACGTTGTTGGTGAACCGTTACCCCGAAC TACGAAAAAGAAGACATCCTGATCATCGGTTCTGGTTCTGGTGAACCAA ATCTCTGGTTTCTATGGCTCAGAAAGCTAAATCTATCGGTGGTACCATCGC TGCTGTTACCATCAACCCGGAATCTACCATCGGTCAGCTGGCTGACATCG TTATCAAAATGCCGGGTTCTCCGAAAGACAAATCTGAAGCTCGTGAACCC ATCCAGCCGATGGGTTCTCTGTTTGAACAGACCCTGCTGCTGTTCTACGAC GCTGTTATCCTGCGTTTTCATGGAAAAAAAAGGTCTGGACACCAAAACCAT GTACGGTCGTCACGCTAACCTGGAATAA</p>
<i>Bacillus sp.</i> B-0618 <i>soxA</i>	<p>ATGCATCATCACCATCACCACCTCTACCCACTTCGACGTTATCGTTGTTGGT GCTGGTTCTATGGGTATGGCTGCTGGTTACCAGCTGGCTAAACAGGGTGT TAAACCCCTGCTGGTTGACGCTTTCGACCCGCCGCACACCAACGGTTCTC ACCACGGTGACACCCGTATCATCCGTCACGCTTACGGTGAAGGTCGTGAA TACGTTCCGCTGGCTCTGCGTTCAGGAACTGTGGTACGAACTGGAAAA AGAAACCCACCACAAAATCTTCACCAAAACCGGTGTTCTGGTTTTCCGGTC CGAAAGGTGAATCTGCTTTCGTTGCTGAAACCATGGAAGCTGCTAAAGAA CACTCTCTGACCGTTGACCTGCTGGAAGGTGACGAAATCAACAAACGTTG GCCGGGTATCACCGTTCGGA AAAACTACAACGCTATCTTCGAACCGAACT CTGGTGTCTGTTCTCTGAAAACCTGCATCCGTGCTTACCGTGAACCTGGCTG AAGCTCGTGGTGCTAAAGTTCTGACCCACACCCGTGTTGAAGACTTCGAC ATCTCTCCGACTCTGTTAAAATCGAAACCGCTAACGGTTCTTACACCGCT GACAAACTGATCGTTTCTATGGGTGCTTGGAACTCTAAACTGCTGTCTAA ACTGAACCTGGACATCCCGCTGCAACCGTACCGTCAGGTTGTTGGTTTTCTT</p>

	<p>CGAATCTGACGAATCTAAATACTCTAACGACATCGACTTCCCGGGTTTCA TGTTGAAAGTTCCGAACGGTATCTACTACGGTTTCCCGTCTTTCCGGTGGTT GCGGTCTGAAACTGGGTACCACACCTTCGGTCAGAAAATCGACCCGGAC ACCATCAACCGTGAATTTGGTGTACCCGGAAGACGAATCTAACCTGCG TGCTTTCCTGGAAGAATACATGCCGGGTGCTAACGGTGAACCTGAAACGTG GTGCTGTTTGCATGTACACCAAACCTGGACGAACACTTCATCATCGAC CTGCACCCGGAACACTCTAACGTTGTTATCGCTGCTGGTTTCTCTGGTCAC GGTTTCAAATTCTCTTCTGGTGTGGTGAAGTTCTGTCTCAGCTGGCTCTG ACCGGTAACCGAACACGACATCTCTATCTTCTCTATCAACCGTCCGGC TCTGAAAGAATCTCTGCAAAAAACCACCATCTAA</p>
<p><i>Corynebacterium glutamicum</i> R CgadH</p>	<p>ATGCATCATCACCATCACCACACCACCGCTGCTCCGCAGGAATTTACCGC TGCTGTTGTTGAAAAATTCGGTCACGAAGTTACCGTTAAAGACATCGACC TGCCGAAACCGGGTCCGAACCAGGCTCTGGTTAAAGTTCTGACCTCTGGT ATCTGCCACACCGACCTGCACGCTCTGGAAGGTGACTGGCCGGTTAAACC GGAACCGCCGTTTCGTTCCGGGTCACGAAGGTGTTGGTGAAGTTGTTGAAC TGGGTCCGGGTGAACACGACGTTAAAGTTGGTGACATCGTTGGTAACGCT TGGCTGTGGTCTGCTTGCAGTACCTGCGAATACTGCATCACCAGGTCGTGA AACCCAGTGCAACGAAGCTGAATACGGTGGTTACACCCAGAACGGTTCTT TCGGTCAGTACATGCTGGTTGACACCCGTTACGCTGCTCGTATCCCGGAC GGTGTGACTACCTGGAAGCTGCTCCGATCCTGTGCGCTGGTGTACCAGTT TACAAAGCTCTGAAAGTTTCTGAAACCCGTCGGGTGAGTTTCATGGTTAT CTCTGGTGTGGTGGTCTGGGTACATCGCTGTTTACGACGCTGCTGCTAT GGGTATGCGTGTATCGCTGTTGACATCGCTGACGACAAACTGGAACCTGG CTCGTAAACACGGTCTGAATTTACCGTTAACGCTCGTAAACGAAGACCCG GGTGAAGCTGTTTACGAAATACACCAACGGTGGTGGTCTCACGGTGTCTGGT TACCGCTGTTACGAAGCTGCTTTCGGTCAGGCTCTGGACATGGCTCGTCCG TGCTGGTACCATCGTTTTCAACGGTCTGCCGCCGGGTGAATTTCCGGCTTC TGTTTTCAACATCGTTTTCAAGGTCTGACCATCCGTGGTTCTCTGGTTGG TACCCGTCAGGACCTGGCTGAAGCTCTGGACTTCTTCGCTCGTGGTCTGAT CAAACCGACCGTTTCTGAATGCTCTCTGGACGAAGTTAACGACGTTCTGG ACCGTATGCGTAACGGTAAAATCGACGGTCTGTTGCTATCCGTTACTAA</p>
<p><i>Bacillus stearothermophilus</i> Bsadh2</p>	<p>ATGCATCATCACCATCACCACAAAGCTGCTGTTGTTAACGAATTTAAAAA AGCTCTGGAAATCAAAGAAGTTGAACGTCGAAACTGGAAGAAGGTGAA GTTCTGGTTAAAATCGAAGCTTGCAGGTTTGGCCACACCGACCTGCACGC TGCTCACGGTGACTGGCCGATCAAACCGAAACTGCCGCTGATCCCGGGTC ACGAAGGTGTTGGTATCGTTGTTGAAGTTGCTAAAGGTGTTAAATCTATC AAAGTTGGTGACCGTGTGGTATCCCGTGGCTGTACTCTGCTTGCAGGTGA ATGCGAATACTGCCTGACCGGTCAGGAAACCCTGTGCCCGCACCAGCTGA ACGGTGGTACTCTGTTGACGGTGGTTACGCTGAATACTGCAAAGCTCCG GCTGACTACGTTGCTAAAATCCCGGACAACCTGGACCCGGTTGAAGTTGC TCCGATCCTGTGCGCTGGTGTACCACCTACAAAGCTCTGAAAGTTTCTGG TGCTCGTCCGGGTGAATGGGTTGCTATCTACGGTATCGGTGGTCTGGGTC ACATCGCTCTGCAATACGCTAAAGCTATGGGTCTGAACGTTGTTGCTGTTG ACATCTCTGACGAAAAATCTAAACTGGCTAAAGACCTGGGTGCTGACATC GCTATCAACGGTCTGAAAGAAGACCCGGTTAAAGCTATCCACGACCAGGT TGGTGGTGTTCACGCTGCTATCTCTGTTGCTGTTAACAAAAAGCTTTCGA ACAGGCTTACCAGTCTGTTAAACGTGGTGGTACCCTGGTTGTTGTTGGTCT GCCGAACGCTGACCTGCCGATCCCGATCTTCGACACCGTTCTGAACGGTG</p>

	<p>TTTCTGTAAAGGTTCTATCGTTGGTACCCGTAAAGACATGCAGGAAGCT CTGGACTTCGCTGCTCGTGGTAAAGTTCGTCCGATAGTTGAAACCGCTGA ACTGGAAGAAATCAACGAAGTTTTCGAACGTATGGAAAAAGGTAAAATC AACGGTCGTATCGTTCTGAAACTGAAAGAAGACTAA</p>
<p><i>Bacillus methanolicus</i> MGA3 <i>Bmmdh3</i></p>	<p>ATGCATCATCACCATCACCACAAAAACACCCAGTCTGCTTTCTACATGCC GTCTGTAAACCTGTTCCGGTGGTCTGGTTCTGTAAACGAAGTTGGTACCCGTCT GGCTGGTCTGGGTGTAAAAAAGCTCTGCTGGTTACCGACGCTGGTCTGC ACTCTCTGGGTCTGTCTGAAAAAATCGCTGGTATCATCCGTGAAGCTGGT GTTGAAGTTGCTATCTTCCCGAAAGCTGAACCGAACCCGACCGACAAAAA CGTTGCTGAAGGTCTGGAAGCTTACAACGCTGAAAACCTGCGACTCTATCG TTACCCTGGGTGGTGGTTCTTCTCACGACGCTGGTAAAGCTATCGCTCTGG TTGCTGCTAACGGTGGTACCATCCACGACTACGAAGGTGTTGACGTTTCT AAAAAACCGATGGTTCCGCTGATCGCTATCAACACCACCGCTGGTACCGG TTCTGAACTGACCAAATTCACCATCATCACCACCGAACCGAACGTAAAGTTA AAATGGCTATCGTTGACAAACACGTTACCCCGACCCGTCTATCAACGAC CCGGAACCTGATGGTTGGTATGCCGCCGTCTCTGACCGCTGCTACCGGTCT GGACGCTCTGACCCACGCTATCGAAGCTTACGTTTCTACCGGTGCTACCCC GATCACCACGCTCTGGCTATCCAGGCTATCAAAATCATCTCTAAATACC TGCCGCGTGCTGTTGCTAACGGTAAAGACATCGAAGCTCGTGAACAGATG GCTTCGCTCAGTCTCTGGCTGGTATGGCTTTCAACAACGCTGGTCTGGGT TACGTTACGCTATCGCTCACCAGCTGGGTGGTTTCTACAACCTCCCGCAC GGTGTGTTGCAACGCTATCCTGCTGCCGCACGTTTGCCGTTTCAACCTGATC TCTAAAGTTGAACGTTACGCTGAAATCGCTGCTTTCCTGGGTGAAAACGT TGACGGTCTGTCTACCTACGAAGCTGCTGAAAAAGCTATCAAAGCTATCG AACGTATGGCTCGTGACCTGAACATCCCGAAAGGTTTCAAAGAAGTGGGT GCTAAAGAAGAAGACATCGAAACCCTGGCTAAAAACGCTATGAACGACG CTTGCGCTCTGACCAACCCGCGTAAACCGAAACTGGAAGAAGTTATCCAG ATCATCAAAAACGCTATGTAA</p>

Supplementary Table 4

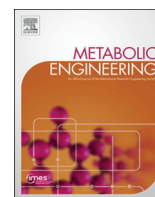
Non-standard reactions used in the flux balanced analysis. ‘ub’, ‘lb’ and ‘c’ stand for upper bound, lower bound and objective function coefficient respectively. The formula is in accordance with the naming convention of the core model and the syntax of the Cobra toolbox. ‘Sarco’ and ‘FA’ stand for sarcosine and formaldehyde respectively.

Reaction	Formula	ub	lb	c
EX_sarco	sarco[c] <=>	- 1000	1000	0
EX_xylose	xylose[c] <=>	- 1000	1000	0
EX_gly	gly[c] <=>	0	1000	0
EX_E4P	e4p[c] <=>	0	1000	0
SOX	sarco[c] + h2o[c] + o2[c] --> FA[c] + gly[c] + h2o2[c]	0	1000	0
CATALASE	2 h2o2[c] --> 2 h2o[c] + o2[c]	0	1000	0
HPS	FA[c] + ru5p-D[c] --> h6p[c]	0	1000	-1
PHI	h6p[c] <=> f6p[c]	- 1000	1000	0
Xyloisomerase	xylose[c] <=> xylulose[c]	- 1000	1000	0
Xukinase	xylulose[c] + atp[c] --> xu5p-D[c] + adp[c] + h[c]	0	1000	0
glyA	ser[c] + thf[c] <=> gly[c] + thf_ch2[c] + h2o[c]	0	1000	0
GCV	gly[c] + nad[c] + thf[c] <=> thf_ch2[c] + nadh[c] + co2[c] + nh4[c]	0	1000	0
serABC	3pg[c] + nad[c] + glu-L[c] + h2o[c] <=> ser[c] + nadh[c] + akg[c] + pi[c] + h[c]	0	1000	0
Biomass	0.462 ser[c] + 1.033 gly[c] + 1.033 thf_ch2[c] + 3.7478 accoa[c] + 59.81 atp[c] + 0.361 e4p[c] + 0.0709 f6p[c] + 0.129 g3p[c] + 0.205 g6p[c] + 0.2557 gln-L[c] + 3.4454 glu-L[c] + 59.35 h2o[c] + 2.051 nad[c] + 13.0279 nadph[c] + 1.7867 oaa[c] + 0.5191 pep[c] + 2.8328 pyr[c] + 0.8977 r5p[c] <=> 59.81 adp[c] + 2.6222 akg[c] + 3.7478 coa[c] + 58.27 h[c] + 2.051 nadh[c] + 13.0279 nadp[c] + 58.31 pi[c] + 1.033 thf[c]	1	1	0



Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/meteng

An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in *E. coli*



Hai He^a, Rune Höper^a, Moritz Dodenhöft^a, Philippe Marlière^{b,**}, Arren Bar-Even^{a,*}

^a Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476, Potsdam-Golm, Germany

^b TESSSI, The European Syndicate of Synthetic Scientists and Industrialists, 81 rue Réaumur, 75002, Paris, France

ARTICLE INFO

Keywords:

Pathway design
Promiscuous enzymes
Formaldehyde assimilation
Serine cycle
Growth selection

ABSTRACT

Engineering biotechnological microorganisms to use methanol as a feedstock for bioproduction is a major goal for the synthetic metabolism community. Here, we aim to redesign the natural serine cycle for implementation in *E. coli*. We propose the homoserine cycle, relying on two promiscuous formaldehyde aldolase reactions, as a superior pathway design. The homoserine cycle is expected to outperform the serine cycle and its variants with respect to biomass yield, thermodynamic favorability, and integration with host endogenous metabolism. Even as compared to the RuMP cycle, the most efficient naturally occurring methanol assimilation route, the homoserine cycle is expected to support higher yields of a wide array of products. We test the *in vivo* feasibility of the homoserine cycle by constructing several *E. coli* gene deletion strains whose growth is coupled to the activity of different pathway segments. Using this approach, we demonstrate that all required promiscuous enzymes are active enough to enable growth of the auxotrophic strains. Our findings thus identify a novel metabolic solution that opens the way to an optimized methylotrophic platform.

1. Introduction

Microbial production of commodity chemicals is limited by feedstock availability and cost. Sugars and starches, despite being commonly used, are not ideal microbial feedstocks as their biotechnological utilization directly competes with human consumption, thus eroding food security (Walker, 2009). Furthermore, the expansion of agricultural cultivation comes at the expense of shrinking natural habitats, hence threatening biodiversity (Fitzherbert et al., 2008). The use of lignocellulosic biomass, while avoiding some of these problems, presents other challenges, including heterogenic composition, difficult processing, and deleterious waste products (Sanderson, 2011). One carbon compounds provide a favorable alternative as they can be produced at high levels without burdening agricultural production and they represent homogenous, easy-to-handle microbial feedstocks (Schrader et al., 2009; Takors et al., 2018; Yishai et al., 2016). Methanol is especially interesting as it is completely water miscible, avoiding mass transfer barriers that constrain the use of gaseous one carbon compounds (e.g., carbon monoxide and methane). Methanol can be produced at low-cost from fossil or renewable methane (Zakaria and Kamarudin, 2016), or it can be produced sustainably and efficiently from CO₂ and electrochemically derived hydrogen (Szyma and Cormos,

2018).

There has been much recent progress in the metabolic engineering of microorganisms that naturally grow on methanol, e.g., *Methylobacterium extorquens* (Marx and Lidstrom, 2004; Schada von Borzyskowski et al., 2015). Still, when compared to biotechnological microorganisms, such as *Escherichia coli*, natural methylotrophs are more difficult to manipulate and engineer (Pfeifenschneider et al., 2017). To alleviate this problem, multiple recent efforts have sought to engineer biotechnological hosts for growth on methanol via one of the naturally occurring methanol assimilation pathways (Wang et al., 2020; Whitaker et al., 2015; Zhang et al., 2017): the ribulose monophosphate (RuMP) cycle (Chen et al., 2018; He et al., 2018; Meyer et al., 2018), the dihydroxyacetone (DHA) cycle (Dai et al., 2017), and the serine cycle (Yu and Liao, 2018). However, these natural pathways might not represent optimal solutions. Better pathways, more efficient in the use of cellular resources and/or more metabolically compatible with the host microorganism, could be designed and implemented. For example, a recent study designed and partially implemented a modified serine cycle in *E. coli*, where some of the natural reactions were replaced with others, better fitting the endogenous metabolism of the host (Yu and Liao, 2018). However, the serine cycle and its modified variants are ATP-inefficient, which results in low biomass and product yields

* Corresponding author.

** Corresponding author.

E-mail addresses: pmarliere@isthmus.fr (P. Marlière), Bar-Even@mpimp-golm.mpg.de (A. Bar-Even).

<https://doi.org/10.1016/j.ymben.2020.03.002>

Received 15 January 2020; Received in revised form 26 February 2020; Accepted 6 March 2020

Available online 10 March 2020

1096-7176/ © 2020 The Author(s). Published by Elsevier Inc. on behalf of International Metabolic Engineering Society. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

(Claassens et al., 2019). In other studies, new-to-nature reactions were demonstrated, including the formolase reaction that condenses formaldehyde, the direct product of methanol oxidation, to glycolaldehyde or DHA (Lu et al., 2019; Wang et al., 2017; Yang et al., 2019), or hydroxyacyl-CoA lyase that condenses formyl-CoA with formaldehyde to generate glycolyl-CoA (Chou et al., 2019). However, the condensation rate and/or affinity for formaldehyde of these reactions are too low to be physiologically relevant.

In this study, we aim to upgrade the serine cycle for implementation in *E. coli*. By applying several design principles – including maximization of biomass yield and thermodynamic favorability as well as replacement of CO₂ fixation with formaldehyde assimilation – we devise the homoserine cycle, which relies on two promiscuous formaldehyde-condensing aldolase reactions. We show that this synthetic route is expected to outperform the serine cycle and its modified variants in multiple key metrics. Moreover, even when compared to the RuMP cycle, the most efficient natural methanol assimilation pathway, the homoserine cycle is expected to support higher yields of products that are derived from acetyl-CoA, including ethanol, acetone, butyrate, butanol, citrate, itaconate, 2-ketoglutarate, and levulinic acid. As such, this cycle can outcompete all the natural aerobic methanol assimilation pathways for a wide array of products.

While establishing growth via the complete homoserine cycle is a challenging task that would require extensive further research, here we demonstrate that *E. coli* native enzymes can promiscuously catalyze all non-natural reactions of the homoserine cycle in a physiologically relevant manner, producing essential cellular building blocks. Overall, our work confirms the feasibility of a novel methanol assimilation pathway that could pave the way for future implementation of highly efficient conversion of this one carbon feedstock into commodity chemicals.

2. Results

2.1. Design of the homoserine cycle

Inspecting the structure of the serine cycle (Fig. 1a and Supplementary Table S1), we identified three key shortcomings: (i) Three of the pathway reactions participate in the central metabolism of the host (the pentose phosphate pathway, glycolysis, and the TCA cycle). This could lead to competition for flux between the pathway reactions and those of central metabolism, thus making regulation highly challenging (Bar-Even, 2016). (ii) The serine cycle is ATP-inefficient, consuming 3 ATP molecules per acetyl-CoA generated. This lowers the biomass and product yields as, rather than being assimilated into biomass, a substantial fraction of the substrate methanol needs to be oxidized to generate the required ATP (Claassens et al., 2019). (iii) For each formaldehyde molecule assimilated by the serine cycle, one CO₂ molecule is fixed. While CO₂ fixation has the advantage of serving as an electron sink for access reducing power (Cotton et al., 2020), it leads to two key problems. First, as carboxylation is thermodynamically unfavorable, it needs to be coupled directly or indirectly to ATP hydrolysis, thus reducing the ATP efficiency of the pathway (Bar-Even et al., 2012). In the case of the serine cycle, glycerate kinase consumes an ATP that is subsequently used to energize carboxylation by phosphoenolpyruvate carboxylase. Second, CO₂ fixation must be accompanied by two reduction steps, in order to bring the carbon to the average oxidation state of carbon in biomass. These extra reduction steps require more of the substrate to be completely oxidized to generate NAD(P)H, thus necessitating a delicate regulation between substrate assimilation and substrate oxidation.

We aimed to design a more applicable version of the serine cycle, which overcomes the shortcomings of the natural pathway. Rather than restricting the design to primary enzymatic reactions, we considered also promiscuous enzyme activities, i.e., side reactions. This dramatically expands the metabolic solution space and thus assists in

identifying optimal pathway structures (Trudeau et al., 2018). To facilitate pathway implementation, we further aimed to employ only *E. coli* enzymes, reduce their overall number, and avoid overlap with central metabolism.

Importantly, we aimed the new pathway to be independent on carboxylation so as to lower the demand for NAD(P)H generation by complete substrate oxidation. A formaldehyde assimilation reaction should ideally replace the existing carboxylation reaction, as, unlike CO₂, formaldehyde is already at the average oxidation state of cellular carbon. Similarly, consumption of ATP should be minimized in order to reduce substrate oxidation and increase biomass (and product) yield. Despite the reduction of ATP consumption, the pathway should be at least as thermodynamically favorable as the serine cycle (Noor et al., 2014).

Following a comprehensive literature search of formaldehyde condensing reactions catalyzed by *E. coli* native enzymes, we identified the homoserine cycle (Fig. 1b and Supplementary Table S1) as a superior design. In the homoserine cycle, glycine is directly condensed with formaldehyde to generate serine. This serine aldolase (SAL) reaction (Fig. 1b) was previously found to be promiscuously catalyzed (*in vitro*) by threonine aldolase (LtaE) (Contestabile et al., 2001). The SAL reaction bypasses the very long, multi-cofactor-dependent, and ATP-inefficient route for formaldehyde assimilation to 5,10-methylene-tetrahydrofolate (CH₂-THF) (Crowther et al., 2008) (Fig. 1b). As within the previously proposed modified serine cycles (Bar-Even, 2016; Yu and Liao, 2018), serine is then deaminated to pyruvate by serine deaminase (SDA in Fig. 1b), bypassing a longer route via glycerate, which further involves the highly toxic intermediate hydroxypyruvate (Kim and Copley, 2012) (Fig. 1b). Despite being rather oxygen-sensitive, serine deaminase was shown to support high flux under aerobic conditions (Yu and Liao, 2018).

Pyruvate is then condensed with formaldehyde to generate the non-native metabolite 4-hydroxy-2-oxobutanoate (HOB) (Bouzon et al., 2017), which is subsequently aminated to homoserine. The first of these reactions – HOB aldolase (HAL in Fig. 1b) – was found to be promiscuously catalyzed by *E. coli* 2-keto-3-deoxy-L-rhamnonate aldolase (RhmA) (Hernandez et al., 2017) as well as by similar aldolases (Wang et al., 2019). The latter reaction – HOB amination (HAT in Fig. 1b) – is supported by numerous aminotransferases (Hernandez et al., 2017; Walther et al., 2018; Zhong et al., 2019) as well as amino acid dehydrogenases such as (engineered) glutamate dehydrogenase (Chen et al., 2015). This route effectively replaces a carboxylation reaction (by phosphoenolpyruvate carboxylase) with a formaldehyde assimilation reaction that provides an alternative way to generate a C₄ intermediate. Homoserine is then metabolized by homoserine kinase (ThrB, HSK in Fig. 1b) and threonine synthase (ThrC, TS in Fig. 1b) to produce threonine. Finally, threonine is cleaved by threonine aldolase (LTA in Fig. 1b, catalyzed by the same LtaE that catalyzes the SAL reaction) to regenerate glycine and produce acetaldehyde which can be further oxidized to acetyl-CoA and assimilated to central metabolism.

2.2. The homoserine cycle could outperform natural methanol assimilation pathways

We compared the homoserine cycle to the serine cycle and the modified serine cycle according to multiple criteria. Assuming formaldehyde as the substrate of the pathways, we find that: (i) the homoserine cycle requires only eight enzymes, which is half the number of enzymes needed for the serine cycle (16) and ~40% fewer enzymes than the modified serine cycle (13, Fig. 1c); also, unlike the other pathways, the homoserine cycle is not dependent on foreign enzymes. (ii) The homoserine cycle consumes a single ATP molecule for the production of acetyl-CoA, while the serine cycle requires 3 ATP molecules and the modified serine cycle needs 4 ATP equivalents (Fig. 1c). (iii) The homoserine cycle does not lead to the net consumption of NAD(P)H, while the other two pathways consume 2 NAD(P)H per acetyl-

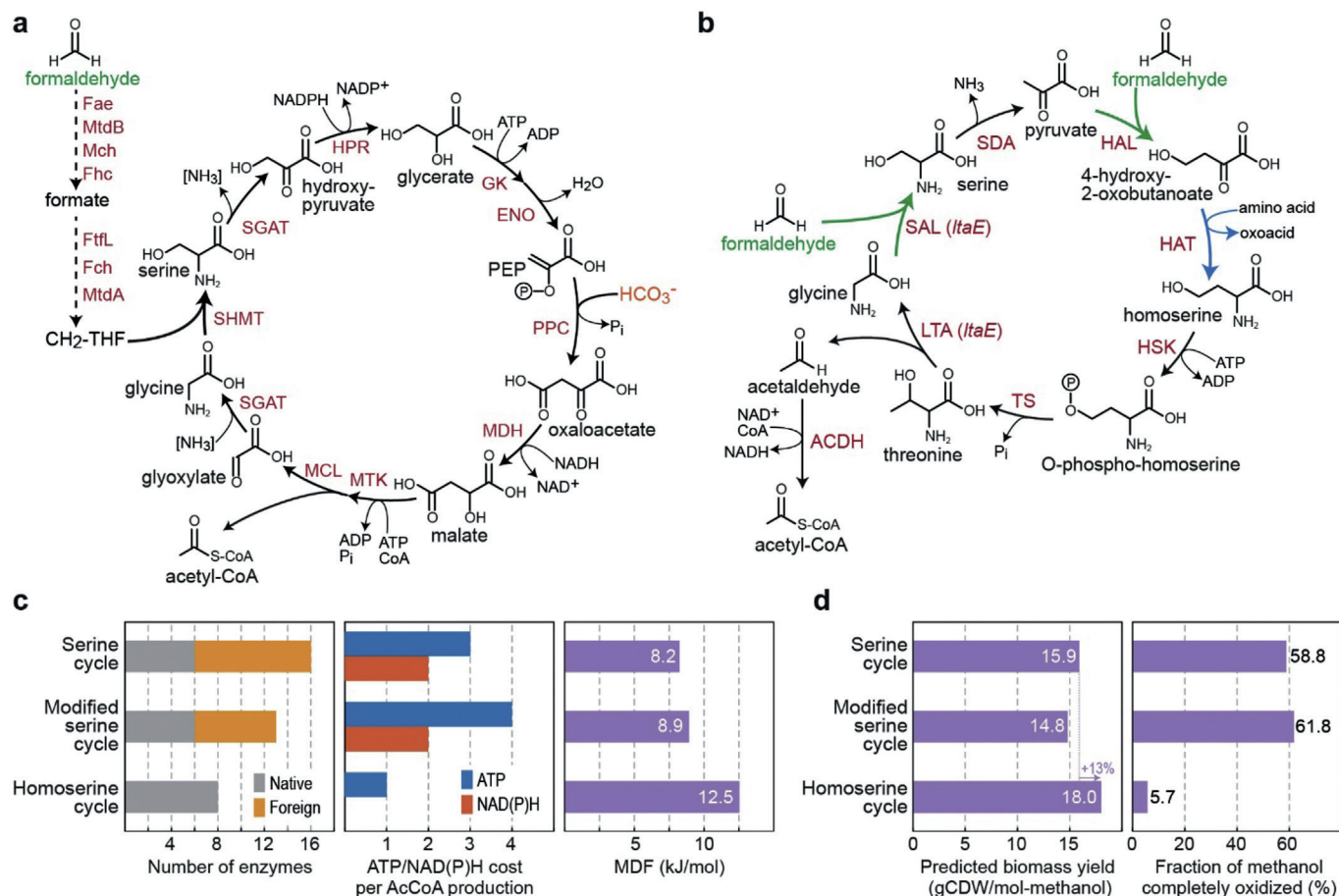


Fig. 1. Natural and synthetic methanol assimilation pathways. (a) The natural serine cycle (formaldehyde as feedstock). Pathway enzymes and reactions are listed in Supplementary Table S1 (b) The synthetic homoserine cycle (formaldehyde as feedstock). The homoserine cycle harbors three promiscuous enzyme activities: serine aldolase (SAL), 4-hydroxy-2-oxobutanoate (HOB) aldolase (HAL), and the HOB aminotransferase (HAT). The other pathway reactions include threonine biosynthesis (homoserine kinase, HSK; threonine synthase, TS) and cleavage (threonine aldolase, LTA) as well as serine deaminase (SDA). Note that both SAL and LTA are catalyzed by the same LtaE enzyme. Acetylating acetaldehyde dehydrogenase (ACDH) can convert the direct pathway product acetaldehyde to acetyl-CoA. (c) The homoserine cycle outperforms the natural serine cycle and the previously suggested modified serine cycle (Yu and Liao, 2018) in terms of simplicity (small number of reactions), resource consumption efficiency and thermodynamic driving force as indicated by the Max-min Driving Force (MDF) criterion (Noor et al., 2014). These analyses were based on formaldehyde as pathway feedstock (Methods). (d) The homoserine cycle is predicted by Flux Balance Analysis to support higher biomass yield on methanol than the natural serine cycle and modified serine cycle. Furthermore, unlike the other pathways, the homoserine cycle is almost independent on the complete oxidation of methanol to provide the cell with reducing power and energy.

CoA (Fig. 1c); as mentioned above, this difference can be attributed to the fact that the homoserine cycle replaces a carboxylation reaction with another formaldehyde assimilation reaction. (iv) Despite its low resource consumption, the homoserine cycle is more energetically favorable than the other two pathways as its Max-min Driving Force – MDF, representing the minimal thermodynamic driving force via the pathway reactions after optimizing metabolite concentrations within a physiological range (Noor et al., 2014) – is substantially higher (Fig. 1c, Supplementary Fig. S1). The higher thermodynamic driving force could assist in pulling methanol oxidation forward, a reaction which is thermodynamically unfavorable and represents a major constraint for establishing synthetic methylotrophy.

We further used Flux Balance Analysis (FBA) to calculate the maximal biomass yield with each of the three pathways using methanol as a substrate (Methods). We found the expected biomass yield of the homoserine cycle to be 13% higher than with the serine cycle, while the modified serine cycle supported an even lower yield (Fig. 1d). Moreover, while the operation of the two other pathways requires > 50% of the methanol to be completely oxidized to provide NAD(P)H and ATP, only ~6% of methanol is completely oxidized during growth via the homoserine cycle (Fig. 1d). This indicates that the establishment of highly efficient formaldehyde and formate oxidation systems is strictly

required for the operation of the serine cycle and its modified variants, while the homoserine cycle could operate without such systems, in which case the biomass yield drops by less than 1%. The ability to bypass formaldehyde and formate oxidation is key for the establishment of this synthetic route, as it circumvents the need for a delicate fine-tuning of the formaldehyde oxidation flux relative to the formaldehyde assimilation flux.

It is commonly argued that the RuMP cycle is the most efficient naturally occurring route for methanol assimilation. Indeed, among all aerobic methanol assimilation pathways, the RuMP cycle supports the highest biomass yield (Claassens et al., 2019). We used FBA to compare the maximal yields of various central metabolites, precursors, and chemicals of interest using methanol as a substrate and either the RuMP cycle or the homoserine cycle as assimilation routes (Methods). While the RuMP cycle was able to sustain higher yields of biomass and phosphosugars, the homoserine cycle supported substantially higher yields of compounds derived from acetyl-CoA, including ethanol, acetone, butyrate, 1-butanol, citrate, itaconate (Zhao et al., 2018), 2-ketoglutarate, levulinic acid (Cheong et al., 2016), and adipic acid (Yu et al., 2014) (Fig. 2a). Both pathways supported similar yields of pyruvate and its derivatives, including lactate, isobutanol, and 2,3-butanediol (Fig. 2a).

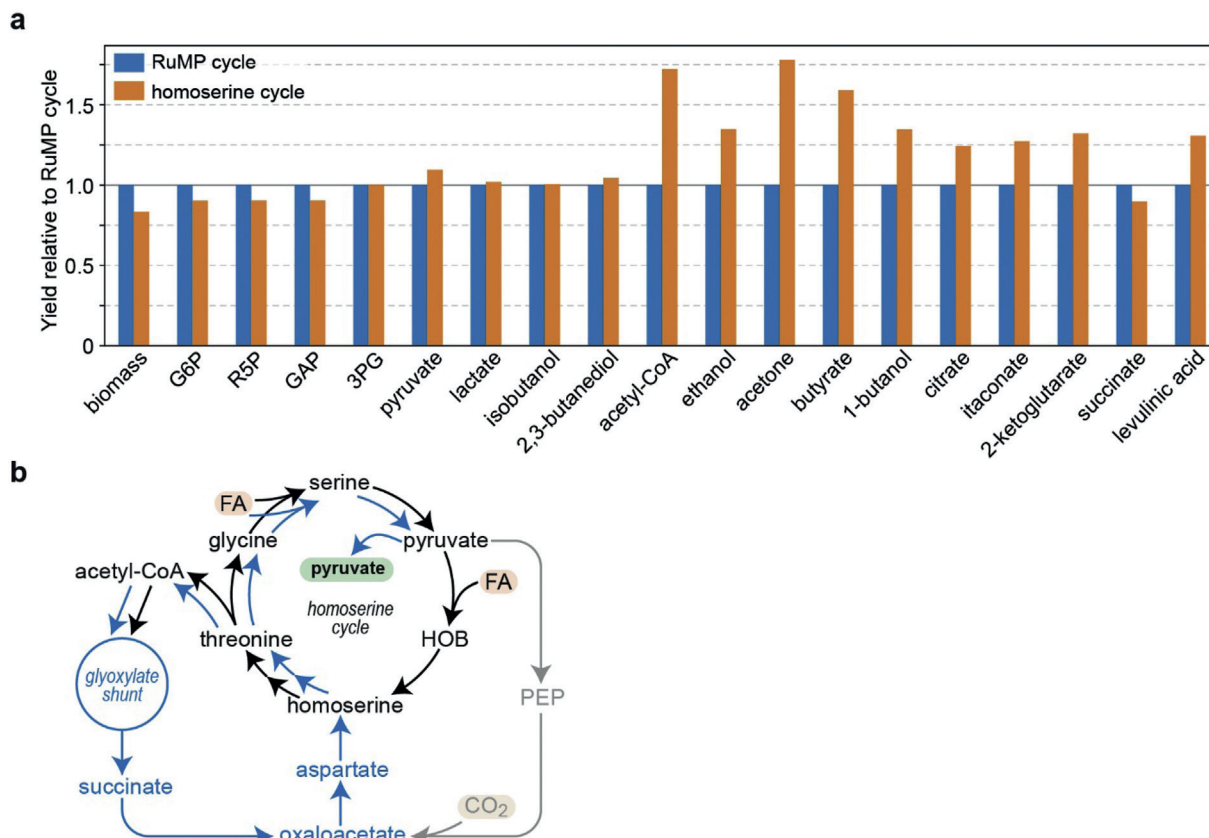


Fig. 2. The homoserine cycle can outperform the RuMP cycle for the production of various compounds. (a) We used Flux Balance Analysis to calculate the maximal yields of various central metabolites, precursors, and products using methanol as carbon source and either the RuMP cycle or the homoserine cycle as assimilation route. The values were normalized to the yields via the RuMP cycle. (b) Production of pyruvate, or other $C_{\geq 3}$ metabolites, via the homoserine cycle involves assimilation of acetyl-CoA via the glyoxylate shunt, followed by conversion of succinate to homoserine, thus reentering the homoserine cycle (blue arrows). Excess reducing power, produced from methanol oxidation, can use CO_2 as an electron sink, where a fraction of the pyruvate molecules is diverted to PEP synthesis and anaplerosis (grey arrows), and subsequently reduced to homoserine. FA corresponds to formaldehyde. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Why does homoserine cycle outcompete the RuMP cycle for so many products? Methanol is a highly reduced carbon source, more reduced than most products. Excess electrons generated during methanol bioconversion are channeled to the respiratory chain and dissipated wastefully. The more electrons are dissipated in this way, the greater the loss in potential yield (Cotton et al., 2020). The RuMP cycle especially suffers from electron overflow as it does not use an electron sink such as CO_2 and further releases CO_2 during the oxidative biosynthesis of acetyl-CoA (Bogorad et al., 2014). The homoserine cycle, on the other hand, generates acetyl-CoA directly without the release of CO_2 . Moreover, while the homoserine cycle is not strictly dependent on carboxylation, it can channel a fraction of its flux via a carboxylation route that serves as sink for excess reducing power: some of the pyruvate, instead of reacting with formaldehyde, can be converted to homoserine via the carboxylation-dependent anaplerotic route, which serves as an electron sink (grey arrows in Fig. 2b). The flexibility of the homoserine cycle in terms of CO_2 use – being generally independent of carboxylation while able to moderately fix CO_2 when necessary – is the main reason of the high yields it can support. Furthermore, conversion of acetyl-CoA, the product of the homoserine cycle, into pyruvate and other $C_{\geq 3}$ compounds does not involve wasteful decarboxylation: succinate, the product of the glyoxylate shunt, can be reintegrated directly into the homoserine cycle to produce pyruvate (blue arrow in Fig. 2b).

Overall, the homoserine cycle seems to be preferable to its counterparts, with the potential to outperform both the serine cycle and the RuMP cycle for the production of a wide array of value-added chemicals. However, as the synthetic pathway relies on promiscuous enzyme

activities, the question remains whether these activities are high enough as to be of physiological relevance. In the rest of the paper, we explore the *in vivo* activities of these reactions.

2.3. Demonstration of the *in vivo* activity of serine aldolase

Most of the reactions of the homoserine cycle correspond to the primary activities of their catalyzing enzymes and thus are less likely to constrain pathway flux. For example, overexpression of serine deaminase, either on a plasmid or from the genome, enabled *E. coli* to use serine as a sole carbon and energy source (Supplementary Fig. S2). However, three pathway reactions – SAL, HAL, and HAT – correspond to promiscuous activities that, while characterized *in vitro*, might not be able to support physiologically relevant fluxes. Hence, we aimed to test each of these promiscuous activities *in vivo* within dedicated gene deletion strains, the growth of which is dependent on the activity of these reactions.

We started by testing the ability of LtaE to catalyze the SAL reaction *in vivo* (Fig. 3a). Towards this aim we constructed two strains auxotrophic for glycine and serine. In both strains the gene encoding for 3-phosphoglycerate dehydrogenase ($\Delta serA$) was deleted. In one strain the gene encoding serine hydroxymethyltransferase was also deleted ($\Delta glyA$) while in the other strain the genes of the glycine cleavage system were deleted ($\Delta gcvTHP$). The growth of these strains required the addition of glycine and serine, as the cellular interconversion of these compounds is blocked (Fig. 3b,c).

We reasoned that if the SAL reaction indeed supports

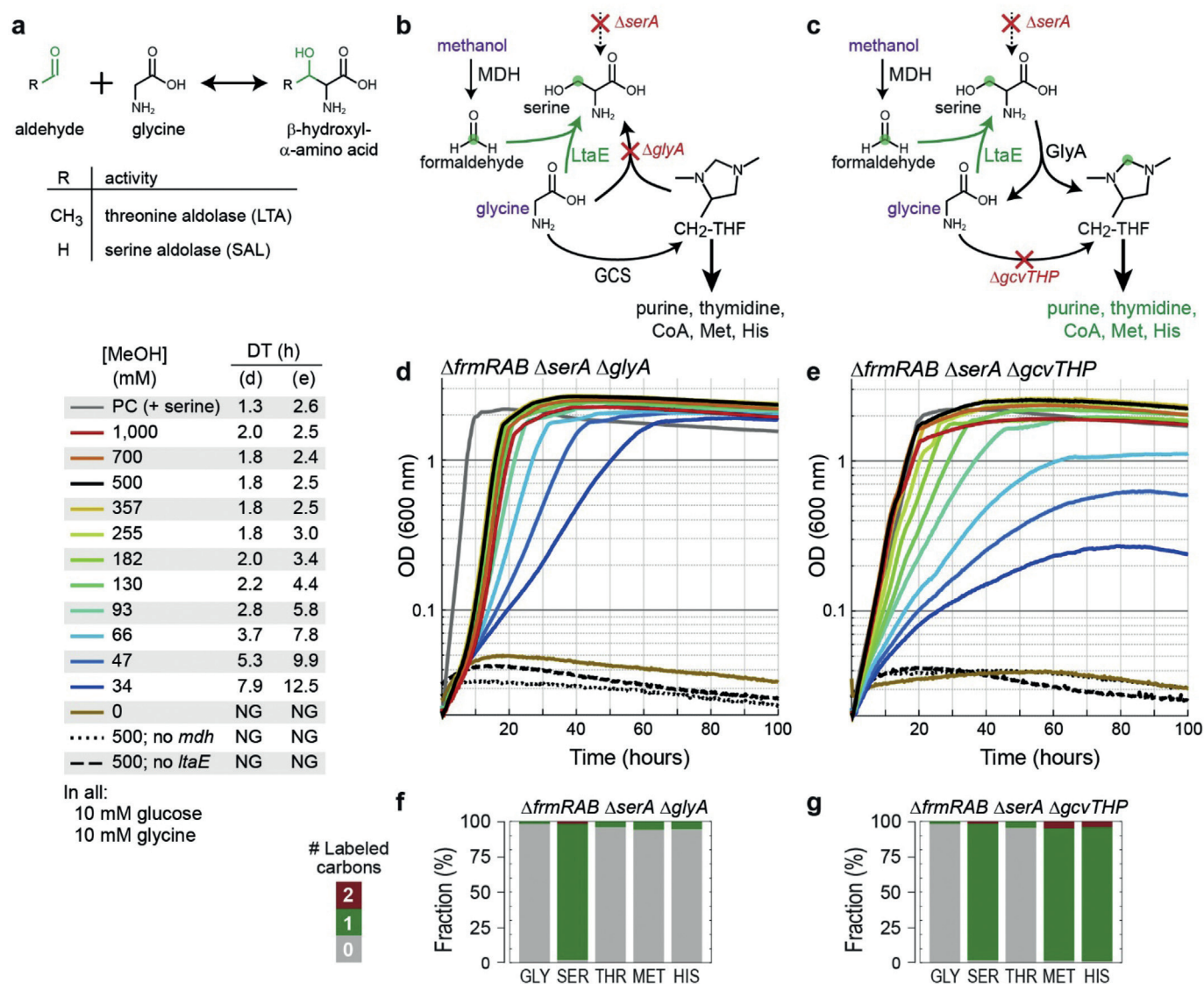


Fig. 3. *In vivo* serine aldolase activity catalyzed by LtaE. (a) LtaE catalyzes aldol condensation between glycine and different aldehydes. Threonine aldolase (LTA) is its primary function, while serine aldolase is a promiscuous activity (Contestabile et al., 2001). (b), (c) Two selection schemes for the *in vivo* activity of the SAL reaction. Carbon sources are shown in purple (glucose not shown) while the formaldehyde moiety is shown in green. (b) $\Delta frmRAB \Delta serA \Delta glyA$ strain in which methanol assimilation is required for the biosynthesis of serine and the cellular C₁ moieties. (c) $\Delta frmRAB \Delta serA \Delta gcvTHP$ strain, in which methanol assimilation is required for the biosynthesis of both serine and the cellular C₁ moieties. (d), (e) Growth with different concentrations of methanol confirm the activity of the SAL reaction of the strains shown in (b) and (c), respectively. In all cases, 10 mM glucose and 10 mM glycine were added to the medium. Each growth curve represents the average of three replicates, which differ from each other by less than 5%. “PC” corresponds to positive control. “DT” corresponds to doubling time. (f), (g) Labeling pattern of proteinogenic glycine (GLY), serine (SER), threonine (THR), methionine (MET) and histidine (HIS), within the strains shown in (b) and (c), respectively, upon feeding with ¹³C-methanol as well as unlabeled glucose and glycine. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

physiologically relevant flux, both strains should be able to grow when methanol dehydrogenase (MDH) and LtaE are overexpressed and serine is replaced with methanol in the medium. In the $\Delta serA \Delta glyA$ strain, the SAL reaction would be responsible for the production of serine (Fig. 3b), which accounts for ~3% of the carbon in biomass (Neidhardt et al., 1990). In the $\Delta serA \Delta gcvTHP$ strain, the SAL reaction would be responsible for the production of both serine and the cellular C₁ moieties (Fig. 3c), together accounting for ~6% of the carbon in biomass (Neidhardt et al., 1990). To avoid formaldehyde oxidation to formate, which might deplete its intracellular pool and constrain its assimilation, the genes encoding for the glutathione-dependent formaldehyde oxidation system were also deleted ($\Delta frmRAB$).

Upon overexpression of MDH and LtaE, we observed growth of both selection strains with glucose as the main carbon source and glycine and methanol as precursors of serine (Fig. 3d,e). This indicates that the

SAL reaction can operate *in vivo* at a physiologically significant rate. Expression of only MDH or only LtaE failed to sustain growth, indicating that the native expression of genomic *ltaE* is too low to support the SAL reaction. The observed growth rate and yield were dependent on the concentration of methanol, where the $\Delta serA \Delta glyA \Delta frmRAB$ strain had higher rates and yields than the $\Delta serA \Delta gcvTHP \Delta frmRAB$ strain on low methanol concentrations. This corresponds to our prediction that the latter strain depends on the SAL reaction to provide a higher fraction of the cellular carbons. Methanol concentrations in the range of 200–1000 mM seem to be optimal, supporting growth rates similar to that of the positive control (in which serine was added to the medium). In all experiments, we added 50 μ M MnCl₂, as Mn²⁺ is a known cofactor of LtaE (Fesko, 2016). Without the additional supplementation of MnCl₂, we observed lower growth rates and yields (Supplementary Fig. S3).

To confirm the activity of the SAL reaction, we conducted ^{13}C -labeling experiments. We cultivated both strains with ^{13}C -methanol as well as unlabeled glucose and glycine. In the $\Delta serA \Delta glyA \Delta frmRAB$ strain we found serine to be entirely singly labeled as expected, while the other amino acids were unlabeled (beyond the natural abundance of ^{13}C , Fig. 3f). As threonine and methionine are derived from oxaloacetate, and hence carry a carbon that originates from CO_2 (i.e., anaplerotic reactions), their lack of labeling indicates that formaldehyde oxidation to CO_2 is negligible, as expected by the deletion of *frmRAB*. In the $\Delta serA \Delta gcvTHP \Delta frmRAB$ strain we found serine, methionine, and histidine to be entirely singly labeled (Fig. 3g). Unlike threonine, both methionine and histidine harbor a carbon derived from THF carrying a C_1 unit – methyl-THF in the case of methionine and formyl-THF in the case of histidine. The labeling of these amino acids thus indicates that all cellular C_1 moieties are derived from methanol. Overall, the labeling results confirm that the SAL reaction provides the sole source of serine in both strains and sole source of C_1 moieties in the $\Delta serA \Delta gcvTHP \Delta frmRAB$ strain.

Next, we tested whether it is possible to omit glycine from the medium, such that it will be produced endogenously via LtaE-dependent threonine cleavage (Fig. 4a). Towards this aim, we deleted, in both selection strains, the genes encoding for threonine dehydrogenase and 2-amino-3-ketobutyrate CoA-ligase ($\Delta kbl\text{-}tdh$), thus blocking the LtaE-independent route of threonine degradation (Fig. 4a) (Yishai et al., 2017). We found that replacing glycine with threonine did not alter the growth of either selection strain (Fig. 4b, note that growth was still strictly dependent on methanol). To enable this growth, the overexpressed LtaE catalyzes two subsequent reactions, first cleaving threonine to glycine and acetaldehyde and then reacting glycine with formaldehyde to produce serine. Hence, LtaE can be regarded as a glycytransferase – transferring a glycine moiety from one small aldehyde (acetaldehyde) to another (formaldehyde).

We wondered whether it was also possible to omit threonine from the medium and rely on native threonine biosynthesis to provide this amino acid as a precursor for glycine and serine (Fig. 4a). Indeed, despite showing reduced growth rates, both selection strains were able to grow with only glucose (as main carbon source) and methanol without the addition of glycine or threonine (green lines in Fig. 4b). This indicates that half of the homoserine cycle is active: homoserine (generated natively from aspartate) is metabolized to threonine via HSK and TS, and LtaE then cleaves threonine to glycine (and acetaldehyde) and condenses glycine with formaldehyde to produce serine (Fig. 4a).

To confirm that, also in the absence of externally provided glycine or threonine, all cellular serine is produced from glycine condensation with formaldehyde, we conducted ^{13}C -labeling experiments. The strains were cultured in the presence of labeled or unlabeled methanol as well as glucose labeled at different carbons (glucose-1- ^{13}C , glucose-2- ^{13}C , and glucose-3- ^{13}C). While the labeling pattern of glycine changed according to the labeled carbon of glucose, cultivation with ^{13}C -methanol always resulted in exactly one more labeled carbon in serine than in glycine (Fig. 4c). This unequivocally confirms the methanol-dependent production of serine from glycine when the latter compound is produced internally from homoserine metabolism.

Our results demonstrate the capability of LtaE to convert threonine to serine *in vivo* by releasing acetaldehyde and assimilating formaldehyde. These findings further confirm the physiologically relevant activity of half of the homoserine cycle, where homoserine metabolism to glycine and serine provided all the biomass requirement of these amino acids as well as cellular C_1 moieties.

2.4. Demonstration of the *in vivo* activity of HOB aldolase and transaminase

After demonstrating the methanol-dependent conversion of homoserine to serine, we aimed to demonstrate methanol-dependent conversion of pyruvate to homoserine. To select for the *in vivo* conversion

of pyruvate to homoserine and threonine via HOB production and amination, we constructed a homoserine auxotroph strain: a deletion of the gene encoding for aspartate-semialdehyde dehydrogenase (Δasd) resulted in a strain capable of growing only when homoserine and diamino-pimelate (DAP) were added to the medium (Cardineau and Curtiss, 1987). In this strain, homoserine is metabolized to methionine, threonine, and isoleucine, while DAP is metabolized to lysine and peptidoglycans. (We note that despite being formally reversible, homoserine dehydrogenase was not able to oxidize homoserine to aspartate-semialdehyde, the precursor of DAP, and hence the addition of the latter intermediate to the medium was required).

We reasoned that, in the presence of methanol and methanol dehydrogenase, the combined activities of HAL and HAT should enable the $\Delta asd \Delta frmRAB$ strain to grow without the addition of homoserine to the medium (Fig. 5a). While only few enzymes were previously shown to catalyze the HAL reaction (Hernandez et al., 2017; Wang et al., 2019), we speculated that similar aldolases could also support this activity, maybe even outperforming those identified before. Hence, we searched for all *E. coli* enzymes (strain MG1655, using EcoCyc (Keseler et al., 2005)) that are known to catalyze an aldolase reaction with pyruvate as a donor (and which might be able to use formaldehyde as an acceptor). Besides RhmA itself (Hernandez et al., 2017), we found six candidate aldolases: GarL, YagE, YjhH, Eda, DgoA, and MhpE (Fig. 5b). RhmA and GarL belong to the structural family of HpcH while mhpE belongs to DmpG family. Both of these families are Type II pyruvate aldolases which use a divalent metal cation for donor binding and enolization (Fang et al., 2019). YagE and YjhH belong to the structural family of DHDPs while Eda and DgoA belong to KDPG family. These families are Type I pyruvate aldolases, using a lysine residue to form a Schiff base with the donor substrate (Fang et al., 2019). We decided to test all of the seven aldolases for their ability to catalyze the HAL reaction *in vivo*. As the HAT reaction is known to be supported by the native aspartate aminotransferase (AspC) (Zhong et al., 2019), a highly expressed protein (Li et al., 2014), and might be further catalyzed by other highly expressed, promiscuous aminotransferases, we hypothesized that no dedicated enzyme overexpression would be required to enable this key reaction.

We found that overexpression of *mdh* together with *rhmA*, *garL*, *yagE*, or *yjhH* enabled growth of the $\Delta asd \Delta frmRAB$ strain when homoserine was replaced with methanol (Fig. 5c). These four aldolase enzymes supported roughly the same growth rates. No growth was observed without methanol, or when methanol dehydrogenase or the aldolase enzymes were overexpressed alone. Growth was not observed when *mdh* was expressed together with *eda*, *dgoA*, or *mhpE*. The relative sequence similarity between RhmA, GarL, YagE, and YjhH (Fig. 5b and Supplementary Fig. S4) might explain why these enzymes, and not the others, were able to support the HAL reaction. Indeed, the structures of RhmA and GarL are almost identical (Supplementary Fig. S5). As anticipated, growth was possible without dedicated overexpression of an aminotransferase enzyme, presumably since the native high expression level of these enzymes is sufficient to support the promiscuous HAT reaction. Genomic overexpression of *aspC* or a mutated version of alanine aminotransferase (*alaC**, the protein product of which was previously shown to catalyze the HAT reaction (Bouzon et al., 2017)) did not alter growth substantially (Supplementary Fig. S6). Similarly, genomic overexpression of *thrBC* did not consistently assist growth (Supplementary Fig. S6). This indicates that the HAT, HSK and TS reactions do not constrain the flux from pyruvate to threonine.

To confirm that homoserine and its downstream products threonine and methionine are produced from pyruvate and methanol via the HAL and HAT reactions, we performed ^{13}C -labeling experiments. Upon cultivation with unlabeled glucose and ^{13}C -methanol, we found threonine and methionine to be completely once labeled, where lysine and aspartate (serving as control) were fully unlabeled (beyond the natural abundance of ^{13}C , Fig. 5d). This confirms that homoserine and threonine are completely derived from pyruvate and methanol.

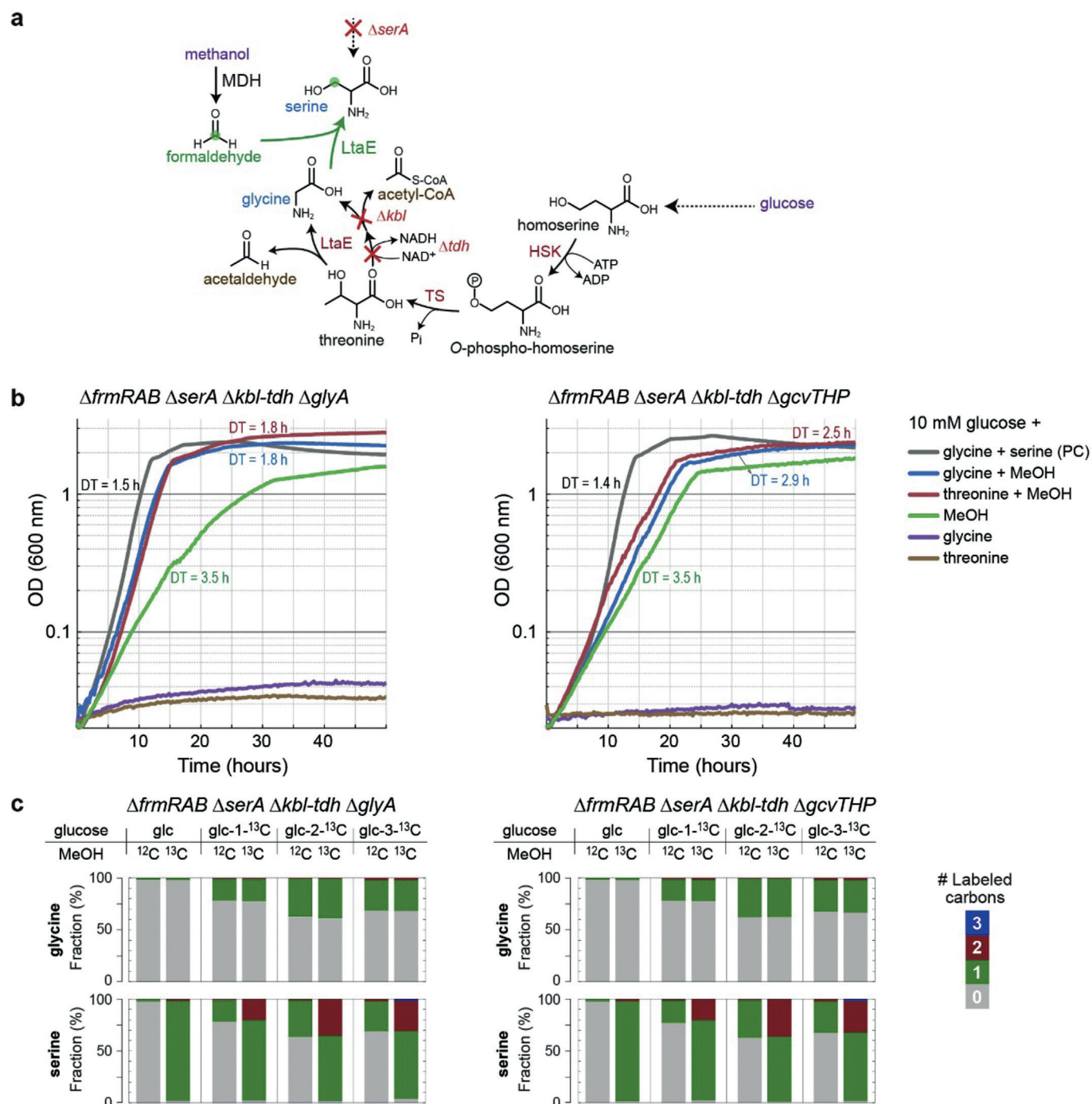


Fig. 4. LtaE can operate as threonine aldolase and serine aldolase simultaneously. (a) Selection scheme for serine and glycine production from threonine and homoserine. (b) The two selection strains, deleted in the LtaE-independent threonine cleavage system ($\Delta kbl\text{-}tdh$), can grow with methanol as serine precursor. Glycine is either provided externally (10 mM, blue lines) or produced internally, either from externally added threonine (10 mM, red lines) or from the internal pool of threonine (green lines). The latter growth confirms that LtaE can catalyze the LTA and SAL reactions simultaneously. In all cases, 10 mM glucose and 500 mM methanol were added. Each growth curve represents the average of three replicates, which differ from each other by less than 5%. “PC” corresponds to positive control. “DT” corresponds to doubling time. (c) Labeling pattern of proteinogenic glycine and serine within the strains shown in (b) upon feeding with glucose (glc) labeled at different carbon as well as labeled or unlabeled methanol. This labeling confirms that all cellular serine is produced from glycine and methanol even when glycine is produced internally from threonine biosynthesis and degradation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Our experiments thus demonstrate the capacity of several *E. coli* endogenous aldolases to generate HOB *in vivo* and the capacity of the native aminotransferase network to convert HOB into homoserine. These findings further confirm the physiologically relevant activity of the “other half” of the homoserine cycle, where pyruvate metabolism to homoserine and threonine provided the biomass requirement of these

amino acids.

3. Discussion

Different metabolic engineering strategies can be used to establish novel modes of growth and production. In the simplest approach, a

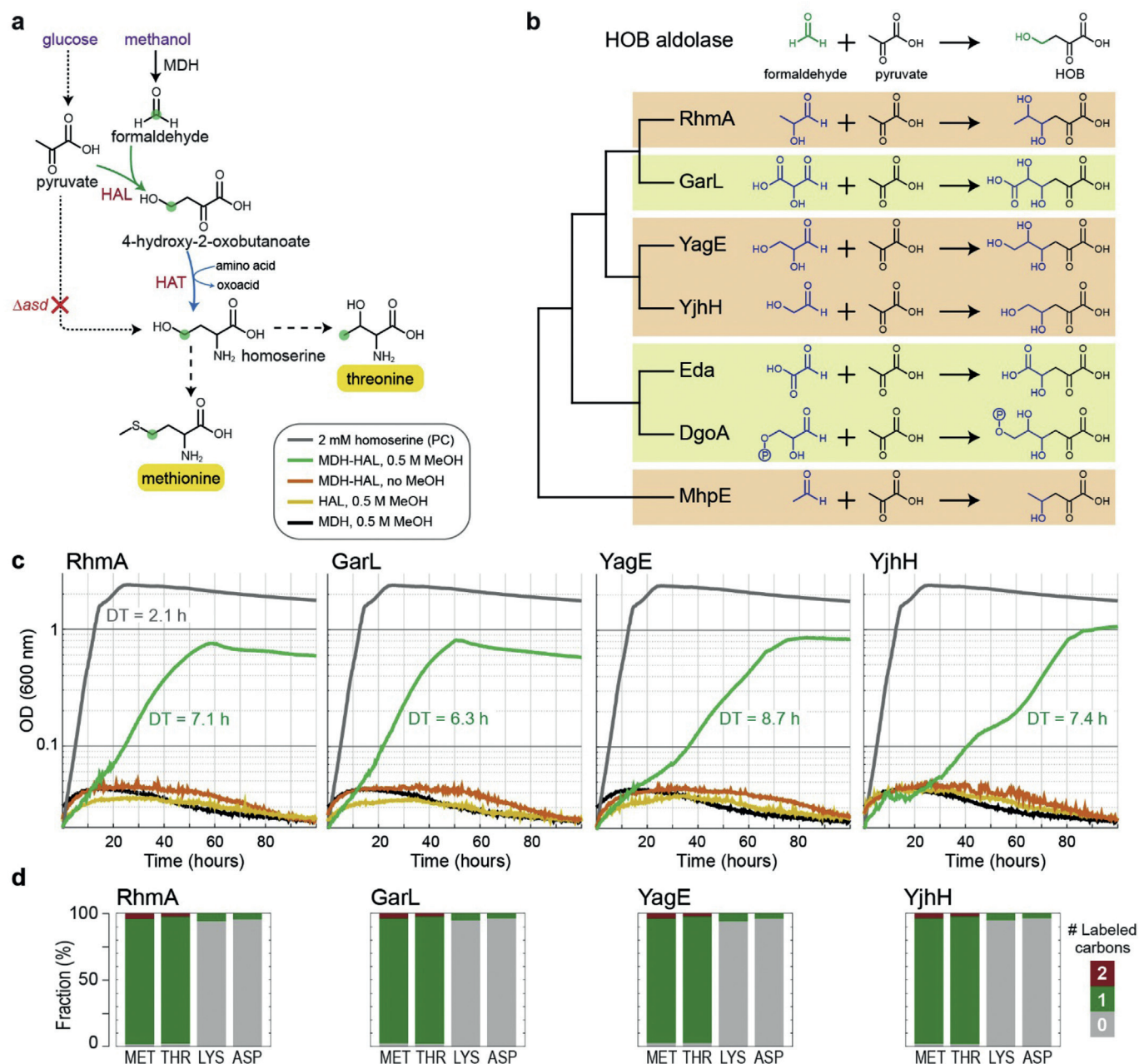


Fig. 5. *In vivo* HOB aldolase and transaminase activities. (a) A selection scheme for the *in vivo* activity of the HAL and the HAT reactions. Carbon sources are shown in purple while the formaldehyde moiety is shown in green. (b) Several *E. coli* enzymes are known to catalyze an aldolase reaction with pyruvate as a donor, and might be able to react with formaldehyde as an acceptor. The sequence similarity of these enzymes is indicated by the schematic tree to the left. YagE and YjhH can catalyze each others reactions. Similarly, Eda and DgoA can catalyze each others reactions. (c) Four aldolases, once overexpressed together with methanol dehydrogenase, support growth of the selection strain. Glucose was added at 10 mM, methanol at 500 mM, diaminopimelate at 0.25 mM, and isoleucine at 1 mM. Each growth curve represents the average of three replicates, which differ from each other by less than 5%. “PC” corresponds to positive control. “DT” corresponds to doubling time. (d) Labeling pattern of proteinogenic methionine (MET), threonine (THR), lysine (LYS), and aspartate (ASP) within the strains shown in (c) upon feeding with unlabeled glucose, diaminopimelate and isoleucine as well as ^{13}C -methanol. The results confirm that all cellular threonine and methionine are derived from HOB aldolase and transaminase reactions. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

metabolic pathway is “copied” from one organism and “pasted” in another, where pathway enzymes are expressed in the new host (Erb et al., 2017). Most studies that try to establish synthetic methylotrophy in heterotrophic hosts follow this strategy by expressing the enzymes of the known methanol assimilation pathways in *E. coli* (Chen et al., 2018; He et al., 2018; Meyer et al., 2018) and other biotechnological organisms (Lessmeier et al., 2015; Withhoff et al., 2015). However, this approach is limited to a small number of possible routes, which might not represent an optimal solution (e.g., in terms of resource consumption

and biomass yield) and might not interact favorably with the endogenous metabolism of the host (e.g., disrupt key cellular fluxes, necessitate complex regulation, or introduce reactive metabolic intermediates).

Alternatively, a mix-and-match of existing enzymes could expand the space of possible pathways: while all the components are known, the combinatorics of their integration can be quite extensive (Erb et al., 2017). The design of the modified serine cycle exemplifies this approach, where key reactions of the natural pathway were replaced with

other known reactions in order to optimize pathway activity in the host (Yu and Liao, 2018). While this approach is considerably more flexible than “copy-paste”, it is still limited by the repertoire of known enzymes and reactions.

In order to fully harness the power of synthetic biology and identify the most promising pathways, it is beneficial to consider also reactions which, while chemically plausible, are not known to be enzymatically catalyzed (Bouzon et al., 2017; Erb et al., 2017). A recent study has demonstrated this approach for the design and *in vitro* realization of synthetic photorespiration pathways which do not release CO₂ and could thus boost photosynthetic carbon fixation (Trudeau et al., 2018). Following such guidelines, previous studies designed novel methanol assimilation pathways, based on direct condensation of formaldehyde molecules (Chou et al., 2019; Lu et al., 2019; Wang et al., 2017; Yang et al., 2019); however, the rates of most of the key reactions remain too low to be of physiological significance.

In this study, we use a similar strategy to identify an optimal methanol assimilation design based on the general structure of the serine cycle. Instead of relying on completely novel reactions, the establishment of which could be highly challenging, we decided to follow a middle-ground approach, where besides known enzymatic reactions we considered characterized promiscuous activities for which at least basic supporting evidence exists.

We identified the homoserine cycle as a superior pathway design. It is short and relies solely on *E. coli* native enzymes. Still, its reactions are “simple” enough to be catalyzed by native enzymes of many different hosts. The pathway does not overlap with central metabolism, thus bypassing possible regulation difficulties. As its consumption of ATP and NAD(P)H is low, it is expected to support high biomass yield. Importantly, by replacing a carboxylation reaction with another formaldehyde condensation reaction, the homoserine cycle should be able to operate effectively without oxidizing formaldehyde, thus avoiding the difficulty of regulating a flux bifurcation point in which formaldehyde (and formate) can either be fully oxidized or assimilated to biomass. We further showed that the homoserine cycle has the potential to outperform even the RuMP cycle for the production of a wide array of chemicals which are derived from acetyl-CoA. Moreover, whereas the synthetic pathway is highly thermodynamically favorable (MDF > 12 kJ/mol), the RuMP cycle is only moderately favorable (MDF ~ 4 kJ/mol). Overall, the homoserine cycle seems to display clear advantages when compared to all of its natural counterparts.

By using dedicated auxotrophic strains we were able to demonstrate that the promiscuous enzyme activities required for the homoserine cycle are sufficiently high to enable the production of biomass building blocks. Quite surprisingly, we identified four native aldolase enzymes that can catalyze the HAL reaction at a sufficient rate to enable growth of the auxotrophic strain. While dedicated overexpression of the promiscuous aldolase enzymes was necessary to establish the required activity, the needed aminotransferase reaction was supported by endogenously expressed aminotransferases.

The findings of this study provide a firm basis for future experiments towards the establishment of methylotrophic growth via the full homoserine cycle. Protein engineering and evolution can be applied to improve the activity of the promiscuous enzymes and thus enhance pathway kinetics. Specifically, while LtaE activity, which generates 10% of cellular carbon, supports a relatively high growth rate, all aldolases catalyzing the HAL reaction support a relatively low growth rate despite providing only 4% of cellular carbon. Hence, engineering these aldolases for better kinetics with formaldehyde would be highly useful.

The establishment of the activity of the full homoserine cycle would probably require long term adaptive evolution under selective conditions (Gresham and Hong, 2015). The auxotrophic strains used in this study would be useful for such evolution, as formaldehyde assimilation via the aldolase reactions must avoid the biosynthesis of serine and homoserine via other routes which would thermodynamically and

kinetically “push” the aldolases in the reverse direction (i.e., formaldehyde-producing). For example, a strain auxotrophic to both serine and homoserine, overexpressing methanol dehydrogenase and both aldolases, could be first cultivated on a limiting amount of glycine and saturating amount of pyruvate and methanol. This would select for the emergence of a strain that could grow without glycine, thus combining the activities of both aldolases. The resulting strain can be then cultivated on a limiting amount of pyruvate and saturating amount of methanol, until a strain capable of growing on methanol as a sole carbon source emerges. For such growth on methanol to arise, a delicate balance between the metabolic fluxes within the homoserine cycle and those that converge to and diverge from the pathway would have to evolve. For example, the fluxes that generate pyruvate – from serine deamination and acetyl-CoA assimilation – would have to balance the fluxes that consume pyruvate, both within the homoserine cycle as well as towards other cellular routes – for example, pyruvate oxidation to acetyl-CoA. Such an evolutionary approach was previously shown to be successful for the implementation of the Calvin Cycle in *E. coli*, where the fluxes that diverge from the cycle were downregulated to match the fluxes within the cycle (Antonovsky et al., 2016; Gleizer et al., 2019). While achieving *E. coli* growth via the homoserine cycle is undoubtedly a challenging task, it holds the promise of vast new production opportunities for the bioindustry.

4. Methods

4.1. Max-min driving force (MDF) analysis

MDF analysis (Noor et al., 2014) was applied to evaluate the thermodynamic feasibility of the homoserine cycle for acetyl-CoA production from formaldehyde. The natural serine cycle and the modified serine cycle (Yu and Liao, 2018) were also analyzed for comparison. eQuilibrator-API (Noor et al., 2013) was used for these analyses. Metabolite concentrations were constrained as described before (Noor et al., 2014) with two changes: (i) formaldehyde upper bound was set to 0.5 mM according to highest concentration *E. coli* can tolerate (He et al., 2018); (ii) glutamate and 2-oxoglutarate were set as cofactors with concentrations of 100 mM and 0.5 mM, respectively (Bennett et al., 2009). pH was assumed to be 7.5 (as *E. coli* cytoplasm) and ionic strength was assumed to be 0.25 M (as recommended (Alberly et al., 2011)). Since energetic calculation using C₁-tetrahydromethanopterin intermediates are not supported by eQuilibrator-API, we used formate as substrate for serine cycle and the modified serine cycle. The scripts and details can be found at <https://gitlab.com/hi-he/wenzhangfujian> within the “2020_Promiscuous aldolases” directory.

4.2. Flux balance analysis

Flux balance analysis was conducted in Python with COBRApy (Ebrahim et al., 2013). New reactions were added to the most updated *E. coli* genome-scale metabolic network iML1515 (Monk et al., 2017) with several curations and changes: (i) transhydrogenase (THD2pp) translocates one proton instead of two (Bizouarn et al., 2005); (ii) homoserine dehydrogenase (HSDy) produces homoserine from aspartate-semialdehyde irreversibly (as we found experimentally, see main text); (iii) GLYCK (glycerate-3P producing glycerate kinase) and POR5 (pyruvate synthase) were removed from the model as their existence in *E. coli* is highly disputable; (iv) since we introduced NAD-dependent formate dehydrogenase, the two quinone-dependent formate dehydrogenase, FDH4pp and FDH5pp, were removed from the model; (v) pyruvate formate lyase (PFL) and 2-oxobutanoate formate lyase (OBTFL) were removed from the model as they operate only under anaerobic condition. We further removed the ATP maintenance reaction (ATPM) as, rather than estimating growth rate, we used FBA to estimate the maximal yield. Biomass yield was calculated by the predicted maximal biomass reaction flux divided by the flux of methanol

Table 1
List of *E. coli* strains and plasmids used in the study.

Strain	Genotype	Source
MG1655	K-12 F ⁻ λ ⁻ <i>ibvG⁻ rfb-50 rph-1</i>	Lab collection
SIJ488	MG1655 Tn7::para-exo-beta-gam; prha-FLP; xylSpm-Iscel	(Jensen et al., 2015)
DH5α	F ⁻ <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r_K m_K⁺), λ⁻</i>	Lab collection
	SIJ488 Δ <i>frmRAB ΔserA ΔglyA</i>	This study
	SIJ488 Δ <i>frmRAB ΔserA ΔgcvTHP</i>	This study
	SIJ488 Δ <i>frmRAB ΔserA ΔglyA Δkbl-tdh</i>	This study
	SIJ488 Δ <i>frmRAB ΔserA ΔgcvTHP Δkbl-tdh</i>	This study
	SIJ488 Δ <i>frmRAB Δasd</i>	This study
PthrBC_D	SIJ488 Δ <i>P_{thrLABC}::CAP-P_{pgi-20} thrBC</i>	This study
PalaC_D	SIJ488 Δ <i>P_{alaC}::CAP-P_{pgi-20} alaC*</i>	This study
PaspC_D	SIJ488 Δ <i>P_{aspC}::CAP-P_{pgi-20} aspC</i>	This study
	SIJ488 Δ <i>frmRAB Δasd ΔP_{thrLABC}::P_{pgi-20} thrBC</i>	This study
	SIJ488 Δ <i>frmRAB Δasd ΔP_{alaC}::P_{pgi-20} alaC*</i>	This study
	SIJ488 Δ <i>frmRAB Δasd ΔP_{aspC}::P_{pgi-20} aspC</i>	This study
	SIJ488 Δ <i>frmRAB Δasd ΔP_{thrLABC}::P_{pgi-20} thrBC ΔP_{alaC}::P_{pgi-20} alaC*</i>	This study
	SIJ488 Δ <i>frmRAB Δasd ΔP_{thrLABC}::P_{pgi-20} thrBC ΔP_{aspC}::P_{pgi-20} aspC</i>	This study
Plasmid	Genes	Source
pZASS	p15A ori; Strep ^R ; P _{pgi-20}	Lab collection
pZASS-mdh	pZASS::C <i>gadHA</i>	(He et al., 2018)
pZASS-ltaE-mdh	pZASS::ltaE, C <i>gadHA</i>	This study
pZASS-ltaE	pZASS::ltaE	This study
pZASS-mdh-rhmA	pZASS::C <i>gadHA</i> , <i>rhmA</i>	This study
pZASS-rhmA	pZASS::rhmA	This study
pZASS-mdh-garL	pZASS::C <i>gadHA</i> , <i>garL</i>	This study
pZASS-garL	pZASS::garL	This study
pZASS-mdh-yagE	pZASS::C <i>gadHA</i> , <i>yagE</i>	This study
pZASS-yagE	pZASS::yagE	This study
pZASS-mdh-yjhH	pZASS::C <i>gadHA</i> , <i>yjhH</i>	This study
pZASS-yjhH	pZASS::yjhH	This study
pZASS-mdh-eda	pZASS::C <i>gadHA</i> , <i>eda</i>	This study
pZASS-eda	pZASS::eda	This study
pZASS-mdh-dgoA	pZASS::C <i>gadHA</i> , <i>dgoA</i>	This study
pZASS-dgoA	pZASS::dgoA	This study
pZASS-mdh-mhpE	pZASS::C <i>gadHA</i> , <i>mhpE</i>	This study
pZASS-mhpE	pZASS::mhpE	This study
pZASS-sdaA	pZASS::sdaA	This study

uptake. While modeling acetyl-CoA production, two unrealistic pathways were blocked by deleting reactions DRPA and PAI2T respectively. The full code, including changes to the model, reactions specific to the methanol assimilation pathways, and the reactions of each production route can be found at <https://gitlab.com/hi-he/wenzhangfujian> within the “2020_Promiscuous aldolases” directory.

4.3. Strains and genomic manipulation

Strains used in this study are listed in Table 1. An *E. coli* MG1655 derived strain SIJ488 (Jensen et al., 2015) was used as the parental strain for genomic modifications. Iterative rounds of λ-Red recombineering (Jensen et al., 2015) or P1 phage transduction (Thomason et al., 2007) were used for gene deletions. For the recombineering, selectable resistance cassettes were generated via PCR – primers 50 bp homologous arms as in (Baba et al., 2006) – using the FRT-PGK-gb2-neo-FRT (Km) cassette (Gene Bridges, Germany) for

kanamycin resistance (Km) and the pKD3 plasmid (GenBank: AY048742) (Datsenko and Wanner, 2000) as a template for chloramphenicol resistance cassettes (CAP). The procedures of the deletion, verification and antibiotic cassette removal are detailed in (Wenk et al., 2018).

A similar strategy was applied to exchange the genomic promoter of target genes. A constitutive strong promoter *pgi-20* (*P_{pgi-20}*) (Braatsch et al., 2008) and a ribosome binding site “C” (AAGTTAAGAGGCAAGA) (Zelcbuch et al., 2013) were constructed downstream of the CAP cassette using primers listed shown in Supplementary Table S2. The synthetic promoter was first introduced to the SIJ488 strain by the recombineering method; P1 transduction was then used to transfer the synthetic promoter into the selection strains. *thrB* (encoding homoserine kinase, HSK) and *thrC* (encoding threonine synthase, TS) are on the same operon with *thrL* and *thrA*. Since *thrL* encodes regulatory peptide and *thrA* is redundant in the Δ*asd* selection strains, *thrLA* was deleted during the promoter exchange of *thrBC*. The point mutations A142P Y275D (Bouzon et al., 2017) were introduced along with the promoter exchange of *alaC* (In this case, the recombineering cassette has the mutated gene downstream the CAP cassette and synthetic promoter). Promoter exchanges were confirmed by sequencing the promoter regions.

4.4. Plasmids construction

All cloning procedures were carried out in *E. coli* DH5α strain. *E. coli* native genes *ltaE*, *rhmA*, *garL*, *yagE*, *yjhH*, *eda*, *dgoA*, *mhpE*, and *SdaA* were cloned from *E. coli* MG1655 genome with the primers shown in Supplementary Table S2. NAD-dependent methanol dehydrogenase (CgAdhA) was taken from *Corynebacterium glutamicum* R after codon optimization (He et al., 2018). Genes were inserted into a pNivC vector downstream of a ribosome binding site “C” (AAGTTAAGAGGCAAGA) (Zelcbuch et al., 2013). The genes were assembled into one operon using BioBrick enzymes: *BcuI*, *Sall*, *NheI* and *XhoI* (FastDigest, Thermo Scientific) (Zelcbuch et al., 2013). Using *EcoRI* and *PstI*, the synthetic operon was then inserted into an overexpression pZASS vector (Wenk et al., 2018) under a constitutive strong promoter *pgi-20* (Braatsch et al., 2008). The final plasmids are listed in Table 1.

4.5. Growth media

LB medium (0.5% yeast extract, 1% tryptone, 1% NaCl) was used for strain engineering and recombinant plasmids cloning. Antibiotics were used at the following concentrations: kanamycin, 50 μg/mL; ampicillin, 100 μg/mL; streptomycin, 100 μg/mL; chloramphenicol, 30 μg/mL. Growth experiments were performed in M9 minimal media (47.8 mM Na₂HPO₄, 22 mM KH₂PO₄, 8.6 mM NaCl, 18.7 mM NH₄Cl, 2 mM MgSO₄ and 100 μM CaCl₂), supplemented with trace elements (134 μM EDTA, 31 μM FeCl₃, 6.2 μM ZnCl₂, 0.76 μM CuCl₂, 0.42 μM CoCl₂, 1.62 μM H₃BO₃, 0.081 μM MnCl₂). Additional 50 μM MnCl₂ was added for all experiments since it improves *in vivo* activity of LtaE (Supplementary Fig. S3). Carbon sources were added according to the strain and the specific experiment (if not stated otherwise in the text): 10 mM glucose, 500 mM methanol, 10 mM glycine, 10 mM serine, 10 mM threonine, 2 mM homoserine, and 1 mM isoleucine. 0.25 mM diaminopimelate (DAP) was supplemented in all media used to cultivate the Δ*asd* strain (Cardineau and Curtiss, 1987).

4.6. Growth experiments

Strains were precultured in 4 mL M9 medium with proper carbon sources and streptomycin. The precultures were harvested and washed three times in M9 medium, then inoculated in M9 media with suitable carbon sources, with a starting OD₆₀₀ of 0.02. 150 μL of culture were added to each well of 96-well microplates (Nunclon Delta Surface, Thermo Scientific). Further 50 μL mineral oil (Sigma-Aldrich) was

added to each well to avoid evaporation (while enabling gas diffusion). The 96-well microplates were incubated at 37°C in microplate reader (BioTek EPOCH 2). The shaking program cycle (controlled by Gen5 v3) had 4 shaking phases, lasting 60 s each: linear shaking followed by orbital shaking, both at an amplitude of 3 mm, then linear shaking followed by orbital shaking both at an amplitude of 2 mm. The absorbance (OD₆₀₀) in each well was monitored and recorded after every three shaking cycles (~16.5 min). Raw data from the plate reader were calibrated to normal cuvette measured OD₆₀₀ values according to OD_{cuvette} = OD_{plate}/0.23. Growth parameters were calculated using MATLAB (MathWorks) based on three technical triplicates – the average values were used to generate the growth curves. Checked in MATLAB, in all cases variability between triplicates measurements were less than 5%.

4.7. Stable isotopic labeling

¹³C-Methanol, glucose-1-¹³C, glucose-2-¹³C, glucose-3-¹³C were purchased from Sigma-Aldrich. Strains were cultivated aerobically in glass tubes on M9 minimal media with the appropriate carbon sources at the concentrations mentioned above (section 4.5: Growth media). Experiments were performed in duplicates which in all cases showed identical results (± 5%). Cells were harvested at the late exponential phase. The equivalent volume of 1 mL of culture at OD₆₀₀ of 1 was harvested and washed by centrifugation. Protein biomass was hydrolyzed with 6 M HCl, at 95°C for 24 h (You et al., 2012). The samples were completely dried under a stream of air at 95°C. Hydrolyzed amino acids were analyzed with UPLC–ESI–MS as previously described (Giavalisco et al., 2011). Chromatography was performed with a Waters Acquity UPLC system (Waters), using an HSS T3 C₁₈ reversed phase column (100 mm × 2.1 mm, 1.8 μm; Waters). 0.1% formic acid in H₂O (A) and 0.1% formic acid in acetonitrile (B) were the mobile phases. The flow rate was 0.4 mL/min and the gradient was: 0 to 1 min – 99% A; 1 to 5 min – linear gradient from 99% A to 82%; 5 to 6 min – linear gradient from 82% A to 1% A; 6 to 8 min – kept at 1% A; 8–8.5 min – linear gradient to 99% A; 8.5–11 min – re-equilibrate. Mass spectra were acquired using an Exactive mass spectrometer (Thermo Scientific) in positive ionization mode, with a scan range of 50.0 to 300.0 m/z. The spectra were recorded during the first 5 min of the LC gradients. Data analysis was performed using Xcalibur (Thermo Scientific). The identification of amino acids was based on retention times and m/z, which were determined by analyzing amino acid standards (Sigma-Aldrich) under the same conditions.

4.8. Molecular phylogenetic analysis

The protein sequences of the aldolases predicted to catalyze the HAL reaction were obtained from UniProt: RhmA/YfaU P76469, GarL P23522, YagE P75682, YjhH P39359, Eda POA955, DgoA Q6BF16 and MhpE P51020. MAFFT v7 (Katoh and Standley, 2013) was used for multiple sequence alignment with default parameters. The aligned sequences were used by MEGA X (Kumar et al., 2018) with Maximum Likelihood method to construct a phylogenetic tree. The bootstrap consensus tree was generated with the setting No. of bootstrap replications to 1000.

CRediT authorship contribution statement

Hai He: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Rune Höper:** Investigation, Methodology. **Moritz Dodenhöft:** Investigation, Methodology. **Philippe Marlière:** Conceptualization, Writing - review & editing. **Arren Bar-Even:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing.

Declaration of competing interest

A.B.E. is co-founder of b.fab, aiming to commercialize microbial growth of C₁ compounds. b.fab was not involved in this study and did not fund it.

Acknowledgements

The authors thanks Anne Michaelis and Pauline Kuschel for experimental assistance, Elad Noor for help in the MDF analysis, and Charlie Cotton, Seohyoung Kim, Steffen Lindner, Nico Claassens, and Andrew Hanson for critical reading of the manuscript. This study was funded by the Max Planck Society. Hai He is funded by the China Scholarship Council.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymben.2020.03.002>.

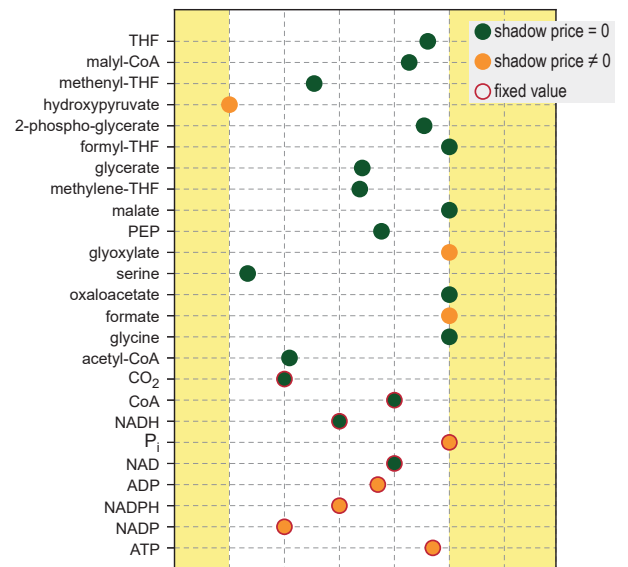
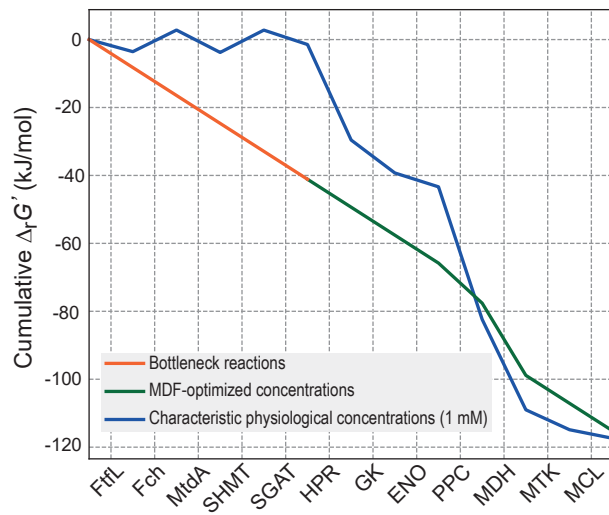
References

- Alberly, R.A., Cornish-Bowden, A., Goldberg, R.N., Hammes, G.G., Tipton, K., Westerhoff, H.V., 2011. Recommendations for terminology and databases for biochemical thermodynamics. *Biophys. Chem.* 155, 89–103. <https://doi.org/10.1016/j.bpc.2011.03.007>.
- Antonovsky, N., Gleizer, S., Noor, E., Zohar, Y., Herz, E., Barenholz, U., Zelcbuch, L., Amram, S., Wides, A., Tepper, N., Davidi, D., Bar-On, Y., Bareia, T., Wernick, D.G., Shani, I., Malitsky, S., Jona, G., Bar-Even, A., Milo, R., 2016. Sugar synthesis from CO₂ in *Escherichia coli*. *Cell* 166, 115–125. <https://doi.org/10.1016/j.cell.2016.05.064>.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H., 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006–2008. <https://doi.org/10.1038/msb4100050>.
- Bar-Even, A., 2016. Formate assimilation: the metabolic architecture of natural and synthetic pathways. *Biochemistry* 55, 3851–3863. <https://doi.org/10.1021/acs.biochem.6b00495>.
- Bar-Even, A., Flamholz, A., Noor, E., Milo, R., 2012. Thermodynamic constraints shape the structure of carbon fixation pathways. *Biochim. Biophys. Acta* 1817, 1646–1659. <https://doi.org/10.1016/j.bbabi.2012.05.002>.
- Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., Rabinowitz, J.D., 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* 5, 593–599. <https://doi.org/10.1038/nchembio.186>.
- Bizouarn, T., van Boxel, G.I., Bhakta, T., Jackson, J.B., 2005. Nucleotide binding affinities of the intact proton-translocating transhydrogenase from *Escherichia coli*. *Biochim. Biophys. Acta* 1708, 404–410. <https://doi.org/10.1016/j.bbabi.2005.04.004>.
- Bogorad, I.W., Chen, C.T., Theisen, M.K., Wu, T.Y., Schlenz, A.R., Lam, A.T., Liao, J.C., 2014. Building carbon-carbon bonds using a biocatalytic methanol condensation cycle. *Proc. Natl. Acad. Sci. U. S. A.* 111, 15928–15933. <https://doi.org/10.1073/pnas.1413470111>.
- Bouzon, M., Perret, A., Loreau, O., Delmas, V., Perchat, N., Weissenbach, J., Taran, F., Marlière, P., 2017. A synthetic alternative to canonical one-carbon metabolism. *ACS Synth. Biol.* 6, 1520–1533. <https://doi.org/10.1021/acssynbio.7b00029>.
- Braatsch, S., Helmark, S., Kranz, H., Koebmann, B., Jensen, P.R., 2008. *Escherichia coli* strains with promoter libraries constructed by Red/ET recombination pave the way for transcriptional fine-tuning. *Biotechniques* 45, 335–337. <https://doi.org/10.2144/000112907>.
- Cardineau, G.A., Curtiss 3rd, R., 1987. Nucleotide sequence of the *asd* gene of *Streptococcus mutans*. Identification of the promoter region and evidence for attenuator-like sequences preceding the structural gene. *J. Biol. Chem.* 262, 3344–3353.
- Chen, C.T., Chen, F.Y., Bogorad, I.W., Wu, T.Y., Zhang, R., Lee, A.S., Liao, J.C., 2018. Synthetic methanol auxotrophy of *Escherichia coli* for methanol-dependent growth and production. *Metab. Eng.* 49, 257–266. <https://doi.org/10.1016/j.ymben.2018.08.010>.
- Chen, Z., Geng, F., Zeng, A.P., 2015. Protein design and engineering of a *de novo* pathway for microbial production of 1,3-propanediol from glucose. *Biotechnol. J.* 10, 284–289. <https://doi.org/10.1002/biot.201400235>.
- Cheong, S., Clomburg, J.M., Gonzalez, R., 2016. Energy- and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions. *Nat. Biotechnol.* 34, 556–561. <https://doi.org/10.1038/nbt.3505>.
- Chou, A., Clomburg, J.M., Qian, S., Gonzalez, R., 2019. 2-Hydroxyacyl-CoA lyase catalyzes acyl condensation for one-carbon bioconversion. *Nat. Chem. Biol.* 15, 900–906. <https://doi.org/10.1038/s41589-019-0328-0>.
- Claassens, N.J., Cotton, C.A.R., Kopljar, D., Bar-Even, A., 2019. Making quantitative sense of electromicrobial production. *Nat Catal* 2, 437–447. <https://doi.org/10.1038/>

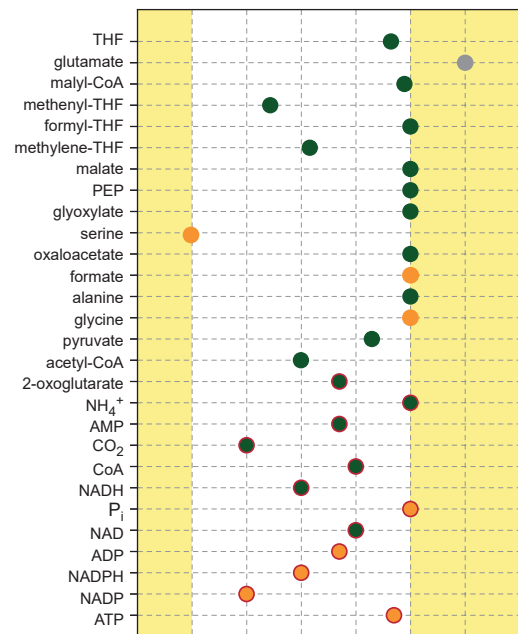
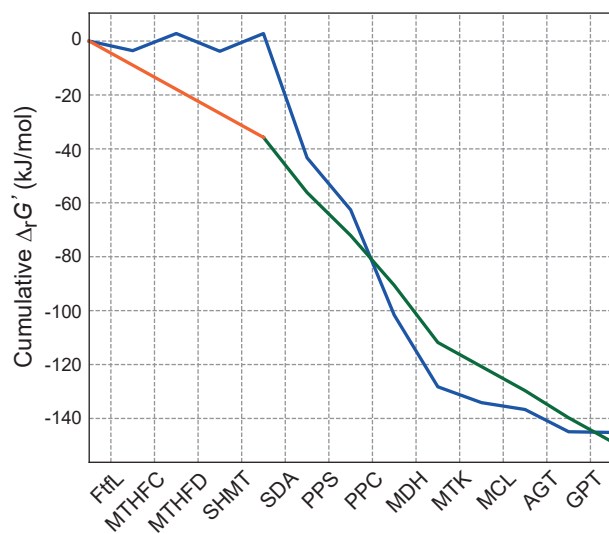
- s41929-019-0272-0.
- Contestabile, R., Paiardini, A., Pascarella, S., di Salvo, M.L., D'Aguianno, S., Bossa, F., 2001. L-Threonine aldolase, serine hydroxymethyltransferase and fungal alanine racemase. A subgroup of strictly related enzymes specialized for different functions. *Eur. J. Biochem.* 268, 6508–6525. <https://doi.org/10.1046/j.0014-2956.2001.02606.x>.
- Cotton, C.A., Claessens, N.J., Benito-Vaquero, S., Bar-Even, A., 2020. Renewable methanol and formate as microbial feedstocks. *Curr. Opin. Biotechnol.* 62, 168–180. <https://doi.org/10.1016/j.copbio.2019.10.002>.
- Crowther, G.J., Kosaly, G., Lidstrom, M.E., 2008. Formate as the main branch point for methylotrophic metabolism in *Methylobacterium extorquens* AM1. *J. Bacteriol.* 190, 5057–5062. <https://doi.org/10.1128/JB.00228-08>.
- Dai, Z., Gu, H., Zhang, S., Xin, F., Zhang, W., Dong, W., Ma, J., Jia, H., Jiang, M., 2017. Metabolic construction strategies for direct methanol utilization in *Saccharomyces cerevisiae*. *Bioresour. Technol.* 245, 1407–1412. <https://doi.org/10.1016/j.biortech.2017.05.100>.
- Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6640–6645. <https://doi.org/10.1073/pnas.120163297>.
- Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyde, D.R., 2013. COBRApy: Constraints-based reconstruction and analysis for Python. *BMC Syst. Biol.* 7, 74. <https://doi.org/10.1186/1752-0509-7-74>.
- Erb, T.J., Jones, P.R., Bar-Even, A., 2017. Synthetic metabolism: metabolic engineering meets enzyme design. *Curr. Opin. Chem. Biol.* 37, 56–62. <https://doi.org/10.1016/j.cbpa.2016.12.023>.
- Fang, J., Hait, D., Head-Gordon, M., Chang, M.C.Y., 2019. Chemoenzymatic platform for synthesis of chiral organofluorines based on Type II aldolases. *Angew. Chem. Int. Ed. Engl.* 58, 11841–11845. <https://doi.org/10.1002/anie.201906805>.
- Fesko, K., 2016. Threonine aldolases: perspectives in engineering and screening the enzymes with enhanced substrate and stereo specificities. *Appl. Microbiol. Biotechnol.* 100, 2579–2590. <https://doi.org/10.1007/s00253-015-7218-5>.
- Fitzherbert, E.B., Struebig, M.J., Morel, A., Daniels, F., Brühl, C.A., Donald, P.F., Phalan, B., 2008. How will oil palm expansion affect biodiversity? *Trends Ecol. Evol.* 23, 538–545. <https://doi.org/10.1016/j.tree.2008.06.012>.
- Giavalisco, P., Li, Y., Matthes, A., Eckhardt, A., Hubberten, H.M., Hesse, H., Segu, S., Hummel, J., Köhl, K., Willmitzer, L., 2011. Elemental formula annotation of polar and lipophilic metabolites using ¹³C, ¹⁵N and ³⁴S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J.* 68, 364–376. <https://doi.org/10.1111/j.1365-3113.2011.04682.x>.
- Gleizer, S., Ben-Nissan, R., Bar-On, Y.M., Antonovsky, N., Noor, E., Zohar, Y., Jona, G., Krieger, E., Shamshoum, M., Bar-Even, A., Milo, R., 2019. Conversion of *Escherichia coli* to generate all biomass carbon from CO₂. *Cell* 179, 1255–1263. <https://doi.org/10.1016/j.cell.2019.11.009>. e12.
- Gresham, D., Hong, J., 2015. The functional basis of adaptive evolution in chemostats. *FEMS Microbiol. Rev.* 39, 2–16. <https://doi.org/10.1111/1574-6976.12082>.
- He, H., Edlich-Muth, C., Lindner, S.N., 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. <https://doi.org/10.1021/acssynbio.8b00093>.
- Hernandez, K., Bujons, J., Joglar, J., Charnock, S.J., de Maria, P.D., Fessner, W.D., Clapes, P., 2017. Combining aldolases and transaminases for the synthesis of 2-Amino-4-hydroxybutanoic acid. *ACS Catal.* 7, 1707–1711. <https://doi.org/10.1021/acscatal.6b03181>.
- Jensen, S.I., Lennen, R.M., Herrgard, M.J., Nielsen, A.T., 2015. Seven gene deletions in seven days: fast generation of *Escherichia coli* strains tolerant to acetate and osmotic stress. *Sci. Rep.* 5, 17874. <https://doi.org/10.1038/srep17874>.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>.
- Keseler, I.M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I.T., Peralta-Gil, M., Karp, P.D., 2005. EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.* 33, D334–D337. <https://doi.org/10.1093/nar/gki108>.
- Kim, J., Copley, S.D., 2012. Inhibitory cross-talk upon introduction of a new metabolic pathway into an existing metabolic network. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2856–E2864. <https://doi.org/10.1073/pnas.1208509109>.
- Kumar, S., Stecher, G., Li, M., Nkay, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. <https://doi.org/10.1093/molbev/msy096>.
- Lessmeier, L., Pfeifenschneider, J., Carnicer, M., Heux, S., Portais, J.C., Wendisch, V.F., 2015. Production of carbon-13-labeled cadaverine by engineered *Corynebacterium glutamicum* using carbon-13-labeled methanol as co-substrate. *Appl. Microbiol. Biotechnol.* 99, 10163–10176. <https://doi.org/10.1007/s00253-015-6906-5>.
- Li, G.W., Burkhardt, D., Gross, C., Weissman, J.S., 2014. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635. <https://doi.org/10.1016/j.cell.2014.02.033>.
- 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., Jiang, H., 2019. Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design. *Nat. Commun.* 10, 1378. <https://doi.org/10.1038/s41467-019-09095-z>.
- Marx, C.J., Lidstrom, M.E., 2004. Development of an insertional expression vector system for *Methylobacterium extorquens* AM1 and generation of null mutants lacking *mtaA* and/or *fch*. *Microbiology* 150, 9–19. <https://doi.org/10.1099/mic.0.26587-0>.
- Meyer, F., Keller, P., Hartl, J., Groninger, O.G., Kiefer, P., Vorholt, J.A., 2018. Methanol-essential growth of *Escherichia coli*. *Nat. Commun.* 9, 1508. <https://doi.org/10.1038/s41467-018-03937-y>.
- Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., Feist, A.M., Palsson, B.O., 2017. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* 35, 904–908. <https://doi.org/10.1038/nbt.3956>.
- Neidhardt, F.C., Ingraham, J.L., Schaechter, M., 1990. Building blocks needed to produce 1g of *E. coli* protoplasm. *Physiology of the Bacterial Cell: A Molecular Approach* 134–143.
- Noor, E., Bar-Even, A., Flamholz, A., Reznik, E., Liebermeister, W., Milo, R., 2014. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput. Biol.* 10, e1003483. <https://doi.org/10.1371/journal.pcbi.1003483>.
- Noor, E., Haraldsdottir, H.S., Milo, R., Fleming, R.M., 2013. Consistent estimation of Gibbs energy using component contributions. *PLoS Comput. Biol.* 9, e1003098. <https://doi.org/10.1371/journal.pcbi.1003098>.
- Pfeifenschneider, J., Brautaset, T., Wendisch, V.F., 2017. Methanol as carbon substrate in the bio-economy: metabolic engineering of aerobic methylotrophic bacteria for production of value-added chemicals. *Biofuel Bioprod Biorefin* 11, 719–731. <https://doi.org/10.1002/bbb.1773>.
- Sanderson, K., 2011. Lignocellulose: a chewy problem. *Nature* 474, S12–S14. <https://doi.org/10.1038/474S012a>.
- Schada von Borzyskowski, L., Remus-Emsermann, M., Weishaupt, R., Vorholt, J.A., Erb, T.J., 2015. A set of versatile brick vectors and promoters for the assembly, expression, and integration of synthetic operons in *Methylobacterium extorquens* AM1 and other alphaproteobacteria. *ACS Synth. Biol.* 4, 430–443. <https://doi.org/10.1021/sb500221v>.
- Schrader, J., Schilling, M., Holtmann, D., Sell, D., Filho, M.V., Marx, A., Vorholt, J.A., 2009. Methanol-based industrial biotechnology: current status and future perspectives of methylotrophic bacteria. *Trends Biotechnol.* 27, 107–115. <https://doi.org/10.1016/j.tibtech.2008.10.009>.
- Szima, S., Cormos, C.-C., 2018. Improving methanol synthesis from carbon-free H₂ and captured CO₂: a techno-economic and environmental evaluation. *J. CO₂ Util.* 24, 555–563. <https://doi.org/10.1016/j.jcou.2018.02.007>.
- Takors, R., Kopf, M., Mampel, J., Bluemke, W., Blombach, B., Eikmanns, B., Bengelsdorf, F.R., Weuster-Botz, D., Dürre, P., 2018. Using gas mixtures of CO, CO₂ and H₂ as microbial substrates: the do's and don'ts of successful technology transfer from laboratory to production scale. *Microb. Biotechnol.* 11, 606–625. <https://doi.org/10.1111/1751-7915.13270>.
- Thomason, L.C., Costantino, N., Court, D.L., 2007. *E. coli* genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.* 79 (1), 1–17. <https://doi.org/10.1002/0471142727.mb0117s79>.
- 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., Bar-Even, A., 2018. Design and *in vitro* realization of carbon-conserving photorespiration. *Proc. Natl. Acad. Sci. U. S. A.* 115, E11455–E11464. <https://doi.org/10.1073/pnas.1812605115>.
- Walker, D.A., 2009. Biofuels, facts, fantasy, and feasibility. *J. Appl. Phycol.* 21, 509–517. <https://doi.org/10.1007/s10811-009-9446-5>.
- Walther, T., Calvayrac, F., Malbert, Y., Alkim, C., Dressaire, C., Cordier, H., Francois, J.M., 2018. Construction of a synthetic metabolic pathway for the production of 2,4-dihydroxybutyric acid from homoserine. *Metab. Eng.* 45, 237–245. <https://doi.org/10.1016/j.ymben.2017.12.005>.
- Wang, C., Ren, J., Zhou, L., Li, Z., Chen, L., Zeng, A.P., 2019. An aldolase-catalyzed new metabolic pathway for the assimilation of formaldehyde and methanol to synthesize 2-Keto-4-hydroxybutyrate and 1,3-propanediol in *Escherichia coli*. *ACS Synth. Biol.* 8, 2483–2493. <https://doi.org/10.1021/acssynbio.9b00102>.
- Wang, X.L., Wang, Y., Liu, J., Li, Q.G., Zhang, Z.D., Zheng, P., Lu, F.P., Sun, J.B., 2017. Biological conversion of methanol by evolved *Escherichia coli* carrying a linear methanol assimilation pathway. *Bioresour. Bioprocess* 4, 41–46. <https://doi.org/10.1186/s40643-017-0172-6>.
- Wang, Y., Fan, L., Tuyishime, P., Zheng, P., Sun, J., 2020. Synthetic methylotrophy: a practical solution for methanol-based biomanufacturing. *Trends Biotechnol.* <https://doi.org/10.1016/j.tibtech.2019.12.013>.
- Wenk, S., Yishai, O., Lindner, S.N., Bar-Even, A., 2018. An engineering approach for re-wiring microbial metabolism. *Methods Enzymol.* 608, 329–367. <https://doi.org/10.1016/bs.mie.2018.04.026>.
- Whitaker, W.B., Sandoval, N.R., Bennett, R.K., Fast, A.G., Papoutsakis, E.T., 2015. Synthetic methylotrophy: engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization. *Curr. Opin. Biotechnol.* 33, 165–175. <https://doi.org/10.1016/j.copbio.2015.01.007>.
- Withhoff, S., Schmitz, K., Niedner, S., Noh, K., Noack, S., Bott, M., Marienhagen, J., 2015. Metabolic engineering of *Corynebacterium glutamicum* for methanol metabolism. *Appl. Environ. Microbiol.* 81, 2215–2225. <https://doi.org/10.1128/AEM.03110-14>.
- 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., Ma, Y., 2019. Systematic design and *in vitro* validation of novel one-carbon assimilation pathways. *Metab. Eng.* 56, 142–153. <https://doi.org/10.1016/j.ymben.2019.09.001>.
- Yishai, O., Goldbach, L., Tenenboim, H., Lindner, S.N., Bar-Even, A., 2017. Engineered assimilation of exogenous and endogenous formate in *Escherichia coli*. *ACS Synth. Biol.* 6 (9), 1722–1731. <https://doi.org/10.1021/acssynbio.7b00086>.
- Yishai, O., Lindner, S.N., Gonzalez de la Cruz, J., Tenenboim, H., Bar-Even, A., 2016. The formate bio-economy. *Curr. Opin. Chem. Biol.* 35, 1–9. <https://doi.org/10.1016/j.cbpa.2016.07.005>.
- You, L., Page, L., Feng, X., Berla, B., Pakrasi, H.B., Tang, Y.J., 2012. Metabolic pathway confirmation and discovery through ¹³C-labeling of proteinogenic amino acids. *J. Vis. Exp.* 59. <https://doi.org/10.3791/3583>.

- Yu, H., Liao, J.C., 2018. A modified serine cycle in *Escherichia coli* converts methanol and CO₂ to two-carbon compounds. Nat. Commun. 9, 3992. <https://doi.org/10.1038/s41467-018-06496-4>.
- Yu, J.L., Xia, X.X., Zhong, J.J., Qian, Z.G., 2014. Direct biosynthesis of adipic acid from a synthetic pathway in recombinant *Escherichia coli*. Biotechnol. Bioeng. 111, 2580–2586. <https://doi.org/10.1002/bit.25293>.
- Zakaria, Z., Kamarudin, S.K., 2016. Direct conversion technologies of methane to methanol: an overview. Renew. Sustain. Energy Rev. 65, 250–261. <https://doi.org/10.1016/j.rser.2016.05.082>.
- Zelcbuch, L., Antonovsky, N., Bar-Even, A., Levin-Karp, A., Barenholz, U., Dayagi, M., Liebermeister, W., Flamholz, A., Noor, E., Amram, S., Brandis, A., Bareia, T., Yofe, I., Jubran, H., Milo, R., 2013. Spanning high-dimensional expression space using ribosome-binding site combinatorics. Nucleic Acids Res. 41, e98. <https://doi.org/10.1093/nar/gkt151>.
- Zhang, W.M., Zhang, T., Wu, S.H., Wu, M.K., Xin, F.X., Dong, W.L., Ma, J.F., Zhang, M., Jiang, M., 2017. Guidance for engineering of synthetic methylotrophy based on methanol metabolism in methylotrophy. RSC Adv. 7, 4083–4091. <https://doi.org/10.1039/c6ra27038g>.
- Zhao, M., Lu, X., Zong, H., Li, J., Zhuge, B., 2018. Itaconic acid production in microorganisms. Biotechnol. Lett. 40, 455–464. <https://doi.org/10.1007/s10529-017-2500-5>.
- Zhong, W., Zhang, Y., Wu, W., Liu, D., Chen, Z., 2019. Metabolic engineering of a homoserine-derived non-natural pathway for the *de novo* production of 1,3-propanediol from glucose. ACS Synth. Biol. 8, 587–595. <https://doi.org/10.1021/acssynbio.9b00003>.

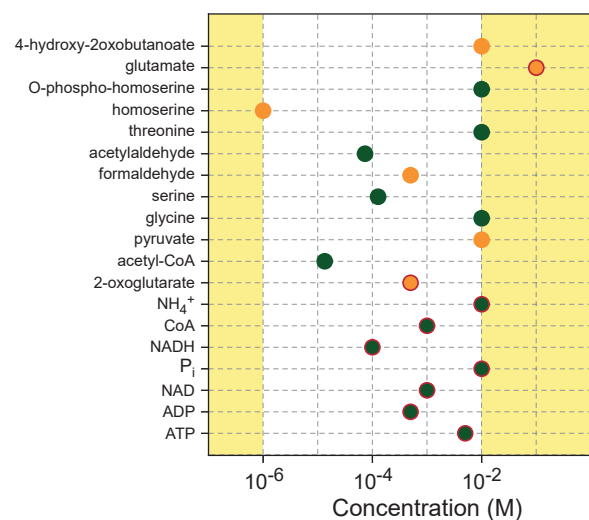
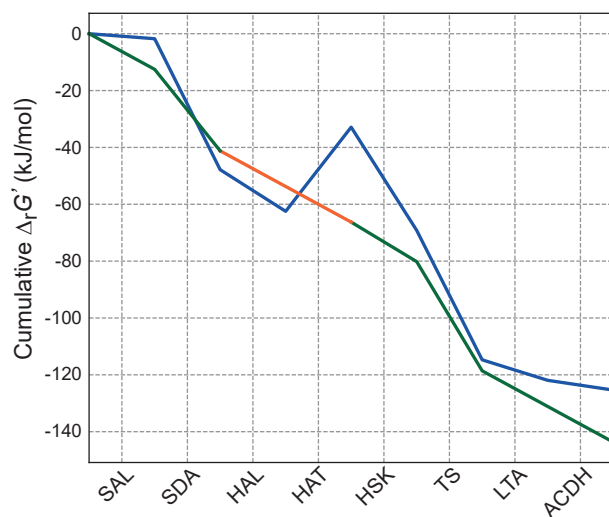
a Serine cycle



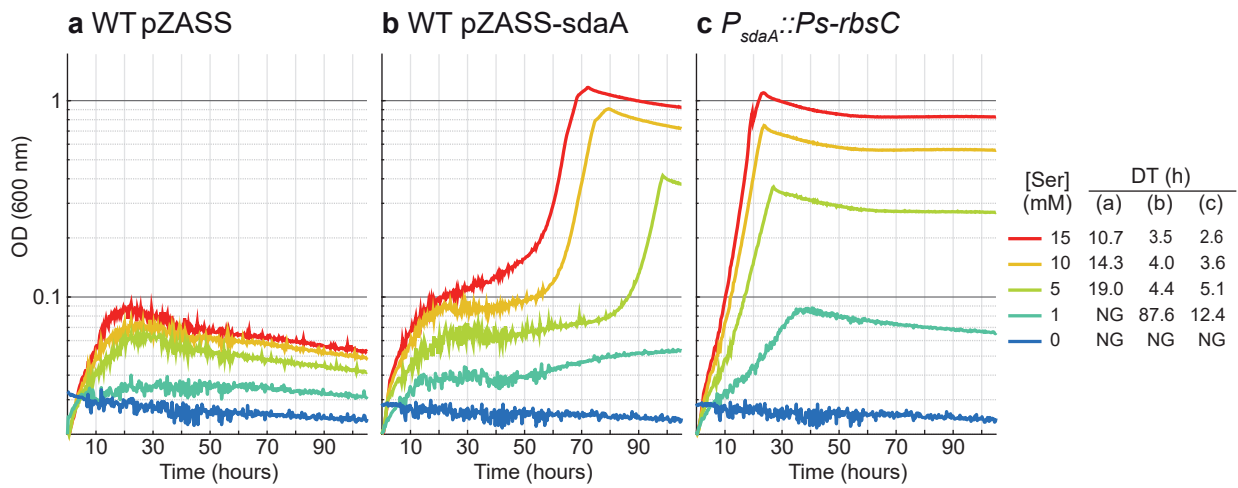
b The modified serine cycle



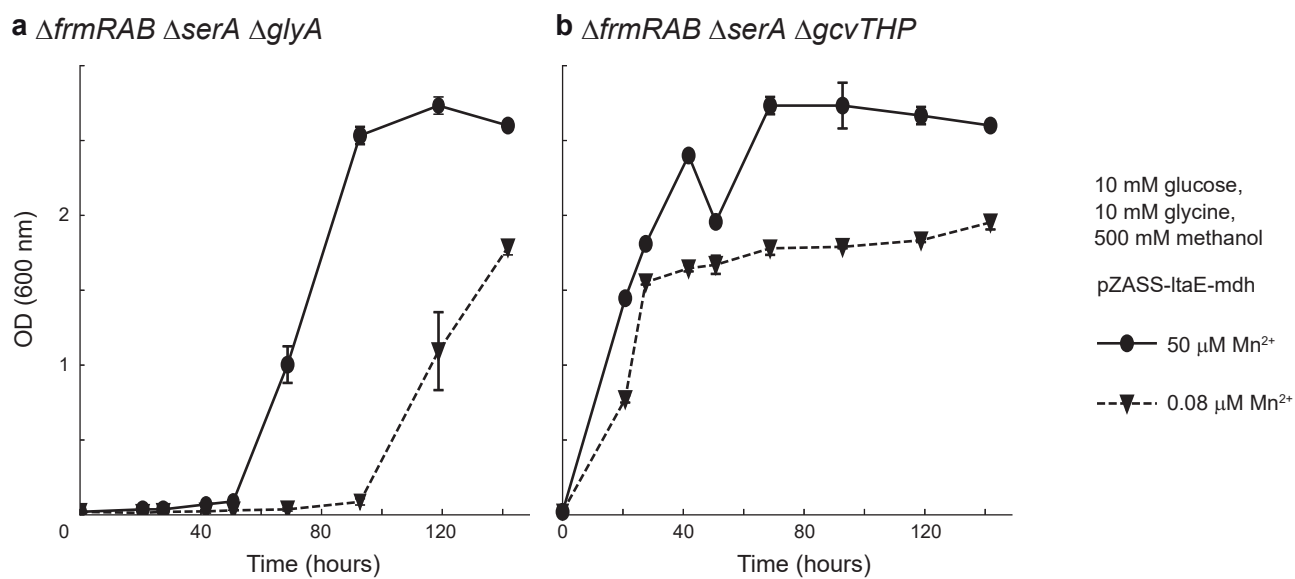
c Homoserine cycle



Supplementary Figure S1. Optimized thermodynamic profiles and Max-min Driving Force (MDF) analysis. Left panels show energetic profile of (a) the serine cycle, (b) the modified serine cycle, and (c) the homoserine cycle. Formate was assumed to be the substrate of the serine cycle and modified serine cycle, while formaldehyde was assumed to be the substrate of the homoserine cycle (see Methods for explanations). Blue lines correspond to $\Delta_rG'^\circ$ values of pathway reactions at pH 7.5. Green lines correspond to Δ_rG' values of pathway reactions after MDF optimization. The orange lines represent the predicted bottleneck reactions. Right panels show the metabolite concentrations, within physiological range (see Methods), after MDF optimization. Cofactors, shown in with a red circle, have fixed concentrations (see Methods). Compound whose concentration limits the driving force, according to the MDF analysis, are shown in orange.

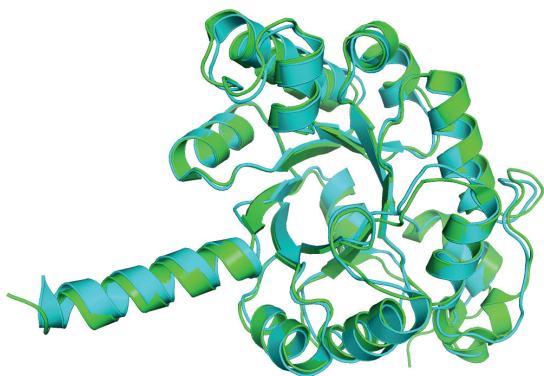


Supplementary Figure S2. *E. coli* uses serine as sole carbon source when *sdaA* is overexpressed on a plasmid or in the genome. Each growth curve represents the average of three replicates, which differ from each other by less than 5%.

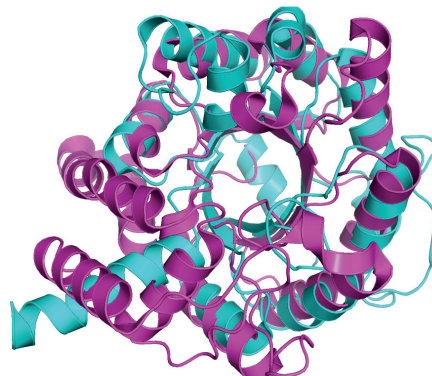


Supplementary Figure S3. Mn^{2+} supplementation improves LtaE dependent growth. While Mn^{2+} is a cofactor of LtaE, M9 medium contains only 0.08 $\mu M Mn^{2+}$. Addition of 50 $\mu M MnCl_2$ increases the growth rate and yield of the LtaE-dependent strains. Error bars represent standard deviations, $n=3$.

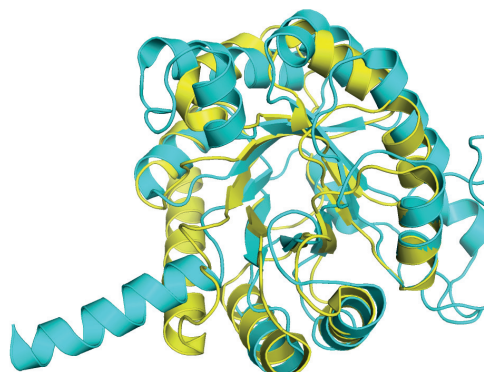
GarL-RhmA



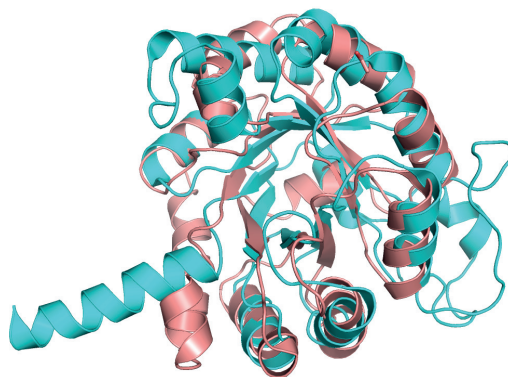
GarL-YagE



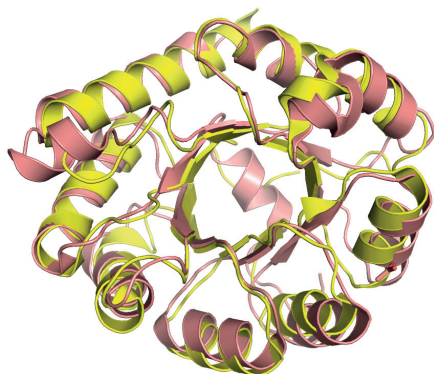
GarL-DgoA



GarL-Eda



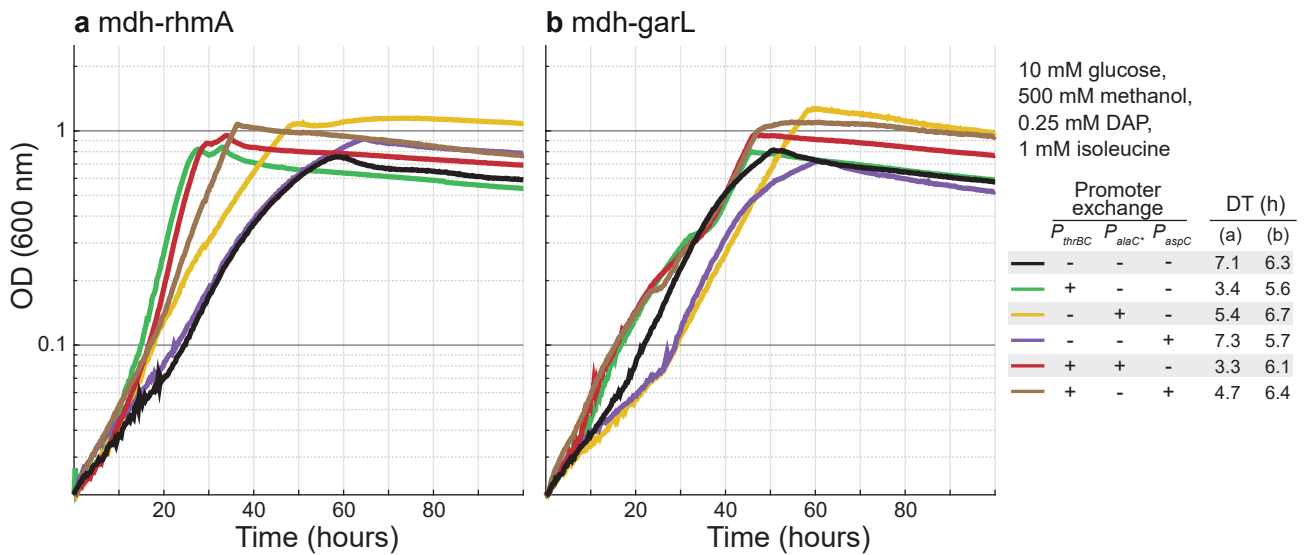
DgoA-Eda



TM-score RMSD (Å)	GarL	RhmA	YagE	DgoA	Eda
GarL (1dxf)		0.9713	0.5448	0.7012	0.6853
RhmA (2vwt)	1.07		0.5354	0.6991	0.6864
YagE (4ptn)	2.84	3.19		0.7722	0.7197
DgoA (2v82)	3.23	3.24	2.48		0.8404
Eda (1eua)	3.36	3.47	2.73	1.81	

Supplementary Figure S5. Protein structure based alignments of HAL candidates. Available protein structures, 1dxf¹ for GarL, 2vwt² for RhmA, 4ptn³ for YagE, 2v82⁴ for DgoA and 1eua⁵ for Eda were obtained from RCSB PDB. Structures was aligned by TM-align⁶ within PyMOL. Results of the alignments, RMSD and TM-score, are shown in the table. The figures were rendered using PyMOL.

1. Izard, T. & Blackwell, N.C. Crystal structures of the metal-dependent 2-dehydro-3-deoxy-galactarate aldolase suggest a novel reaction mechanism. *Embo j* **19**, 3849-3856 (2000).
2. Rea, D. et al. Crystal structure and functional assignment of YfaU, a metal ion dependent class II aldolase from *Escherichia coli* K12. *Biochemistry-US* **47**, 9955-9965 (2008).
3. Manicka, S. et al. Crystal structure of YagE, a putative DHDPS-like protein from *Escherichia coli* K12. *Proteins* **71**, 2102-2108 (2008).
4. Walters, M.J. et al. Characterization and crystal structure of *Escherichia coli* KDPGal aldolase. *Bioorganic & medicinal chemistry* **16**, 710-720 (2008).
5. Allard, J., Grochulski, P. & Sygusch, J. Covalent intermediate trapped in 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase structure at 1.95-Å resolution. *Proc. Natl Acad. Sci. USA* **98**, 3679-3684 (2001).
6. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* **33**, 2302-2309 (2005).



Supplementary Figure S6. Effects of genomic overexpression of the enzyme of the homoserine cycle. ThrB (HSK), ThrC (TS), AlaC* (AlaC A142P Y275D) or AspC were overexpressed by exchanging their native promoters with synthetic promoters within the selection strain $\Delta frmRAB \Delta asd$ (see Methods and Supplementary Table S2). Each growth curve represents the average of three replicates, which differ from each other by less than 5%.

Supplementary Table S1 Pathway reactions



RxnID	Enzyme	Gene	Rxn formula
Homoserine cycle			
MeDH	methanol dehydrogenase (NAD)	<i>adhA</i> (<i>C.g.</i>)	methanol + NAD ⁺ \rightleftharpoons formaldehyde + NADH + H ⁺
SAL	serine aldolase	<i>ltaE</i> (<i>E.c.</i>)	glycine + formaldehyde \rightleftharpoons serine
SDA	serine deaminase	<i>sdaA</i> (<i>E.c.</i>)	serine \rightleftharpoons pyruvate + NH ₃
HAL	HOB aldolase	<i>rhmA</i> (<i>E.c.</i>)/ <i>garL</i> (<i>E.c.</i>)/ <i>yagE</i> (<i>E.c.</i>)/ <i>yjhH</i> (<i>E.c.</i>)	pyruvate + formaldehyde \rightleftharpoons 4-hydroxy-2-oxobutanoate
HAT	HOB amino-transferase	<i>alaC</i> (<i>E.c.</i>)/ <i>aspC</i> (<i>E.c.</i>)	4-hydroxy-2-oxobutanoate + glutamate \rightleftharpoons homoserine + 2-oxoglutarate
HSK	homoserine kinase	<i>thrB</i> (<i>E.c.</i>)	homoserine + ATP \rightleftharpoons O-phospho-homoserine + ADP
TS	threonine synthase	<i>thrC</i> (<i>E.c.</i>)	O-phospho-homoserine + H ₂ O \rightleftharpoons threonine + Pi
LTA	threonine aldolase	<i>ltaE</i> (<i>E.c.</i>)	threonine \rightleftharpoons acetaldehyde + glycine
ACDH	acetaldehyde dehydrogenase (acetylating)	<i>adhE</i> (<i>E.c.</i>)/ <i>mhpF</i> (<i>E.c.</i>)/ <i>eute</i> (<i>E.c.</i>)	acetaldehyde + NAD ⁺ + CoA \rightleftharpoons acetyl-CoA + NADH + H ⁺
GLDH	glutamate dehydrogenase	<i>gldA</i> (<i>E.c.</i>)	2-oxoglutarate + NH ₄ ⁺ + NADPH \rightleftharpoons glutamate + NADP ⁺ + H ₂ O
FDH	formate dehydrogenase	<i>fdh</i>	formate + NAD ⁺ \rightleftharpoons CO ₂ + NADH + H ⁺
Serine cycle			
MeDH	methanol dehydrogenase (NAD)	<i>mdh</i>	methanol + NAD ⁺ \rightleftharpoons formaldehyde + NADH + H ⁺
Fae	formaldehyde-activating enzyme	<i>fae</i>	formaldehyde + H ₄ MPT \rightleftharpoons 5,10-methylene-H ₄ MPT + H ₂ O
MtdB	methylene-tetrahydromethanopterin dehydrogenase	<i>mtdB/mtDA</i>	5,10-methylene-H ₄ MPT + NAD ⁺ \rightarrow 5,10-methenyl-H ₄ MPT + NADH + H ⁺
Mch	methenyl-H ₄ MPT cyclohydrolase	<i>mch</i>	5,10-methenyl-H ₄ MPT + H ₂ O \rightarrow 5-formyl-H ₄ MPT
Fhc	formyltransferase/hydrolase complex	<i>fhc</i>	5-formyl-H ₄ MPT + H ₂ O \rightarrow formate + H ₄ MPT
FtfL	formate-tetrahydrofolate ligase	<i>fthfl</i>	formate + THF + ATP \rightleftharpoons 10-formyl-THF + ADP + Pi
Fch	methenyltetrahydrofolate cyclohydrolase	<i>fch</i>	10-formyl-THF \rightleftharpoons 5,10-methyl-THF + H ₂ O
MtdA	methylenetetrahydrofolate dehydrogenase (NADP)	<i>mtdA</i>	5,10-methyl-THF + NADPH + H ⁺ \rightleftharpoons 5,10-methylene-THF + NADP ⁺
SHMT	serine hydroxymethyltransferase	<i>glyA</i> (<i>E.c.</i>)	5,10-methylene-THF + glycine + H ₂ O \rightleftharpoons serine + THF
SGAT	serine-glyoxylate amino-transferase	<i>sgaA</i>	serine + glyoxylate \rightleftharpoons hydroxypyruvate + glycine
HPR	hydroxypyruvate reductase (NADH)	<i>ghrA</i> (<i>E.c.</i>)	hydroxypyruvate + NADH + H ⁺ \rightleftharpoons D-glycerate + NAD ⁺
GK	glycerate kinase	<i>garK</i> (<i>E.c.</i>)	D-glycerate + ATP \rightleftharpoons 2-phospho-D-glycerate + ADP
ENO	enolase	<i>eno</i> (<i>E.c.</i>)	2-phospho-D-glycerate \rightleftharpoons H ₂ O + phosphoenolpyruvate
PPC	phosphoenolpyruvate carboxylase	<i>ppc</i> (<i>E.c.</i>)	H ₂ O + CO ₂ + phosphoenolpyruvate \rightleftharpoons Pi + oxaloacetate
MDH	malate dehydrogenase	<i>maldh</i> (<i>E.c.</i>)	NADH + oxaloacetate + H ⁺ \rightleftharpoons NAD ⁺ + malate
MTK	malate thiokinase	<i>mtk</i>	ATP + CoA + malate \rightleftharpoons ADP + Pi + malylyl-CoA
MCL	malylyl-CoA lyase	<i>mcl</i>	malylyl-CoA \rightleftharpoons acetyl-CoA + glyoxylate
FDH	formate dehydrogenase	<i>fdh</i>	formate + NAD ⁺ \rightleftharpoons CO ₂ + NADH + H ⁺
The modified serine cycle (Yu and Liao 2018)			
MeDH	methanol dehydrogenase (NAD)	<i>mdh</i> (<i>C.n.</i>)	methanol + NAD ⁺ \rightleftharpoons formaldehyde + NADH + H ⁺
FALDH	formaldehyde dehydrogenase	<i>fdhA</i> (<i>P.p.</i>)	formaldehyde + NAD ⁺ \rightleftharpoons formate + NADH
FtfL	formate-tetrahydrofolate ligase	<i>fthfl</i> (<i>M.t.</i>)	formate + THF + ATP \rightleftharpoons 10-formyl-THF + ADP + Pi
MTHFC	methenyltetrahydrofolate cyclohydrolase	<i>mthfs</i> (<i>M.t.</i>)	10-formyl-THF \rightleftharpoons 5,10-methyl-THF + H ₂ O
MTHFD	methylenetetrahydrofolate dehydrogenase (NADP)	<i>mthfs</i> (<i>M.t.</i>)	5,10-methyl-THF + NADPH + H ⁺ \rightleftharpoons 5,10-methylene-THF + NADP ⁺
SHMT	glycine hydroxymethyltransferase	<i>glyA</i> (<i>E.c.</i>)	5,10-methylene-THF + glycine + H ₂ O \rightleftharpoons serine + THF
SDA	serine deaminase	<i>sdaA</i> (<i>C.n.</i>)	serine \rightleftharpoons pyruvate + NH ₃
PPS	phosphoenolpyruvate synthase	<i>ppsA</i> (<i>E.c.</i>)	H ₂ O + ATP + pyruvate \rightleftharpoons Pi + AMP + phosphoenolpyruvate
PPC	phosphoenolpyruvate carboxylase	<i>ppc</i> (<i>E.c.</i>)	H ₂ O + CO ₂ + phosphoenolpyruvate \rightleftharpoons Pi + oxaloacetate
MDH	malate dehydrogenase	<i>maldh</i> (<i>E.c.</i>)	NADH + oxaloacetate + H ⁺ \rightleftharpoons NAD ⁺ + malate
MTK	malate thiokinase	<i>mtk</i> (<i>M.c.</i>)	ATP + CoA + malate \rightleftharpoons ADP + Pi + malylyl-CoA
MCL	malylyl-CoA lyase	<i>mcl</i> (<i>M.e.</i>)	malylyl-CoA \rightleftharpoons acetyl-CoA + glyoxylate
AGT	alanine-glyoxylate transaminase	<i>AGX1</i> (<i>S.c.</i>)	alanine + glyoxylate \rightleftharpoons pyruvate + glycine
GPT	alanine transaminase	<i>alaA</i> (<i>E.c.</i>)/ <i>alaC</i> (<i>E.c.</i>)	glutamate + pyruvate \rightleftharpoons 2-oxoglutarate + alanine
GLDH	glutamate dehydrogenase	<i>gldA</i> (<i>E.c.</i>)	2-oxoglutarate + NH ₄ ⁺ + NADPH \rightleftharpoons glutamate + NADP ⁺ + H ₂ O
FDH	formate dehydrogenase	<i>fdh</i>	formate + NAD ⁺ \rightleftharpoons CO ₂ + NADH + H ⁺

Supplementary Table S2 List of oligo primers used in this study

Primer	Sequence (5'→3')
rhmA_F	atgcatcatcaccatcaccacaacgcattattaagcaatccc
rhmA_R	gcgctagctcaataactacctttatgc
yagE_F	atgcatcatcaccatcaccacaccgcagtcgccgttggtc
yagE_R	gctagcattagcaaaagcttgagctgtg
yjhH_F	atgcatcatcaccatcaccacaaaaaattcagecggcattattcc
yjhH_R	gctagcattagactggtaaaatgccct
yjhH_C	taataaagggttgacgggctg
yjhH_D	cagcccgtcaaccctttatta
yjhH_A	gtaaccattgttgacgggcgag
yjhH_B	ctcggccgtcaacaatggttac
dgoA_F	atgcatcatcaccatcaccacacagtggaactaaactcc
dgoA_R	gctagcatcattgcaactgcctctcg
dgoA_B	tcccgtgaactccccacaatg
dgoA_C	cattgtggggagttcagcggga
eda_A	gaatgcatcatcaccatcaccacaaaaactggaaaacaagtgcagaatcaatcctgaccac
eda_B	gaacggaccgcgcaatcgccttgacgggctttcac
eda_C	gtgaaagccctgcaagcgattgcgggtccgttc
eda_D	cgctagctctagattacagcttagcgccttctacagcttcacg
mhpE_F	atgcatcatcaccatcaccacaaacggtaaaaaactttatatctcggacg
mhpE_R	gctagcattatttgttgcgcagatc
ltaE_F	atgcatcatcaccatcaccacattgatttacgcagtgataccgttaccgacc
ltaE_B	gctgcgtagcttgcagagcattaac
ltaE_C	gtaaatgctctgcaagactacgcagc
ltaE_R	ctcttactgccccgatcaacgctagcttaacgcgccaggaatgcacgccag
garL_F	gtaagaggcaagaatgcataataacgatgtttcccga
garL_R	ccgcgctagctctagattatttttaaggtatcagccagt
sdaA_F	atgcatcatcaccatcaccacattagtctattcgacatgtttaaggtgg
sdaA_R	ctcttactgccccgatcaacgctagcttagtcacactggactttgattgcc
alaC_F	atgcatcatcaccatcaccacgctgacactgccctgaa
alaC_C	gcgcggtgattccaggggcgcaggtta
alaC_B	tacctgcgccctggaatcaccgcgc
alaC_E	gctatcacgatgacggcacctttacg
alaC_D	cgtaaaggtgccgtcatcgtgatagc
alaC_R	gctagcttattccgcgttttcgtgaa
PalaC_R	agtaaaccgtcggcacggaacatc
PalaC_F	ctctatgataggtaacctgaaggctgatgaccagcagggcgttttgaggaattaaccctcactaaagggcg
PthrBC_R	aaccgacgctcatattggcactggaagccggggcataaaacttaaccattctgcctcttaactttaaag
PthrBC_F	caatgttgaccgtttgctgcatgatattgaaaaaatacaccataaaattaaccctcactaaagggcg
PaspC_F	ggtcctgttttttatacttccagagcaatctcacgcttgcaaaaacaattaaccctcactaaagggcg
PaspC_R	gccaggcccagaatcgggtcggcaggagcggcggtaatgttctcaaacattctgccttactttaaag
PsdaA_F	ccgcaggcataatcgtgagctggcgtgcaaattggtgtgaaaccctgaaattaaccctcactaaagggcg
PsdaA_R	gaagatgagggaaccaatccccacttaaacatgtcgaatagactaatcactctgcctcttaactttaaag

Article

In Vivo Rate of Formaldehyde Condensation with Tetrahydrofolate

Hai He ¹, Elad Noor ², Perla A. Ramos-Parra ^{3,4}, Liliana E. García-Valencia ^{3,4}, Jenelle A. Patterson ⁵, Rocío I. Díaz de la Garza ^{3,4}, Andrew D. Hanson ⁵ and Arren Bar-Even ^{1,*}

- ¹ Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany; he@mpimp-golm.mpg.de
- ² Institute of Molecular Systems Biology, ETH Zürich, Otto-Stern-Weg 3, 8093 Zürich, Switzerland; noor@imsb.biol.ethz.ch
- ³ Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Ave. Eugenio Garza Sada 2501, Monterrey N.L. 64849, Mexico; perlaramos@tec.mx (P.A.R.-P.); liligarcia@tec.mx (L.E.G.-V.); rociodia@tec.mx (R.I.D.d.I.G.)
- ⁴ Tecnológico de Monterrey, Centro de Biotecnología-FEMSA, Ave. Eugenio Garza Sada 2501, Monterrey N.L. 64849, Mexico
- ⁵ Horticultural Sciences Department, University of Florida, Gainesville, FL 32611, USA; patterson.j@ufl.edu (J.A.P.); adha@ufl.edu (A.D.H.)
- * Correspondence: bar-even@mpimp-golm.mpg.de; Tel.: +49 331 567-8910

Received: 29 December 2019; Accepted: 11 February 2020; Published: 12 February 2020



Abstract: Formaldehyde is a highly reactive compound that participates in multiple spontaneous reactions, but these are mostly deleterious and damage cellular components. In contrast, the spontaneous condensation of formaldehyde with tetrahydrofolate (THF) has been proposed to contribute to the assimilation of this intermediate during growth on C₁ carbon sources such as methanol. However, the in vivo rate of this condensation reaction is unknown and its possible contribution to growth remains elusive. Here, we used microbial platforms to assess the rate of this condensation in the cellular environment. We constructed *Escherichia coli* strains lacking the enzymes that naturally produce 5,10-methylene-THF. These strains were able to grow on minimal medium only when equipped with a sarcosine (*N*-methyl-glycine) oxidation pathway that sustained a high cellular concentration of formaldehyde, which spontaneously reacts with THF to produce 5,10-methylene-THF. We used flux balance analysis to derive the rate of the spontaneous condensation from the observed growth rate. According to this, we calculated that a microorganism obtaining its entire biomass via the spontaneous condensation of formaldehyde with THF would have a doubling time of more than three weeks. Hence, this spontaneous reaction is unlikely to serve as an effective route for formaldehyde assimilation.

Keywords: one-carbon metabolism; spontaneous reaction; auxotrophy; serine cycle; phenotypic phase plane

1. Introduction

Most metabolic conversions are catalyzed by dedicated enzymes or by promiscuous enzymes that evolved to catalyze other reactions [1]. Nevertheless, some reactions occur spontaneously at a sufficient rate, making enzymatic catalysis unnecessary. Spontaneous reactions seem to be prevalent in specialized metabolism, where reaction rate and specificity are of lesser importance [2]. Yet, spontaneous reactions also take place in central metabolism. For example, proline biosynthesis involves the spontaneous cyclization of glutamate 5-semialdehyde to pyrroline 5-carboxylate [3]; leucine biosynthesis involves the spontaneous decarboxylation of 2-isopropyl-3-oxosuccinate; and

formaldehyde detoxification in various organisms involves the spontaneous condensation of this reactive molecule with glutathione [4].

In fact, the high reactivity of formaldehyde makes it a prime candidate for participating in spontaneous reactions. Most of these reactions are unproductive and deleterious; for example, the reaction of formaldehyde with the side-chains of arginine, cysteine, histidine, and lysine within proteins results in their inactivation and cross-linking [5]. In other cases, spontaneous reactions of formaldehyde might play a physiological role. An especially interesting case is methylotrophic microorganisms that use methanol as a sole carbon source [6]. As formaldehyde is a key intermediate in methanol assimilation, its condensation with other molecules is essential for the conversion of the C_1 feedstock into multi-carbon metabolites. Some of the formaldehyde condensation reactions in methylotrophic organisms were speculated to be spontaneous. For example, condensation of formaldehyde with tetrahydromethanopterin, which plays a key role in the methylotrophic serine cycle [7], was once thought to proceed only spontaneously, but later an enzyme catalyzing this reaction was identified ('formaldehyde-activating enzyme', Fae) [8].

Another spontaneously occurring formaldehyde condensation reaction occurs with the universal C_1 carrier, tetrahydrofolate (THF) [9]. This reaction was speculated to contribute to formaldehyde assimilation via the serine cycle [10–12]. However, a later study cast doubt on the physiological role of this reaction, which seems to proceed at a rate too slow to be of importance [7]. Still, although the kinetics of the spontaneous condensation of formaldehyde and THF have been measured *in vitro* [13], the *in vivo* reaction rate—that is, reaction kinetics under physiological conditions and concentrations—have never been measured. It therefore remains unclear whether this spontaneous reaction can play a physiological role in formaldehyde metabolism and assimilation.

Here, we construct microbial platforms to directly assess the *in vivo* condensation rate of formaldehyde and THF. We generate an *Escherichia coli* strain deleted in the two enzymes that produce 5,10-methylene-THF, the key metabolic precursor of the cellular C_1 moieties (Figure 1). Using this strain, we show that 5,10-methylene-THF can be produced *in vivo* via the condensation of formaldehyde and THF, but only at high formaldehyde concentrations and supporting only low growth rates. We then use computational modeling to derive the *in vivo* condensation rate from the measured growth rates. From the rate of this spontaneous reaction we deduce that a microbe depending solely on the condensation with THF for formaldehyde assimilation could not exceed the very low growth rate of 0.0012 h^{-1} (i.e., a doubling time of 580 h or 24 days). Hence, while the spontaneous condensation of formaldehyde and THF is feasible under *in vivo* conditions, its slow rate practically rules out a substantial contribution to growth on C_1 feedstocks.

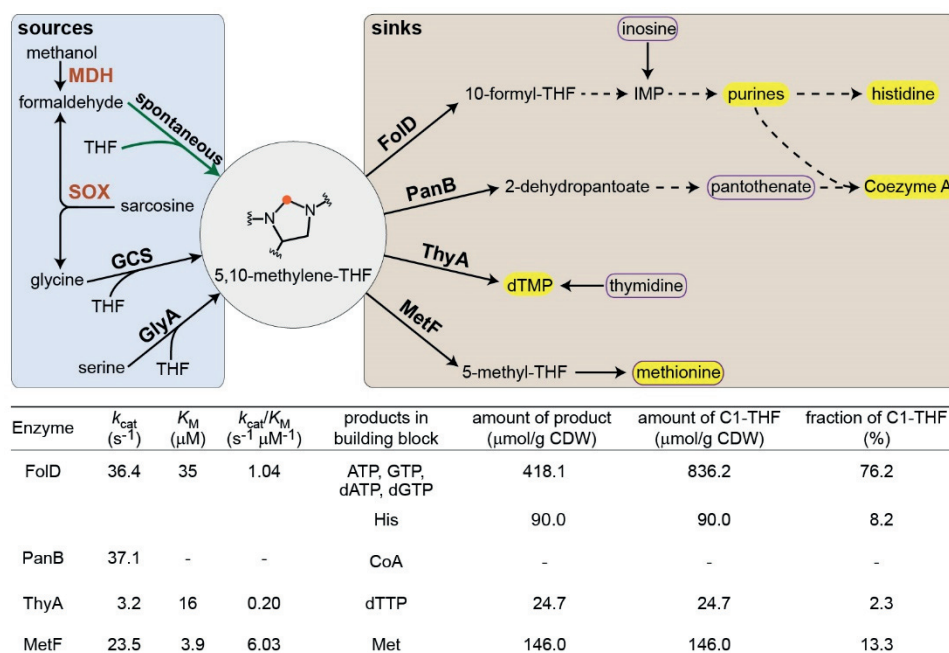


Figure 1. Metabolic sources and sinks for 5,10-methylene-THF. C₁-derived biomass building blocks are shown within yellow frame, while compounds added to the medium to relieve the C₁ auxotrophy are shown within a purple box. GCS corresponds to the glycine cleavage system. Formaldehyde is produced via the activity of either methanol dehydrogenase (MDH) or sarcosine oxidase (SOX). The kinetic parameters of the enzymes utilizing 5,10-methylene-THF [14–17] are shown together with the metabolic requirements for their products [18].

2. Results

In *E. coli*, as in many other organisms, the essential cellular C₁ moieties are derived from 5,10-methylene-THF (Figure 1). This metabolite is produced either via serine cleavage, as catalyzed by serine hydroxymethyltransferase (encoded by *glyA* in *E. coli*), or via glycine cleavage, as catalyzed by the glycine cleavage system (GCS, encoded by *gcvTHP* in *E. coli*) [19]. We constructed a $\Delta glyA \Delta gcvTHP$ strain, auxotrophic for glycine and all cellular C₁ moieties [20]. This strain could grow on a minimal medium only if provided with glycine and several compounds whose biosynthesis requires a C₁ moiety (Figure 2a, growth rate of $0.25 h^{-1} \pm 0.007$, SD with $n = 3$): inosine, as a source of purines and histidine; pantothenate, as a source of coenzyme A; thymidine; and methionine (boxed in purple in Figure 1).

We reasoned that the in vivo condensation with THF would enable formaldehyde to serve as a precursor for 5,10-methylene-THF, making the addition of the C₁-derived compounds redundant (Figure 1). As direct addition of formaldehyde to the medium is restricted to very low concentrations, due to its toxicity, we decided to explore the addition of compounds that can be converted to formaldehyde in situ [21]. We explored both methanol and sarcosine (i.e., *N*-methyl-glycine) as precursors of formaldehyde. Methanol can be oxidized to formaldehyde via the activity of methanol dehydrogenase (MDH), although the reaction rate is highly constrained by unfavorable thermodynamics and kinetics [21]. On the other hand, sarcosine can be converted irreversibly to formaldehyde (and glycine) via the activity of sarcosine oxidase (SOX), an enzyme characterized by a high rate [22]. Indeed, in a previous study, we found that while the methanol-dependent system enabled in vivo formaldehyde accumulation up to 50 μM only, the sarcosine-dependent system supported formaldehyde accumulation to concentrations higher than 500 μM [21]. Such formaldehyde accumulation relies on the disruption of the glutathione-dependent formaldehyde-oxidation pathway ($\Delta frmRAB$), which otherwise depletes the pool of this reactive intermediate [21].

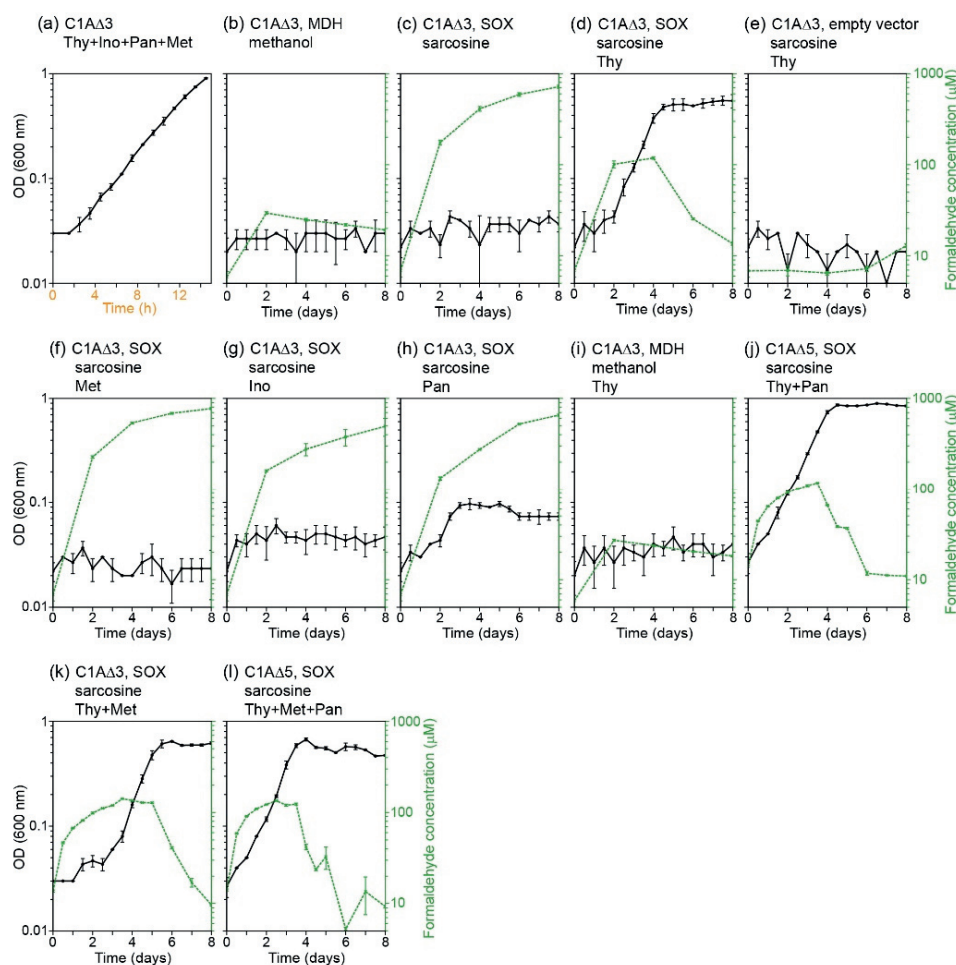


Figure 2. Growth of the C1AΔ3 and C1AΔ5 strains with formaldehyde as source of 5,10-methylene-THF. The strain and carbon sources used in each experiment are shown above each box. Note that the scale of x-axis in panel (a) is different from the others. C1AΔ3 corresponds to $\Delta glyA \Delta gcvTHP \Delta frmRAB$, C1AΔ5 corresponds to $\Delta glyA \Delta gcvTHP \Delta frmRAB \Delta panB \Delta ltaE$, MDH to methanol dehydrogenase, and SOX to sarcosine oxidase (Table 1). Strains were cultivated on a minimal medium with 10 mM glucose, supplemented either with 1 mM sarcosine or with 700 mM methanol + 1 mM glycine, and further supplemented (as indicated in the title) with 0.3 mM thymidine (Thy), 0.3 mM methionine (Met), 0.3 mM inosine (Ino), or 6 μM pantothenate (Pan). Solid black lines correspond to OD₆₀₀, while dashed green lines indicate formaldehyde concentrations. Error bars represent standard deviations with n = 3.

We therefore constructed a $\Delta glyA \Delta gcvTHP \Delta frmRAB$ strain, which we termed C1AΔ3 (C₁ auxotroph with three gene/operon deletions), in which we expressed either MDH or SOX (Table 1). Neither enzyme was able to support growth without the added C₁-derived compounds when supplemented with glycine and methanol or sarcosine, respectively (Figure 2b,c). However, the C1AΔ3 strain expressing SOX was able to grow when supplemented with sarcosine and thymidine (Figure 2d). This indicates that the in vivo condensation of formaldehyde and THF can generate 5,10-methylene-THF for the biosynthesis of all C₁-derived compounds except thymidine. Growth was not possible when SOX was not expressed (Figure 2e), confirming that thymidine itself cannot provide 5,10-methylene-THF for growth.

To unequivocally confirm that 5,10-methylene-THF is derived from formaldehyde, we conducted a ¹³C-labeling experiment. We fed the strain with thymidine and sarcosine-(methyl-¹³C) and measured the labeling profile of proteinogenic amino acids. As expected, we found methionine and histidine, both harboring a carbon that is derived from 5,10-methylene-THF [20], to be singly labeled. Glycine, serine, and threonine, serving as a control, were unlabeled (Figure 3).

Table 1. List of *E. coli* strains and plasmids.

Strain	Genotype	Source
SIJ488	K-12 MG1655 Tn7::para-exo-beta-gam; prha-FLP; xylSpm-IsceI	[23]
DH5 α	F ⁻ <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20</i> ϕ 80dlacZ Δ M15 Δ (<i>lacZYA-argF</i>)U169, <i>hsdR17</i> (r _K ⁻ m _K ⁺), λ -	Lab collection
Frm	SIJ488 Δ <i>frmRAB</i>	[21]
<i>gcv</i> _don	K-12 MG1655 Δ <i>gcvTHP::Km</i>	Lab collection
<i>ltaE</i> _don	K-12 BW25113 Δ <i>ltaE::Km</i>	[24]
<i>panB</i> _don	K-12 BW25113 Δ <i>panB::Km</i>	[24]
<i>glyA</i> _don	K-12 BW25113 Δ <i>glyA::Km</i>	[24]
C1A Δ 3	SIJ488 Δ <i>frmRAB</i> Δ <i>gcvTHP</i> Δ <i>glyA::Km</i>	This study
C1A Δ 5	SIJ488 Δ <i>frmRAB</i> Δ <i>gcvTHP</i> Δ <i>ltaE</i> Δ <i>panB</i> Δ <i>glyA::Km</i>	This study
Plasmid	Genes	Source
pZASS	p15A ori; Strep ^R ; P _{pgi-20}	Lab collection
pZASS- <i>soxA</i>	pZASS:: <i>soxA</i> ; sarcosine oxidase	[21]
pZASS-C <i>gadhA</i>	pZASS:: <i>CgadhA</i> ; methanol dehydrogenase	[21]

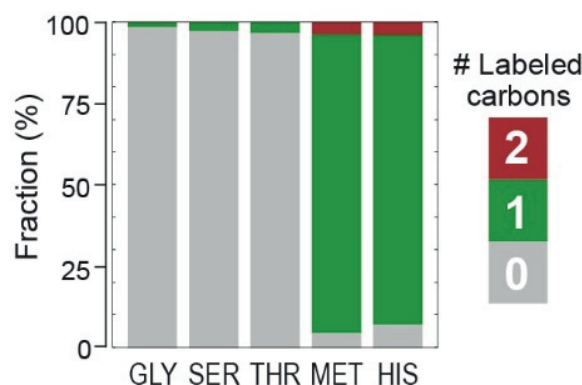


Figure 3. ¹³C-labeling confirms that 5,10-methylene-THF is produced solely from formaldehyde. Labeling pattern of proteinogenic glycine (GLY), serine (SER), threonine (THR), methionine (MET), and histidine (HIS) upon feeding the C1A Δ 3 strain with unlabeled glucose, glycine, thymidine, as well as sarcosine-(methyl-¹³C). Methionine and histidine contain a carbon derived from 5,10-methylene-THF and thus are fully singly labeled, while glycine, serine, and threonine serve as a control.

Unlike with thymidine, the addition of methionine, inosine, or pantothenate separately did not support growth even in the presence of sarcosine (Figure 2f–h). Why is the addition of thymidine necessary for growth while the other C₁-derived compounds can be produced from formaldehyde? As thymidine biosynthesis requires a considerably lower amount of 5,10-methylene-THF than that required for the biosynthesis of purines, histidine, and methionine (Figure 1), the reason cannot be related to an especially higher metabolic burden of diverting 5,10-methylene-THF towards the production of thymidine. Instead, the kinetics of the different enzymes that use 5,10-methylene-THF may explain the requirement for thymidine supplementation. Among these enzymes, thymidylate synthase (ThyA) has the lowest reported *k*_{cat} and *k*_{cat}/*K*_M values with respect to 5,10-methylene-THF (Figure 1). Hence, it is probably outcompeted by the other enzymes for the use of 5,10-methylene-THF produced from formaldehyde, rendering the cell auxotrophic for thymidine and thus non-viable [25].

The concentration of formaldehyde differed substantially according to the conditions. In conditions which did not support growth, the continuous oxidation of sarcosine by the cells present in the inoculation resulted in a monotonic increase in formaldehyde concentration, reaching 1000 μ M (Figure 2c,f–h). However, in the growing culture, formaldehyde concentration plateaued at 100 μ M during exponential growth, and then decreased during the stationary phase (Figure 2d). This indicates active metabolism of formaldehyde by the cells.

Notably, the strain expressing MDH was not able to grow when supplemented with glycine, methanol and thymidine (Figure 2i). As a formaldehyde concentration of $\geq 100 \mu\text{M}$ seems to be necessary for a sufficiently high rate of its spontaneous condensation with THF (Figure 2d), the inability to grow with methanol could be explained by the low formaldehyde concentration observed with methanol oxidation, 20–30 μM (Figure 2b,i).

It is possible that the condensation of formaldehyde and THF is catalyzed promiscuously by an *E. coli* enzyme, thus masking the spontaneous reaction. Reviewing *E. coli* native enzymes, we identified two that might be able to catalyze the condensation reaction promiscuously: (i) 3-methyl-2-oxobutanoate hydroxymethyltransferase (PanB), some variants of which are known to condense formaldehyde and 3-methyl-2-oxobutanoate to give 2-dehydropantoate [26], which could then donate a C_1 moiety to THF via the reversal of the major enzyme activity. (ii) Threonine aldolase (LtaE), an enzyme evolutionarily related to serine hydroxymethyltransferase, which might be able to promiscuously accept formaldehyde and THF [27].

We therefore constructed a $\Delta\text{glyA } \Delta\text{gcvTHP } \Delta\text{frmRAB } \Delta\text{panB } \Delta\text{ltaE}$ strain, which we termed C1A Δ 5 (C_1 auxotroph with five gene/operon deletions, Table 1). Due to the deletion of *panB*, this strain requires the addition of pantothenate to the medium regardless of the presence of formaldehyde. The C1A Δ 5 strain expressing SOX displayed the same growth phenotype as the C1A Δ 3 strain when cultivated on sarcosine and thymidine (and pantothenate in the case of the former strain, Figure 2j). This indicates that the observed growth is unlikely to be due to promiscuous enzyme catalysis but rather corresponds to spontaneous condensation of formaldehyde and THF (although we cannot completely rule out the existence of an uncharacterized enzyme catalyzing this reaction).

Since the growth rate observed with formaldehyde serving as a precursor of 5,10-methylene-THF and downstream metabolites (Figure 2d,j) was substantially lower than that observed when all C_1 -derived compounds were added to the medium (Figure 2a), it is clear that the spontaneous condensation of formaldehyde and THF limits growth. Therefore, there should exist a direct relationship between the formaldehyde condensation flux and the growth rate, such that the former can be deduced from the latter. Moreover, assuming a fixed rate of formaldehyde condensation with THF, the growth rate should be proportional to the metabolic requirement for 5,10-methylene-THF (further assuming that C_1 -THF is not “wasted”, e.g., via the activity of 10-formyl-THF deformylase). Importantly, in a previous study, we found that the addition of $\leq 5 \text{ mM}$ sarcosine to a ΔfrmRAB strain expressing SOX did not impair growth [21], thus confirming that the accumulated formaldehyde in this case is not toxic and does not influence the observed growth rate.

The growth rate observed without the addition of pantothenate, $0.037 \text{ h}^{-1} \pm 0.007$ (SD with $n = 9$), was not significantly different from that observed with pantothenate, $0.035 \text{ h}^{-1} \pm 0.005$ (SD with $n = 9$; p -value > 0.4 , t -test). This is expected from the minor contribution of coenzyme A to biomass, such that the metabolic requirement for 5,10-methylene-THF changes only negligibly with the addition of pantothenate. Overall, the growth rate observed with thymidine, with or without pantothenate, was $0.036 \text{ h}^{-1} \pm 0.006$ (SD with $n = 18$).

To derive the formaldehyde condensation rate from this growth rate we used a metabolic model of *E. coli* [28], from which we deleted the reactions that natively produce 5,10-methylene-THF and added a reaction that condenses formaldehyde with THF (Methods). We assumed glucose, glycine, formaldehyde, and thymidine as carbon sources and used Flux Balance Analysis (FBA) to calculate the phenotypic phase plane [29] of the formaldehyde condensation reaction. The slope of the lower bound of the phenotypic phase plane represents the expected relationship between the growth rate and the formaldehyde condensation rate: $0.7 \text{ mmol gCDW}^{-1}$ (Figure 4a; adding pantothenate barely changed the phenotypic phase plane, due to the negligible contribution of coenzyme A to biomass). By intersecting the phenotypic phase plane with the measured growth rate, we derived a formaldehyde condensation rate of $0.025 \text{ mmol gCDW}^{-1} \text{ h}^{-1} \pm 0.004$ (Figure 4a). We note that this value corresponds to the net reaction rate, that is, the rate of the condensation reaction minus the dissociation rate of 5,10-methylene-THF.

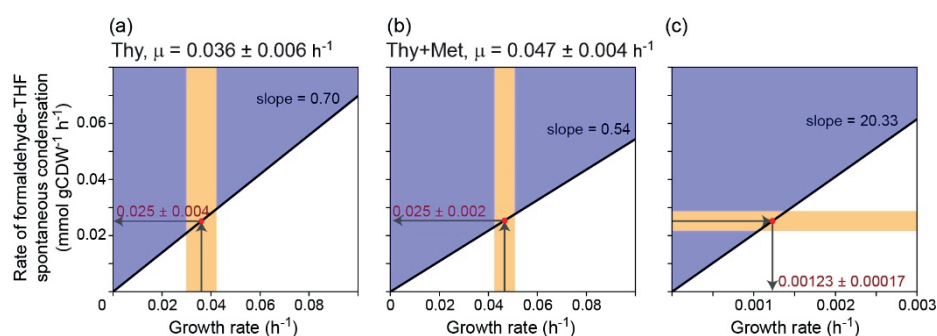


Figure 4. Computational modeling of the relationship between the growth rate and the rate of the spontaneous condensation of formaldehyde and THF. Shown are the phenotypic phase planes of the formaldehyde condensation reaction, when (a) thymidine (Thy) is added to the medium; or when (b) thymidine and methionine (Thy + Met) are added to the medium. (c) Phenotypic phase plane of the formaldehyde condensation reaction when this reaction provides the sole source of carbon via the serine cycle.

To check the robustness of this derivation, we decided to repeat it, cultivating the SOX expressing strains, both C1AΔ3 and C1AΔ5, in the same conditions as before, while further adding methionine to the medium (Figure 2k, Figure 2l, respectively). The addition of methionine is expected to substantially decrease the metabolic requirement for 5,10-methylene-THF, thus supporting a higher growth rate. Indeed, the growth rate under these conditions, $0.047 \text{ h}^{-1} \pm 0.004$ (SD with $n = 18$) was significantly higher than that observed without the addition of methionine (p -value $< 10^{-5}$, t -test; as before, addition or omission of pantothenate did not significantly change the growth rate, p -value > 0.1 , t -test).

We used FBA again to calculate the phenotypic phase plane of the formaldehyde condensation reaction when methionine is also added to the medium. We found that the correlation factor between the growth rate and the formaldehyde condensation rate (i.e., the slope of the phenotypic phase plane, Figure 4b) decreased to $0.54 \text{ mmol gCDW}^{-1}$, as expected from the lower metabolic requirement for 5,10-methylene-THF when methionine is added to the medium. By intersecting the new phenotypic phase plane with the growth rate measured with methionine addition, we derived a formaldehyde condensation flux of $0.025 \text{ mmol gCDW}^{-1} \text{ h}^{-1} \pm 0.002$ (Figure 4b), identical to what was previously calculated.

The fact that we derived an identical formaldehyde condensation rate using different growth conditions that are characterized by different growth rates validates our approach and assumptions, especially confirming that the condensation rate does not change between conditions and that the growth rate is directly related to the metabolic requirement for 5,10-methylene-THF.

To better understand the metabolic adaptation to the novel 5,10-methylene-THF biosynthesis route, we quantified the intracellular concentrations of THF and its C₁ derivatives by UPLC (Methods). We found the total THF pool in the C1AΔ3 strain to be almost 3-fold higher than in a WT strain (Figure 5). This might reflect a feedback mechanism, in which the low rate at which C₁-THFs are produced led the cell to increase the biosynthesis of the cofactor. Interestingly, the fraction of unbound THF does not seem to be higher in the C1AΔ3 strain, as [THF] + [5,10-methylene-THF] (which our analysis cannot separate) represents $\leq 10\%$ of the total THF pool both in the WT strain and in the C1AΔ3 strain. However, the THF pool in the C1AΔ3 strain is strongly shifted towards a more reduced state (Figure 5). While the oxidized 5-formyl-THF, 10-formyl-THF, and 5,10-methenyl-THF constitute 66% of the total THF pool in the WT strain, they represent only 13% of the total in the C1AΔ3 strain. In contrast, the reduced 5-methyl-THF constitutes only 27% of the THF pool in the WT strain, but represents 77% of the THF pool in the C1AΔ3 strain.

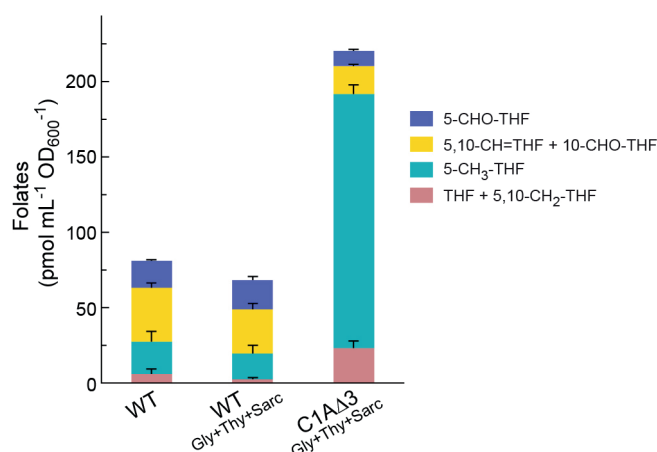


Figure 5. Intracellular content of THF and its C₁ derivatives. Strains were cultivated on a minimal medium with 10 mM glucose, supplemented with 1 mM glycine (Gly), 0.3 mM thymidine (Thy), and 1 mM sarcosine when indicated. Error bars represent standard errors, with $n = 3$.

How does the in vivo derived condensation flux compare with that expected using the in vitro measured parameters? Using the previously measured kinetic parameters, the rate of 5,10-methylene-THF formation equals $k_{\text{forward}} \times [\text{formaldehyde}] \times [\text{THF}] \times [\text{H}^+]$, where $k_{\text{forward}} = 8.7 \times 10^8 \text{ s}^{-1} \text{ M}^{-2}$ [13]. Even if we assume that the THF/5,10-methylene-THF pool (Figure 5) exists only as free THF, such that $[\text{THF}] = 23 \text{ pmol mL}^{-1} \text{ OD}_{600}^{-1} = 18 \text{ }\mu\text{M}$ (considering culture biomass of $0.39 \text{ mgCDW mL}^{-1} \text{ OD}_{600}^{-1}$ and intracellular biomass of $300 \text{ mgCDW mL}^{-1}$ [30]), further assuming $[\text{formaldehyde}] = 100 \text{ }\mu\text{M}$ (Figure 2d,j–l) and $[\text{H}^+] \approx 32 \text{ nM}$ (pH 7.5), we get a 5,10-methylene-THF formation rate of only $0.00058 \text{ mmol gCDW}^{-1} \text{ h}^{-1}$, that is, 43-fold lower than the in vivo derived rate. Since $[\text{THF}]$ should be lower than $[\text{THF}] + [5,10\text{-methylene-THF}]$, and the reverse reaction (i.e., 5,10-methylene-THF cleavage) further lowers the net formaldehyde condensation flux, the in vivo condensation flux seems likely to be 2–3 orders of magnitude higher than that expected from the in vitro parameters. While the intracellular concentration of formaldehyde might be somewhat higher than that measured in the medium, it is highly unlikely that this discrepancy accounts for the 2–3 orders of magnitude difference. Hence, it is rather clear that under cellular conditions, the condensation rate of formaldehyde with THF is substantially higher than that measured in simple model in vitro systems, further emphasizing the importance of the analysis performed here.

Next, we asked what would be the growth rate of an organism growing on formaldehyde (or a precursor thereof) using the serine cycle and assuming that 5,10-methylene-THF is produced solely via the spontaneous condensation of formaldehyde with THF. To assess this rate, we added to the model of *E. coli* the formaldehyde-THF condensation reaction as well as the specific reactions of the serine cycle (e.g., malyl-CoA synthetase and lyase, serine transaminase, see Methods). Again, we calculated the phenotypic phase plane of the formaldehyde condensation reaction, assuming formaldehyde to be the sole carbon source. As expected, the correlation factor between the growth rate and the formaldehyde condensation rate was much higher than before, $>20 \text{ mmol gCDW}^{-1}$ (Figure 4c, note the different scaling of the x-axis). By intersecting the phenotypic phase plane with the formaldehyde condensation rate calculated above, we derived a maximal growth rate of $0.0012 \text{ h}^{-1} \pm 0.0002$. (In fact, this growth rate can also be estimated directly from the growth rate of the C1AΔ3 strain, which uses formaldehyde as a source of the non-thymidine C₁ moieties (Figure 2d,j, 0.036 h^{-1}): as 2.3% of the carbons in biomass are derived from non-thymidine C₁ moieties [18], the expected growth rate when the condensation reaction provides 100% of the carbons is expected to be $0.036 \times (2.3\%/100\%) = 0.0008 \text{ h}^{-1}$, very close to the value derived by the computational modeling.) This growth rate is equivalent to a doubling time of 580 h, i.e., 24 days. Obviously, such a growth rate is impractical for most microorganisms, ruling out the spontaneous condensation of formaldehyde and THF as an effective route for formaldehyde assimilation.

3. Discussion

In this study, we constructed microbial platforms to assess the *in vivo* rate of the spontaneous condensation of formaldehyde with THF. We were able to select for this reaction as the sole source of 5,10-methylene-THF and downstream metabolites, but only when (i) sarcosine oxidation sustained a high concentration of formaldehyde and (ii) thymidine was added to the medium. The failure to obtain growth with methanol as a source of formaldehyde might be attributed to the use of an NAD-dependent MDH, which strongly favors formaldehyde reduction over methanol oxidation. A quinone-dependent MDH or an O₂-dependent methanol oxidase, both strongly favoring methanol oxidation, is expected to enable growth with methanol just as SOX enabled growth with sarcosine. The strict necessity for thymidine supplementation could be attributed to the poor kinetics of the enzyme that generates it from 5,10-methylene-THF. This kinetic constraint does not seem to hamper thymidine biosynthesis in wild-type *E. coli* presumably because the biosynthesis of 5,10-methylene-THF is not limiting. In our strain, the low production rate of 5,10-methylene-THF likely generates a “competition” between its various metabolic sinks (Figure 1), such that the kinetically poor ThyA cannot sustain sufficiently high flux to support growth.

As we demonstrated, a microorganism obtaining its entire biomass via the spontaneous condensation of formaldehyde with THF would be constrained to a very low growth rate, having a doubling time of more than three weeks. However, it is important to remember that the spontaneous condensation rate might differ between organisms. For example, it could be improved by increasing the cellular concentration of THF or operating at a higher formaldehyde concentration. Moreover, changing the cellular conditions—temperature, pH, ionic strength, and concentrations of different ions—might alter the rate at which formaldehyde reacts with THF. Whether such changes could substantially increase the spontaneous production rate of 5,10-methylene-THF, thus enabling a more reasonable growth rate, remains to be seen. It will be interesting to use adaptive laboratory evolution [31] to try to increase the spontaneous condensation rate. For example, the *E. coli* strains established here could be cultivated in a turbidostat regime [32] selecting for changes in cellular conditions that support a higher 5,10-methylene-THF production rate, thus increasing the growth rate. Alternatively, it is at least possible that a variant of the formaldehyde activating enzyme [8] has evolved, or could be evolved, to accept THF rather than tetrahydromethanopterin, thus enabling a dramatic increase in the rate of 5,10-methylene-THF production from formaldehyde.

Despite the kinetic barrier, assimilation of formaldehyde via its spontaneous condensation with THF might be advantageous in conditions that favor high biomass yield rather than growth rate. This is due to two reasons. First, microorganisms that grow on methanol via the serine cycle use a long, tetrahydromethanopterin-dependent route (Figure 6). Bypassing this route via the direct condensation of formaldehyde and THF saves the need for the costly biosynthesis of multiple proteins as well the additional cofactor. Moreover, the direct condensation route saves the ATP consumed to energize the condensation of formate with THF (Figure 6). Hence, while unlikely, we cannot rule out the possibility of a slowly growing methylotrophic microorganism, living in a resource-deprived environment, which operates the serine cycle via the spontaneous formaldehyde-THF condensation in order to save cellular resources and maximize biomass yield. We leave it to future studies to explore the possible existence of such an interesting growth mode.

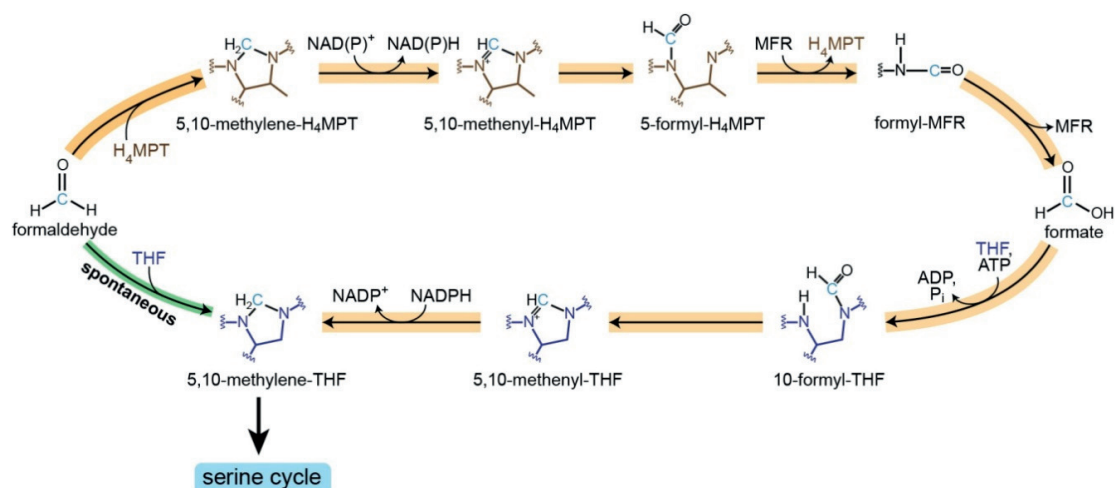


Figure 6. Spontaneous formaldehyde THF condensation bypasses the long and ATP costly tetrahydromethanopterin (H₄MPT)-dependent route.

4. Materials and Methods

4.1. Strains and Genomic Modifications

All strains used in this study are listed in Table 1. An *E. coli* K-12 MG1655 derived strain SIJ488 [23] was used as the parental strain for genomic modifications and further experiments. Multiple-gene knockouts were obtained by iterative rounds of P1 phage transduction [33] and antibiotic resistant gene cassette flip-out. Donor lysates of the transduction were prepared using “_don” strains. Gene knockouts were confirmed by PCR after antibiotic selection. Antibiotic cassette was removed by inducing Flippase [23]. Colonies growing only on plates without antibiotic were further confirmed by PCR of the cassette removal. Codon optimized methanol dehydrogenase (MDH) from *Corynebacterium glutamicum* R and sarcosine oxidase (SOX) from *Bacillus* sp. B-0618 were overexpressed in a pZASS plasmid, which has a p15A origin, a constitutive promoter pgi-20, and a streptomycin selection marker; gene sequences and cloning details were provided previously in ref. [21].

4.2. Media and Growth Conditions

LB medium (1% NaCl, 0.5% yeast extract, and 1% tryptone) was used for genomic strain modifications. Antibiotics were used at the following concentrations: kanamycin, 50 µg/mL and streptomycin, 100 µg/mL. Growth experiments were performed in M9 minimal media (47.8 mM Na₂HPO₄, 22 mM KH₂PO₄, 8.6 mM NaCl, 18.7 mM NH₄Cl, 2 mM MgSO₄, and 100 µM CaCl₂), supplemented with trace elements (134 µM EDTA, 31 µM FeCl₃·6H₂O, 6.2 µM ZnCl₂, 0.76 µM CuCl₂·2H₂O, 0.42 µM CoCl₂·2H₂O, 1.62 µM H₃BO₃, and 0.081 µM MnCl₂·4H₂O). Additional 50 µM MnCl₂ was added to relieve possible oxidative stress from sarcosine oxidation [34]. 10 mM glucose was the main carbon source and 1 mM of glycine was used to relieve the glycine auxotroph. 1 mM sarcosine or 700 mM of methanol was used as the formaldehyde source, C₁ supplements were added according to the specific experiment: 0.3 mM of thymidine, 0.3 mM of inosine, 0.3 mM of methionine, and 6 µM of pantothenate [20]. We note that the very high concentration of methanol was required for formaldehyde production because methanol oxidation is thermodynamically unfavorable and must be pushed forward. On the other hand, sarcosine oxidation is highly favorable, such that even 1 mM of this precursor is sufficient to sustain the efficient production of formaldehyde.

Strains were precultured in 4 mL M9 minimal medium with all C₁ supplements, as well as glucose and glycine. The precultures were harvested and washed three times in M9 medium, then inoculated into 4 mL M9 media with the different carbon sources as explained for each experiment at a starting OD₆₀₀ of 0.02. The same preculture was used to inoculate the three experimental replicates. The

cultures were incubated at 37 °C, 200 rpm shaken (GFL Orbital Shaker 3017, Burgwedel, Germany). Cell densities were measured by photometer (DiluPhotometer OD600, IMPLLEN, Munich, Germany) at 600 nm. Cell densities at exponential phase were fitted to $N = N_0 * e^{\mu * t}$ using LMFIT on python, where N_0 and N are cell densities at start and at time t , and μ is growth rate. Student's t -test was used for the statistical analysis.

4.3. Formaldehyde Concentration Measurement

At the indicated time points, 140 μ L of culture were collected and centrifuged for 10 min at 3300 \times g, 4 °C. Formaldehyde concentration in culture supernatants was determined by Nash assay [35]. 125 μ L of culture supernatant was mixed with 125 μ L of Nash reagent (2 M of ammonium acetate, 20 mM of acetylacetone, and 50 mM of acetic acid) in 96-well plate (Brandplates, Wertheim, Germany), and incubated at 37 °C for 1 h. The absorption at 412 nm was measured for each sample using plate reader (Infinite M200 pro, Tecan, Männedorf, Switzerland). A standard curve showed that the assay was linear at the range from 1 mM to 0 mM formaldehyde in M9.

4.4. Isotopic-labeling Experiments

For stationary isotope tracing of proteinogenic amino acids, C1A Δ 3 strain was grown in M9 with 10 mM of glucose, 0.3 mM of thymidine, and 1 mM of either unlabeled sarcosine or methyl-¹³C labeled sarcosine (Sigma-Aldrich, Steinheim, Germany). Experiments were performed in duplicate. Cells were harvested at the late exponential phase (OD 0.5). The 2 mL of culture was harvested and washed by centrifugation. Protein biomass was hydrolyzed with 6 M HCl, at 95 °C for 24 h [36]. The samples were completely dried under a stream of N₂ at 95 °C.

Hydrolyzed amino acids were analyzed with ultra-performance liquid chromatography coupled with electrospray mass spectrometry (UPLC–ESI–MS) as previously described [37]. Chromatography was performed with a Waters Acquity UPLC system (Waters), using an HSS T3 C₁₈ reversed phase column (100 mm \times 2.1 mm, 1.8 μ m; Waters). Of formic acid 0.1% in H₂O (A) and 0.1% of formic acid in acetonitrile (B) were the mobile phases. The flow rate was 0.4 mL/min and the gradient was 0–1 min—99% A; 1–5 min—linear gradient from 99% A to 82%; 5–6 min—linear gradient from 82% A to 1% A; 6–8 min—kept at 1% A; 8–8.5 min—linear gradient to 99% A; and 8.5–11 min—re-equilibrate. Mass spectra were acquired using an Exactive mass spectrometer (Thermo Scientific, Bremen, Germany) in positive ionization mode, with a scan range of 50.0–300.0 m/z . The spectra were recorded during the first 5 min of the LC gradients. Data analysis was performed using Xcalibur (Thermo Scientific, Bremen, Germany). Determination of retention times was performed by analyzing amino-acid standards (Sigma-Aldrich) under the same conditions.

4.5. Flux Balance Analysis

For Flux Balance Analysis and phenotypic phase plane calculations, we used the COBRApy package [38]. All scripts are available as Jupyter notebooks at <https://gitlab.com/hi-he/wenzhangfujian> under the directory of “2020_formaldehyde condensation”.

Adjustments made to *i*ML1515 model [28]:

- Reactions POR5 and GLYCK were removed from the model since they are not present in *E. coli* cells.
- The stoichiometry of the reaction THD2pp was changed, as the translocation uses only one proton instead of two.
- PFL and OBTFI were removed, since they are catalyzed by oxygen sensitive enzymes and therefore not active in aerobic conditions.
- Based on gathered evidence, FTHFLi was removed. This reaction is especially important in our case, since it could allow charging of THF with formate directly.

- We changed the directionality of three reactions in the nucleotide biosynthesis pathway (FTHFD, GARFT, and AICART) to be irreversible, in order to avoid nucleotides degradation to charge THF, since that is very unlikely to occur.
- We changed the thymidine periplasmic symporter reaction (THYMt3pp) to be reversible, based on our experimental evidence (adding thymidine to the media rescued growth in certain conditions).
- We knocked out the glycine cleavage system reaction and the GlyA reaction (GLYCL and GHMT2r).
- We added the formaldehyde-THF spontaneous condensation (THFSPONT) reaction: $\text{thf_c} + \text{fald_c} \rightarrow \text{mlthf_c} + \text{h2o_c}$.

We made further adjustments to the model to enable growth on formaldehyde via the serine cycle and ethylmalonyl-CoA cycle:

- Removed the following reactions to avoid potential carbon fixation cycles: GLXCL, GLYCL, THRD, THRA2, THRA, SERD_L.
- Added NAD-dependent formate dehydrogenase (FDH), enabling the utilization of the formate that is created from formaldehyde as source of reducing power.
- Added the crotonoyl-CoA carboxylase reaction (EtMaCoA), required for the ethylmalonyl-CoA cycle: $\text{b2coa_c} + \text{nadh_c} + \text{co2_c} + \text{q8_c} + \text{h2o_c} \rightarrow \text{glx_c} + \text{ppcoa_c} + \text{nadh_c} + \text{q8h2_c}$.
- Added the propionyl-CoA carboxylase reaction (MMCDr), as required for recycling the propionyl-CoA, which was created by EtMaCoA: $\text{ppcoa_c} + \text{atp_c} + \text{hco3_c} \rightarrow \text{mmcoa_S_c} + \text{adp_c} + \text{h_c} + \text{pi_c}$.
- Made the methylmalonyl-CoA mutase reaction (MMM) reversible so that the methylmalonyl-CoA (mmcoa__S_c) can be recycled.
- Added the reactions of the serine cycle:
 - serine-hydroxypyruvate amino-transferase (SGAT1): $\text{ser_L_c} + \text{akg_c} \rightarrow \text{glu_L_c} + \text{hpyr_c}$.
 - glycine-glyoxylate amino-transferase (SGAT2): $\text{gly_c} + \text{akg_c} \rightarrow \text{glx_c} + \text{glu_L_c}$.
 - malate thiokinase (MTK): $\text{mal_L_c} + \text{atp_c} + \text{coa_c} \rightarrow \text{malylcoa_c} + \text{adp_c} + \text{pi_c}$.
 - malyl-CoA lyase (MCL): $\text{malylcoa_c} \rightarrow \text{accoa_c} + \text{glx_c}$.

4.6. Folate Profiling

75 mL of each culture was incubated in 250 mL Erlenmeyer flask. At the late exponential phase, cells of each strain equivalent to approximately 5 mg total protein [30] were harvested by centrifugation (3200× g, 10 min, 4 °C) and resuspended in 10 mL of 50 mM HEPES-CHES (2-(N-cyclohexylamino)ethanesulfonic acid; pH 7.8 ± 0.05) containing 2% (*w/v*) ascorbic acid and 10 mM β-mercaptoethanol (extraction buffer). The cells were disrupted using a FastPrep Instrument (MP Biomedicals), boiled for 10 min, and centrifuged (16,000× g, 10 min, 4 °C). The supernatants were collected and treated with AtGGH2, as described by Ramos-Parra et al. [39]. Folates were purified by affinity chromatography and analyzed by UPLC-coupled electrochemical detector (CoulArray Model 5600A; ESA, Chelmsford, MA) as described previously [40]. Separation of folate derivatives was achieved with a Prodigy ODS(2) column (150 mm × 3.2 mm, 5 μm particle size, Phenomenex, Torrance, CA, USA) using a nonlinear gradient of mobile phase A (28 mM K₂HPO₄ and 59 mM H₃PO₄) and B (75% phase A and 25% acetonitrile). Detection was achieved with an electrochemical cell adjusted to 100, 200, 300, and 400 mV, and quantification was performed using commercial standards (Schircks, Jona, Bauma, Switzerland).

Author Contributions: Conceptualization, H.H., J.A.P., A.D.H. and A.B.-E.; methodology, H.H., E.N., P.A.R.-P., L.E.G.-V., R.I.D.d.I.G., J.A.P., A.D.H. and A.B.-E.; software, H.H. and E.N.; formal analysis, H.H., E.N., R.I.D.d.I.G., and A.B.-E.; investigation, H.H. P.A.R.-P., L.E.G.-V., and A.B.-E.; data curation, H.H. R.I.D.d.I.G., and A.B.-E.; writing—original draft preparation, H.H. and A.B.-E.; writing—review and editing, H.H., E.N., J.A.P., P.A.R.-P., L.E.G.-V., R.I.D.d.I.G., A.D.H. and A.B.-E.; visualization, H.H.; supervision, A.B.-E.; project administration, H.H. and A.B.-E.; funding acquisition, A.B.-E. All authors have read and agreed to the published version of the manuscript.

Funding: H.H. and A.B.-E. are funded by the Max Planck Society. H.H. is also funded by the China Scholarship Council. J.A.P. and A.D.H. are funded by the US National Science Foundation grant MCB-1611711. P.A.R.-P., L.G.-V. and R.I.D.d.I.G. are funded by CONACyT (Grant 243058) and Tecnológico de Monterrey Research Fund 0020209E01.

Acknowledgments: The authors thank Anne Michaelis and Pauline Kuschel for experimental assistance. The authors thank Charlie Cotton and Seohyoung Kim for critical review of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. A.B.-E. is cofounder of b.fab, exploring the commercialization of microbial bioproduction using C₁ feedstocks. The company was not involved in any way in performing or funding this study.

References

1. Khersonsky, O.; Tawfik, D.S. Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **2010**, *79*, 471–505. [[CrossRef](#)]
2. Bar-Even, A.; Salah Tawfik, D. Engineering specialized metabolic pathways—is there a room for enzyme improvements? *Curr. Opin. Biotechnol.* **2013**, *24*, 310–319. [[CrossRef](#)]
3. Delauney, A.J.; Verma, D.P.S. Proline Biosynthesis and Osmoregulation in Plants. *Plant J.* **1993**, *4*, 215–223. [[CrossRef](#)]
4. Dekant, W.; Vamvakas, S. Glutathione-dependent bioactivation of xenobiotics. *Xenobiotica* **1993**, *23*, 873–887. [[CrossRef](#)]
5. Metz, B.; Kersten, G.F.; Hoogerhout, P.; Brugghe, H.F.; Timmermans, H.A.; de Jong, A.; Meiring, H.; ten Hove, J.; Hennink, W.E.; Crommelin, D.J.; et al. Identification of formaldehyde-induced modifications in proteins: Reactions with model peptides. *J. Biol. Chem.* **2004**, *279*, 6235–6243. [[CrossRef](#)]
6. Chistoserdova, L.; Lidstrom, M.E. Aerobic methylotrophic prokaryotes. In *The Prokaryotes: Prokaryotic Physiology and Biochemistry*; Springer: Berlin, Heidelberg, Germany, 2013; pp. 267–285.
7. Crowther, G.J.; Kosaly, G.; Lidstrom, M.E. Formate as the main branch point for methylotrophic metabolism in *Methylobacterium extorquens* AM1. *J. Bacteriol.* **2008**, *190*, 5057–5062. [[CrossRef](#)]
8. Vorholt, J.A.; Marx, C.J.; Lidstrom, M.E.; Thauer, R.K. Novel formaldehyde-activating enzyme in *Methylobacterium extorquens* AM1 required for growth on methanol. *J. Bacteriol.* **2000**, *182*, 6645–6650. [[CrossRef](#)]
9. Blakley, R.L. The reaction of tetrahydropteroylglutamic acid and related hydropteridines with formaldehyde. *Biochem. J.* **1959**, *72*, 707–715. [[CrossRef](#)]
10. Large, P.J.; Quayle, J.R. Microbial growth on C₁ compounds. 5. Enzyme activities in extracts of *Pseudomonas* AM1. *Biochem. J.* **1963**, *87*, 386–396. [[CrossRef](#)]
11. Heptinstall, J.; Quayle, J.R. Pathways leading to and from serine during growth of *Pseudomonas* AM1 on C₁ compounds or succinate. *Biochem. J.* **1970**, *117*, 563–572. [[CrossRef](#)]
12. Marx, C.J.; Van Dien, S.J.; Lidstrom, M.E. Flux analysis uncovers key role of functional redundancy in formaldehyde metabolism. *PLoS Biol.* **2005**, *3*, e16. [[CrossRef](#)]
13. Kallen, R.G.; Jencks, W.P. The mechanism of the condensation of formaldehyde with tetrahydrofolic acid. *J. Biol. Chem.* **1966**, *241*, 5851–5863.
14. Dev, I.K.; Harvey, R.J. A complex of N₅,N₁₀-methylene tetrahydrofolate dehydrogenase and N₅,N₁₀-methenyl tetrahydrofolate cyclohydrolase in *Escherichia coli*. Purification, subunit structure, and allosteric inhibition by N₁₀-formyl tetrahydrofolate. *J. Biol. Chem.* **1978**, *253*, 4245–4253.
15. Powers, S.G.; Snell, E.E. Ketopantoate hydroxymethyltransferase. II. Physical, catalytic, and regulatory properties. *J. Biol. Chem.* **1976**, *251*, 3786–3793.
16. Dev, I.K.; Yates, B.B.; Leong, J.; Dallas, W.S. Functional role of cysteine-146 in *Escherichia coli* thymidylate synthase. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 1472–1476. [[CrossRef](#)]

17. Guenther, B.D.; Sheppard, C.A.; Tran, P.; Rozen, R.; Matthews, R.G.; Ludwig, M.L. The structure and properties of methylenetetrahydrofolate reductase from *Escherichia coli* suggest how folate ameliorates human hyperhomocysteinemia. *Nat. Struct. Biol.* **1999**, *6*, 359–365. [[CrossRef](#)]
18. Neidhardt, F.C.; Ingraham, J.L.; Schaechter, M. Building blocks needed to produce 1 g of *E. coli* protoplasm. In *Physiology of the Bacterial Cell: A Molecular Approach*; Sinauer Associates: Sunderland, MA, USA, 1990; pp. 134–143.
19. Maden, B.E. Tetrahydrofolate and tetrahydromethanopterin compared: Functionally distinct carriers in C₁ metabolism. *Biochem. J.* **2000**, *350*, 609–629. [[CrossRef](#)]
20. Yishai, O.; Goldbach, L.; Tenenboim, H.; Lindner, S.N.; Bar-Even, A. Engineered Assimilation of Exogenous and Endogenous Formate in *Escherichia coli*. *ACS Synth. Biol.* **2017**, *6*, 1722–1731. [[CrossRef](#)]
21. He, H.; Edlich-Muth, C.; Lindner, S.N.; Bar-Even, A. Ribulose Monophosphate Shunt Provides Nearly All Biomass and Energy Required for Growth of *E. coli*. *ACS Synth. Biol.* **2018**, *7*, 1601–1611. [[CrossRef](#)]
22. Hassan-Abdallah, A.; Zhao, G.; Chen, Z.W.; Mathews, F.S.; Schuman Jorns, M. Arginine 49 is a bifunctional residue important in catalysis and biosynthesis of monomeric sarcosine oxidase: A context-sensitive model for the electrostatic impact of arginine to lysine mutations. *Biochemistry* **2008**, *47*, 2913–2922. [[CrossRef](#)]
23. Jensen, S.I.; Lennen, R.M.; Herrgard, M.J.; Nielsen, A.T. Seven gene deletions in seven days: Fast generation of *Escherichia coli* strains tolerant to acetate and osmotic stress. *Sci. Rep.* **2015**, *5*, 17874. [[CrossRef](#)]
24. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K.A.; Tomita, M.; Wanner, B.L.; Mori, H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2006**, *2*, 2006–2008. [[CrossRef](#)]
25. Guzman, E.C.; Martin, C.M. Thymineless death, at the origin. *Front. Microbiol.* **2015**, *6*, 499. [[CrossRef](#)]
26. Sugantino, M.; Zheng, R.; Yu, M.; Blanchard, J.S. *Mycobacterium tuberculosis* ketopantoate hydroxymethyltransferase: Tetrahydrofolate-independent hydroxymethyltransferase and enolization reactions with alpha-keto acids. *Biochemistry* **2003**, *42*, 191–199. [[CrossRef](#)]
27. Contestabile, R.; Paiardini, A.; Pascarella, S.; di Salvo, M.L.; D’Aguanno, S.; Bossa, F. L-Threonine aldolase, serine hydroxymethyltransferase and fungal alanine racemase. A subgroup of strictly related enzymes specialized for different functions. *Eur. J. Biochem.* **2001**, *268*, 6508–6525. [[CrossRef](#)]
28. Monk, J.M.; Lloyd, C.J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **2017**, *35*, 904–908. [[CrossRef](#)]
29. Edwards, J.S.; Ramakrishna, R.; Palsson, B.O. Characterizing the metabolic phenotype: A phenotype phase plane analysis. *Biotechnol. Bioeng.* **2002**, *77*, 27–36. [[CrossRef](#)]
30. Milo, R.; Jorgensen, P.; Moran, U.; Weber, G.; Springer, M. BioNumbers—The database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **2010**, *38*, D750–D753. [[CrossRef](#)]
31. Dragosits, M.; Mattanovich, D. Adaptive laboratory evolution—Principles and applications for biotechnology. *Microb. Cell Fact.* **2013**, *12*, 64. [[CrossRef](#)]
32. Flegr, J. Two Distinct Types of Natural Selection in Turbidostat-like and Chemostat-like Ecosystems. *J. Theor. Biol.* **1997**, *188*, 121–126. [[CrossRef](#)]
33. Thomason, L.C.; Costantino, N.; Court, D.L.E. *E. coli* genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.* **2007**, *79*, 1–17. [[CrossRef](#)]
34. Imlay, J.A. The mist metallation of enzymes during oxidative stress. *J. Biol. Chem.* **2014**, *289*, 28121–28128. [[CrossRef](#)] [[PubMed](#)]
35. Nash, T. The colorimetric estimation of formaldehyde by means of the Hantzsch reaction. *Biochem. J.* **1953**, *55*, 416–421. [[CrossRef](#)]
36. You, L.; Page, L.; Feng, X.; Berla, B.; Pakrasi, H.B.; Tang, Y.J. Metabolic pathway confirmation and discovery through ¹³C-labeling of proteinogenic amino acids. *J. Vis. Exp.* **2012**, e3583. [[CrossRef](#)]
37. Giavalisco, P.; Li, Y.; Matthes, A.; Eckhardt, A.; Hubberten, H.M.; Hesse, H.; Segu, S.; Hummel, J.; Kohl, K.; Willmitzer, L. Elemental formula annotation of polar and lipophilic metabolites using ¹³C, ¹⁵N and ³⁴S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J.* **2011**, *68*, 364–376. [[CrossRef](#)]
38. Ebrahim, A.; Lerman, J.A.; Palsson, B.O.; Hyduke, D.R. COBRApy: CONstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **2013**, *7*, 74. [[CrossRef](#)]

39. Ramos-Parra, P.A.; Urrea-López, R.; Díaz de la Garza, R.I. Folate analysis in complex food matrices: Use of a recombinant *Arabidopsis* γ -glutamyl hydrolase for folate deglutamylation. *Food Res. Int.* **2013**, *54*, 177–185. [[CrossRef](#)]
40. Ramos-Parra, P.A.; García-Salinas, C.; Hernández-Brenes, C.; Díaz de la Garza, R.I. Folate Levels and Polyglutamylation Profiles of Papaya (*Carica papaya* cv. Maradol) during Fruit Development and Ripening. *J. Agric. Food Chem.* **2013**, *61*, 3949–3956. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Research Article

Thioprolinone formation as a driver of formaldehyde toxicity in *Escherichia coli*

Jenelle A. Patterson^{1,*},  Hai He^{2,*}, Jacob S. Folz^{3,*}, Qiang Li⁴, Mark A. Wilson⁵, Oliver Fiehn³, Steven D. Bruner⁴, Arren Bar-Even² and  Andrew D. Hanson¹

¹Department of Horticultural Sciences, University of Florida, Gainesville, FL, U.S.A.; ²Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm, Germany; ³West Coast Metabolomics Center, University of California Davis, Davis, CA, U.S.A.; ⁴Chemistry Department, University of Florida, Gainesville, FL, U.S.A.; ⁵Department of Biochemistry and Redox Biology Center, University of Nebraska, Lincoln, NE, U.S.A

Correspondence: Andrew D. Hanson (adha@ufl.edu)

Formaldehyde (HCHO) is a reactive carbonyl compound that formylates and cross-links proteins, DNA, and small molecules. It is of specific concern as a toxic intermediate in the design of engineered pathways involving methanol oxidation or formate reduction. The interest in engineering these pathways is not, however, matched by engineering-relevant information on precisely why HCHO is toxic or on what damage-control mechanisms cells deploy to manage HCHO toxicity. The only well-defined mechanism for managing HCHO toxicity is formaldehyde dehydrogenase-mediated oxidation to formate, which is counter-productive if HCHO is a desired pathway intermediate. We therefore sought alternative HCHO damage-control mechanisms via comparative genomic analysis. This analysis associated homologs of the *Escherichia coli* *pepP* gene with HCHO-related one-carbon metabolism. Furthermore, deleting *pepP* increased the sensitivity of *E. coli* to supplied HCHO but not other carbonyl compounds. PepP is a proline aminopeptidase that cleaves peptides of the general formula X-Pro-Y, yielding X + Pro-Y. HCHO is known to react spontaneously with cysteine to form the close proline analog thioprolinone (thiazolidine-4-carboxylate), which is incorporated into proteins and hence into proteolytic peptides. We therefore hypothesized that certain thioprolinone-containing peptides are toxic and that PepP cleaves these aberrant peptides. Supporting this hypothesis, PepP cleaved the model peptide Ala-thioprolinone-Ala as efficiently as Ala-Pro-Ala *in vitro* and *in vivo*, and deleting *pepP* increased sensitivity to supplied thioprolinone. Our data thus (i) provide biochemical genetic evidence that thioprolinone formation contributes substantially to HCHO toxicity and (ii) make PepP a candidate damage-control enzyme for engineered pathways having HCHO as an intermediate.

Introduction

Formaldehyde (HCHO) is a highly reactive, toxic, and mutagenic carbonyl compound that all organisms produce in minor amounts during normal metabolism [1–3]. It is also a major and mandatory intermediate in natural and engineered (‘synthetic’) methylotrophic and formatotrophic pathways [1,4–6], and in synthetic pathways to which methanol supplies reducing power but not carbon [7] (Figure 1).

The toxicity and mutagenicity of HCHO lie in its ability to react spontaneously with amino and thiol groups. With amines such as the 6-amino group of peptidyl lysine, HCHO forms a carbinolamine that can either be oxidized to an *N*-formyl derivative or undergo dehydration, yielding a labile Schiff base that can form cross-links (methylene bridges) with various amino acids and nucleobases [8,9] (Figure 1). With thiols, notably free and peptidyl cysteine, HCHO forms the cyclic derivative thioprolinone (thiazolidine-4-carboxylate) [10,11] (Figure 1). All of these reactions except oxidation and cross-linking are reversible, with the HCHO condensation direction being thermodynamically preferred. Although it is clear that the above general types of damage reactions underlie the toxic and mutagenic effects of HCHO, the specific reactions that actually drive these effects and their relative

*These authors contributed equally to this work.

Received: 5 March 2020

Revised: 14 April 2020

Accepted: 17 April 2020

Accepted Manuscript online:

17 April 2020

Version of Record published:

15 May 2020

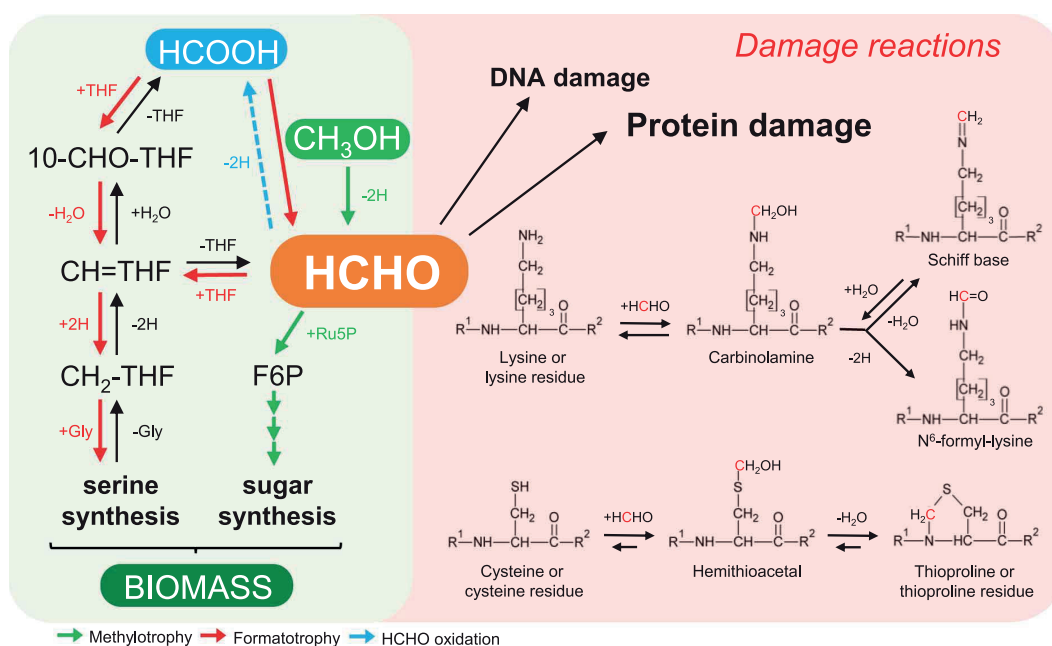


Figure 1. Formaldehyde as a central one-carbon metabolite and driver of damage reactions.

Formaldehyde (HCHO) participates in methylotrophy and formatotrophy pathways but reacts spontaneously with amino and thiol groups, e.g. those in the side chains of lysine and cysteine, respectively. The Schiff base formed with lysine undergoes various cross-linking reactions with proteins and nucleic acids. HCHO can be detoxified by oxidation to formate (HCOOH) but this route is incompatible with methylotrophy or formatotrophy pathways in which HCHO is an obligatory intermediate. R¹, H or peptide chain; R², OH or peptide chain; Ru5P, ribulose 5-phosphate; THF, tetrahydrofolate, 10-CHO-, 10-formyl; CH=, 5,10-methenyl; CH₂-, 5,10-methylene.

importance *in vivo* remain poorly defined [10]. The least poorly defined toxic effect is probably the N⁶-formylation of lysine residues in certain histones, which is inferred to impair histone function [12].

The archetypal defense against HCHO toxicity is to oxidize HCHO to non-toxic formate via glutathione-dependent or -independent dehydrogenases [1–3] (Figure 1). This detoxification strategy is counterproductive if HCHO is a pathway substrate, as in methylotrophic or formatotrophic pathways, because it leads to loss of the substrate. Cells operating methylotrophic or formatotrophic pathways must therefore cope with HCHO because they cannot simply detoxify it by dehydrogenation [13]. A major coping mechanism is fine-tuning the expression of HCHO-producing and -consuming enzymes [13,14], but this delicate flux-balancing act entails non-negligible steady-state HCHO levels as well as possible HCHO spikes due to upshifts in methanol availability. It is consequently worth identifying other mechanisms that cells use to control HCHO damage. The only known damage-control mechanism of this type appears to be deformylation of N⁶-formyl lysine residues by bacterial and mammalian sirtuins [15].

The research we report here was based on two ideas: that there are HCHO damage-control enzymes left to discover [16] and that such enzymes are valuable components to have in hand when designing and building metabolic pathways that involve HCHO [17]. After applying comparative genomics to identify a candidate HCHO damage-control enzyme, we obtained genetic and biochemical evidence that *Escherichia coli* PepP, an Xaa-Pro aminopeptidase, mitigates downstream consequences of the presence of HCHO-generated thioproline residues in proteins. That deleting *pepP* increases HCHO sensitivity provides genetic evidence that thioproline formation is one of the drivers of HCHO toxicity.

Materials and methods

Bioinformatics

Comparative genomics analyses used STRING [18] (<https://string-db.org/>) and SEED [19] (<http://pubseed.theseed.org/>) databases and tools. For both databases, we used the ‘guilt by association’ principle [20], i.e. to

find candidate HCHO damage-control genes we searched for genes that consistently cluster on the chromosome with genes encoding enzymes that catalyze reactions leading to or from HCHO in KEGG maps 00680 (methane metabolism) and 00670 (one carbon pool by folate). Full results of the SEED analysis of 1127 representative bacterial genomes are encoded in the SEED subsystem named 'PepP-C1'. Phylogenetic trees were drawn with MEGA (www.megasoftware.net).

Chemicals

Phusion® High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA) was used for PCR reactions. All other chemicals were from Sigma–Aldrich, unless otherwise stated. HCHO solution (37%, with methanol stabilizer) was from Fisher Scientific (Cat. No. BP531-500). Tripeptides used for *in vitro* assays and growth experiments (H-Ala-Pro-Ala-OH, H-Ala-Thz-Ala-OH trifluoroacetate, and H-Glu(Thz-Gly-OH)-OH trifluoroacetate) were synthesized by Bachem (U.K.) Ltd. (St. Helens, U.K.); their purities (>98%) were validated by HPLC. N⁶-Formyl-L-lysine was from TCI Chemicals (Cat. No. F0136). Formylated Boc-L-Lys-AMC was synthesized from Boc-L-Lys-AMC (Peptide Solutions LLC, Tucson, AZ, Cat. No. BAA3650.0100) using a modification of a published procedure [21]. To a solution of Boc-L-Lys-AMC (50 mg, 0.12 mmol, 1.0 equiv.) and a catalytic amount of DIPEA (0.01 mmol) in dry dimethylformamide (2 ml) was added 2,2,2-trifluoroethyl formate (61 mg, 0.48 mmol, 4.0 equiv.) and the reaction mixture was stirred at 22°C for 12 h. The mixture was concentrated *in vacuo* and purified by silica gel column chromatography (CH₂Cl₂:MeOH, 20/1 v/v) to give the product (36 mg, 67%). ¹H NMR (300 MHz, DMSO-*d*₆): δ 10.43 p.p.m. (s, 1H), 7.98 (s, 2H), 7.78 (d, 1H, *J* = 3 Hz), 7.73 (d, 1H, *J* = 9 Hz), 7.50 (d, 1H, *J* = 9 Hz), 7.14 (d, 1H, *J* = 9 Hz), 6.26 (s, 1H), 4.06 (m, 1H), 3.06 (m, 2H), 2.40 (s, 3H), 1.62–1.57 (m, 2H), 1.42–1.38 (m, 11H), 1.29–1.27 (m, 2H). LC–MS (ESI) *m/z* 430.2 [M–H][–].

Bacterial strains and expression constructs

Keio collection wild type background (BW25113), *ΔpepP*, and *ΔyhbO* strains [22] were the gift of V. de Crecy-Lagard (University of Florida). Primers are listed in Supplementary Table S1. All constructs were sequence-verified. For *ΔpepP* mutant complementation, the *E. coli* PepP coding sequence was PCR-amplified (primers 1 and 2) from BW25113 genomic DNA prepared using a GeneJET Genomic DNA Purification Kit (Thermo Scientific, Cat No. K0721) according to the manufacturer's instructions. Amplicons were digested with *EcoRI* and *SalI* (New England Biolabs) and ligated into pBAD24. P1 phage transduction of the *ΔyhbO*::Kan^r locus from the Keio *ΔyhbO* donor into the recipient BW25113 strain was performed according to [23]. P1 phage was isolated from strain EST877 P1cml, clr100 (the gift of L.N. Csonka, Purdue University), which was precultured in 2 ml Luria Broth (LB) with 30 μg/ml chloramphenicol at 30°C. The preculture was subcultured (1 : 100) in 50 ml LB and grown at 30°C with shaking until OD₆₀₀ ~ 0.8 before shifting the temperature to 40°C for 35 min, then 37°C for 1 h. The cells were lysed by adding 0.5 ml chloroform; after 10 min, the phage-containing supernatant was collected, filtered (0.45 μm) and stored at 4°C. Transduced cells were selected on M9 minimal medium plates containing 50 μg/ml kanamycin at 37°C for 48 h, and positive strains were sequence-verified.

Media and culture conditions

Growth experiments were performed in M9 minimal medium (47.8 mM Na₂HPO₄, 22 mM KH₂PO₄, 8.6 mM NaCl, 18.7 mM NH₄Cl, 2 mM MgSO₄ and 100 μM CaCl₂), supplemented with trace elements (134 μM EDTA, 31 μM FeCl₃, 6.2 μM ZnCl₂, 0.76 μM CuCl₂, 0.42 μM CoCl₂, 1.62 μM H₃BO₃, 0.081 μM MnCl₂) and 22 mM glucose, unless otherwise specified. NH₄Cl was omitted for the nitrogen source experiments. For liquid growth curves, strains were precultured in 4 ml M9 medium containing 10 mM glucose. The precultures were harvested and washed three times in M9 medium, then inoculated in M9 medium containing 10 mM glucose plus specified additions, with a starting OD₆₀₀ of 0.005. One hundred and fifty μl of culture was added to each well of 96-well microplates (Nunclon Delta Surface, Thermo Scientific), then 50 μl mineral oil (Sigma–Aldrich) was added to each well to avoid evaporation (while enabling gas diffusion). The 96-well microplates were incubated at 37°C in a microplate reader (BioTek EPOCH 2). The shaking program cycle (controlled by Gen5 v3) had four shaking phases, lasting 60 s each: linear shaking followed by orbital shaking, both at an amplitude of 3 mm, then linear shaking followed by orbital shaking both at an amplitude of 2 mm. The OD₆₀₀ in each well was monitored and recorded after every three shaking cycles (~16.5 min). Raw data from the plate reader were calibrated to normal cuvette-measured OD₆₀₀ values according to OD_{cuvette} = OD_{plate}/0.23.

For HCHO insult experiments, 50 ml of M9 medium was inoculated with BW25113 and $\Delta pepP$ overnight precultures to a starting OD₆₀₀ of 0.01 and incubated with shaking at 37°C until OD₆₀₀ reached 0.2. Pilot experiments showed that adding HCHO to cultures to a final concentration of 1 mM caused a cessation of growth after which BW25113 and $\Delta pepP$ strains resumed growth after 13 h and 20 h, respectively. For untreated samples, cells were collected by centrifugation at OD₆₀₀ of 0.2, flash frozen and stored at –80°C; HCHO-treated samples were similarly collected 2 h after the insult began.

M9 medium plates containing 1% agar were used to test growth on solid medium. When applicable, 400 μM formaldehyde (or 60 mM acetaldehyde, 5 mM glyoxal, or 0.5 mM methylglyoxal) were added to the medium after cooling to <45°C. Overnight M9 medium precultures were used to inoculate (1 : 100) 2 ml M9 medium cultures, which were grown at 37°C until OD₆₀₀ of 1.0 before ten-fold serially diluting and spotting 3.5 μl aliquots on the plates. Plates were incubated at 37°C for 16–20 h before imaging.

To test Ala-Pro-Ala or Ala-thioprolino-Ala as sole nitrogen source, M9 plates were prepared as above except that NH₄Cl was omitted. Overnight M9 medium precultures of BW25113 and Keio $\Delta pepP$ strains were used to inoculate (1 : 100) 10 ml M9 medium cultures, which were grown at 37°C until OD₆₀₀ of ~1.0. Cells from each strain were 10-fold diluted into water agar solution (0.75% w/v) cooled to ~40°C, and 5 ml was overlaid on plates prewarmed to 37°C. Once cooled, 10 mm disks of sterile No. 2 Whatman filter paper were imbibed with 10 μl of 25 mg/ml alanine or 50 mg/ml tripeptide solution, then applied to the agar-overlaid plates. The plates were incubated at 37°C for 24 h before imaging.

Production and purification of recombinant PepP

For PepP protein expression, the coding sequence was PCR-amplified (primers 3 and 4) from wild type BW25113 genomic DNA, adding an N-terminal His₆-tag as well as 3' and 5' sequences complementary to the pET15b multiple cloning site. Following pET15b PCR amplification (primers 5 and 6), the vector and insert amplicons were gel-purified (GeneJet Gel Extraction Kit, Thermo Scientific Cat No. K0691) according to the manufacturer's instructions and inserted using circular polymerase extension cloning [24]. To produce recombinant PepP protein, *E. coli* BL21 (DE3) RIPL harboring the expression plasmid was grown in 100 ml LB medium at 37°C until an OD₆₀₀ of 0.8, after which the cultures were supplemented with 1.0 mM isopropyl-β-D-thiogalactoside and incubated a further 2 h at 22°C. Cells were collected by centrifugation, frozen in liquid N₂, and stored at –80°C. Pellets were resuspended in 1 ml ice-cold lysis buffer (50 mM potassium phosphate, pH 8.0, 300 mM NaCl, and 10 mM imidazole), and sonicated (Fisher Scientific Ultrasonic Dismembrator, model 150E) using five 12 s pulses at 70% power, cooling on ice for 30 s between pulses. Following centrifugation (14 000 g, 5 min, 4°C), the supernatant was applied to a 0.1 ml column of Ni²⁺-NTA Superflow resin (Qiagen Cat No. 30410) and protein was purified using the manufacturer's protocol. The purified protein was desalted using a PD MiniTrap G-25 column (GE Healthcare), eluting with 50 mM potassium phosphate, pH 7.5, containing 10% (v/v) glycerol; aliquots (50 μl) were frozen in liquid N₂ and stored at –80°C. Protein concentration was determined by Bradford dye-binding assays [25].

Enzyme assays

For all activity assays, PepP protein was pre-incubated with 0.5 mM MnCl₂ for 15 min on ice. To assay deformylase activity, 40 μl reactions containing 2 μg PepP, 0.5 mM MnCl₂, 50 mM potassium phosphate, pH 7.5, and 5 mM N⁶-formyl-L-lysine or 2 mM formylated BOC-Lys-AMC (diluted from a 30 mM stock solution in dimethylsulfoxide) were incubated at 22°C for 16 h. To quantify formate release, 0.05 units of formate dehydrogenase (Megazyme, Bray, Ireland, Cat. No. E-FDHCB) and 1 mM NAD were added (final reaction volume of 80 μl, 22°C). Absorbance at 340 nm for N⁶-formyl-L-lysine and 365 nm for formylated BOC-Lys-AMC was recorded every 15 s for 5 min. Dimethylsulfoxide at the concentration present in the assays had no effect on formate dehydrogenase.

Aminopeptidase activity against glutathione and γ-Glu-thioprolino-Gly was assayed 22°C in 80 μl reactions containing 2 μg PepP, 0.5 mM MnCl₂, 50 mM HEPES, pH 8.8, 1 mM NAD, and 0.02 units L-glutamate dehydrogenase (Sigma–Aldrich Cat. No. G2626). Reactions were started by adding 2 mM substrate. Assays using Ala-Pro-Ala and Ala-thioprolino-Ala as substrate contained 2 μg PepP, 0.5 mM MnCl₂, 50 mM Tris–HCl, pH 8.8, 1 mM NAD, and 0.05 units alanine dehydrogenase (Sigma–Aldrich Cat. No. 73063), and were started by adding 0.1–3 mM substrate. Absorbance at 340 nm was recorded every 10 s for 5 min for all aminopeptidase activity assays. Three independent assays at six substrate concentrations were performed to determine

enzyme kinetics, which were calculated by fitting data to the Michaelis–Menten equation using Prism software (GraphPad Software, Inc., La Jolla, CA).

LC–MS analysis of protein-bound N^6 -formyl lysine

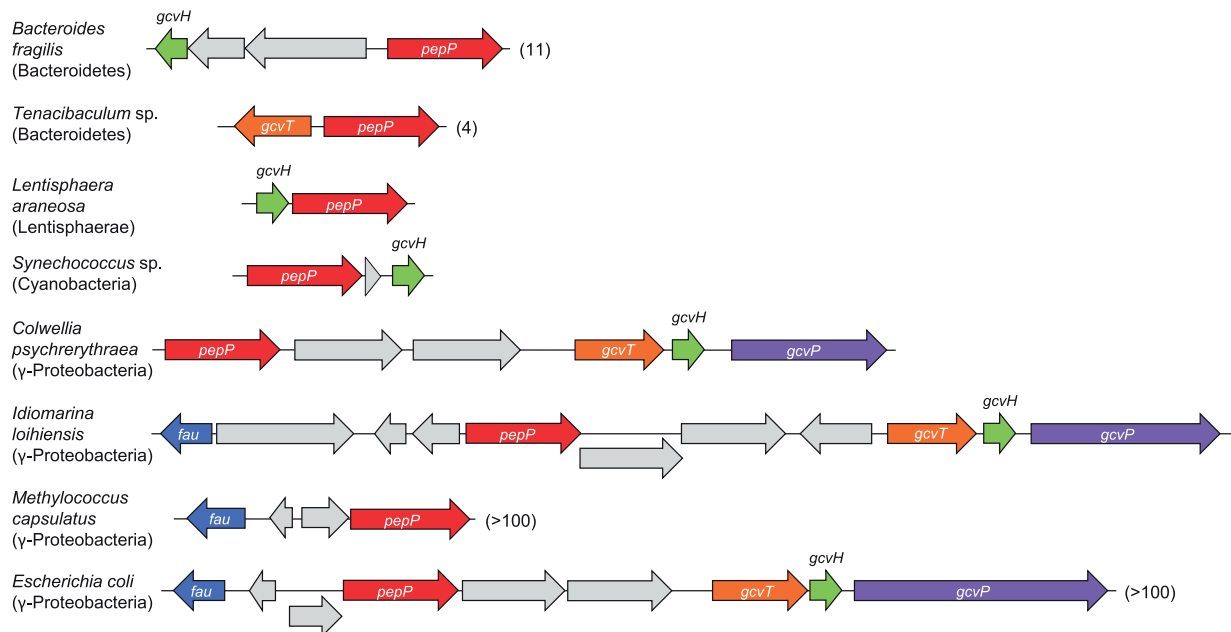
Cell pellets were resuspended in 1 ml PBS buffer (10 mM sodium phosphate, pH 7.4, 137 mM NaCl, 2.7 mM KCl), sonicated as above, and centrifuged to clear. Soluble proteins were precipitated in 80% (v/v) acetone and harvested by centrifugation (5000 g, 3 min, 4°C); the pellets were air-dried overnight at 22°C and stored at –20°C. The pelleted proteins were mixed with 200 μ l 1 M NaOH and heated at 50°C until dissolved (briefly vortexing as necessary) before adding 400 μ l 100 mM ammonium bicarbonate buffer, pH 8.5 (brought to pH 8.5 with 5 M HCl). Samples were then desalted on PD MiniTrap G-25 columns, eluting with 100 mM ammonium bicarbonate buffer, pH 8.5, and protein content was determined as above. To digest proteins to free amino acids, *Streptomyces griseus* protease [11] was added (5 μ g per 50 μ g protein) and the mixtures were incubated at 37°C for ~16 h. The digests were vacuum-dried and stored at –20°C. Digests were dissolved by adding 0.5 ml HPLC grade 80 : 20 acetonitrile/water (v/v), vortexing for 30 s, sonicating for 5 min, and vortexing for another 30 s. From this solution 0.1 ml was aliquoted into a clean 1.5 ml microcentrifuge tube and dried under vacuum. Dried samples were resuspended in 0.04 ml of run solvent consisting of 80 : 20 acetonitrile/water (v/v) with deuterium labeled internal standards (IS) (Supplementary Table S2). To resuspend, solvent was added, samples were vortexed for 30 s, sonicated for 5 min, vortexed again for 30 s, centrifuged (16 000 g, 2 min, room temperature), and transferred to an LC–MS vial with glass insert. Samples were stored at 4°C in an autosampler and analyzed using a Vanquish LC system (Thermo Fisher Scientific), coupled to a 6500+ QTrap mass spectrometer (SCIEX) operated in multiple reaction monitoring (MRM) mode. Standard curves of analytes were diluted in LC–MS grade 80 : 20 acetonitrile/water (v/v), dried under vacuum, and resuspended in the same run solvent as samples. A Waters Acquity UPLC BEH Amide column (150 mm \times 2.1 mm, 1.7 μ m particle size) with an Acquity VanGuard BEH Amide pre-column (5 mm \times 2.1 mm, 1.7 μ m particle size) was held at 45°C throughout analysis. Mobile phase (A) was LC–MS grade water and (B) was 95:5 LC–MS grade acetonitrile/water (v/v). Both mobile phases were modified to 10 mM ammonium formate and 0.125% formic acid. The mobile phase flow was 0.4 ml/min and the gradient was 100% B from 0–2 min, brought to 70% B between 2 and 7.7 min, brought to 40% B between 7.7 and 9.5 min, returned to 100% B between 9.5 and 12.75 min and maintained at 100% B until 17 min. Five microlitres of sample was injected and needle wash solution was 1 : 1 acetonitrile/water (v/v). Collision gas was set to ‘Medium’, IonSpray voltage was 5500 V and temperature was 250°C. Linear 5–7 point calibration curves of the ratio of analyte to IS with 1/ \times weighting were fit for lysine (IS: D8 lysine), N^6 -formyl lysine (IS: D5 threonine), and cystine (IS: D8 lysine) in MultiQuant version 3.0.3 (SCIEX). The background rate of N^6 -formylation of lysine during analysis was 0.03% determined from a calibration curve of pure lysine standard spiked into the matrix. Background formylation was corrected for by subtracting $0.0003 \times [\text{lysine}]$ from the measured concentrations of N^6 -formyl lysine in samples.

Results and discussion

Comparative genomics links PepP with one-carbon metabolism

We first used comparative genomics analysis [26] with the STRING [18] and SEED [19] databases to find candidate genes that could play a role in controlling HCHO damage. We searched for genes that are consistently clustered on the chromosome with genes of one-carbon metabolism, in which HCHO is central. These one-carbon metabolism ‘guide genes’ included genes encoding enzymes that produce or consume HCHO (e.g. methanol and formaldehyde dehydrogenases) and genes of tetrahydrofolate (THF)-mediated one-carbon metabolism, which is both a disposal route for HCHO and a potential HCHO source via spontaneous dissociation of 5,10-methylene-THF [27]. The analysis detected genes encoding homologs of *E. coli* aminopeptidase PepP clustered with genes encoding subunits of the glycine cleavage complex, which generates a THF-bound HCHO unit [2], and genes for 5-formyl-THF cyclo-ligase, which recycles a one-carbon folate byproduct of serine hydroxymethyltransferase, another enzyme that generates a THF-bound HCHO unit [28] (Figure 2A). These clustering arrangements are unlikely to be due to chance alone because they involve different HCHO-related genes and occur in many genomes from four phyla (Figure 2A). Moreover, the clustering specifically involves the PepP aminopeptidase family, but not related aminopeptidase families such as Ypff (Figure 2B).

A



B

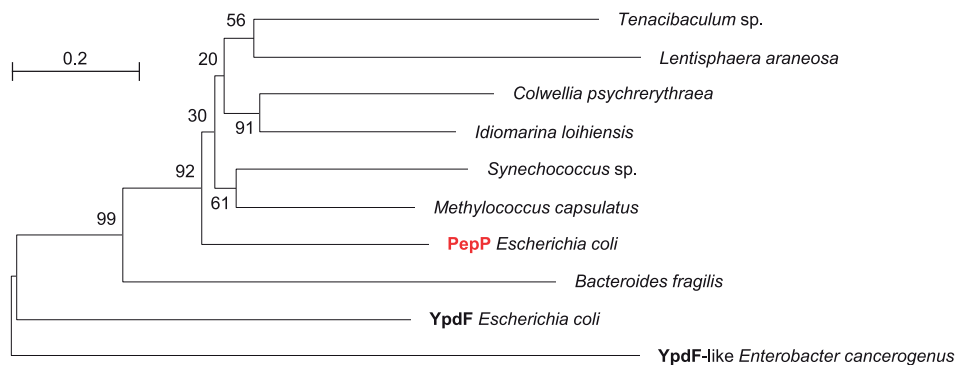


Figure 2. Bioinformatic analysis of PepP homologs from diverse bacteria.

(A) PepP (Xaa-Pro aminopeptidase) homologs from four bacterial phyla cluster on the chromosome with genes encoding enzymes of one-carbon metabolism: 5-formyl-THF cyclo-ligase (*fau*) and the T, H, and P subunits of the glycine cleavage complex (*gcvT*, *gcvH*, and *gcvP*, respectively). Gray genes encode proteins whose function is either unrelated to one-carbon metabolism or unknown. Values in parentheses indicate the number of genomes in the SEED PepP-C1 subsystem that have similar clusters to the example shown. (B) Neighbor-joining tree (500 bootstrap replicates) confirming that the proteins encoded by the *pepP* homologs above belong to the same clade as *E. coli* PepP, and not to other aminopeptidase clades (YpdF or YpdF-like clades). Bootstrap values are shown for branch support. Evolutionary distances are in units of amino acid substitutions per site.

PepP contributes to formaldehyde resistance in *E. coli*

The above genomic evidence linking PepP with one-carbon metabolism led us to compare the HCHO resistance of an *E. coli* *pepP* deletion strain with that of the parent BW25113 strain (henceforth referred to as wild type). The $\Delta pepP$ strain was markedly more sensitive to HCHO at high concentrations (400–500 μM) when cultured in liquid or solid medium (Figure 3A,B). To confirm that the *pepP* deletion causes the HCHO-sensitive phenotype (i.e. that no mutations elsewhere in the genome are involved) we complemented the $\Delta pepP$ strain with a plasmid-borne *E. coli* *pepP* gene (Figure 3A). The methanol stabilizer present in the HCHO reagent (37% HCHO, 10–15% methanol) did not detectably affect the growth of the wild type or

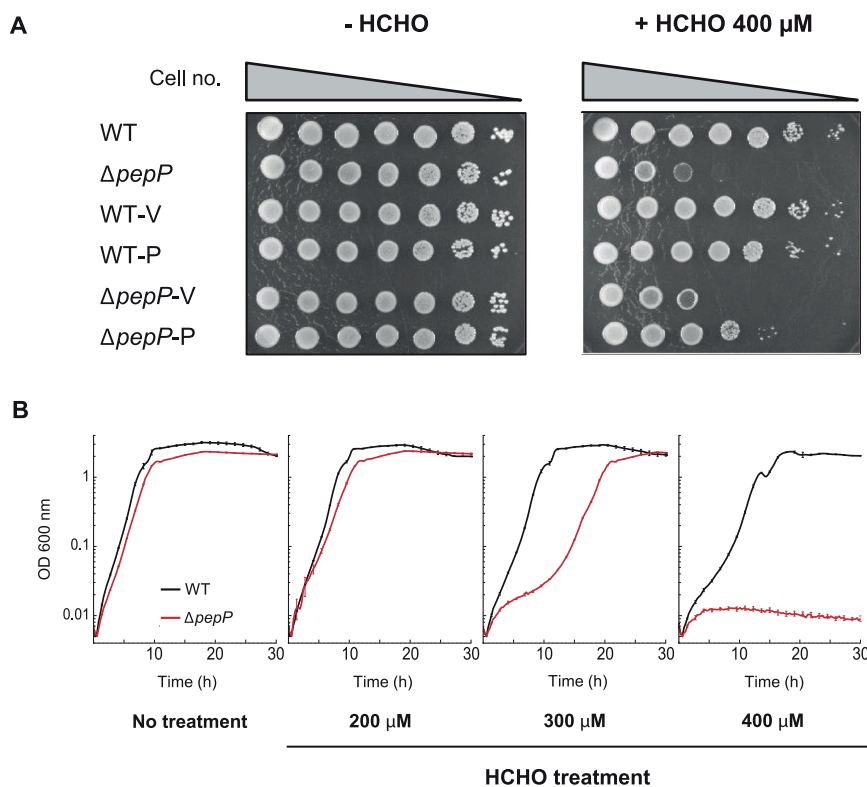


Figure 3. Growth assays implicate *E. coli* PepP in resistance to formaldehyde.

(A) *E. coli* BW25113 wild type (WT) and $\Delta pepP$ strains harboring the pBAD24 vector alone (V) or containing *pepP* (P) were cultured on M9 minimal medium containing 0.4% glucose and 0.02% arabinose, minus or plus 400 μ M HCHO. Overnight liquid cultures of each strain were ten-fold serially diluted and 3.5 μ l aliquots were spotted on the plates. Images were captured after incubation at 37°C for 1.5 d. (B) Wild type and $\Delta pepP$ strains were cultured in 150 μ l of M9 liquid medium containing 10 mM glucose and the indicated concentrations of HCHO. Data are means \pm S.E. ($n = 3$).

$\Delta pepP$ strains when added to plates at the concentration that would have accompanied the HCHO (Supplementary Figure S1A).

To determine whether PepP impacts resistance to carbonyl compounds besides HCHO, we compared the growth of the parent and $\Delta pepP$ strains in the presence of inhibitory concentrations of acetaldehyde, glyoxal, or methylglyoxal. The $\Delta pepP$ deletant showed little or no increase in sensitivity to these carbonyls (Supplementary Figure S1B), showing that PepP action is essentially specific to HCHO. The slightly increased sensitivity to methylglyoxal may be attributable to the HCHO present as a contaminant (up to ~ 17 mol %) in commercial methylglyoxal [29]. We also tested whether the HCHO-sensitivity of the $\Delta pepP$ strain depends on the nature or concentration of the carbon source. Similar sensitivity was observed on eight diverse carbon sources, which included fermentable and non-fermentable substrates taken up by the phosphotransferase system or other carriers (Supplementary Figure S2). Sensitivity was also similar at various concentrations of glucose (Supplementary Figure S3). The HCHO-sensitivity of the $\Delta pepP$ strain is thus a robust phenotype.

Pepp does not deformylate N^6 -formyl lysine *in vitro* or *in vivo*

E. coli PepP (Xaa-Pro aminopeptidase, EC 3.4.11.9) cleaves the N-terminal amino acid from a wide range of peptides whose second residue is proline; the reaction catalyzed is thus $X\text{-Pro-Y}_n \rightarrow X + \text{Pro-Y}_n$ (where n can be up to ~ 10) [30]. Because N^6 -formylation of lysine residues is a known HCHO-driven protein damage reaction [11,12] and because some peptidases have amidase activity [31,32], we tested whether PepP cleaves the amide bond in N^6 -formyl lysine (i.e. deformylates lysine) *in vitro* or *in vivo*. For *in vitro* tests, N^6 -formyl lysine and its N- and C-terminally blocked derivative Boc- N^6 -formyl Lys-AMC were incubated with recombinant *E. coli* PepP and a coupled spectrophotometric assay was used to measure formate release. No activity was

detected with either substrate. The activity detection limit ($0.6 \text{ nmol mg}^{-1} \text{ protein min}^{-1}$) was $<0.01\%$ of the activity of PepP against the canonical peptide substrate Ala-Pro-Ala (see below). To extend the study from model substrates to proteins *in vivo* we measured N^6 -formyl lysine levels in proteins extracted from wild type and $\Delta pepP$ cells before and after treatment with a nonlethal, bacteriostatic concentration of HCHO (Figure 4). HCHO-treated wild type and $\Delta pepP$ cells had modestly higher N^6 -formyl lysine levels than untreated cells, consistent with the known formylating activity of HCHO [11,12]. However, N^6 -formyl lysine levels in $\Delta pepP$ cells did not differ significantly from those in wild type cells either before HCHO treatment (when cells were growing) or after HCHO treatment (when cells were not growing). The lack of effect of deleting *pepP* agrees with the *in vitro* evidence that PepP does not have lysine-deformylating activity and hence confirms that this activity cannot explain the role of PepP in HCHO resistance.

A deductive hypothesis connecting PepP to formaldehyde toxicity

It is not obvious how an aminopeptidase like PepP could impact HCHO toxicity unless it has substantial de-formylase activity, which the above evidence indicates it does not. However, a plausible multi-part hypothesis can be proposed, based on four facts: (i) thioproline is a close analog of proline, with similar bond lengths and angles [33]; (ii) thioproline forms spontaneously *in vivo* from the condensation of HCHO and free or protein-bound cysteine [10,11]; (iii) thioproline is charged on $tRNA^{Pro}$ and enters proteins [33,34]; and (iv) PepP is specific for proline as the second residue in its peptide substrates [30]. Our hypothesis is as follows: HCHO buildup causes the formation of thioproline-containing proteins whose degradation produces thioproline-containing peptides, some of which are toxic; PepP is necessary for hydrolysis of these peptides, enabling release of free thioproline (Figure 5A). Depending on the relative rates of proline dehydrogenase (PutA) and thioproline incorporation into $tRNA$, some of the released thioproline will be metabolized to cysteine and formate [35,36] (Figure 5A, solid blue arrows) while another part will be reincorporated into protein (Figure 5A, dashed blue arrow). In either case, the core of the hypothesis (Figure 5A, box) is that certain thioproline peptides are toxic and that PepP is essential for their removal.

This hypothesis explains why deleting PepP does not increase sensitivity to acetaldehyde or glyoxal, nor probably to methylglyoxal (Supplementary Figure S1). Acetaldehyde reacts with free cysteine to form the

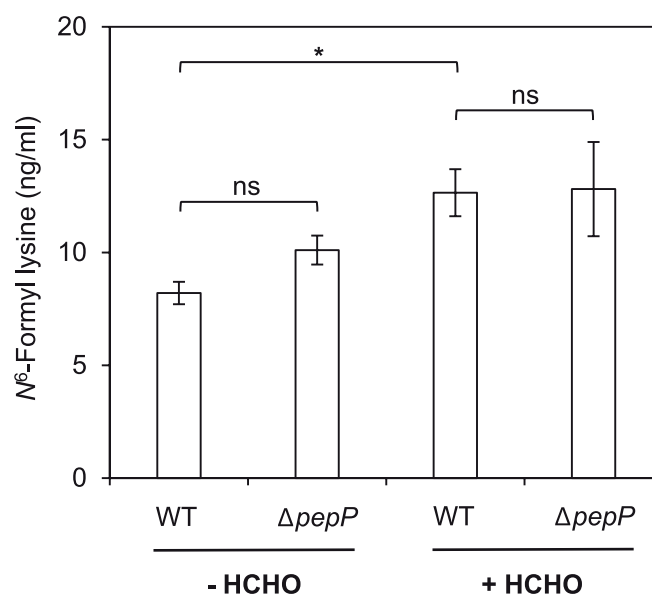


Figure 4. N^6 -Formyl lysine levels in proteins from wild type and $\Delta pepP$ strains.

E. coli BW25113 wild type (WT) and $\Delta pepP$ strains were grown in M9 minimal medium containing 0.4% glucose until OD_{600} reached 0.4. Control (–HCHO) cells were then harvested; HCHO-treated (+HCHO) cells were cultured for another 2 h after adding HCHO (final concentration 1 mM for BW25113, 0.5 mM for $\Delta pepP$) and then harvested. Extracted proteins were hydrolyzed with *S. griseus* protease; the hydrolysate was analyzed by LC–MS for N^6 -formyl lysine. Data are means \pm S.E. ($n = 6$). Significance was determined by Student's *t*-test. * $P < 0.05$; ns, non-significant.

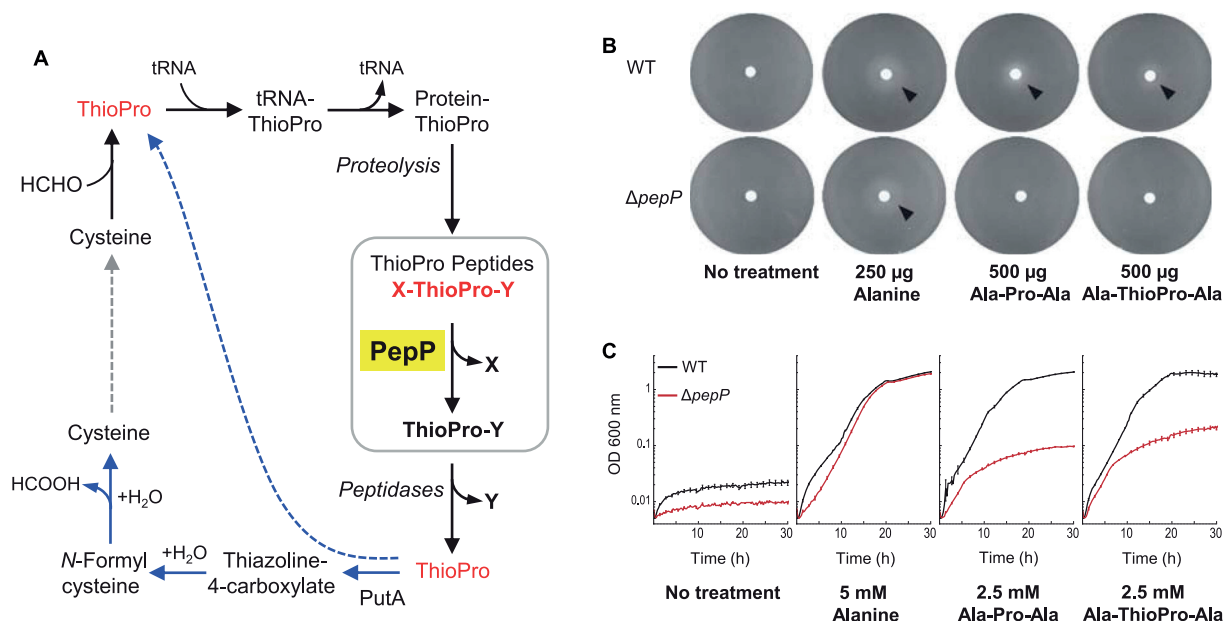


Figure 5. A hypothesis connecting PepP to formaldehyde toxicity and evidence supporting it.

(A) Schematic representation of the hypothesis. Compounds inferred or known to be toxic are in red font. The core parts of the hypothesis—the toxicity of thioproline-containing peptides and their cleavage by PepP—are boxed in gray. ThioPro, thioproline; X and Y represent any amino acid residue. (B) Growth of lawns of BW25113 wild type (WT) and $\Delta pepP$ cells on plates of M9 minimal medium without N, plus 0.4% glucose. The central disc contained the additions indicated; alanine was included as a positive control. Arrowheads mark growth halos around the discs. (C) Growth of wild type and $\Delta pepP$ cells in liquid M9 minimal medium without N, plus 10 mM glucose and the additions indicated. Data are means \pm S.E. ($n = 3$).

thioproline analog 2-methylthioproline [37] but this amino acid is almost surely not charged to tRNA and incorporated into proteins [38], and acetaldehyde has not been reported to react with peptidyl-cysteine; hence there would be no 2-methylthioproline-containing peptides for PepP to cleave. Glyoxal and methylglyoxal cannot react with cysteine to form thiazolidine adducts (i.e. thioproline-type derivatives) [39], so again there would be no thioproline-like peptides for PepP to act on.

The hypothesis also makes two testable predictions, namely that: (i) PepP readily hydrolyzes peptides with the general formula X-thioproline-Y to X + thioproline-Y (X and Y being any amino acid); and (ii) deleting *pepP* increases sensitivity to supplied thioproline because this is incorporated into proteins [34] that are cleaved to peptides that are toxic. Our next step was to test these predictions.

Experimental tests of hypothesis predictions

We first tested the activity of purified recombinant PepP against the thioproline-containing peptide Ala-thioproline-Ala using its natural counterpart Ala-Pro-Ala as a benchmark (Ala-Pro-Ala is known to be a good PepP substrate [30,40]). Activity was assayed via a coupled spectrophotometric assay that measured release of alanine. As pilot experiments showed similar activity against both peptides, we proceeded to a full kinetic characterization (Table 1 and Supplementary Figure S4). Our K_M and k_{cat} values for the natural substrate Ala-Pro-Ala were consistent with those previously reported for this peptide [40]. As predicted, Ala-thioproline-Ala was a good substrate, giving K_M and k_{cat} values that did not differ statistically from those for Ala-Pro-Ala. More generally, the K_M and k_{cat} values for Ala-thioproline-Ala were within the range reported for PepP acting on various peptides [30]. PepP thus acted on a model thioproline-containing peptide just as if it were a regular proline-containing peptide.

To confirm that PepP cleaves Ala-thioproline-Ala *in vivo* as well as *in vitro* we exploited the ability of *E. coli* to use alanine as sole nitrogen source [41]. We reasoned that if PepP is the sole or main peptidase that cleaves the N-terminal amino acid from X-thioproline-Y and X-proline-Y peptides, and if such peptides are not hydrolyzed by carboxypeptidase activity [42], then wild type *E. coli* should be able to use Ala-thioproline-Ala or

Table 1 Kinetic parameters of recombinant *E. coli* PepP

Substrate	V_{\max} ($\mu\text{mol min}^{-1} \text{mg}^{-1}$)	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($\text{s}^{-1} \text{M}^{-1}$)
Ala-Pro-Ala	16.4 ± 1.6	14.0 ± 1.3	1.10 ± 0.26	1.27×10^4
Ala-thioprolino-Ala	16.2 ± 0.9	13.8 ± 0.9	1.01 ± 0.15	1.37×10^4

A coupled spectrophotometric assay was used. Values are means \pm S.E. ($n = 3$). The K_M value for Ala-Pro-Ala is close to the reported value (0.77 mM) [40] and the k_{cat} value is consistent with that reported previously (85 s^{-1}) [40], allowing for differences in assay pH and temperature. All values for Ala-thioprolino-Ala fall in the 95% confidence intervals for Ala-Pro-Ala values, i.e. are statistically the same.

Ala-Pro-Ala as sole nitrogen source and the ΔpepP strain should not. This proved to be the case (Figure 5B). Plates containing a lawn of cells and a central disc charged with either tripeptide gave growth halos around the disc with wild type cells but not with ΔpepP cells. Comparable results were obtained in liquid culture, i.e. the ΔpepP deletion strain lost most of its capacity to use Ala-thioprolino-Ala or Ala-Pro-Ala as sole nitrogen source (Figure 5C). The residual growth of the ΔpepP deletant was presumably due to a low level of alanine release by another peptidase(s). Note that the similar growth patterns of wild type and ΔpepP cells on Ala-Pro-Ala or Ala-thioprolino-Ala show that Ala-thioprolino-Ala was not toxic at the concentration used (2.5 mM). Note also that the non-toxicity of this particular tripeptide does not invalidate our hypothesis that some of the many possible thioproline-containing peptides are toxic, as further discussed below. Finally, we compared the responses of the wild type and ΔpepP strains to supplied thioproline. As predicted, the ΔpepP strain was more sensitive, becoming progressively more inhibited than wild type as the thioproline concentration was increased from 2 mM to 4 mM (Figure 6).

Ancillary experimental tests

Given the role of PutA in detoxifying thioproline (Figure 5A) [35], we compared the effect of deleting *putA* on sensitivity to HCHO to that of deleting *pepP*, using a concentration of HCHO (400 μM) that inhibits growth of the ΔpepP strain (Supplementary Figure S5). At this HCHO concentration, the ΔputA strain showed no growth defect. This result implies that the thioproline-containing peptides that PepP hydrolyzes are more toxic than thioproline itself. PepP thus appears to be more critical than PutA in the chain of reactions that cope with protein-bound thioproline (Figure 5A).

Because glutathione is ~ 100 -fold more abundant in *E. coli* than free cysteine [43,44] and reacts readily with HCHO to form S-(hydroxymethyl)glutathione and other adducts [45] we tested whether PepP can hydrolyze the adduct γ -Glu-thioprolino-Gly. This compound could conceivably form by spontaneous reaction of HCHO

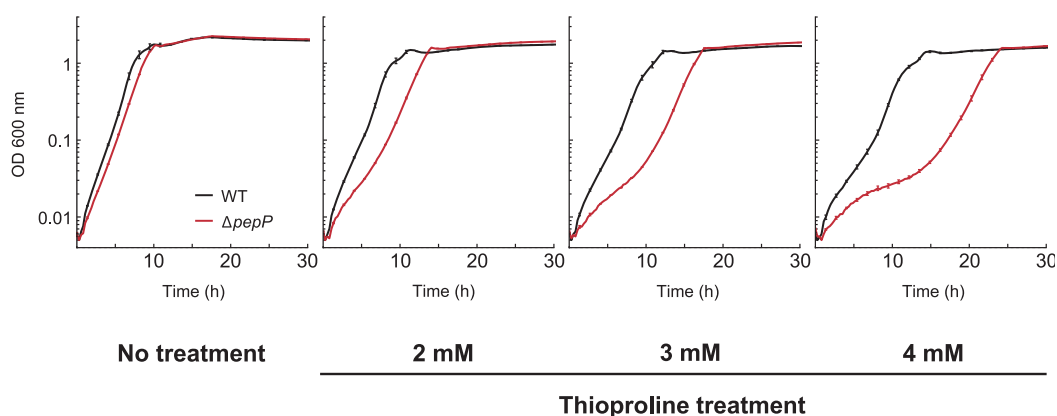


Figure 6. Deleting *pepP* increases sensitivity to supplied thioproline.

BW25113 wild type (WT) and ΔpepP strains were cultured in 150 μl of M9 liquid medium containing 10 mM glucose and the indicated concentrations of thioproline. Data are means \pm S.E. ($n = 3$).

with glutathione's cysteine residue, although it is not one of the reported HCHO adducts of glutathione [45]. γ -Glu-thioproline-Gly was incubated with PepP and glutamate release was assayed by a coupled spectrophotometric procedure. No activity was detected; the detection limit ($2 \text{ nmol mg}^{-1} \text{ protein min}^{-1}$) was $<0.01\%$ of the activity of PepP against Ala-Pro-Ala.

The Keio collection $\Delta yhbO$ strain carries a *pepP* nonsense mutation

In the course of this work we found by genome sequencing that the Keio collection $\Delta yhbO$ strain carries a nonsense mutation in the *pepP* ORF that changes codon 285 from CAG (glutamine) to TAG (stop). The resulting truncated PepP protein is evidently non-functional because the Keio $\Delta yhbO$ strain had the same HCHO sensitivity as the $\Delta pepP$ strain (Supplementary Figure S6). A $\Delta yhbO$ strain lacking the *pepP* nonsense mutation (constructed by P1 phage transduction from the BW25113 Keio parent strain) showed no increase in HCHO sensitivity (Supplementary Figure S6), confirming that the HCHO sensitivity of the *pepP*- $\Delta yhbO$ strain is due solely to the PepP defect.

Conclusions

The core parts of our hypothesis connecting the Xaa-Pro aminopeptidase PepP with HCHO resistance are (i) that certain thioproline-containing peptides having thioproline as second residue are toxic, and (ii) that PepP can cleave Xaa-thioproline peptides (Figure 5A). Part (ii) is validated by the evidence presented above. Part (i) is a strong inference—and will likely remain so because it is so challenging to test. With thioproline at position two, there are 20^2 (400) possible tripeptides, 20^3 (8000) possible tetrapeptides, and so forth up to 20^9 (512 billion) possible decapeptides. Assuming, as for other peptides [46,47], that toxicity is restricted to a tiny fraction of the total sequence space, it would be necessary to screen an intractable number of peptides of various sizes and sequences in order to identify the few that are toxic. Moreover, these peptides would have to be supplied exogenously but might not be taken up because bacterial peptide transporters have various degrees of substrate specificity [48], or might be hydrolyzed before uptake by peptidases released by dead cells [49].

Although toxicity testing of thioproline-containing peptides is infeasible, it is interesting to ask what the basis of their toxicity could be. We suggest the following general explanation, based on the findings that certain non-natural oligopeptides inhibit *E. coli* growth [50] and that thioproline-containing polypeptides can differ greatly in physicochemical properties from their natural proline-containing counterparts [34]. Perhaps certain thioproline-containing peptides simply differ enough from their harmless natural analogs to be toxic via the same undefined mechanisms as other alien oligopeptides [50].

In sum, our genetic and biochemical findings on PepP provide strong evidence that the formation of thioproline-containing peptides is a substantial driver of HCHO toxicity. From an engineering standpoint, this evidence makes PepP a potential damage-control component [17] for use in designing and building pathways in which HCHO is an intermediate (Figure 1). Although the damage-control pathway involving PepP (Figure 5A) ultimately converts HCHO to formate, the flux via this pathway is constrained by the rate of proteolysis and is hence almost certainly too low to constitute a major drain on the HCHO pool. From a physiological standpoint, the evidence that PepP deals with a toxic consequence of HCHO formation—which happens in normal metabolism [1–3]—raises a question: Is cleaving thioproline peptides the primary role of PepP? If so, the well-characterized activity of PepP against proline peptides would be of secondary importance. That the $\Delta pepP$ strain has no growth defect on minimal medium (Figure 3A,B) fits with this possibility. So too does the inference [51] that *Lactococcus lactis* PepP does not have a major role in nitrogen nutrition and probably has some other function.

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript. A.B.-E. is co-founder of b.fab, which aims to commercialize microbial growth on C_1 compounds. b.fab was not involved in this study and did not fund it.

Funding

This research was supported by U.S. National Science Foundation awards MCB-1611711 to A.D.H. and MCB-1611846 to O.F., by Hatch Project FLA-HOS-005796, and by an endowment from the C.V. Griffin Sr. Foundation. A.B.-E. and H.H. are funded by the Max Planck Society. H.H. is also funded by the China Scholarship Council.

Author Contribution

A.D.H., A.B.-E., J.A.P., H.H., and M.A.W. conceived and designed the research; J.A.P. and H.H. performed experiments with supervision from A.D.H. and A.B.-E.; Q.L. and S.D.B. synthesized substrates; J.S.F. and O.F. carried out metabolomic analyses; A.D.H. made computational gene clustering analyses; J.A.P. and A.D.H. wrote the article.

Acknowledgements

We thank Laszlo N. Csonka for his insightful advice on bacterial genetics, and Michael J. Ziemak for his contribution to advancing this work.

Abbreviations

HCHO, formaldehyde; LB, Luria Broth; THF, tetrahydrofolate.

References

- Chen, N.H., Djoko, K.Y., Veyrier, F.J. and McEwan, A.G. (2016) Formaldehyde stress responses in bacterial pathogens. *Front. Microbiol.* **7**, 257 <https://doi.org/10.3389/fmicb.2016.00257>
- Hanson, A.D. and Roje, S. (2001) One-carbon metabolism in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 119–137 <https://doi.org/10.1146/annurev.arplant.52.1.119>
- Dringen, R., Brandmann, M., Hohnholt, M.C. and Blumrich, E.M. (2015) Glutathione-dependent detoxification processes in astrocytes. *Neurochem. Res.* **40**, 2570–2582 <https://doi.org/10.1007/s11064-014-1481-1>
- Cotton, C.A., Claassens, N.J., Benito-Vaquerizo, S. and Bar-Even, A. (2019) Renewable methanol and formate as microbial feedstocks. *Curr. Opin. Biotechnol.* **62**, 168–180 <https://doi.org/10.1016/j.copbio.2019.10.002>
- 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 <https://doi.org/10.1016/j.cbpa.2016.07.005>
- Kim, S.J., Yoon, J., Im, D.K., Kim, Y.H. and Oh, M.K. (2019) Adaptively evolved *Escherichia coli* for improved ability of formate utilization as a carbon source in sugar-free conditions. *Biotechnol. Biofuels* **12**, 207 <https://doi.org/10.1186/s13068-019-1547-z>
- Schroer, K., Luef, K.P., Hartner, F.S., Glieder, A. and Pscheidt, B. (2010) Engineering the *Pichia pastoris* methanol oxidation pathway for improved NADH regeneration during whole-cell biotransformation. *Metab. Eng.* **12**, 8–17 <https://doi.org/10.1016/j.ymben.2009.08.006>
- Metz, B., Kersten, G.F., Hoogerhout, P., Brugghe, H.F., Timmermans, H.A., de Jong, A. et al. (2004) Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J. Biol. Chem.* **279**, 6235–6243 <https://doi.org/10.1074/jbc.M310752200>
- Hoffman, E.A., Frey, B.L., Smith, L.M. and Auble, D.T. (2015) Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J. Biol. Chem.* **290**, 26404–26411 <https://doi.org/10.1074/jbc.R115.651679>
- Kamps, J.J.A.G., Hopkinson, R.J., Schofield, C.J. and Claridge, T.D.W. (2019) How formaldehyde reacts with amino acids. *Commun. Chem.* **2**, 126 <https://doi.org/10.1038/s42004-019-0224-2>
- Liu, J., Chan, K.K. and Chan, W. (2016) Identification of protein thiazolidination as a novel molecular signature for oxidative stress and formaldehyde exposure. *Chem. Res. Toxicol.* **29**, 1865–1871 <https://doi.org/10.1021/acs.chemrestox.6b00271>
- Wisniewski, J.R., Zougman, A. and Mann, M. (2008) N^6 -Formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. *Nucleic Acids Res.* **36**, 570–577 <https://doi.org/10.1093/nar/gkm1057>
- Van Spanning, R.J.M., de Vries, S. and Harms, N. (2000) Coping with formaldehyde during C1 metabolism of *Paracoccus denitrificans*. *J. Mol. Catal. B: Enzym.* **8**, 37–50 [https://doi.org/10.1016/S1381-1177\(99\)00065-X](https://doi.org/10.1016/S1381-1177(99)00065-X)
- Jakobsen, Ø.M., Benichou, A., Flickinger, M.C., Valla, S., Ellingsen, T.E. and Brautaset, T. (2006) Upregulated transcription of plasmid and chromosomal ribulose monophosphate pathway genes is critical for methanol assimilation rate and methanol tolerance in the methylotrophic bacterium *Bacillus methanolicus*. *J. Bacteriol.* **188**, 3063–3072 <https://doi.org/10.1128/JB.188.8.3063-3072.2006>
- Seidel, J., Klockenbusch, C. and Schwarzer, D. (2016) Investigating deformylase and deacylase activity of mammalian and bacterial sirtuins. *ChemBioChem* **17**, 398–402 <https://doi.org/10.1002/cbic.201500611>
- Bouzon, M., Perret, A., Loreau, O., Delmas, V., Perchat, N., Weissenbach, J. et al. (2017) A synthetic alternative to canonical one-carbon metabolism. *ACS Synth. Biol.* **6**, 1520–1533 <https://doi.org/10.1021/acssynbio.7b00029>
- Sun, J., Jeffries, J.G., Henry, C.S., Bruner, S.D. and Hanson, A.D. (2017) Metabolite damage and repair in metabolic engineering design. *Metab. Eng.* **44**, 150–159 <https://doi.org/10.1016/j.ymben.2017.10.006>
- Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J. et al. (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 <https://doi.org/10.1093/nar/gky1131>
- Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M. et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**, 5691–5702 <https://doi.org/10.1093/nar/gki866>
- Hanson, A.D., Pribat, A., Waller, J.C. and de Crécy-Lagard, V. (2009) ‘Unknown’ proteins and ‘orphan’ enzymes: the missing half of the engineering parts list—and how to find it. *Biochem. J.* **425**, 1–11 <https://doi.org/10.1042/BJ20091328>
- McClure, J.J., Inks, E.S., Zhang, C., Peterson, Y.K., Li, J., Chundru, K. et al. (2017) Comparison of the deacylase and deacetylase activity of zinc-dependent HDACs. *ACS Chem. Biol.* **12**, 1644–1655 <https://doi.org/10.1021/acscchembio.7b00321>
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M. et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 <https://doi.org/10.1038/msb4100050>

- 23 Thomason, L.C., Costantino, N. and Court, D.L. (2007) *E. coli* genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.* **79**, 1.17.1–1.17.8 <https://doi.org/10.1002/0471142727.mb0117s79>
- 24 Quan, J. and Tian, J. (2009) Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* **4**, e6441 <https://doi.org/10.1371/journal.pone.0006441>
- 25 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 [https://doi.org/10.1016/0003-2697\(76\)90527-3](https://doi.org/10.1016/0003-2697(76)90527-3)
- 26 de Crécy-Lagard, V., Haas, D. and Hanson, A.D. (2018) Newly-discovered enzymes that function in metabolite damage-control. *Curr. Opin. Chem. Biol.* **47**, 101–108 <https://doi.org/10.1016/j.cbpa.2018.09.014>
- 27 Kallen, R.G. and Jencks, W.P. (1966) The mechanism of the condensation of formaldehyde with tetrahydrofolic acid. *J. Biol. Chem.* **241**, 5851–5863 PMID:5954363
- 28 Stover, P. and Schirch, V. (1990) Serine hydroxymethyltransferase catalyzes the hydrolysis of 5,10-methenyltetrahydrofolate to 5-formyltetrahydrofolate. *J. Biol. Chem.* **265**, 14227–14233 PMID:2201683
- 29 Pourmotabbed, T. and Creighton, D.J. (1986) Substrate specificity of bovine liver formaldehyde dehydrogenase. *J. Biol. Chem.* **261**, 14240–14244 PMID:3771532
- 30 Yoshimoto, T., Orawski, A.T. and Simmons, W.H. (1994) Substrate specificity of aminopeptidase P from *Escherichia coli*: comparison with membrane-bound forms from rat and bovine lung. *Arch. Biochem. Biophys.* **311**, 28–34 <https://doi.org/10.1006/abbi.1994.1204>
- 31 Bompard-Gilles, C., Villeret, V., Davies, G.J., Fanuel, L., Joris, B., Frère, J.M. et al. (2000) A new variant of the Ntn hydrolase fold revealed by the crystal structure of L-aminopeptidase D- α -esterase/amidase from *Ochrobactrum anthropi*. *Structure* **8**, 153–162 [https://doi.org/10.1016/S0969-2126\(00\)00091-5](https://doi.org/10.1016/S0969-2126(00)00091-5)
- 32 Komeda, H., Hariyama, N. and Asano, Y. (2006) L-Stereoselective amino acid amidase with broad substrate specificity from *Brevundimonas diminuta*: characterization of a new member of the leucine aminopeptidase family. *Appl. Microbiol. Biotechnol.* **70**, 412–421 <https://doi.org/10.1007/s00253-005-0068-9>
- 33 Budisa, N., Minks, C., Medrano, F.J., Lutz, J., Huber, R. and Moroder, L. (1998) Residue-specific bioincorporation of non-natural, biologically active amino acids into proteins as possible drug carriers: structure and stability of the per-thiaproline mutant of annexin V. *Proc. Natl Acad. Sci. U.S.A.* **95**, 455–459 <https://doi.org/10.1073/pnas.95.2.455>
- 34 Liu, J., Hao, C., Wu, L., Madej, D., Chan, W. and Lam, H. (2020) Proteomic analysis of thioproline misincorporation in *Escherichia coli*. *J. Proteomics* **210**, 103541 <https://doi.org/10.1016/j.jprot.2019.103541>
- 35 Unger, L. and DeMoss, R.D. (1966) Metabolism of a proline analogue, L-thiazolidine-4-carboxylic acid, by *Escherichia coli*. *J. Bacteriol.* **91**, 1564–1569 <https://doi.org/10.1128/JB.91.4.1564-1569.1966>
- 36 Jeelani, G., Sato, T., Soga, T., Watanabe, H. and Nozaki, T. (2014) Mass spectrometric analysis of L-cysteine metabolism: physiological role and fate of L-cysteine in the enteric protozoan parasite *Entamoeba histolytica*. *mBio* **5**, e01995 <https://doi.org/10.1128/mBio.01995-14>
- 37 Liu, J., Meng, X. and Chan, W. (2016) Quantitation of thioprolines in grape wine by isotope dilution-liquid chromatography-tandem mass spectrometry. *J. Agric. Food Chem.* **64**, 1361–1366 <https://doi.org/10.1021/acs.jafc.5b05604>
- 38 Saravanan Prabhu, N., Ayyadurai, N., Deepankumar, K., Chung, T., Lim, D. and Yun, H. (2012) Reassignment of sense codons: designing and docking of proline analogs for *Escherichia coli* prolyl-tRNA synthetase to expand the genetic code. *J. Mol. Catal. B* **78**, 57–64 <https://doi.org/10.1016/j.molcatb.2012.02.004>
- 39 Jeffryes, J.G., Colastani, R.L., Elbadawi-Sidhu, M., Kind, T., Niehaus, T.D., Broadbelt, L.J. et al. (2015) MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. *J. Cheminform.* **7**, 44 <https://doi.org/10.1186/s13321-015-0087-1>
- 40 Graham, S.C., Lilley, P.E., Lee, M., Schaeffer, P.M., Kralicek, A.V., Dixon, N.E. et al. (2006) Kinetic and crystallographic analysis of mutant *Escherichia coli* aminopeptidase P: insights into substrate recognition and the mechanism of catalysis. *Biochemistry* **45**, 964–975 <https://doi.org/10.1021/bi0518904>
- 41 Varricchio, F. (1969) Control of glutamate dehydrogenase synthesis in *Escherichia coli*. *Biochim. Biophys. Acta* **177**, 560–564 [https://doi.org/10.1016/0304-4165\(69\)90319-5](https://doi.org/10.1016/0304-4165(69)90319-5)
- 42 Yaron, A., Mlynar, D. and Berger, A. (1972) A dipeptidocarboxypeptidase from *E. coli*. *Biochem. Biophys. Res. Commun.* **47**, 897–902 [https://doi.org/10.1016/0006-291X\(72\)90577-3](https://doi.org/10.1016/0006-291X(72)90577-3)
- 43 Wheldrake, J.F. (1967) Intracellular concentration of cysteine in *Escherichia coli* and its relation to repression of the sulphate-activating enzymes. *Biochem. J.* **105**, 697–699 <https://doi.org/10.1042/bj1050697>
- 44 Smirnova, G.V., Tyulenev, A.V., Bezmaternykh, K.V., Muzyka, N.G., Ushakov, V.Y. and Oktyabrsky, O.N. (2019) Cysteine homeostasis under inhibition of protein synthesis in *Escherichia coli* cells. *Amino Acids* **51**, 1577–1592 <https://doi.org/10.1007/s00726-019-02795-2>
- 45 Hopkinson, R.J., Barlow, P.S., Schofield, C.J. and Claridge, T.D.W. (2010) Studies on the reaction of glutathione and formaldehyde using NMR. *Org. Biomol. Chem.* **8**, 4915–4920 <https://doi.org/10.1039/c0ob00208a>
- 46 Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Open Source Drug Discovery Consortium et al. (2013) *In silico* approach for predicting toxicity of peptides and proteins. *PLoS ONE* **8**, e73957 <https://doi.org/10.1371/journal.pone.0073957>
- 47 Wang, G., Li, X. and Wang, Z. (2015) APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* **44**, D1087–D1093 <https://doi.org/10.1093/nar/gkv1278>
- 48 Garai, P., Chandra, K. and Chakravorty, D. (2017) Bacterial peptide transporters: messengers of nutrition to virulence. *Virulence* **8**, 297–309 <https://doi.org/10.1080/21505594.2016.1221025>
- 49 Lazdunski, A.M. (1989) Peptidases and proteases of *Escherichia coli* and *Salmonella typhimurium*. *FEMS Microbiol. Rev.* **5**, 265–276 [https://doi.org/10.1016/0168-6445\(89\)90035-1](https://doi.org/10.1016/0168-6445(89)90035-1)
- 50 Walker, J.R., Roth, J.R. and Altman, E. (2001) An *in vivo* study of novel bioactive peptides that inhibit the growth of *Escherichia coli*. *J. Pept. Res.* **58**, 380–388 <https://doi.org/10.1034/j.1399-3011.2001.00897.x>
- 51 Matos, J., Nardi, M., Kumura, H. and Monnet, V. (1998) Genetic characterization of *pepP*, which encodes an aminopeptidase P whose deficiency does not affect *Lactococcus lactis* growth in milk, unlike deficiency of the X-prolyl dipeptidyl aminopeptidase. *Appl. Environ. Microbiol.* **64**, 4591–4595 <https://doi.org/10.1128/AEM.64.11.4591-4595.1998>

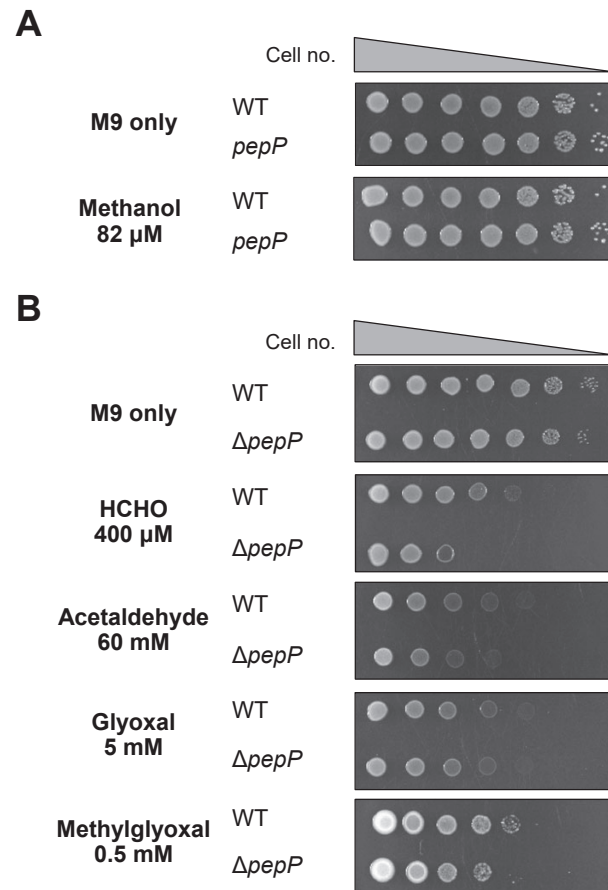


Figure S1. Lack of effect of ablating *pepP* on sensitivity to methanol and small carbonyls.

E. coli BW25113 (WT) and *pepP* mutant cells were cultured on M9 medium containing 0.4% glucose and the indicated concentrations of methanol or small carbonyl compounds. Overnight liquid cultures of each strain were ten-fold serially diluted and 3.5- μ l aliquots were spotted on the plates. Images were captured after incubation at 37°C for 1.5 d.

(A) Methanol. The concentration (82 μ M) is that estimated to be present as a contaminant when the standard HCHO concentration (400 μ M) is used. The *pepP* null mutant carried a deletion (Δ *yhbO*) that does not affect HCHO sensitivity (see Supplementary Figure S6).

(B) Acetaldehyde, glyoxal, or methylglyoxal. The *pepP* null mutant was the Δ *pepP* strain.

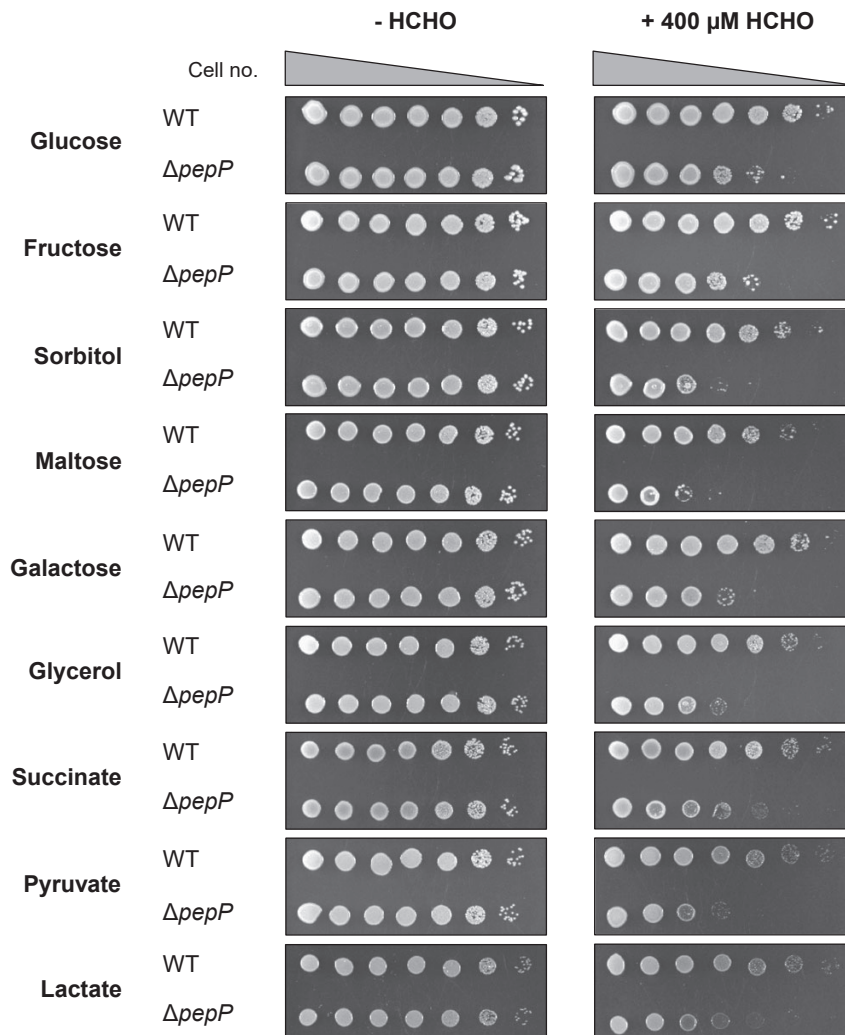


Figure S2. HCHO-sensitivity of the $\Delta pepP$ strain is independent of carbon source.

E. coli BW25113 (WT) and its $\Delta pepP$ derivative were cultured on M9 minimal medium containing 0.4% of the nine carbon sources shown, minus or plus 400 μ M HCHO. Overnight liquid cultures of each strain were ten-fold serially diluted and 3.5- μ l aliquots were spotted on the plates. Images were captured after incubation at 37°C for 1.5 d or, for plates containing lactate or pyruvate, 2 d.

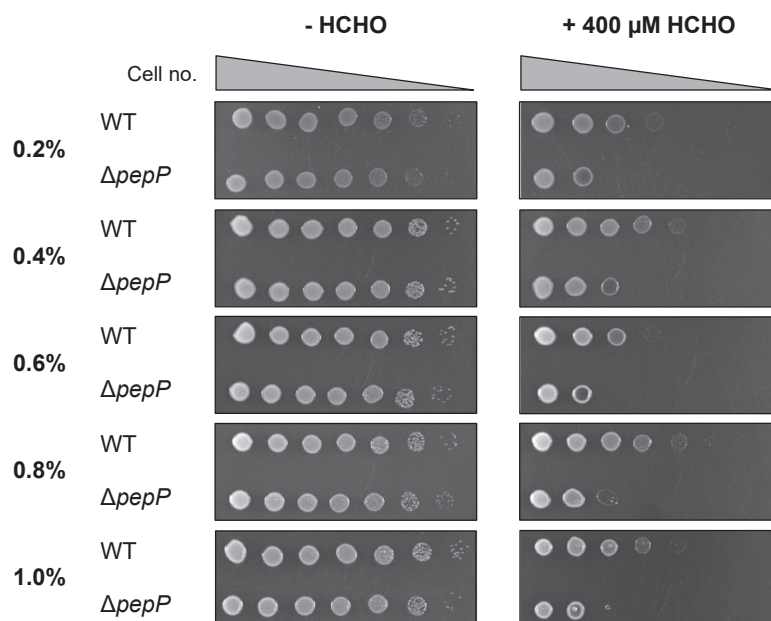


Figure S3. HCHO-sensitivity of the $\Delta pepP$ strain is independent of glucose concentration.

E. coli BW25113 (WT) and its $\Delta pepP$ derivative were cultured on M9 minimal medium containing glucose at the indicated concentrations, minus or plus 400 μ M HCHO. Overnight liquid cultures of each strain were ten-fold serially diluted and 3.5- μ l aliquots were spotted on the plates. Images were captured after incubation at 37°C for 1.5 d.

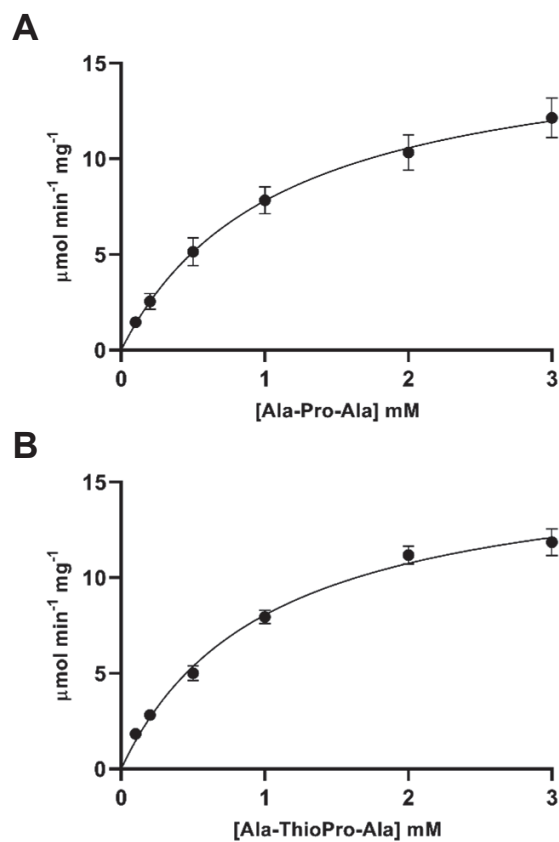


Figure S4. Primary data from kinetic experiments.

K_m and V_{max} values were determined by varying the concentration of Ala-Pro-Ala (**A**) or Ala-thiopropine-Ala (**B**) in coupled spectrophotometric assays containing 50 mM Tris-HCl, pH 8.8, 0.5 mM $MnCl_2$, 1 mM NAD, 0.05 units of alanine dehydrogenase, and 2 μg PepP. The data were analyzed by nonlinear regression using GraphPad Prism. Data are means \pm s.e. ($n = 3$).

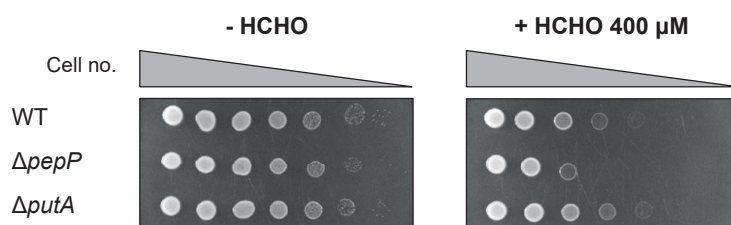


Figure S5. Growth assays indicate that *E. coli* PutA does not contribute to HCHO resistance.

E. coli BW25113 (WT), $\Delta pepP$ and $\Delta putA$ strains were cultured on M9 minimal medium containing 0.4% glucose, minus or plus 400 μ M HCHO. Overnight liquid cultures of each strain were ten-fold serially diluted and 3.5- μ l aliquots were spotted on the plates. Images were captured after incubation at 37°C for 1.5 d.

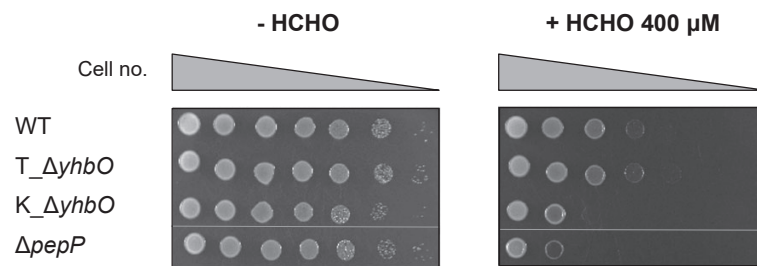


Figure S6. A *pepP* mutation explains the formaldehyde sensitivity of the Keio $\Delta yhbO$ strain.

E. coli BW25113 (WT), Keio $\Delta yhbO$ (K_ΔyhbO), P1 phage-transduced $\Delta yhbO$ (T_ΔyhbO), and Keio $\Delta pepP$ strains were cultured on M9 minimal medium containing 0.4% glucose, minus or plus 400 μM HCHO. Overnight liquid cultures of each strain were ten-fold serially diluted and 3.5-μl aliquots were spotted on the plates. Images were captured after incubation at 37°C for 1.5 d.

Table S1. Primers for PCR amplification.

Primer	Name	Sequence
<i>E. coli</i> ΔpepP complementation		
1	FP_pepP_EcoRI	AAAGAATTCACCATGAGTGAGATATCCCGGC
2	RP_pepP_Sall	TTTGTCGACTCATTGCTTTCTCGCAGC
PepP protein expression		
3	FP_pepP_NC-term	ATGAGTGAGATATCCCG
4	RP_pepP_N-term	CTTTGTTAGCAGCCGGATCATTGCTTTCTCGC
5	RP_p15_N-term	CGGGATATCTCACTCATGCCGCTGCTGTGATGATGATG
6	FP_p15_NC-term	TCCGGCTGCTAACAAAGC

Table S2. MRM parameters for each targeted metabolite.

Q1 and Q3 refer to the mass transitions monitored for each analyte. RT is the chromatographic retention time in minutes, DP, EP, CE, and CXP stand for declustering potential, entrance potential, collision energy, and cell exit potential respectively, and are measured in volts.

Compound	Q1 (Da)	Q3 (Da)	RT (min)	DP	EP	CE	CXP
Lysine	147.1	130.1	9.67	6	10	21	12
<i>N</i> ⁶ -formyl-formyl lysine	175.1	112.1	8.36	26	10	19	12
Cystine	241	152	10.04	31	10	17	16
D5 Threonine (IS)	125.1	79.1	8.44	16	10	15	10
D8 Lysine (IS)	155.2	138.2	9.67	31	10	13	14

Overarching Discussion

C1 feedstocks are of special interest for a future green and sustainable bioindustry. Formaldehyde as a versatile intermediate plays an important role as a link between the abiotic and the biotic worlds. As the application of natural methylotrophs is still limited (Kalyuzhnaya et al., 2015; Wang et al., 2020), great efforts were made to establish synthetic methylotrophy. Especially the RuMP pathway, the highest efficient natural pathway, has been subject to extensive engineering in *E. coli* (Bennett et al., 2018; Chen et al., 2018; Gonzalez et al., 2018; He et al., 2018; Meyer et al., 2018; Müller et al., 2015; Rohlhill et al., 2020; Wang et al., 2019; Whitaker et al., 2017), *C. glutamicum* (Lessmeier et al., 2015; Tuyishime et al., 2018; Witthoff et al., 2015) and yeasts (Dai et al., 2017; Espinosa et al., 2019; Vartiainen et al., 2019). Additionally, many synthetic pathways were designed and tested *in vivo*, such as the modified serine cycle (Yu and Liao, 2018), the FLS pathway (Wang et al., 2017), the HACL-based pathway (Chou et al., 2019), the SACA pathway (Lu et al., 2019) and the reductive glycine pathway (Kim et al., 2020). This thesis aimed at engineering formaldehyde assimilation in the model microbe *E. coli*, exploring various metabolic aspects of synthetic formaldehyde assimilation. Although C1 dependent growth was achieved in the case of the RuMP pathway (Chen et al., 2018; He et al., 2018; Meyer et al., 2018; Tuyishime et al., 2018) and the reductive glycine pathway (Kim et al., 2020), there are still various challenges towards fast growth solely on C1 feedstocks via formaldehyde.

***In vivo* formaldehyde production**

C1 feedstocks can be transformed to formaldehyde as an intermediate for their assimilation. Because formaldehyde is highly toxic, its production *in situ* is advantageous to control the concentration of this reactive intermediate. Natural utilization of methanol, as well as methane, proceeds via formaldehyde and the transformation is catalyzed by methanol dehydrogenase (MDH). However the conversion efficiency is highly constrained by poor kinetics and thermodynamic properties of the NAD-dependent MDH.

In order to test the key enzymes of the RuMP pathway (Chapter 1), the efficient formaldehyde production was achieved by using sarcosine oxidation. Formaldehyde accumulation upon sarcosine oxidation is an order of magnitude higher than with methanol oxidation. This strategy was also applied in Chapter 3, where only a high concentration of formaldehyde from sarcosine oxidation can be used to test the *in vivo* rate of spontaneous formaldehyde THF condensation.

However, C1 substrates rather than sarcosine are the promising feedstocks. Therefore, efficient enzymes for *in situ* formaldehyde formation from C1 feedstocks need to be explored. Methanol oxidation via NAD-dependent MDHs is thermodynamically unfavorable, indicated by a positive reaction Gibbs energy ($\Delta_r G'$). Alternatively, pyrroloquinoline quinone (PQQ)-dependent MDHs or O₂-dependent methanol oxidases (MODs) catalyze a thermodynamically favorable methanol oxidation also with better kinetics (Whitaker et al., 2015). Therefore, they can potentially be used to produce formaldehyde efficiently. But their biotechnological utilization is yet to be established. Moreover, unlike the NAD-dependent MDHs where electrons from methanol are transferred to NADH, electrons in MODs and PQQ-dependent MDHs catalyzing methanol oxidation are transferred directly or indirectly to O₂, thus wasting electrons. As electron conservation is essential for more reduced products, such as butanol, butyrate and isoprene, the application of PQQ-dependent MDHs and MODs will limit the product spectrum. For this reason, NAD-dependent MDHs are more useful for industrial utilization.

NAD-dependent MDH from *Bacillus stearothermophilus* was found to be more active than its counterparts from *B. methanolicus* MGA3 and *C. glutamicum* R (Chapter 1). It indicates that MDHs from new sources or broad substrate specificity alcohol dehydrogenases could potentially be selected for better kinetics. Alternatively, enzyme activities can be improved by mutagenesis (Wu et al., 2016), phage-assisted noncontiguous evolution (Roth et al., 2019), or engineering a microenvironment for enzyme activities via fusion with supercharged proteins (Abdallah et al., 2019). On the other hand, the intracellular ratio of [NAD⁺]/[NADH] could be increased, towards the favorable direction of MDHs, by disrupting the TCA cycle (Meyer et al., 2018; Rohlhill et al., 2020) or introducing an “NADH sink”, for instance a lactate dehydrogenase (Price et al., 2016). It is also possible to employ thermophilic hosts as NAD-dependent methanol oxidation becomes more favorable at high temperatures (Whitaker et al., 2015).

To identify better MDHs, the colorimetric reaction with NASH reagent (Nash, 1953) can be used for high throughput screening. While this chemical reaction detects the extracellular formaldehyde concentration, formaldehyde dependent strains constructed in this thesis or strains reported by Rohlhill et al. (2017) as well as Woolston et al. (2018b) can be employed as formaldehyde biosensors to directly reflect intracellular formaldehyde production.

Conversion of formate to formaldehyde *in vivo* is interesting, as although being a promising feedstock, formate itself is an inert substrate, there are a limited number of formate-fixing

reactions (Bar-Even, 2016). However, the conversion to the reactive formaldehyde is challenging, since the reduction potential of formate to formaldehyde is very low. Formate must be activated if NAD(P)H is to be used as an electron donor (Bar-Even, 2016). Formate activation with CoA or phosphate forming formyl-CoA or formyl-phosphate was suggested, but the efficient enzymes are not available and need further study (Bar-Even, 2016; Siegel et al., 2015). Since formate can be efficiently condensed with THF by formate-THF ligase (Kim et al., 2020; Yishai et al., 2017), cleavage of the derived methylene-THF can be used to produce formaldehyde. But this spontaneous cleavage reaction is too slow, deduced from the results in Chapter 3 showing that the condensation direction of this reversible reaction is slow. Thus enzymes catalyzing such reaction can be explored in this pathway for the efficient transformation of formate to formaldehyde. And the formaldehyde biosensors as mentioned above can be applied to assess the enzymes or pathways.

A more valuable strategy is to assemble formaldehyde-producing enzymes with downstream enzymes forming a substrate channel, as it is not essential to have perfect individual enzymes. Since cascade reactions work as a whole, they can integrate a thermodynamically unfavorable reaction but still be overall thermodynamically favorable (Noor et al., 2014; Wheeldon et al., 2016). In this way, the thermodynamic barrier of formaldehyde producing enzymes, e.g. NAD-dependent MDHs, can be overcome. It has already been shown that assembling or fusing NAD-dependent MDHs with HPS and PHI of the RuMP pathway lead to higher methanol conversion rates (Fan et al., 2018; Price et al., 2016).

Formaldehyde assimilation

A high formaldehyde concentration will promote its assimilation. Selection strains for the RuMP shunt in Chapter 1 and the methylene-THF auxotrophic strain in Chapter 3 did not grow when formaldehyde concentration was produced from methanol, but only grew when high formaldehyde concentrations were sustained from sarcosine oxidation. A high formaldehyde concentration is especially critical for those “new-to-nature” reactions that are catalyzed by enzymes with poor kinetics and that were characterized *in vitro* using high formaldehyde concentrations, such as FLS (Siegel et al., 2015), GALS (Lu et al., 2019) and HACL (Chou et al., 2019).

In most cases, the native formaldehyde detoxification system, such as the glutathione-dependent formaldehyde dehydrogenase system (FrmAB) in *E. coli*, should be knocked-out to avoid wasting substrate by oxidation and to elevate intracellular formaldehyde concentrations.

Such necessity was also found in studies where mutations occurred during evolution of *E. coli* strains with an engineered RuMP pathway (Chen et al., 2018; Meyer et al., 2018). However, for the reductive glycine pathway, the serine cycle and its variants, formaldehyde is first oxidized to formate to be fixed. A fraction of these formate need to be completely oxidized to CO₂ to meet the electron and energy requirements. In order to maximize production while minimizing the amount of resources invested, a fine tuning system to control substrate oxidation and substrate assimilation is vital.

Genetic circuits or metabolic switches (Andrews et al., 2018; Purnick and Weiss, 2009) can be applied to regulate the expressions of genes responsible for formaldehyde production, oxidation and assimilation. The formaldehyde-inducible promoter P_{frm} of *E. coli* can be used to balance the expression of genes responsible for formaldehyde production and consumption. An array of P_{frm} variants developed by Rohlhill et al (2017) can be used as a toolbox. A methanol-inducible expression system was also developed in *E. coli* (Ganesh et al., 2017), and thus can be used to construct sequential genetic circuits with P_{frm} for methanol assimilation. As a result of its reactivity, formaldehyde is highly toxic. Formaldehyde concentrations higher than 0.5 mM in the medium are lethal to WT *E. coli* (Chapter 1). Therefore, such regulation systems can be used to resolve the challenge of requiring high level formaldehyde concentration while avoiding toxicity. In addition, PepP (Chapter 4) and other enzymes/routes can serve as damage-control components to improve the fitness of strains toward formaldehyde.

Apart from system level genetic design, enzyme engineering is a strategic approach. Formaldehyde assimilation enzymes with low activity can be improved by protein engineering to increase enzyme specificity and kinetics. Metabolic channeling between enzymes of producing and consuming formaldehyde is particularly interesting, because it can affect the driving force of the pathway reactions as mentioned above, and also decrease the diffusion of substrates (Wheeldon et al., 2016), thus relieve the effects of formaldehyde toxicity.

Metabolic network

In order to achieve high yields and productivity, pathways with higher theoretical efficiency have attracted more interest, e.g. the RuMP pathway. While a single pathway or a single host not necessarily fits all application cases (Chapter 2), diverse platforms should be developed adapting different scopes in order to make the production feasible and economical. However, constructing a synthetic pathway in a host is still constrained by the insufficient knowledge of the metabolic network. Massive whole genome sequencing and various tools of bio-informatics

and system biology can partially fill the knowledge gap. A good example is given in Chapter 4, which applied a comparative genomics strategy. Nevertheless, a stepwise approach possess high value in metabolic engineering by dividing the pathway into modular segments for bottlenecks identification and optimization. This enables the scheme of the “design-build-test-learn” cycle.

This thesis applied this strategy: in Chapter 1 the RuMP cycle was broken down into three modules. The results from implementing the key enzymes HPS and PHI in selection strains with increasing selection pressure demonstrated that they are of sufficiently active to support almost all requirements for growth. In contrast, the main bottleneck of establishing a full RuMP cycle is Ru5P regeneration. The same conclusion was also demonstrated by Woolston et al. (2018a). In Chapter 3, the spontaneous formaldehyde and THF condensation reaction, which was proposed to contribute to C1 assimilation, was assessed separately and exhibited too little activity to be effective. In Chapter 2, instead of directly testing the full activity of the novel designed homoserine cycle, the feasibility of the pathway was demonstrated *in vivo* by confirming all segment activities.

To realize a fully functional synthetic pathway, modular integration is a major challenge due to the needs of flux balance between modules. Compared to linear pathways, like the reductive glycine pathway, an autocatalytic cycle especially requires proper regulation to avoid draining pathway intermediates. Additionally, overlaps with core metabolism can make pathway engineering more daunting, since innate regulation has evolved for native growth conditions. Genetic circuits and metabolic switches, e.g. the formaldehyde-inducible promoter P_{frm} (Rohllhill et al., 2020), are highly useful rational strategies. However, due to the lack of knowledge and proper toolboxes, adaptive laboratory evolution (ALE) is required. It has been successfully applied in establishing the Calvin cycle in *E. coli* (Gleizer et al., 2019) and improving the growth rate the reductive glycine pathway (Kim et al., 2020) as well as methanol-dependent growth via the RuMP pathway (Chen et al., 2018; Meyer et al., 2018; Tuyishime et al., 2018). It is also particularly important to perform ALE to improve the fitness and robustness of engineered strains for industrial utilization.

Closing remarks

Although the Calvin cycle and the reductive glycine pathway were recently established synthetically, there is a still long way to go towards a bioeconomy fed on C1 feedstocks. Formaldehyde equipped with its unique features is an excellent intermediate. Facilitated by its

ability to participate in diverse reactions, various C1 assimilation pathways via formaldehyde are to be constructed for different industrial usages. This thesis used modularity and selection metabolic engineering strategies to explore and engineer formaldehyde assimilation pathways in the model organism *E. coli*. While a fully functional RuMP cycle or homoserine cycle has yet to be established, the module activities of pathways (naturally existing or newly designed) were demonstrated and bottlenecks were identified. The combinatorial engineering approach of rational design and harnessing the power of nature (i.e. ALE) has the potential to address the challenges and to establish efficient C1-fed microbial cell factories.

References (overarching introduction and discussion)

- Abdallah, W.; Chirino, V.; Wheeldon, I.; Banta, S. Catalysis of Thermostable Alcohol Dehydrogenase Improved by Engineering the Microenvironment through Fusion with Supercharged Proteins. *ChemBioChem* **2019**, *20*, 1827–1837.
- Albert, J.; Lüders, D.; Bösmann, A.; Guldi, D.M.; Wasserscheid, P. Spectroscopic and electrochemical characterization of heteropoly acids for their optimized application in selective biomass oxidation to formic acid. *Green Chem.* **2014**, *16*, 226–237.
- Andrews, L.B.; Nielsen, A.A.K.; Voigt, C.A. Cellular checkpoint control using programmable sequential logic. *Science* **2018**, *361*.
- Averesch, N.J.H.; Kracke, F. Metabolic Network Analysis of Microbial Methane Utilization for Biomass Formation and Upgrading to Bio-Fuels. *Front. Energy Res.* **2018**, *6*.
- Bar-Even, A. Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways. *Biochemistry* **2016**, *55*, 3851–63.
- Bar-Even, A.; Noor, E.; Flamholz, A.; Milo, R. Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. *Biochim. Biophys. Acta* **2013**, *1827*, 1039–47.
- Bar-Even, A.; Noor, E.; Lewis, N.E.; Milo, R. Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8889–94.
- Bennett, R.K.; Gonzalez, J.E.; Whitaker, W.B.; Antoniewicz, M.R.; Papoutsakis, E.T. Expression of heterologous non-oxidative pentose phosphate pathway from *Bacillus methanolicus* and phosphoglucose isomerase deletion improves methanol assimilation and metabolite production by a synthetic *Escherichia coli* methylotroph. *Metab. Eng.* **2018**, *45*, 75–85.
- Bogorad, I.W.; Chen, C.T.; Theisen, M.K.; Wu, T.Y.; Schlenz, A.R.; Lam, A.T.; Liao, J.C. Building carbon-carbon bonds using a biocatalytic methanol condensation cycle. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15928–33.
- Bogorad, I.W.; Lin, T.S.; Liao, J.C. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature* **2013**, *502*, 693–7.
- Brautaset, T.; Jakobsen, O.M.; Josefsen, K.D.; Flickinger, M.C.; Ellingsen, T.E. *Bacillus methanolicus*: a candidate for industrial production of amino acids from methanol at 50°C. *Appl Microbiol Biotechnol* **2007**, *74*, 22–34.
- Burgard, A.P.; Pharkya, P.; Maranas, C.D. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **2003**, *84*, 647–57.
- Calvin, M.; Benson, A.A. The Path of Carbon in Photosynthesis. *Science* **1948**, *107*, 476–480.
- Chen, C.T.; Chen, F.Y.; Bogorad, I.W.; Wu, T.Y.; Zhang, R.; Lee, A.S.; Liao, J.C. Synthetic methanol auxotrophy of *Escherichia coli* for methanol-dependent growth and production. *Metab. Eng.* **2018**, *49*, 257–266.
- Chen, N.H.; Djoko, K.Y.; Veyrier, F.J.; McEwan, A.G. Formaldehyde Stress Responses in Bacterial Pathogens. *Front. Microbiol.* **2016**, *7*, 257.
- Chou, A.; Clomburg, J.M.; Qian, S.; Gonzalez, R. 2-Hydroxyacyl-CoA lyase catalyzes acyloin condensation for one-carbon bioconversion. *Nat. Chem. Biol.* **2019**, *15*, 900–906.
- Claassens, N.J.; Cotton, C.A.R.; Kopljar, D.; Bar-Even, A. Making quantitative sense of electromicrobial production. *Nat. Catal.* **2019a**, *2*, 437–447.
- Claassens, N.J.; He, H.; Bar-Even, A. Synthetic Methanol and Formate Assimilation Via Modular Engineering and Selection Strategies. *Curr. Issues Mol. Biol.* **2019b**, *33*, 237–248.
- Claassens, N.J.; Sanchez-Andrea, I.; Sousa, D.Z.; Bar-Even, A. Towards sustainable feedstocks: A guide to electron donors for microbial carbon fixation. *Curr. Opin. Biotechnol.* **2018**, *50*, 195–205.

- Clomburg, J.M.; Crumbley, A.M.; Gonzalez, R. Industrial biomanufacturing: The future of chemical production. *Science* **2017**, *355*.
- Comer, A.D.; Long, M.R.; Reed, J.L.; Pfleger, B.F. Flux Balance Analysis Indicates that Methane Is the Lowest Cost Feedstock for Microbial Cell Factories. *Metab. Eng. Commun.* **2017**, *5*, 26–33.
- Cotton, C.A.R.; Claassens, N.J.; Benito-Vaquerizo, S.; Bar-Even, A. Renewable methanol and formate as microbial feedstocks. *Curr. Opin. Biotechnol.* **2020**, *62*, 168–180.
- Dai, Z.; Gu, H.; Zhang, S.; Xin, F.; Zhang, W.; Dong, W.; Ma, J.; Jia, H.; Jiang, M. Metabolic construction strategies for direct methanol utilization in *Saccharomyces cerevisiae*. *Bioresour. Technol.* **2017**, *245*, 1407–1412.
- Desmons, S.; Fauré, R.; Bontemps, S. Formaldehyde as a Promising C₁ Source: The Instrumental Role of Biocatalysis for Stereocontrolled Reactions. *ACS Catal.* **2019**, *9*, 9575–9588.
- Devi, G.K.; Chozhavendhan, S.; Jayamuthunagai, J.; Bharathiraja, B.; Praveen kumar, R. Conversion of Biomass to Methanol and Ethanol. In *Horizons in Bioprocess Engineering*; Pogaku, R., Ed.; Springer International Publishing: Cham, **2019**; pp. 61–72.
- Dijkhuizen, L. Methanol, a potential feedstock for biotechnological processes. *Trends Biotechnol.* **1985**, *3*, 262–267.
- Dürre, P.; Eikmanns, B.J. C₁-carbon sources for chemical and fuel production by microbial gas fermentation. *Curr. Opin. Biotechnol.* **2015**, *35*, 63–72.
- Erb, T.J.; Zarzycki, J. A short history of RubisCO: the rise and fall (?) of Nature’s predominant CO₂ fixing enzyme. *Curr. Opin. Biotechnol.* **2018**, *49*, 100–107.
- Espinosa, M.I.; Valgepea, K.; Gonzalez-Garcia, R.A.; Scott, C.; Pretorius, I.S.; Marcellin, E.; Paulsen, I.T.; Williams, T.C. Native and synthetic methanol assimilation in *Saccharomyces cerevisiae*. *bioRxiv* **2019**, <https://doi.org/10.1101/717942>.
- Fan, L.; Wang, Y.; Tuyishime, P.; Gao, N.; Li, Q.; Zheng, P.; Sun, J.; Ma, Y. Engineering Artificial Fusion Proteins for Enhanced Methanol Bioconversion. *ChemBioChem* **2018**, *19*, 2465–2471.
- Fei, Q.; Guarneri, M.T.; Tao, L.; Laurens, L.M.L.; Dowe, N.; Pienkos, P.T. Bioconversion of natural gas to liquid fuel: Opportunities and challenges. *Biotechnol. Adv.* **2014**, *32*, 596–614.
- Fuchs, G. Alternative Pathways of Carbon Dioxide Fixation: Insights into the Early Evolution of Life? *Annu. Rev. Microbiol.* **2011**, *65*, 631–658.
- Ganesh, I.; Vidhya, S.; Eom, G.T.; Hong, S.H. Construction of Methanol-Sensing *Escherichia coli* by the Introduction of a *Paracoccus denitrificans* MxaY-Based Chimeric Two-Component System. *J. Microbiol. Biotechnol.* **2017**, *27*, 1106–1111.
- Gassler, T.; Sauer, M.; Gasser, B.; Egermeier, M.; Troyer, C.; Causon, T.; Hann, S.; Mattanovich, D.; Steiger, M.G. The industrial yeast *Pichia pastoris* is converted from a heterotroph into an autotroph capable of growth on CO₂. *Nat. Biotechnol.* **2020**, *38*, 210–216.
- Gleizer, S.; Ben-Nissan, R.; Bar-On, Y.M.; Antonovsky, N.; Noor, E.; Zohar, Y.; Jona, G.; Krieger, E.; Shamshoum, M.; Bar-Even, A.; et al. Conversion of *Escherichia coli* to Generate All Biomass Carbon from CO₂. *Cell* **2019**, *179*, 1255-1263.e12.
- Gonzalez, J.E.; Bennett, R.K.; Papoutsakis, E.T.; Antoniewicz, M.R. Methanol assimilation in *Escherichia coli* is improved by co-utilization of threonine and deletion of leucine-responsive regulatory protein. *Metab. Eng.* **2018**, *45*, 67–74.
- Hassan, S.S.; Williams, G.A.; Jaiswal, A.K. Lignocellulosic Biorefineries in Europe: Current State and Prospects. *Trends Biotechnol.* **2019**, *37*, 231–234.
- Haynes, C.A.; Gonzalez, R. Rethinking biological activation of methane and conversion to liquid fuels. *Nat. Chem. Biol.* **2014**, *10*, 331–9.

- He, H.; Edlich-Muth, C.; Lindner, S.N.; Bar-Even, A. Ribulose Monophosphate Shunt Provides Nearly All Biomass and Energy Required for Growth of *E. coli*. *ACS Synth. Biol.* **2018**, *7*, 1601–1611.
- Hwang, I.Y.; Nguyen, A.D.; Nguyen, T.T.; Nguyen, L.T.; Lee, O.K.; Lee, E.Y. Biological conversion of methane to chemicals and fuels: technical challenges and issues. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 3071–3080.
- IEA. Putting CO₂ to Use: Creating value from emissions. *IEA: Paris* **2019**.
- Kalyuzhnaya, M.G.; Puri, A.W.; Lidstrom, M.E. Metabolic engineering in methanotrophic bacteria. *Metab. Eng.* **2015**, *29*, 142–152.
- Kamps, J.J.A.G.; Hopkinson, R.J.; Schofield, C.J.; Claridge, T.D.W. How formaldehyde reacts with amino acids. *Commun. Chem.* **2019**, *2*, 126.
- Kim, H.J.; Huh, J.; Kwon, Y.W.; Park, D.; Yu, Y.; Jang, Y.E.; Lee, B.R.; Jo, E.; Lee, E.J.; Heo, Y.; et al. Biological conversion of methane to methanol through genetic reassembly of native catalytic domains. *Nat. Catal.* **2019**, *2*, 342–353.
- Kim, S.; Lindner, S.N.; Aslan, S.; Yishai, O.; Wenk, S.; Schann, K.; Bar-Even, A. Growth of *E. coli* on formate and methanol via the reductive glycine pathway. *Nat. Chem. Biol.* **2020**, *16*, 538–545.
- Latif, H.; Zeidan, A.A.; Nielsen, A.T.; Zengler, K. Trash to treasure: production of biofuels and commodity chemicals via syngas fermenting microorganisms. *Curr. Opin. Biotechnol.* **2014**, *27*, 79–87.
- Lee, O.K.; Hur, D.H.; Nguyen, D.T.N.; Lee, E.Y. Metabolic engineering of methanotrophs and its application to production of chemicals and biofuels from methane. *Biofuels Bioprod. Biorefining-Biofpr* **2016**, *10*, 848–863.
- 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.; Jang, Y.S. A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* **2019**, *2*, 18–33.
- Lessmeier, L.; Pfeifenschneider, J.; Carnicer, M.; Heux, S.; Portais, J.C.; Wendisch, V.F. Production of carbon-13-labeled cadaverine by engineered *Corynebacterium glutamicum* using carbon-13-labeled methanol as co-substrate. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 10163–76.
- Liao, J.C.; Mi, L.; Pontrelli, S.; Luo, S. Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nat. Rev. Microbiol.* **2016**, *14*, 288–304.
- Ljungdahl, L.; Wood, H.G. Incorporation of C¹⁴ From Carbon Dioxide into Sugar Phosphates, Carboxylic Acids, and Amino Acids by *Clostridium thermoaceticum*. *J. Bacteriol.* **1965**, *89*, 1055–1064.
- Lu, X.; Liu, Y.; Yang, Y.; Wang, S.; Wang, Q.; Wang, X.; Yan, Z.; Cheng, J.; Liu, C.; Yang, X.; et al. Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design. *Nat. Commun.* **2019**, *10*, 1378.
- Marmiesse, L.; Peyraud, R.; Cottret, L. FlexFlux: combining metabolic flux and regulatory network analyses. *BMC Syst. Biol.* **2015**, *9*, 93.
- Marx, C.J.; Chistoserdova, L.; Lidstrom, M.E. Formaldehyde-Detoxifying Role of the Tetrahydromethanopterin-Linked Pathway in *Methylobacterium extorquens* AM1. *J. Bacteriol.* **2003**, *185*, 7160–7168.
- Meyer, F.; Keller, P.; Hartl, J.; Groninger, O.G.; Kiefer, P.; Vorholt, J.A. Methanol-essential growth of *Escherichia coli*. *Nat. Commun.* **2018**, *9*, 1508.
- Müller, J.E.N.; Meyer, F.; Litsanov, B.; Kiefer, P.; Potthoff, E.; Heux, S.; Quax, W.J.; Wendisch, V.F.; Brautaset, T.; Portais, J.C.; et al. Engineering *Escherichia coli* for methanol conversion. *Metab. Eng.* **2015**, *28*, 190–201.
- Müller, V. New Horizons in Acetogenic Conversion of One-Carbon Substrates and Biological Hydrogen Storage. *Trends Biotechnol.* **2019**, *37*, 1344–1354.

- Nash, T. The colorimetric estimation of formaldehyde by means of the Hantzsch reaction. *Biochem. J.* **1953**, *55*, 416–21.
- Nielsen, J.; Keasling, J.D. Engineering Cellular Metabolism. *Cell* **2016**, *164*, 1185–1197.
- Nikolaidis, P.; Poullikkas, A. A comparative overview of hydrogen production processes. *Renew. Sustain. Energy Rev.* **2017**, *67*, 597–611.
- Noor, E.; Bar-Even, A.; Flamholz, A.; Reznik, E.; Liebermeister, W.; Milo, R. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput. Biol.* **2014**, *10*, e1003483.
- Ochsner, A.M.; Sonntag, F.; Buchhaupt, M.; Schrader, J.; Vorholt, J.A. *Methylobacterium extorquens*: methylotrophy and biotechnological applications. *Appl Microbiol Biotechnol* **2015**, *99*, 517–34.
- Olah, G.A. Towards oil independence through renewable methanol chemistry. *Angew. Chem. Int. Ed.* **2013**, *52*, 104–7.
- Olah, G.A. Beyond oil and gas: the methanol economy. *Angew. Chem. Int. Ed.* **2005**, *44*, 2636–9.
- Patterson, B.D.; Mo, F.; Borgschulte, A.; Hillestad, M.; Joos, F.; Kristiansen, T.; Sunde, S.; van Bokhoven, J.A. Renewable CO₂ recycling and synthetic fuel production in a marine environment. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 12212–12219.
- Portnoy, V.A.; Bezdán, D.; Zengler, K. Adaptive laboratory evolution—harnessing the power of biology for metabolic engineering. *Curr. Opin. Biotechnol.* **2011**, *22*, 590–594.
- Price, J.V.; Chen, L.; Whitaker, W.B.; Papoutsakis, E.; Chen, W. Scaffoldless engineered enzyme assembly for enhanced methanol utilization. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 12691–12696.
- Purnick, P.E.; Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 410–22.
- Ritala, A.; Hakkinen, S.T.; Toivari, M.; Wiebe, M.G. Single Cell Protein-State-of-the-Art, Industrial Landscape and Patents 2001-2016. *Front. Microbiol.* **2017**, *8*, 2009.
- Rohllhill, J.; Gerald Har, J.R.; Antoniewicz, M.R.; Papoutsakis, E.T. Improving synthetic methylotrophy via dynamic formaldehyde regulation of pentose phosphate pathway genes and redox perturbation. *Metab. Eng.* **2020**, *57*, 247–255.
- Rohllhill, J.; Sandoval, N.R.; Papoutsakis, E.T. Sort-Seq Approach to Engineering a Formaldehyde-Inducible Promoter for Dynamically Regulated *Escherichia coli* Growth on Methanol. *ACS Synth. Biol.* **2017**, *6*, 1584–1595.
- Ross, M.O.; MacMillan, F.; Wang, J.; Nisthal, A.; Lawton, T.J.; Olafson, B.D.; Mayo, S.L.; Rosenzweig, A.C.; Hoffman, B.M. Particulate methane monooxygenase contains only mononuclear copper centers. *Science* **2019**, *364*, 566–570.
- Roth, T.B.; Woolston, B.M.; Stephanopoulos, G.; Liu, D.R. Phage-Assisted Evolution of *Bacillus methanolicus* Methanol Dehydrogenase 2. *ACS Synth. Biol.* **2019**, *8*, 796–806.
- Sander, R. Compilation of Henry's law constants (version 4.0) for water as solvent. *Atmospheric Chem. Phys.* **2015**, *15*, 4399–4981.
- Sanderson, K. Lignocellulose: A chewy problem. *Nature* **2011**, *474*, S12–S14.
- Schiel-Bengelsdorf, B.; Dürre, P. Pathway engineering and synthetic biology using acetogens. *FEBS Lett.* **2012**, *586*, 2191–2198.
- Schrader, J.; Schilling, M.; Holtmann, D.; Sell, D.; Filho, M.V.; Marx, A.; Vorholt, J.A. Methanol-based industrial biotechnology: current status and future perspectives of methylotrophic bacteria. *Trends Biotechnol.* **2009**, *27*, 107–15.
- Schug, Z.T. Formaldehyde Detoxification Creates a New Wheel for the Folate-Driven One-Carbon “Bi”-cycle. *Biochemistry* **2018**, *57*, 889–890.
- Schwander, T.; Borzyskowski, L.S. von; Burgener, S.; Cortina, N.S.; Erb, T.J. A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* **2016**, *354*, 900–904.

- Shishodia, S.; Zhang, D.; El-Sagheer, A.H.; Brown, T.; Claridge, T.D.W.; Schofield, C.J.; Hopkinson, R.J. NMR analyses on *N*-hydroxymethylated nucleobases – implications for formaldehyde toxicity and nucleic acid demethylases. *Org. Biomol. Chem.* **2018**, *16*, 4021–4032.
- 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.; et al. Computational protein design enables a novel one-carbon assimilation pathway. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 3704–9.
- Sushkevich, V.L.; Palagin, D.; Ranocchiari, M.; van Bokhoven, J.A. Selective anaerobic oxidation of methane enables direct synthesis of methanol. *Science* **2017**, *356*, 523–527.
- Takors, R.; Kopf, M.; Mampel, J.; Bluemke, W.; Blombach, B.; Eikmanns, B.; Bengelsdorf, F.R.; Weuster-Botz, D.; Dürre, P. Using gas mixtures of CO, CO₂ and H₂ as microbial substrates: the do's and don'ts of successful technology transfer from laboratory to production scale. *Microb. Biotechnol.* **2018**, *11*, 606–625.
- Tuyishime, P.; Wang, Y.; Fan, L.; Zhang, Q.; Li, Q.; Zheng, P.; Sun, J.; Ma, Y. Engineering *Corynebacterium glutamicum* for methanol-dependent growth and glutamate production. *Metab. Eng.* **2018**, *49*, 220–231.
- Vartiainen, E.; Blomberg, P.; Ilmén, M.; Andberg, M.; Toivari, M.; Penttilä, M. Evaluation of synthetic formaldehyde and methanol assimilation pathways in *Yarrowia lipolytica*. *Fungal Biol. Biotechnol.* **2019**, *6*, 27.
- Venkata Mohan, S.; Modestra, J.A.; Amulya, K.; Butti, S.K.; Velvizhi, G. A Circular Bioeconomy with Biobased Products from CO₂ Sequestration. *Trends Biotechnol.* **2016**, *34*, 506–519.
- Villadsen, J.; Nielsen, J.; Lidén, G. Bioreaction Engineering Principles. 3rd ed. *Springer US* **2011**. <https://doi.org/10.1007/978-1-4419-9688-6>.
- Wang, W.-H.; Himeda, Y.; Muckerman, J.T.; Manbeck, G.F.; Fujita, E. CO₂ Hydrogenation to Formate and Methanol as an Alternative to Photo- and Electrochemical CO₂ Reduction. *Chem. Rev.* **2015**, *115*, 12936–12973.
- Wang, X.; Wang, X.; Lu, X.; Ma, C.; Chen, K.; Ouyang, P. Methanol fermentation increases the production of NAD(P)H-dependent chemicals in synthetic methylotrophic *Escherichia coli*. *Biotechnol. Biofuels* **2019**, *12*, 17.
- Wang, X.L.; Wang, Y.; Liu, J.; Li, Q.G.; Zhang, Z.D.; Zheng, P.; Lu, F.P.; Sun, J.B. Biological conversion of methanol by evolved *Escherichia coli* carrying a linear methanol assimilation pathway. *Bioresour. Bioprocess.* **2017**, *4*, 41.
- Wang, Y.; Fan, L.; Tuyishime, P.; Zheng, P.; Sun, J. Synthetic Methylophony: A Practical Solution for Methanol-Based Biomanufacturing. *Trends Biotechnol.* **2020**, <https://doi.org/10.1016/j.tibtech.2019.12.013>.
- Wenk, S.; Yishai, O.; Lindner, S.N.; Bar-Even, A. An Engineering Approach for Rewiring Microbial Metabolism. In *Methods in Enzymology*; Methods in Enzymology; Academic Press, **2018**; Vol. 608, pp. 329–367.
- Wheeldon, I.; Minter, S.D.; Banta, S.; Barton, S.C.; Atanassov, P.; Sigman, M. Substrate channelling as an approach to cascade reactions. *Nat. Chem.* **2016**, *8*, 299–309.
- Whitaker, W.B.; Jones, J.A.; Bennett, R.K.; Gonzalez, J.E.; Vernacchio, V.R.; Collins, S.M.; Palmer, M.A.; Schmidt, S.; Antoniewicz, M.R.; Koffas, M.A.; et al. Engineering the biological conversion of methanol to specialty chemicals in *Escherichia coli*. *Metab. Eng.* **2017**, *39*, 49–59.
- Whitaker, W.B.; Sandoval, N.R.; Bennett, R.K.; Fast, A.G.; Papoutsakis, E.T. Synthetic methylotrophy: engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization. *Curr. Opin. Biotechnol.* **2015**, *33*, 165–75.
- Witthoff, S.; Schmitz, K.; Niedenfuhr, S.; Noh, K.; Noack, S.; Bott, M.; Marienhagen, J. Metabolic engineering of *Corynebacterium glutamicum* for methanol metabolism. *Appl. Environ. Microbiol.* **2015**, *81*, 2215–25.

- Woolston, B.M.; King, J.R.; Reiter, M.; Van Hove, B.; Stephanopoulos, G. Improving formaldehyde consumption drives methanol assimilation in engineered *E. coli*. *Nat. Commun.* **2018a**, *9*, 2387.
- Woolston, B.M.; Roth, T.; Kohale, I.; Liu, D.R.; Stephanopoulos, G. Development of a formaldehyde biosensor with application to synthetic methylotrophy. *Biotechnol. Bioeng.* **2018b**, *115*, 206–215.
- Wu, T.Y.; Chen, C.T.; Liu, J.T.; Bogorad, I.W.; Damoiseaux, R.; Liao, J.C. Characterization and evolution of an activator-independent methanol dehydrogenase from *Cupriavidus necator* N-1. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 4969–83.
- Yang, D.; Cho, J.S.; Choi, K.R.; Kim, H.U.; Lee, S.Y. Systems metabolic engineering as an enabling technology in accomplishing sustainable development goals. *Microb. Biotechnol.* **2017**, *10*, 1254–1258.
- Yang, H.; Kaczur, J.J.; Sajjad, S.D.; Masel, R.I. Electrochemical conversion of CO₂ to formic acid utilizing Sustainion™ membranes. *J. CO₂ Util.* **2017**, *20*, 208–217.
- Yang, J.G.; Sun, S.S.; Men, Y.; Zeng, Y.; Zhu, Y.M.; Sun, Y.X.; Ma, Y.H. Transformation of formaldehyde into functional sugars via multi-enzyme stepwise cascade catalysis. *Catal. Sci. Technol.* **2017**, *7*, 3459–3463.
- Yang, X.; Yuan, Q.; Luo, H.; Li, F.; Mao, Y.; Zhao, X.; Du, J.; Li, P.; Ju, X.; Zheng, Y.; et al. Systematic design and *in vitro* validation of novel one-carbon assimilation pathways. *Metab. Eng.* **2019**, *56*, 142–153.
- Yishai, O.; Goldbach, L.; Tenenboim, H.; Lindner, S.N.; Bar-Even, A. Engineered Assimilation of Exogenous and Endogenous Formate in *Escherichia coli*. *ACS Synth. Biol.* **2017**, *6*, 1722–1731.
- Yishai, O.; Lindner, S.N.; Gonzalez de la Cruz, J.; Tenenboim, H.; Bar-Even, A. The formate bio-economy. *Curr. Opin. Chem. Biol.* **2016**, *35*, 1–9.
- Yu, H.; Liao, J.C. A modified serine cycle in *Escherichia coli* converts methanol and CO₂ to two-carbon compounds. *Nat. Commun.* **2018**, *9*, 3992.
- Yurimoto, H.; Kato, N.; Sakai, Y. Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. *Chem Rec* **2005**, *5*, 367–75.
- Zakaria, Z.; Kamarudin, S.K. Direct conversion technologies of methane to methanol: An overview. *Renew. Sustain. Energy Rev.* **2016**, *65*, 250–261.
- Zelcbuch, L.; Antonovsky, N.; Bar-Even, A.; Levin-Karp, A.; Barenholz, U.; Dayagi, M.; Liebermeister, W.; Flamholz, A.; Noor, E.; Amram, S.; et al. Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.* **2013**, *41*, e98.
- Zeng, A. New bioproduction systems for chemicals and fuels: Needs and new development. *Biotechnol. Adv.* **2019**, *37*, 508–518.
- Zhou, X.; Liu, R.; Sun, K.; Chen, Y.; Verlage, E.; Francis, S.A.; Lewis, N.S.; Xiang, C. Solar-Driven Reduction of 1 atm of CO₂ to Formate at 10% Energy-Conversion Efficiency by Use of a TiO₂ -Protected III–V Tandem Photoanode in Conjunction with a Bipolar Membrane and a Pd/C Cathode. *ACS Energy Lett.* **2016**, *1*, 764–770.

Acknowledgements

This overseas PhD study was quite an adventure for me at the beginning. I experienced a completely different life in a different world during this period. My utmost gratitude is to my supervisor Arren Bar-Even, who let me join his research group and most importantly shared your deep knowledge and understanding of the metabolic world with me. I believe it was a unique and fruitful collaboration for me.

Also the life and study in this new world could not go so smoothly without the help from the Bar-Even lab, especially Steffen Lindner, Patrick Behrens, my fellows Sebastian Wenk, Armin Kubis and Liliana Calzadiaz Ramirez. Thanks to Charles Cotton who is always be already for language questions. Special gratitude is also to my students and interns: Pauline Kuschel, Sophia Borowski, Moritz Dodenhöft, Rune Höper and Emmanouil Avtziannis, thank you for your work and assistance. Big thanks to all of you.

The study was funded by China Scholarship Council, thanks for giving me this opportunity. At the same time, my sincere thanks are to my friends Liu Yaqin and She Xiqi, who co-signed the contract without any hesitation.

Finally, to my parents, thank you for your unconditional and endless support and sacrifice.