Engineering the Reductive Glycine Pathway : A Promising Synthetic Metabolism Approach for C1-Assimilation

One-Carbon Feedstocks for Sustainable Bioproduction

Claassens, Nico J.; Satanowski, Ari; Bysani, Viswanada R.; Dronsella, Beau; Orsi, Enrico et al

https://doi.org/10.1007/10_2021_181

Adv Biochem Eng Biotechnol (2022) 180: 299–350
https://doi.org/10.1007/10_2021_181

Published online: 2 April 2022

Engineering the Reductive Glycine Pathway: A Promising Synthetic Metabolism Approach for C1-Assimilation

Nico J. Claassens, Ari Satanowski, Viswanada R. Bysani, Beau Dronsella, Enrico Orsi, Vittorio Rainaldi, Suzan Yilmaz, Sebastian Wenk, and Steffen N. Lindner

Contents

N. J. Claassens (✉) and S. Yilmaz
Laboratory of Microbiology, Wageningen University and Research, Wageningen, The Netherlands
e-mail: nico.claassens@wur.nl

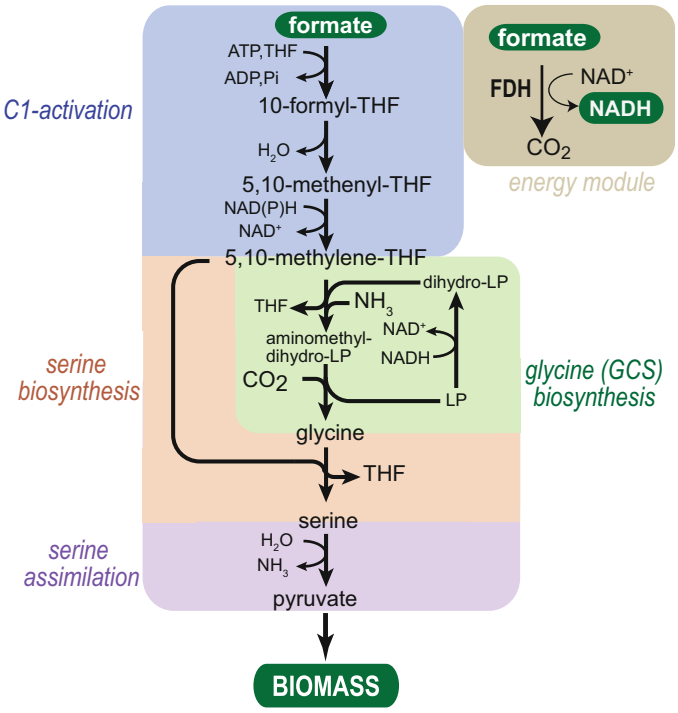
A. Satanowski, V. R. Bysani, B. Dronsella, E. Orsi, V. Rainaldi, and S. Wenk
Systems and Synthetic Metabolism Group, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

S. N. Lindner
Systems and Synthetic Metabolism Group, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

Department of Biochemistry, Charité Universitätsmedizin, Berlin, Germany

Abstract In recent years the reductive glycine pathway (rGlyP) has emerged as a promising pathway for the assimilation of formate and other sustainable C1-feedstocks for future biotechnology. It was originally proposed as an attractive "synthetic pathway" to support formatotrophic growth due to its high ATP efficiency, linear structure, and limited overlap with native pathways in most microbial hosts. Here, we present the current state of research on this pathway including breakthroughs on its engineering. Different variants of the rGlyP are discussed, including its core module for formate to glycine conversion, as well as varying modules for substrate conversion to formate, and glycine assimilation routes. Very recently, the rGlyP has been successfully implemented for synthetic formatotrophic growth, as well as for growth on methanol, in some bacterial hosts. We discuss the engineering strategies employed in these studies, including growth-coupled selection of functional pathway modules. We also compare the rGlyP to other natural and synthetic C1-assimilation pathways. Finally, we provide an outlook on open challenges and opportunities for the rGlyP, including its engineering into more biotechnological hosts, as well as the still-to-be realized production of value-added chemicals via this pathway. We expect that further research on the rGlyP will support the efficient use of sustainable C1-substrates in bioproduction.

Graphical Abstract



Keywords Formate, Glycine cleavage/synthase system, Metabolic engineering, Methanol, One-carbon substrates, Reductive glycine pathway, Synthetic biology

1 Introduction

The emissions of CO_2 and other greenhouse gases are causing global warming, which is set to have catastrophic consequences for nature and humanity [1], making this one of the largest and most urgent global challenges. Hence, it is imperative that fossil-based production processes, which are major contributors to CO_2 emissions, will be replaced by alternative chemical production processes based on atmospheric CO_2 [2–4]. While plant-based, photosynthetic conversion of CO_2 to biomass is an abundant process in nature, this process has a low energy conversion efficiency, making plant biomass unlikely to fully support a sustainable carbon economy [5–7].

A highly promising alternative is provided by emerging, efficient chemical technologies that allow for capture and reduction of CO_2 using electrical power generated from renewable energy [8]. Chemical technologies can utilize electricity to capture and reduce CO_2 (and H_2O) into small, reduced molecules, such as

hydrogen (H_2), carbon monoxide (CO), formate, and methanol at high energetic efficiencies and specificities [4, 9–11]. Subsequently, biological conversion of these compounds by microbial cell factories into desired products seems a highly attractive route for sustainable production, as biological processes can operate at mild conditions and with high product specificity for a broad range of products [12, 13]. Hence, combining the powers of chemical and biological catalysis, either in an integrated electro-microbial reactor system or in separated electrochemical systems and bioreactors [14–16], seems a strong approach for a future, circular carbon economy [10, 11].

Among the compounds that can be efficiently produced from CO_2 (e.g., by electrochemical reduction), formate is a highly suitable feedstock for biotechnological processes as it is soluble and non-volatile [17]. The reduced gasses hydrogen (H_2) and carbon monoxide (CO) can also be generated by direct electrochemical reduction, even at a higher electrochemical energy conversion efficiency than formate. However, the gaseous nature of H_2 and CO, as well as their notorious explosiveness and toxicity, respectively, provides some challenges for efficient and safe operation of the bioreactors in which microbial conversion should take place. Another interesting soluble one-carbon compound is methanol, which is more volatile than formate, but has a higher degree of reduction, which decreases the amount of substrate needed for production compared to formate. Methanol cannot be produced at very high efficiency by direct electrochemical reduction so far. Yet, it can be produced efficiently in a two-step process of electrochemical hydrogen and CO generation and subsequent thermocatalytic conversion of H_2/CO to methanol at a high overall energy conversion efficiency of ~50% [18].

In nature, a range of microorganisms is found that can utilize formate, methanol, as well as H_2 and CO. Many of these organisms can utilize these reduced molecules as their sole energy source, while they obtain their carbon solely from the reduced one-carbon compounds and/or CO_2. Microorganisms can use a number of metabolic pathways for the assimilation of these compounds. However, several of these natural pathways are energetically relatively inefficient in terms of ATP consumption, such as the Calvin-Benson-Bassham (CBB) cycle and the Serine cycle, which limits conversion yields for the biological catalysis [10, 19]. A highly ATP-efficient natural pathway is the reductive acetyl-CoA pathway (rAcP, also known as Wood-Ljungdahl pathway), however it can only operate under anaerobic conditions and has a limited product spectrum [20]. So, several natural pathways for one-carbon (C1) assimilation suffer from limitations for bioproduction. Another notorious challenge for the biological conversion of C1-feedstocks is that many natural organisms growing on these substrates are difficult to cultivate and/or difficult to engineer genetically, limiting their potential in biotechnological applications beyond their natural products.

Hence, there has been an increasing interest in the last decade to design and engineer efficient "synthetic" C1-assimilation pathways into well-amenable model microbial hosts, such as *Saccharomyces cerevisiae, Escherichia coli,* or *Corynebacterium glutamicum* [17, 21, 22]. A very promising, "synthetic" pathway for C1-assimilation is the reductive glycine pathway (rGlyP), which was proposed

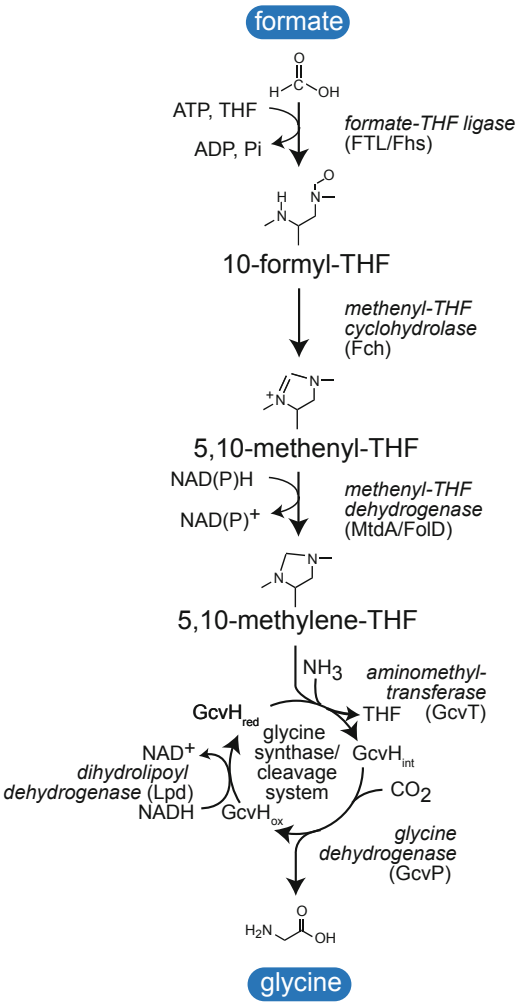


Fig. 1 Core module of the reductive glycine pathway. This module converts formate to glycine and is present in each variant of the rGlyP. The module consists of 2 submodules, the first C1-modules activates formate with tetrahydrofolate (THF) and subsequently reduces it to 5,10-methylene-THF. The second module (GCS) consists of the reversible glycine synthase/cleavage system and converts methylene-THF with CO_2, ammonia, and reducing power (NADH) to glycine

initially as a synthetic formate-assimilation pathway by Arren Bar-Even et al. in 2013 [23]. The rGlyP is a short and linear pathway with relatively little overlap with central carbon metabolism, while it is based on ubiquitous, oxygen-tolerant natural enzymes and has a high ATP efficiency. Hence, it was proposed as the most favorable route for engineering synthetic formatotrophy in model microbes [23] (Fig. 1). Already some decades ago, a variant of the rGlyP was observed in natural purine-degraders, which utilize this pathway to reduce CO_2, allowing redox-balancing during fermentative growth on highly reduced purine substrates [24–26]. However, until very recently it was never observed that this pathway could mediate growth on C1-substrates. A series of metabolic engineering efforts in recent years cumulated in the realization of formatotrophic growth via this pathway, first in *E. coli* and shortly after in *Cupriavidus necator* [27–29]. The potential of the rGlyP

for efficient C1-assimilation is not only limited to growth on formate, as variants of the rGlyP can also support growth on methanol or CO_2/H_2. The implementation of the full rGlyP also led to the first demonstration of synthetic methylotrophic growth using methanol as the sole source of carbon and energy in *E. coli* [27]. Shortly after the demonstration of fully functional, engineered rGlyP versions, another variant of the rGlyP was reported to operate as a natural route supporting autotrophic growth on CO_2 and H_2 [30].

In this chapter we first provide an overview of the core metabolic structure of the rGlyP and several of its variants. We will discuss the modular metabolic engineering approaches that were recently highly instrumental for the successful realization of engineered rGlyP versions in some organisms, and will likely support the realization in industrially more relevant hosts in the coming years. Finally, we compare the potential of the rGlyP with other natural and synthetic alternatives for C1-assimilation and provide an outlook on potential biotechnological applications of this promising and versatile C1-assimilation pathway.

2 The Modules and Variants of the Reductive Glycine Pathway

We define the rGlyP not as a single pathway, but rather as a family of pathway variants that all share a common core, which performs the conversion of formate and CO_2 into glycine [31]. First, we will discuss this core module, after which we will discuss several variant modules via which glycine can be further converted into biomass and products. Next, we will consider some alternative substrate utilization modules for the rGlyP, which can convert other one-carbon compounds, i.e. methanol and CO_2, to its core substrate formate.

2.1 The Core Module for Formate to Glycine Conversion

In the core of the rGlyP, formate, together with CO_2, is converted to glycine in a sequence of biochemical reactions (Fig. 1). First, formate is activated by an ATP-consuming ligase reaction with the co-factor tetrahydrofolate (THF). This reaction is carried out by the formate-THF-ligase (FTL), which first performs a kinase reaction that makes formyl-phosphate, after which it formylates THF with formyl-phosphate to result in 10-formyl-THF [32]. This enzyme is the only formate-assimilating enzyme found to act in formatotrophic pathways in nature [33], and it is also a part of the Serine cycle and the rAcP, which can both directly assimilate formate. Whereas all natural microorganisms operating these two pathways encode FTL, many organisms relevant to biotechnology, including *E. coli*, lack this enzyme activity, thus requiring heterologous expression for the engineering of the rGlyP.

As the next step, 10-formyl-THF is converted to 5,10-methenyl-THF, by a cyclohydrolase reaction, after which 5,10-methenyl-THF is reduced to 5,10-methylene-THF (further referred to as methylene-THF) in a dehydrogenase step. This dehydrogenase reaction generally uses NADPH, but NADH can also be used by some methylene-THF dehydrogenases. However, NADH generally provides less thermodynamic driving force in the reducing direction, favoring the use of NADPH for this step. Both the cyclohydrolase and dehydrogenase activities are present in almost all forms of life. Both activities are often catalyzed by a bifunctional enzyme or by a trifunctional enzyme that also includes FTL activity, the latter is often found in eukaryotes [34].

In many organisms the natural role of the bifunctional methylene-THF dehydrogenase/cyclohydrolase is to oxidize methylene-THF coming from glycine or serine conversion towards 10-formyl-THF, which is required for, e.g., purine biosynthesis. The rGlyP, however, requires flux through these reactions in the reductive direction. Native dehydrogenase/cyclohydrolase enzymes in a given host may thus not always be optimal for this, as for example the native bifunctional FolD in *E. coli* is inhibited by 10-formyl-THF [35]. Methylene THF-dehydrogenases and cyclohydrolases that operate in the reductive direction are, for instance, found in prokaryotes operating the Serine cycle or rAcP. Hence, enzymes from methylotrophs utilizing the Serine cycle or acetogens harboring the rAcP were selected as heterologous candidates for engineering the rGlyP. For example, recent works demonstrating successful operation of the rGlyP core utilized monofunctional cyclohydrolase (Fch) and dehydrogenase (MtdA/FolD) originating from the methylotroph *Methylobacterium extorquens* [27, 28] or the acetogen *Clostridium ljungdahlii* [36].

In the next step of the rGlyP core pathway, the C1-molecule methylene-THF is converted to the two-carbon amino acid glycine by the addition of CO_2. This reaction is mediated by the glycine cleavage/synthase system (GCS), which is a multi-enzyme system that catalyzes the reversible conversion of methylene-THF with CO_2, NH_3, and NADH to form glycine (Fig. 1) [37]. The GCS includes four different proteins, performing different activities: (1) lipoyl bearing-carrier protein (H-protein/GcvH), a central shuttle protein that covalently binds the co-factor lipoic acid; (2) dihydrolipoyl dehydrogenase (L-protein/Lpd), which reduces the lipoyl moiety attached to the H-protein using NADH; (3) aminomethyltransferase (T-protein/GcvT), which condenses ammonia and methylene-THF onto the reduced lipoyl moiety, upon release of THF; and lastly (4) glycine dehydrogenase (P-protein/GcvP), which incorporates CO_2 and releases glycine using a pyridoxal 5-phosphate (PLP) co-factor. Apart from releasing glycine, the P-protein also regenerates the oxidized lipoyl moiety of the H-protein that can re-enter the GCS cycle for a subsequent reduction by the L-protein.

The GCS is present in many organisms throughout the tree of life and is most frequently used for the cleavage of glycine, to generate methylene-THF [38–40], which is a key C1-precursor for several biosynthetic processes and related to the folate and methionine cycles. Methylene-THF serves as a direct C1-precursor in the biosynthesis of Coenzyme A (CoA) and thymidine. Methylene-THF can be further oxidized to formyl-THF, which is used for the formylation of methionine-tRNA and

as precursor in the purine biosynthesis pathway, which provides adenine and guanine nucleotides, as well as the amino acid histidine. Methylene-THF can also be reduced to methyl-THF, which is a precursor for methionine biosynthesis. Methyl-THF serves also as a substrate in the methionine cycle, to regenerate the co-factor S-adenosylmethionine (SAM), which is used for the methylation of, e.g. DNA, RNA, and proteins and hence plays key regulatory roles.

Apart from its function to supply methylene-THF, the GCS reaction can also proceed in the reverse, reductive direction ("synthase direction"). The thermodynamics of the overall GCS reaction in the synthase direction ($\Delta rG'^{m} = -4.9$ kJ/mol at pH 7.5, ionic strength 0.25 and all reactants at 1 mM [41]), dictate that this reaction is reversible. The glycine synthase reaction is favored at higher CO_2 and ammonia concentrations, which leads to the generally observed dependence on elevated CO_2 (above ambient concentrations) to enable growth via the rGlyP. Operation of the GCS in the synthase direction is also observed in nature during rGlyP activity as an electron sink or as an autotrophic pathway [25, 30, 42]. The positive effect of increased ammonia concentrations on the GCS reaction in the synthase direction was observed for *Desulfovibrio desulfuricans,* for which higher ammonia concentrations increased the autotrophic growth rate via the rGlyP [30].

Until now, organisms in which the rGlyP was engineered mostly relied on the overexpression of native GCS enzymes. These were naturally performing the cleavage reaction, but proved to support the reverse activity within the rGlyP context. The T, H, and P proteins seem to require overexpression for high rGlyP activity, however, several studies demonstrated that overexpression of the L-protein does not seem necessary [27, 28, 43]. Here, it is worth to note that the L-protein serves as a dehydrogenase within other enzyme systems, i.e. the pyruvate dehydrogenase and alpha-ketoglutarate dehydrogenase [44]. As both of these systems often carry high flux in central metabolism, this may explain why native L-protein levels are already sufficient to support the GCS activity in the rGlyP. Relatedly, in many prokaryotes the gene encoding the L-protein resides in the same genomic locus as pyruvate dehydrogenase or alpha-ketoglutarate dehydrogenase, and less frequently with the T, H, and P proteins, where these latter three mostly co-localize in a single operon or locus.

It is also noteworthy that in bacteria in which a native rGlyP was identified, the P-protein activity is executed by two different subunits (GcvPa and GcvPb), which are both encoded within the GCS operon [30, 42, 45]. This may suggest that the dual subunit architecture of the P-protein may be beneficial for the glycine synthase direction. However, this has not been studied so far and the dual subunit P-protein is also found in many other organisms that likely run the glycine cleavage reaction, such as *Bacillus subtilis*. As the GCS is a core catalytic complex of the rGlyP, the comparison of different variants for heterologous expression and their kinetic characterization to identify fast candidates deserves more attention. However, it is complicated to characterize the kinetics of the GCS reactions in vitro and hence only little reliable kinetic data are available so far [46].

2.2 Glycine Conversion Modules via Serine

The glycine produced from formate and CO_2 by the core module of the rGlyP can be further assimilated into central metabolism via several glycine assimilation modules (Fig. 2). In these routes, glycine is converted to either serine, glyoxylate or acetyl-phosphate and a range of downstream pathways can assimilate these compounds further. These diverse glycine assimilation modules also have different ATP and reducing equivalent requirements, to generate, for example, the key metabolic intermediate pyruvate (Table 1). Here, we will discuss potential modules, of which some have been observed in nature or have been realized by metabolic engineering, while some are so far only theoretical designs [47]. We note that there are more potential pathways for glycine assimilation than discussed here, however, these rely on the co-assimilation of glycine or its derivatives with other metabolites (e.g., glycine with succinyl-CoA via 5-aminolevulinic acid synthetase or glyoxylate with acetyl-CoA via malate synthase [48, 49]). The requirement for these co-substrates means that these pathways cannot support full growth via the rGlyP.

In the initial design of the rGlyP, it was proposed that glycine is converted to the C3 amino acid serine via a hydroxymethyltransferase [23]. This enzyme condenses glycine with a methylene-THF (generated from another formate) to yield serine. Serine/glycine hydroxymethyltransferase (GlyA/SHMT) catalyzes a reversible reaction, which is found in many organisms. It is often involved in the biosynthesis of glycine and the generation of methylene-THF, but it can also be used in reverse to synthesize serine from glycine and methylene-THF ($\Delta rG'^{m} = 6.7$ kJ/mol at pH 7.5, ionic strength 0.25 and all reactants at 1 mM [41]).

To further assimilate serine into central metabolism it can be deaminated to pyruvate, releasing the ammonia that was initially fixed by the GCS. The conversion of serine to pyruvate is carried out by a PLP-co-factor dependent serine deaminase (SdaA), also known as serine dehydratase. In this reaction, first a hydroxyl group is eliminated by the enzyme from serine, leading to the intermediate aminoacrylate, which is deaminated to pyruvate via non-enzymatic hydrolysis [50].

The combination of SHMT and serine deaminase leads to the shortest glycine assimilation route, with a relatively low energy consumption. This module is so far the default variant for the engineering of the rGlyP, e.g. in *E. coli* and *C. necator*. Like SHMT, serine deaminase is a ubiquitous enzyme and for engineering of the rGlyP usually native variants of both enzymes are overexpressed [27–29].

Another possible variant of glycine conversion via serine proceeds via serine aminotransferase to hydroxypyruvate by the activity of PLP-dependent serine: glyoxylate aminotransferase (SGAT) or serine:pyruvate aminotransferase (SPAT), which donate the amino group of serine to the ketoacids glyoxylate or pyruvate, respectively generating glycine or alanine [51, 52]. Hydroxypyruvate can be further reduced to glycerate, which after phosphorylation can enter central metabolism in glycolysis in the form of 2- or 3-phospho-glycerate (further generally referred to as P-glycerate). The overall ATP and NAD(P)H requirements of this serine amino-transferase route seem very similar to the serine deaminase variant (Table 1).

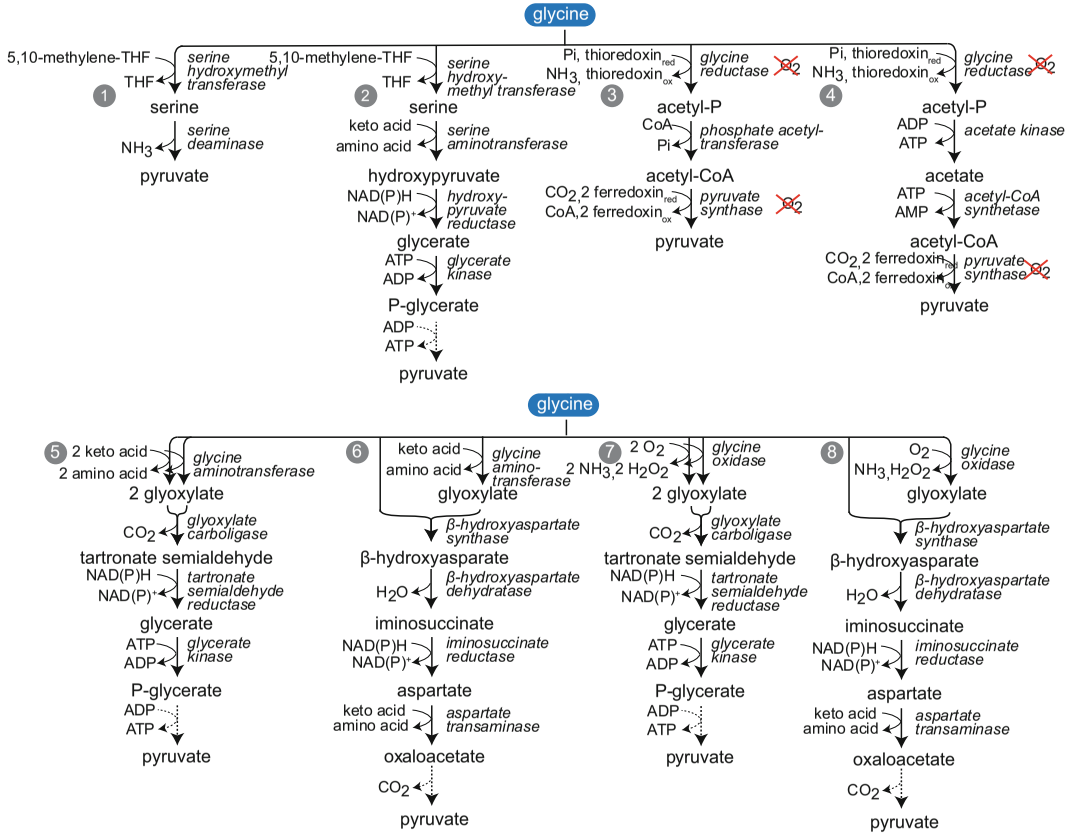


Fig. 2 Overview of different assimilation modules for glycine. The eight possible modules have different stoichiometries and redox and ATP efficiencies, which are compared in Table 1. Oxygen-sensitive enzymes are indicated with crossed O_2 symbols

Table 1 Overview for different variants of the rGlyP and their theoretical resource cost for the biosynthesis of pyruvate

Glycine converting reaction	Serine hydroxymethyltransferase		Glycine reductase		Glycine aminotransferase		Glycine oxidase	
Assimilation route to central metabolism	Serine deaminase	Serine aminotransferase	Phosphate acetyltransferase/ pyruvate synthase	Acetyl-CoA synthetase/ pyruvate synthase	Glycerate	Hydroxyaspartate	Glycerate	Hydroxyaspartate
ATP per pyruvate	2	2	1	2	2[a]	2[a]	2[b]	2[b]
Red. equivalents per pyruvate	3	3[c]	4	4	3[c]	3[c]	5	4[c]
Formate assimil. per pyruvate	2	2	1	1	2	2	2	2
CO_2 per pyruvate	1	1	1	2	1	1	1	1
Total formate per pyruvate	6	6	6	6	6	6	8	≤ 7
Natural, theoretical, or engineered (references)	Eng. [27–29]	Theor.	Theor.	Nat. [30]	Theor.	Theor.	Eng. [28]	Theor.
Variant # in Fig. 2	1	2	3	4	5	6	7	8

For the calculation of total formate consumption it is assumed that the oxidation of one formate to one reducing equivalent (e.g., NADH) can generate 2 ATP during respiration (assuming a P/O ratio of 2 [55, 56]). Numbers of reducing equivalents consumed reflect electron pairs (which can be either 1 NADH, 1 NADPH or 2 reduced ferredoxins/thioredoxins). Extra ATP investment needed to potentially generate NADPH from NADH is not included in the ATP and formate consumption

[a]In addition to [b], these variants may save ATP costs due to the assimilation of NH_3 in the pathway, which can save potential ATP expenses for NH_3 assimilation elsewhere, but this does not affect the ATP consumption for pyruvate biosynthesis

[b]These pathway variants can lead to overall lower ATP consumption when considering biomass or specific compounds other than pyruvate as the product of interest (e.g., metabolites derived from phosphoglycerate or oxaloacetate, respectively)

[c]These pathways consume extra reducing equivalents on top of the numbers reported per pyruvate, but these are retained in the generation of amino acids and can be utilized or released elsewhere in metabolism and hence are not included here

However, for the serine aminotransferase route it is relevant to consider that the amino group transferred to glycine or alanine may be further transferred via other aminotransferases to glutamate and other amino acids. In this way the amino group assimilated by the GSC can be retained in the cellular amino acid pool. This could possibly save some ATP, as cells may otherwise have to fix ammonia via the ATP-dependent glutamine synthetase. However, this will not save energy in many cases, since ammonia is generally fixed via the ATP-independent glutamate dehydrogenase in presence of higher ammonia concentrations, which are likely needed to drive the GCS in the rGlyP [53].

Alternatively, some ATP could be saved via the serine aminotransferase route in biosynthesis, as this route enters central metabolism via P-glycerate instead of pyruvate, which saves the additional ATP required for the gluconeogenic conversion of pyruvate to P-glycerate and higher glycolytic metabolites such as sugars. So overall, this serine aminotransferase route is potentially slightly more energy-efficient than the canonical serine deaminase route [19]. However, it requires more enzymatic steps, which may provide some additional complications for its operation and engineering, for example due to the potential toxicity of hydroxypyruvate [54] or the promiscuous activity of hydroxypyruvate reductase with glyoxyate [52].

2.3 Glycine Conversion Modules via Glyoxylate

As an alternative to the conversion via serine, glycine can otherwise be deaminated to glyoxylate, either via glycine transamination or glycine oxidation (Fig. 2). Glycine can be transaminated to glyoxylate, e.g. via glutamate:glyoxylate aminotransferase, alanine:glyoxylate aminotransferase, or aspartate:glyoxylate aminotransferase [57]. Alternatively, glycine can be converted to glyoxylate by glycine oxidation with oxygen, leading to the formation of ammonia and hydrogen peroxide [58]. Unlike the transamination, the oxidation of glycine "wastes" electrons by transferring them directly to oxygen, which reduces the potential yields of this pathway variant (Table 1). During the initial engineering of the rGlyP in *C. necator*, a route via native glycine oxidation emerged, which led to low biomass yields. By overexpressing the native SHMT and serine deaminase the yield and growth rate of the engineered strain was improved substantially, proving the superiority of the SHMT route over the glycine oxidase route [28]. However, the glycine aminotransferase route could potentially rival the serine routes, as will be explained below.

When glycine is converted to glyoxylate by an oxidase or aminotransferase, a subsequent glyoxylate assimilation route towards central metabolism is required. A canonical route in bacteria for glyoxylate assimilation is the glycerate pathway, which first condenses two glyoxylate molecules to tartronate semialdehyde with the release of one CO_2. Tartronate semialdehyde can be further reduced to glycerate and phosphorylated to P-glycerate to enter central metabolism. In the first *C. necator* strains engineered with the rGlyP, in which glycine oxidase was active, the generated glyoxylate was shown to be assimilated via this glycerate route [28].

Another interesting alternative for the assimilation of glyoxylate is the β-hydroxyaspartate cycle [47, 59, 60]. This route can co-assimilate one glycine molecule with one glyoxylate molecule by condensing them via β-hydroxyasparate synthase. Then, β-hydroxyaspartate can be further hydrolyzed to iminosuccinate and reduced to aspartate. The latter amino acid can be transaminated to oxaloacetate to enter central metabolism. This route provides a particular advantage if oxaloacetate or other related C4 precursors are needed for the production of a compound of interest.

When the glycine aminotransferase would be combined with either the glycerate route or the β-hydroxyaspartate cycle for assimilation, this could lead to rGlyP operation with a potentially higher theoretical efficiency than the canonical serine deaminase route, due to possible energy savings from the ammonia assimilation, as discussed before for the serine aminotransferase route. More importantly, the generation of either P-glycerate or oxaloacetate, rather than pyruvate, could save energy in the synthesis of biomass or product synthesis starting from these compounds. If pyruvate is generated via the canonical serine deaminase route, the generation of P-glycerate or oxaloacetate from pyruvate requires extra ATP equivalents for gluconeogenesis or anaplerosis. However, so far no study has demonstrated the operation of the rGlyP in concert with glycine aminotransferase.

2.4 Glycine Conversion Modules via Acetyl-Phosphate

Another option for the conversion of glycine is provided by the oxygen-sensitive enzyme complex glycine reductase. This multi-subunit enzyme can reduce glycine to acetyl-phosphate by using thioredoxins as redox co-factors. Next, acetyl-phosphate can be converted by phosphate acetyltransferase to acetyl-CoA, which can be reductively carboxylated to pyruvate by the oxygen-sensitive pyruvate synthase (Fig. 2). This route is highly energy-efficient and can in theory generate pyruvate from the assimilation of only one formate molecule, unlike two for all the other rGlyP variants, since one additional CO_2 is assimilated by pyruvate synthase. This reduces the ATP-requirement for pyruvate synthesis to 1, as only one ATP needs to be spent to ligate formate to THF (Table 1).

Glycine reductase is likely carrying most flux coming from the rGlyP core in some natural anaerobes that operate the rGlyP pathway [24–26, 30, 42]. However, proteomics analysis on the autotrophic growth of *Desulfovibrio desulfuricans,* which solely depends on the rGlyP, surprisingly suggested that it probably converts acetyl-P to acetyl-CoA via acetate kinase (generating 1 ATP) and acetyl-CoA synthetase (consuming 2 ATP), rather than directly via the phosphate acetyltransferase. This alternative assimilation route for acetyl-P consumes an additional ATP, slightly decreasing its theoretical yield on formate, making its ATP consumption equal to the canonical, oxygen-tolerant serine deaminase variant (Table 1). The more ATP-efficient variant via glycine reductase and phosphate acetyltransferase is the most efficient known variant of the rGlyP, but its engineering may be challenging for

many hosts as it would operate under low thermodynamic driving force and involves two oxygen-sensitive enzyme complexes: glycine reductase and pyruvate synthase.

2.5 Modules for Supply of Reducing Power

As the rGlyP assimilates the highly oxidized substrates formate and CO_2, it generally requires an input of reducing power (in the form of NAD(P)H), as well as energy (in the form of ATP), to generate more reduced molecules for biomass and product biosynthesis.

Reducing power can be supplied from different sources: (1) via oxidation of formate by formate dehydrogenase; (2) via feedstock oxidation *en route* to formate when a more reduced C1-compound, like methanol, serves as the carbon source (see below); (3) by oxidation of an additional inorganic electron donor, such as H_2, CO, or phosphite [30, 61, 62]; (4) by oxidation of a (highly) reduced organic electron donor, such as purine-substrates or glycerol; or (5) by light-driven splitting of water such as in oxygenic photosynthesis.

When an organism grows aerobically via the rGlyP on formate and CO_2 as sole carbon sources, formate can also serve as the sole energy source. In that case a major part of the formate needs to be oxidized. For example, in the biosynthesis of pyruvate via the canonical serine deaminase route, theoretically 1/3 of the formate is assimilated while 2/3 of the formate needs to be oxidized (Table 1). This theoretical calculation still excludes additional ATP consumption for "cellular maintenance," which would further increase the proportion of formate that needs to be oxidized. Formate can be oxidized to CO_2 to generate reduced redox co-factors, typically NADH, by formate dehydrogenase (FDH). FDHs are divided into two main categories: the first consists of small, single subunit, metal-independent enzymes; the second category is characterized by several subunits and a complex catalytic core containing rare metals such as molybdenum or tungsten [63]. The simpler metal-independent FDHs generally have lower catalytic rates ($<10\ s^{-1}$), which may impose a significant expression burden if all energy in the cell needs to be regenerated through this enzyme activity [23, 64]. Their metal-dependent counterparts can catalyze formate oxidation at much higher rates ($>100\ s^{-1}$) [63]. However, their multi-subunit nature and the correct incorporation of the metal co-factor make their heterologous expression more complex.

Recently published demonstrations of full formatotrophic growth via the rGlyP in *E. coli* relied on heterologous expression of simple metal-independent FDHs from *Pseudomonas* sp. 101, *Arabidopsis thaliana,* or *Candida boidinii* [27, 29]. Here, it was indeed observed that mutations increasing their expression levels boost growth, showing that this enzyme is a bottleneck [27]. Unlike *E. coli*, which has no native NAD^+-dependent formate dehydrogenase, *C. necator* harbors a multi-subunit, molybdenum- and NAD^+-dependent FDH, which naturally enables formatotrophic growth via the CBB cycle. Hence, the engineering of the rGlyP in *C. necator* did not require heterologous FDH expression [28].

Canonically, in addition to NADH (for reduction in the GCS), the rGlyP requires NADPH for the reduction of 5,10-methenyl-THF, which is the preferred co-factor for most dehydrogenase enzymes performing this reduction. As most FDHs generate NADH, a substantial part of the reducing power needs to be transferred from NADH to NADPH. This transfer leads to some energy consumption, as the NADPH/$NADP^+$ redox pair is normally maintained at a more reduced state in cells than the NADH/NAD^+ pair. This electron transfer is usually performed by membrane-bound transhydrogenases, which consume energy in the form of a proton-gradient, coming at the expense of ATP generation [65]. Thus, a more energy-efficient solution consists of directly generating NADPH during oxidation of the feedstock/electron donor, by utilizing $NADP^+$-dependent enzymes. The thermodynamic driving force for the FDH reaction, especially with higher formate concentrations, is sufficient to reduce $NADP^+$ instead of NAD^+. Hence, future implementations of the rGlyP could possibly realize higher yields by implementing natural or engineered $NADP^+$-dependent FDH variants. This was already attempted by Bang et al. by including an $NADP^+$-dependent FDH from *A. thaliana*, but this work still led to low yields on formate [29]. Notably, some recent studies demonstrated successful protein engineering of metal-independent, bacterial FDH variants with high affinity for $NADP^+$ and formate, reaching higher catalytic efficiency than natural or previously engineered variants [66, 67].

In case a variant of the rGlyP is employed in which glycine is converted via glycine reductase and pyruvate synthase, thioredoxin and ferredoxin regeneration is also required. Ferredoxins have a more negative redox potential, requiring dedicated ferredoxin-dependent FDHs or electron coupling/bifurcation systems that allow the reduction of this redox co-factor.

When instead of formate, CO_2 is used as a sole carbon source for the rGlyP, an inorganic electron donor is needed. Among the inorganic electron donors, H_2 is of particular interest as this feedstock can be highly efficiently generated with electrical energy via water electrolysis. For the conversion of H_2 to cellular reducing power, fast oxygen-tolerant hydrogenases are available that catalyze oxidation of H_2 with $NAD(P)^+$ as a direct electron acceptor [68]. Such enzymes are naturally present in certain bacteria like *C. necator*, however, they are composed of several subunits, have complex catalytic cores, and require auxiliary maturation factors for their correct assembly. As such, heterologous expression in other hosts has proven difficult so far. In nature the supply of redox power from H_2 to the rGlyP was already demonstrated during autotrophic growth on H_2 in *D. desulfuricans* [30].

An interesting option for the use of H_2 as energy source would be the mixotrophic use of H_2 as electron donor and formate as carbon source for the rGlyP. This would prevent the oxidation of the major part of the formate and avoid the resulting net release of CO_2.

The reducing power for the rGlyP could also be supplied by an organic carbon source as is the case for certain natural purine-degrading organisms. This concept could also be used to boost the anaerobic production of oxidized compounds from highly reduced substrates, such as glycerol, as recently proposed [69] and also discussed in detail in another chapter in this book chapter Hong et al.

2.6 Modules for Supply of ATP in Aerobic and Anaerobic Conditions

ATP regeneration can be achieved by various means depending on the growth conditions. In aerobic conditions, molecular oxygen serves as a terminal electron acceptor and allows for abundant production of ATP via respiration. When using aerobic respiration, the cell is thus able to satisfy demands of ATP for assimilation and/or bioproduction as long as reducing power is available in sufficient amounts. It must be noted that too much respiratory flux may even lead to excess ATP generation, decreasing theoretical yields of the highly ATP-efficient rGlyP.

In contrast, under anaerobic conditions, the ATP supply is generally constrained, when either a finite amount of a terminal electron acceptor is available (i.e., anaerobic respiration with, e.g., nitrate or sulfate) or no external electron acceptor besides CO_2 is available to the cell at all (i.e., fermentation). Operation of the rGlyP under anaerobic conditions, or any ATP-consuming assimilatory pathway for that matter, thus involves additional challenges for cellular ATP regeneration. However, the rGlyP can, given its high ATP efficiency, still be feasible in anaerobic conditions. This is exemplified by the recently demonstrated use of the rGlyP for autotrophic growth in *D. desulfuricans* under anaerobic conditions. In this organism, sulfate is used as a terminal electron acceptor [30].

In contrast to its operation with electron acceptors like O_2 and sulfate, the rGlyP has, to date, not been shown to allow full growth under fermentative conditions, i.e. when external electron acceptors are entirely absent and only CO_2 is available. In these conditions, a C1-assimilation pathway is only feasible if it can generate ATP rather than consuming it. This mode of autotrophic growth is known in nature only for acetogenic and methanogenic organisms operating the rAcP. The rAcP and an associated set of energy-conservation mechanisms allow these organisms to generate ATP in the process of converting CO_2 into acetate or other short acids and alcohols. To generate small, yet sufficient, amounts of ATP from this reduction of CO_2, acetogens use the so-called electron-bifurcation mechanisms, which lead to the build-up of proton or sodium gradients over the membrane to drive the synthesis of ATP via ATP synthase (chemiosmosis) [70, 71]. Since the glycine-reductase/phosphate acetyltransferase variant of the rGlyP, like the rAcP, only requires one ATP for the formation of one acetyl-CoA from formate and CO_2 (given the appropriate redox investment), this variant may also allow for growth without an electron acceptor. It is intriguing to explore whether natural or engineered glycine-reductase variants of the rGlyP could allow growth without an electron acceptor, by taking advantage of some electron-bifurcating energy-conservation mechanisms as found for the rAcP. Such growth could be interesting for anaerobic biotechnological applications, as an alternative or complementary pathway to the already well-demonstrated use of the rAcP in acetogenic cell factories [72]. In fact, genes encoding all enzymes of the rGlyP have been found in certain acetogens and it has been claimed that the glycine-reductase/phosphate acetyltransferase variant of the rGlyP may operate in parallel to the rAcP in the acetogen *Clostridium drakei*, thus at

least partially contributing to their net carbon fixation [42]. Also, transplantation of the rGlyP genes of *C. drakei* into the acetogen *Eubacterium limosum,* which natively lacks these genes, suggested that cooperation of both rAcP and rGlyP in one organism may improve the growth rate and acetate production rate of acetogens [42].

2.7 CO_2 Reduction as an Alternative Substrate Module

In the sections above, we have discussed the rGlyP primarily as a pathway for assimilation of formate or other reduced C1 compounds available to the organism. However, the formate utilized by the rGlyP can, in principle, also be generated from CO_2 by the microbe itself (Fig. 3). Such a pathway for biosynthesis of glycine from CO_2 alone has been proposed multiple times to play a potential role in natural or synthetic autotrophy [24–26, 62, 73–75]. The formate required for the rGlyP could be generated by reduction of CO_2 via an FDH, i.e. operating in the reverse of the direction discussed above (Fig. 3) [75].

Even though the oxidative direction of the FDH reaction is preferred under standard biological conditions with NAD^+ as the redox co-factor ($\Delta rG'^m = -14.4$ kJ/mol at pH 7.5, ionic strength 0.25 and all reactants at 1 mM [41]), the reaction is reversible. High concentrations of CO_2 as well as low concentrations of formate and a relatively high NADH/NAD^+ ratio can favor this direction. Especially the high NADH/NAD^+ ratio forms a practical hurdle, as typically this redox couple is dominantly present in its oxidized NAD^+-form. To circumvent this, the reductive direction of FDH can be driven by other redox co-factors that have a more negative reduction potential, e.g. ferredoxin (or NADPH due to its different physiological ratio). FDH enzymes dependent on these co-factors are available in nature or could be engineered [63, 76]. In nature, CO_2 reduction by FDH is found in autotrophs utilizing the rAcP. They employ, for example, ferredoxin-dependent FDHs, or hydrogen-dependent CO_2 reductase (HDCR), the latter being an FDH variant that can be driven directly by H_2 [77].

Apart from thermodynamics, the kinetics of the FDH reaction can also be limiting. Metal-dependent (using molybdenum or tungsten), multi-subunit FDHs

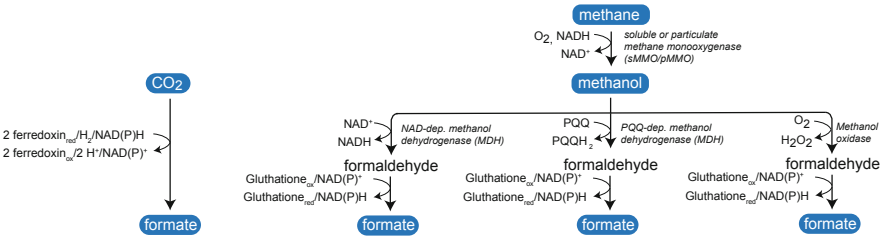


Fig. 3 Overview of alternative substrates modules. These modules could serve to convert alternative C1-substrates (methane, methanol, CO_2) to formate, which can next be assimilated in the rGlyP core (Fig. 1). PQQ: Pyrroloquinoline quinone

have generally higher catalytic rates, which is especially relevant for the thermodynamically challenging, reductive direction. In vitro measurements on metal-dependent FDHs have indicated decent turnover rates are feasible in the CO_2 reduction direction. The oxygen-tolerant molybdenum-dependent FDH of *C. necator*, which is naturally used in the oxidative direction, was shown to have a k_{cat} of 4.8 s^{-1} in the reductive direction [78]. This is roughly one order of magnitude lower compared to the oxidative direction but still above the average range of the k_{cat} for most plant Rubisco variants, which is the carboxylase in the CBB cycle [75]. Faster kinetic rates were measured for the hydrogen-utilizing FDH (HDCR) from *Acetobacterium woodii* ($k_{cat} = 28\ s^{-1}$) [77] and the FDH from *D. desulfuricans* ($k_{cat} = 47\ s^{-1}$) [79], for which the redox co-factor is not completely clear. The former HCDR is used for CO_2 reduction during operation of the rAcP for autotrophic growth of the acetogen *A. woodii* [77]. In addition, this HCDR was recently expressed functionally in *E. coli* leading to formate production from H_2/CO_2, and could potentially be combined with the rGlyP [80]. However, HDCR is oxygen-sensitive and could probably only support the rGlyP in anaerobic conditions. The FDH of *D. desulfuricans* is indeed active in the reductive direction, as autotrophic growth via the rGlyP on CO_2 and H_2 was recently demonstrated [30].

The recent demonstration in *D. desulfuricans* for the first time proved the rGlyP as a CO_2 fixation pathway in nature. So far, autotrophic growth of a strain engineered with the rGlyP has not been demonstrated. A challenge to overcome is the (heterologous) overexpression of a multi-subunit metal-dependent formate dehydrogenase, which will be complex in, for example, *E. coli,* where so far in vivo functionality of only slower metal-independent FDHs was demonstrated. Also, synthetic autotrophic growth via the rGlyP will require another inorganic electron donor, such as H_2 or direct uptake of external electrons, which will also require the heterologous expression of a hydrogenase or electron transfer system that can generate the required reducing equivalents (e.g., ferredoxin or NAD(P)H). The establishment of a synthetic rGlyP for CO_2 fixation would be very promising for biotechnology. This pathway is potentially much more ATP-efficient than many other natural or synthetic CO_2 pathways, albeit it will require elevated CO_2 concentrations [75, 81].

2.8 An Extension Module for Methanol Assimilation

Apart from supporting formatotrophic and autotrophic growth, the rGlyP can also be extended for the growth on more reduced C1-substrates, which then first need to be converted to formate (or methylene-THF). Among the reduced C1-substrates methanol stands out as a highly attractive C1-feedstock, as was already discussed in the introduction. Methanol can be converted into formate in two steps to feed into the core of the rGlyP (Fig. 3). In the first step methanol is oxidized to formaldehyde, this highly toxic intermediate then needs to be further oxidized in a second oxidation step into formate.

The first oxidation can, for example, be catalyzed by an NAD-dependent methanol dehydrogenase (MDH). This enzyme converts methanol to formaldehyde and retains the reducing power from this oxidation step within NADH. The generated NADH can be used to support further reductions in the rGlyP and the rest of metabolism, or part of the NADH can be oxidized to generate ATP in oxidative respiration. Apart from the NAD-dependent MDHs, there are also pyrroloquinoline quinone (PQQ) dependent MDHs, as found in some gram-negative methylotrophic bacteria [82, 83]. These PQQ-MDHs are interesting to consider for synthetic methylotrophy as they are kinetically and thermodynamically less constrained than their NAD-dependent counterparts, while they still retain some reducing power in reduced quinones, which can be used for ATP regeneration in respiration. However, successful heterologous expression of PQQ-MDHs, which reside in the periplasm, has not been shown yet. In addition, there are oxygen-dependent methanol oxidases as found in methylotrophic yeast species [84]. They may be interesting candidates to consider for engineering methylotrophy as they have much better kinetics than the NAD-MDHs, but they generate toxic H_2O_2 (yeasts express them in peroxisomes) and they dissipate all available reducing power directly.

Therefore, NAD-MDHs are the most promising option for methanol conversion for the rGlyP, at least in terms of theoretical yields [19], but this reaction brings some thermodynamic and kinetic challenges. The thermodynamics of this reaction are constrained ($\Delta rG'^{m} = +30.5$ kJ per mol at pH 7.5, ionic strength 0.25 and all reactants at 1 mM [41]). Hence, the oxidation requires, for example, high methanol concentrations and/or a high NAD^+/NADH ratio. In nature, some thermophilic bacteria are known to operate this enzyme for methylotrophic growth, such as *Bacillus methanolicus* [85]. Some of the thermophilic NAD-MDHs were prime candidates of engineering methanol to formaldehyde conversion, also in mesophilic hosts, such as *E. coli* [86]. Trying to overcome some of the kinetic limitations, a mesophilic NAD^+-dependent, broad substrate alcohol dehydrogenase from *C. necator* N-1 was successfully engineered to have better kinetic performance with methanol [87], and was recently shown to successfully convert methanol to formaldehyde in *E. coli* to support growth on methanol via the ribulose monophosphate (RuMP) cycle [88, 89].

The oxidation of methanol generates first formaldehyde, which is a highly reactive and hence generally toxic intermediate, which can be further oxidized to formate to connect to the core of the rGlyP. The thermodynamically more favorable oxidation of formaldehyde to formate can be performed by several formaldehyde dehydrogenases, which either donate electrons directly to NAD^+ or use an intermediate electron shuttle, such as glutathione or mycothiol, which eventually also donate the electrons to generate NAD(P)H [83].

In addition to the full implementation of the rGlyP for formatotrophic growth in *E. coli,* Kim et al. also attempted to demonstrate growth on methanol via the rGlyP [27]. Hence, they tested the heterologous expression of several thermophilic and mesophilic NAD-dependent MDHs, as well as the expression of NAD-dependent formaldehyde dehydrogenases. They observed slow, but fully methanol-dependent growth by the expression of one of the tested NAD-MDHs, i.e. the thermophilic

MDH from *Bacillus methanolicus*. Heterologous expression of two different NAD-dependent formaldehyde dehydrogenases from two *Pseudomonas* strains did not further improve this growth. This suggests that formaldehyde may have been converted well enough by the native *E. coli* glutathione-dependent formaldehyde dehydrogenase (FrmAB). Even though this work showed the principle of methylotrophic growth via the rGlyP, both yield (4.2 gCDW/mol methanol) and growth rate (doubling time of 54 h) were far from their theoretical optima, requiring further optimization [27].

Further optimization of the methanol conversion module may potentially lead to fast growth on methanol. The demonstration of fast methylotrophic growth (doubling time of 8 h) of an extensively engineered and evolved *E. coli* strain growing via the RuMP cycle is an encouraging result [88], showing that synthetic methylotrophic growth via the rGlyP may also be within reach.

2.9 Possible Module for Methane Conversion

Certain natural methylotrophs can grow not only on methanol, but they can also use the more reduced methane as a substrate. Methane is also considered an interesting feedstock for biotechnology, though mostly derived from fossil natural gas, also biogas can be a possible source. To allow growth on methane via the rGlyP, on top of the methanol conversion module, a methane monooxygenase reaction is needed (Fig. 3).

Methane monooxygenase enzymes oxidize methane with molecular oxygen, and therefore constrain the process to aerobic environments and result in the "wasteful" transfer of two electron pairs to molecular oxygen, also requiring an additional "wasteful" investment of reducing power in the form of NAD(P)H. Methane monooxygenase exists in nature in two different forms: one is soluble (sMMO) and found in the cytoplasm, the other is a particulate (pMMO) membrane-bound enzyme with a periplasmic domain. The heterologous expression of this category of enzymes has proven very challenging because of their complexity [90]. However, recently a heavily engineered, heterologous pMMO was successfully purified from *E. coli*, with kinetic in vitro parameters comparable to the native version of the enzyme [91]. Even more recently it was reported in a preprint that sMMO from *Methylococcus capsulatus* could be functionally expressed in *E. coli* and oxidize methane in vivo to methanol by co-expression of both *M. capsulatus* and *E. coli* GroESL chaperones [92].

These developments bring the engineering of synthetic methanotrophy closer and could potentially be integrated with the rGlyP. However, it should be noted that MMOs are rather slow because of kinetic barriers due to poor solubility and low chemical reactivity of methane.

3 Metabolic Engineering Approaches and Achievements

The implementation of a synthetic substrate assimilation route like the rGlyP in, for example, *E. coli,* which cannot naturally grow on C1-feedstocks, requires a profound change in central carbon metabolism. Simply overexpressing the required pathway enzymes is likely not sufficient to achieve any or efficient growth on C1-substrates. Firstly, properly functional (heterologous) enzyme variants have to be identified for different steps of the pathway, and their expression levels likely need fine-tuning to enable sufficient flux without overburdening the cell. In addition, interactions between the native metabolic regulation and the engineered pathway can result in unforeseen disruption of metabolic fluxes [93]. All these engineering challenges can be more effectively addressed by splitting up the construction of the engineered pathway in a stepwise implementation of several smaller metabolic modules [93].

This modular approach to pathway integration can be combined with a growth-coupled selection strategy. Module functionality can be coupled to growth of an engineered strain, by testing the module in a "selection strain" that is auxotrophic to the product of the module [94]. The construction of an auxotrophic selection strain involves strategic gene deletions to knock out native production of essential metabolites. These metabolites can then either be included in the medium to sustain growth of the auxotroph ("relaxing conditions") or it can be tested whether an engineered pathway module can sufficiently provide the essential metabolite (in "stressing conditions"). The rGlyP can be split up into several modules (Fig. 4a). These modules and combinations of them can be tested in different auxotrophic selection strains (Fig. 4b–h). Dependent on the proportion of biomass that is dependent on the module to be tested, different amounts of flux through the synthetic module are required. Demand for higher flux leads to a higher "demand" on the module. If an expressed module construct leads to low growth rates in a low "demand" selection strain, this is an indication that this module needs to be optimized. This can, for example, be done by screening for more enzyme variants or optimizing expression levels by promoter or RBS tuning. Once these issues have been tackled, the module can be tested for higher selection "demand," also in combination with other modules.

Another key strategy for the engineering of the rGlyP is the use of adaptive laboratory evolution (ALE) [27, 28, 95]. When the operation of a module or the whole pathway is coupled to a growth phenotype this provides an excellent chance for evolutionary "tinkering."

Here, we will discuss a series of selection schemes and strains that can be useful to test individual and combined modules for stepwise engineering of the rGlyP (Fig. 4). Many of these selection schemes were highly instrumental in the establishment of the pathway in *E. coli* and some other organisms recently, as will be discussed in more detail (overview of all engineering works so far in Table 2). Some of these selection schemes will likely prove to be an effective approach for the establishment of the rGlyP in more organisms. We will furthermore discuss complementary ^{13}C-based approaches to demonstrate module and pathway functionality. Furthermore, we will discuss a few studies which did not use growth-coupled, modular selection for engineering the rGlyP, which led to varying results and levels of success.

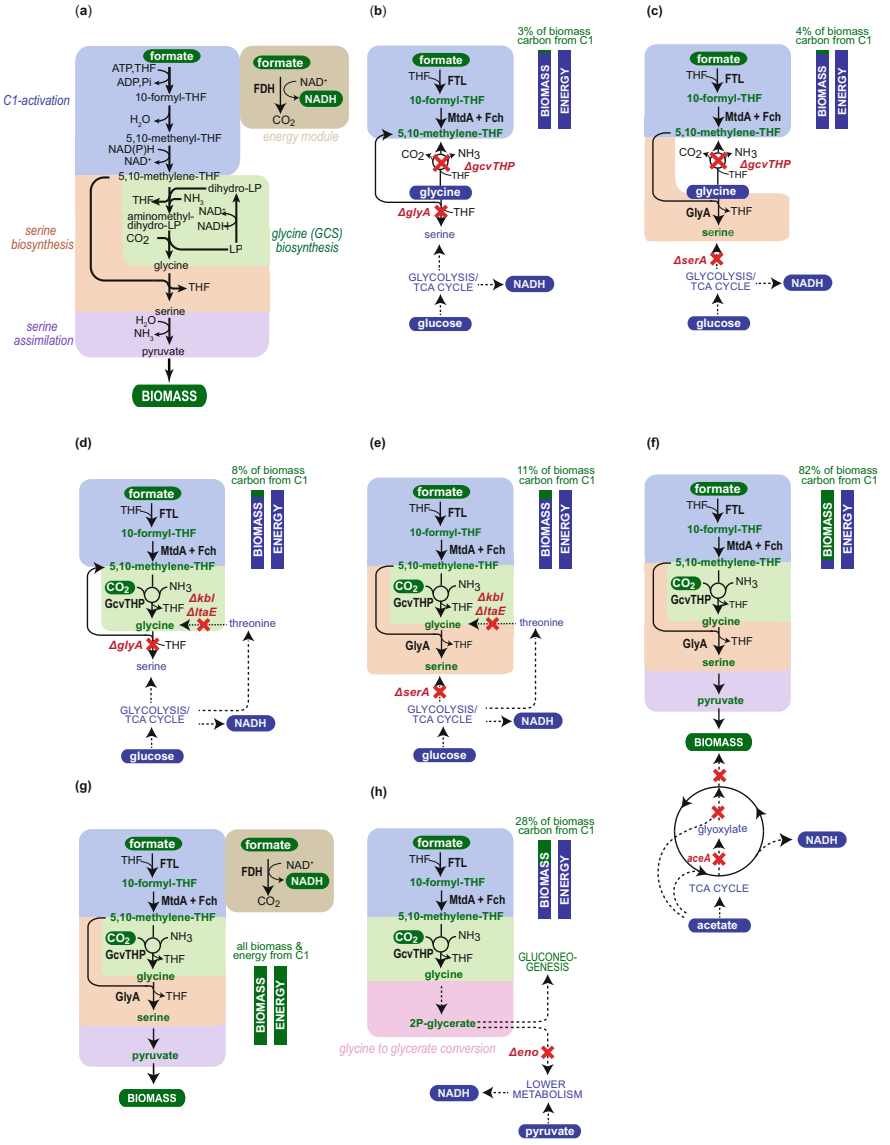


Fig. 4 Different selection schemes for different modules and combinations of modules of the rGlyP. (**a**) Overview of the different modules of the rGlyP core and canonical assimilation via serine deaminase. Different selection strain designs require activity of one or more rGlyP modules for biosynthesis of (**b**): C1, (**c**) C1 + β-carbon serine, (**d**) C1 + glycine, (**e**) C1 + glycine+serine, (**f**) almost all biomass except acetyl-CoA, and (**g**) all biomass and energy. All these were used previously (see Table 2). (**h**) is a proposed scheme to select for a pathway variant generating glycerate rather than pyruvate. The expected flux demand for biomass and energy supply via the modules from formate is indicated for each selection scheme based on *E. coli* biomass composition [96]. Figure is an adaption from [93]

Table 2 Overview of published studies on engineering the rGlyP

Host species	Study reference	Main achievements	Strain name	Selection for	Deletions for selection	C1-module	GCS-module	Glycine assim. module
E. coli	[97]	C1 + serine module demonstrated, first growth-coupled, modular engineering effort on rGlyP		C1	*ΔgcvTHP, ΔglyA*	*ftl* (*M. extorquens*, plasmid), *folD* (*E. coli*, native)		–
				C1 + serine (β-carbon)	*ΔserA, ΔgcvTHP*	*ftl* (*M. extorquens*, plasmid), *folD* (*E. coli*, plasmid)		*glyA* (*E. coli*, plasmid)
E. coli	[36]	C1 + GCS + serine module demonstrated in partial auxotroph, strain also integrated with electrochemical formate production	*7*	C1 + glycine (partial)	*ΔglyA*	*fhs, fchA, folD* (*C. ljungdahlii*, plasmid)	*gcvTHP* (*E. coli*, plasmid)	
			6	C1+ glycine + serine (partial)	*ΔserA*	*fhs, fchA, folD* (*C. ljungdahlii* plasmid)	*gcvTHP* (*E. coli*, plasmid)	*glyA* (*E. coli*, native)
			9	C1 + glycine +serine (partial)	*ΔserA*	*fhs, fchA, folD* (*C. ljungdahlii*, plasmid)	*gcvTHP* (*E. coli*, plasmid)	*glyA* (*E. coli*, native), *sdaA* (*E. coli*, genomic)
E. coli	[98]	C1 + GCS + serine module demonstrated in full auxotroph		C1 + glycine	*ΔglyA, Δkbl, ΔltaE, ΔaceA*	*ftl, fch, mtdA* (*M. extorquens*, plasmid)	*gcvTHP* (*E. coli*, plasmid)	
				C1+ glycine + serine	*ΔserA, Δkbl, ΔltaE, ΔaceA*	*ftl, fch, mtdA* (*M. extorquens*, plasmid)	*gcvTHPL* (*E. coli*, plasmid)	

(continued)

Table 2 (continued)

Host species	Study reference	Main achievements	Strain name	Selection for	Deletions for selection	C1-module	GCS-module	Glycine assim. module
E. coli	[102]	C1 + GCS + serine module achieved via FTL overexpression and ALE	G4670	C1+ glycine + serine	*ΔserA, Δtdh, Δkbl, ΔltaE, ΔaceBAK*	*ftfL* (*C. kluyveri*, genomic), *folD* (*E. coli*, native)	*gcvTHPL* (*E. coli*, native)	glyA (*E. coli*, native)
E. coli	[100]	C1 + GCS + serine module achieved, and energy module expression attempted	RG6	C1+ glycine + serine (partial)	*ΔserA*	*ftl, fch, mtdA* (*M. extorquens*, plasmid)	*gcvTHPL* (*E. coli*, genomic and plasmid)	*glyA* (*E. coli*, plasmid)
			RG8	C1+ glycine + serine (partial)	*ΔserA*	*ftl, fch, mtdA* (*M. extorquens*, plasmid)	*gcvTHPL* (*E. coli*, genomic and plasmid)	*glyA* (*E. coli*, plasmid)
S. cerevisiae	[43]	C1 + glycine module achieved in industrially relevant yeast species		C1 + glycine	ΔSHM12, ΔGLY1, ΔAGX1	MIS1 (*S. cerevisiae*, plasmid)	GCV123 (*S. cerevisiae*, plasmid)	
E. coli	[27]	First time rGlyP fully established by rational expression and ALE leading to fastest demonstrated formatotrophic growth via rGlyP, as well as slow methylotrophic growth	gC1M gC2M gC3M	Biomass (except acetyl-CoA)	*ΔserA, Δkbl, ΔltaE, ΔaceA*	*ftl, fch, mtdA* (*M. extorquens*, genome)	*gcvTHP* (*E. coli*, genome)	*sdaA, glyA* (*E. coli*, genome)
			K4	All biomass + energy	*ΔserA, Δkbl, ΔltaE, ΔaceA*	*ftl, fch, mtdA* (*M. extorquens*, genome)	*gcvTHP* (*E. coli*, genome)	*sdaA, glyA* (*E. coli*, genome)
			K4e	All biomass + energy	*ΔserA, Δkbl, ΔltaE, ΔaceA*	*ftl, fch, mtdA* (*M. extorquens*, genome)	*gcvTHP* (*E. coli*, genome)	*sdaA, glyA* (*E. coli*, genome)

E. limosum	[42]	Introduction of GCS module in acetogen with rAcP boosts autotrophic growth, likely via rGlyP	GSRP	No selection	–	*fdh, fhs, fchA, folD* (*E. limosum*, native)	*gcvTHPaPb* (*C. drakei*, plasmid)	*grdABCDEX* (*C. drakei*, plasmid)
Synechocystis sp. PCC6803	[101]	FTL expression in cyanobacterium to boost CO_2 fixation with formate CO-assimilation. No convincing pathway/module activity		No selection	–	*ftl* (*M. extorquens*, genome)		
C. necator	[28]	Formatotrophic growth is rGlyP fully established by rational expression and ALE in natural formatotroph deleted for native CBB cycle Highest yield demonstrated via rGlyP, but still slightly lower than CBB cycle		C1 + glycine	*ΔglyA, Δkbl, ΔltaE*	*ftl, fch, mtdA* (*M. extorquens*, plasmid)	*gcvTHP* (*C. necator*, plasmid)	–
			CRG1	All biomass +energy	*ΔccbSLc2, ΔccbSLp*	*ftl, fch, mtdA* (*M. extorquens*, plasmid)	*gcvTHP* (*C. necator*, genome)	*dadA6, gcl, tsr* (C. necator, native)
			CRG4	All biomass +energy	*ΔccbSLc2, ΔccbSLp*	f*tl, fch, mtdA* (*M. extorquens,* plasmid)	*gcvTHP* (*C. necator,* genome)	*sdaA, glyA* (*C. necator*, plasmid)
E. coli	[29]	Second demonstration of formatotrophy via rGlyP in *E. coli* by rational engineering. Additional rational targets, but yield and growth rate relatively low. Highest ODs achieved for rGlyP (by fed-batch feeding)	FC8	All biomass +energy	*ΔserA*	*ftl, fch, mtdA* (*M. extorquens*, plasmid)	*gcvTHPL* (*E. coli*, genomic)	*sdaA, glyA* (*E. coli*, native)
C. pasteurianum	[69]	GCS expression in anaerobic glycerol-consumer to improve redox balance and yield of fermentation. No convincing pathway/module activity	GCSY1	No selection	–		*gcvTHPaPbL* (*G. acidurici*, plasmid)	

(continued)

Table 2 (continued)

Study reference	Energy module	Additional mutations (rational or ALE)	Doubling time (h)	Yield (gCDW/mol formate	Key labelling results
[97]			2.3		^{13}C-formate → 96% +1 methionine (from methylene-THF)
			1.6		^{13}C-formate gives 95% +1-labelled serine
[36]			25*		^{13}C-formate + $^{13}CO_2$ → 41% total ^{13}C in glycine, 14% in alanine
			10*		^{13}C-formate + $^{13}CO_2$ → 36% total ^{13}C in serine, 18% in alanine
			20*		^{13}C-formate + $^{13}CO_2$ → 37% total ^{13}C in serine, 17% in alanine
[98]			1.6		^{13}C-formate + $^{13}CO_2$ → 94% total ^{13}C in glycine
			1.7		^{13}C-formate + $^{13}CO_2$ → 87% total ^{13}C in serine
[102]		Rational: Δ*gcvR*, Δ*kdgK* ALE: many, including duplication *folD* and *ftl*	3		^{13}C-formate → glutathione demonstrates glycine + serine from formate (data see paper)
[100]		Rational: Δ*gcvR*, Δ*pflB*	10*		^{13}C-formate → 98% +2 serine; (13-formate → 15% +2 alanine)
	fdh (NADH, *C. boidinii*, plasmid)	Rational: Δ*gcvR*, Δ*pflB*	3*		
[43]			6.9		^{13}C-formate + $^{13}CO_2$ → 60% total ^{13}C in glycine
[27]			13		
	fdh (NADH, *Pseudomonas sp. 101*, genome)		70	2.3	
	fdh (NADH, *Pseudomonas sp. 101*, genome)	ALE: e.g., promoter_*pntAB*: up, *5'UTR_fdh*: up	8		^{13}C-formate + $^{13}CO_2$ → 98% total ^{13}C in alanine

[42]	Hydrogenase (*E. limosum*, native)		96		
[101]				–	^{13}C-formate → 6% +1 histidine
[28]		Rational: *ΔphaC1*	5.1		^{13}C-formate → 82% +1 glycine
	fdh (NADH, *C. necator*, native)	Rational *ΔphaC1 ALE:* e.g. *ΔccbRc2,* promoter_*gcvTHP* (plasmid): down	56		
	fdh (NADH, *C. necator*, native)	Rational *ΔphaC1*. ALE: incl. *ΔccbRc2,* promoter *gcvTHP* (genome): up	12	2.6	^{13}C-formate +$^{13}CO_2$ → 97% total ^{13}C in alanine
[29]	*fdh* (NADH, *C. boidinii,* plasmid), *fdh-mut* (*A. thaliana,* NADPH, plasmid)	Rational: *ΔgcvR, ΔpflB,ΔppsR, prom*_promoter_*ppsA: up, ΔpurT*	158	2.2	Not done for FC8 (only FC5)
[69]					No labelling data available for engineered strain

For each study key selection schemes (if applied) are given, including knockouts to generate these selection strains. It is indicated which specific genes contribute to module or pathway operation. For the listing of these genes: plasmid indicates that the specified gene(s) are expressed from a plasmid, native means they are not overexpressed and the module operation relies on natural expression, genomic means that the genes are inserted into the chromosome or overexpressed by promoter engineering. Additional rational modifications and mutations other than these made for pathway expression and selection schemes are separately mentioned. In the same "Additional mutations" column some key mutations are indicated that were found in strain optimization by Adaptive Laboratory Evolution (ALE) if applied. Doubling times and yields reported for growth of the described strains are included. * denote that the growth rate is estimated from a growth curve. Key labelling data show key pathway intermediates or derivative (alanine from pyruvate, methionine from methylene-THF, histidine from formyl-THF) which were measured after growth on either ^{13}C-formate, $^{13}CO_2$ or both as indicated. Percentages close to 100% indicate that almost all flux is coming as intended from the module or rGlyP. For detailed labelling data and conditions, we refer to the original papers. Abbreviations (including metabolic roles of some if not discussed elsewhere in this chapter): *aceA* (isocitrate lyase), *aceB* (malate synthase, consumed glyoxylate coming from *aceA* with acetyl-CoA), *aceK* (isocitrate dehydrogenase kinase/phosphatase, occurs in operon with *aceAB*, as it controls flux into glyoxylate shunt, by inactivating/activating isocitrate dehydrogenase by phosphorylation and de-phoshorylation, the latter enzyme competes with aceA for isocitrate substrate), AGX1(alanine:glyoxylate aminotransferase *S. cerevisiae*), *ccbSLc2/p* (small and large subunit of Rubisco on chromosome 2 and megaplasmid in *C. necator,* which deletion blocks formatotrophic growth via the CBB cycle), *dadA6* (glycine oxidase which converts glycine to glyoxylate in *C. necator*), *fch(A)* (methenyl tetrahydrofolate cyclohydrolase), *fdh* (formate dehydrogenase), *ftl/fhs* (formate-tetrahydrofolate ligase), *folD* (bifunctional methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase), *gcl* (glyoxylate carboligase, can assimilate glyoxylate from glycine oxidation), *gcvR* (transcriptional regulator of *gcvTHP*, likely a repressor that if deleted upregulates *gcvTHP* expression constitutively in *E. coli*), *gcvTHP(aPb)(L)* (glycine cleavage system

(continued)

Table 2 (continued)

T, H, P, and L-protein PaPb denotes heterodimeric P protein), GCV123 (glycine cleavage/synthase system T, H, and P-protein in *S. cerevisiae*), *glyA* (serine hydroxymethyltransferase), GLY1(threonine aldolase *S. cerevisiae*), *grdABCDEX* (glycine reductase enzyme complex), MIS1 (trifunctional formate-tetrahydrofolate ligase/methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase), *kbl* (glycine acetyltransferase), *kdgK* (2-dehydro-3-deoxygluconokinasewas not deleted for a functional reason but used as a locus to insert heterologous *ftl*), *ltaE/A* (threonine aldolase), *mtdA* (methylene-tetrahydrofolate dehydrogenase), *phaC1* (polyhydroxybutyrate polymerase, blocks PHB formation in *C. necator*, which could be an unwanted by-product if the formatotrophic strains are later engineered for products), *pflB* (pyruvate formate lyase, could theoretically be a way that *E. coli* could internally generate formate from pyruvate, but only functions in anaerobic conditions), *pntAB* (membrane-bound proton-translocating transhydrogenase that can convert NADPH to NADH), *ppsA* (phosphoenol pyruvate synthetase, this was upregulated by promoter engineering to boost gluconeogenic flux from pyruvate upward), *ppsR* (phosphoenol pyruvate synthetase regulatory protein, this regulator was deleted to boost gluconeogenic flux from pyruvate upward), *purT* (phosphoribosylglycinamide formyltransferase, could indirectly counteract FTL activity by being involved in a reverse pathway) *sdaA* (serine deaminase), *serA* (phosphoglycerate dehydrogenase), SHM12 (mitochondrial and cytoplasmic serine hydroxymethyltransferase 1 and 2 from *S. cerevisiae*), *tdh* (threonine dehydrogenase, can if knocked out block threonine to glycine conversion, alternative to knocking out *kbl*), *tsr* (tartronate semialdehyde reductase, can contribute to glyoxylate assimilation)

3.1 Selection Schemes for the C1, GCS and Serine Modules

Four pathway modules have been defined for a modular selection of the serine deaminase variant of the rGlyP (Fig. 4a) [93]. To select for the first C1-activation module of the rGlyP (formate to methylene-THF) a "low selection pressure" auxotrophic strain was designed (Fig. 4b). Methylene-THF, as well as the other intermediate of the C1-module, 10-formyl-THF, are essential metabolites, as they are precursors for the production of purines and thymidine (required for nucleotide formation), coenzyme A, histidine, methionine and the formylation of methionyl-tRNA. In total, these two C1-species account for ~3% of *E. coli* biomass [96]. For testing the C1-module in *E. coli*, C1-auxotrophy (methylene-THF + formyl-THF) was first created by deletion of the GCS (Δ*gcvTHP*) and serine hydroxymethyltransferase (Δ*glyA*). This abolished the native routes for methylene-THF formation from glycine and serine, and this strain could only grow on minimal medium with glucose when several C1-derived compounds were supplemented (methionine, thymidine, inosine, and pantothenate) [97]. In addition to the C1-supplements, the *E. coli* C1-auxotroph was typically fed with glycine, as most glycine is natively synthesized from serine in *E. coli*, although the alternative pathways via threonine cleavage were left intact. Alternative to C1-supplementation, this strain should be able to grow if formate would be converted by the C1-module of the rGlyP, for which *E. coli* natively lacks at least the FTL enzyme. Overexpression of the FTL from *M. extorquens* from a plasmid and feeding formate (with glucose and glycine) rescued the growth of this selection strain, showing the functionality of this first module [97]. This selection strain provided a first proof-of-principle for the first module of the pathway in *E. coli*.

To select for activity of the serine biosynthesis in combination with the C1-module, a C1 + serine-auxotroph can be created, in which the beta-carbon of serine is derived from methylene-THF of the C1-module and the rest from supplemented glycine (Fig. 4c). To generate this selection scheme in *E. coli,* native synthesis of serine from the glycolytic intermediate 3P-glycerate was abolished by deletion of 3P-glycerate dehydrogenase (Δ*serA*), as well as the deletion of the GCS (*Δ*g*cvTHP*). These deletions cause that this strain can only grow on minimal medium when supplemented with glycine and serine. The *E. coli* C1 + serine auxotroph was then used to test the combined expression of the C1-module and serine biosynthesis-module (GlyA), by testing growth on formate and glucose, while still supplementing glycine. This selection scheme demands slightly higher flux from the C1-module than the C1-auxotroph (~4% vs. 3%). This selection allowed for the discovery of a bottleneck in the production of 10-formyl-THF and methylene-THF, as this selection strain could not grow well anymore with expression of merely FTL. Growth was improved by overexpression of the native bifunctional methylene-THF dehydrogenase/methenyl-THF cyclohydrolase (*folD*) [97].

Another possible selection scheme is glycine+C1-auxotrophy, which can be used to select for both C1-module and GCS-module activity (Fig. 4d). In the previous selection schemes, glycine was supplemented to compensate for lack of glycine

synthesis from serine, but this does not render the strain fully auxotroph to glycine because glycine can also be produced from threonine. In order to obtain a complete glycine auxotroph, both conversion from serine (*ΔglyA*) and cleavage of threonine to glycine must be disrupted (Δ*kbl*, Δ*ltaE*). Glycine can possibly also be generated by aminotransferases converting glyoxylate to glycine by accepting the amino group from another amino acid. As there are many potential aminotransferases this route was knocked out in *E. coli* by knocking out glyoxylate biosynthesis via an isocitrate lyase deletion (*ΔaceA*) [98]. Also, for the engineering of the rGlyP in *C. necator* and *S. cerevisiae* a glycine auxotroph was created. In *C. necator* isocitrate lyase was not knocked out, but it was still observed that this strain could barely grow on minimal medium without glycine rendering it a useful selection strain [28]. In *S. cerevisiae* one annotated alanine-glyoxylate aminotransferase was knocked out on top of other glycine biosynthesis genes to prevent the aminotransferase route [43].

In the *E. coli* C1 + glycine auxotroph it was realized that the previously engineered C1-module (consisting of *M. extorquens* FTL and native bifunctional FolD) in combination with overexpression of the native GCS was not sufficient to allow growth of the glycine auxotroph (requiring ~8% of biomass coming from the module) [98]. Only when also expressing heterologous, monofunctional methylene-THF dehydrogenase and cyclohydrolase (MtdA and Fch) from *M. extorquens* the growth of the glycine auxotroph on formate and glucose could be realized. Another *E. coli* study which utilized an incomplete glycine auxotroph (only *ΔglyA*) found that the C1-module from the acetogen *C. ljungdahlii* could also function to generate part of the glycine pool [36]. This partial auxotroph was not very suitable to demonstrate module functionality by growth, but it could be used to demonstrate module functionality by measuring (partial) ^{13}C-labelling of glycine. Another study used several tight selection strains, including a full *E. coli* glycine+C1-auxotroph to perform ALE. This study only heterologously expressed FTL from *Clostridium kluyveri* from the genome. After ALE, this led to growth via the C1 + glycine modules, and required several mutations, including duplications of the native *E. coli* FolD and the heterologous FTL.

In *C. necator* the glycine auxotroph was the first selection strain used and a full heterologous C1-module from *M. extorquens* and the native GCS overexpression was included from the start. The selection strain was used to select properly performing promoter-strength combinations driving operons on two different plasmids for both modules (C1 plasmid and GCS plasmid) [28]. Also, in *S. cerevisiae* the engineering of the rGlyP was started directly with a glycine auxotroph to select for both the C1-module and GCS-module, which were successfully realized by overexpressing native genes from plasmids [43]. Overall, the glycine+C1-auxotroph seems a promising starting selection strain to establish the C1 + GCS modules in one-go. In case this fails to work, one can resort to previously discussed selection schemes that first select for the C1-module separately.

To engineer variants of the pathway that assimilate glycine via serine, a selection scheme is useful which selects for the combined C1, GCS, and serine-modules in a serine+glycine+C1-auxotroph (Δ*kbl* Δ*ltaE* Δ*serA*) (Fig. 4e). Unlike all previous discussed selection schemes, which were “dead-end” selection schemes, this is an

open-end selection. This is because none of the knockouts in this serine+glycine +C1-auxotroph are involved in the pathway. Previous selection schemes all involved the knockout of enzymes that are involved in the rGlyP, i.e. GlyA or GCS, which need to be restored or overexpressed later on to engineer the full pathway. This is an advantage as this serine+glycine+C1-auxotroph could be more easily expanded to later test the whole pathway for formatotrophic growth by adding further modules and optimization.

The serine+glycine+C1-selection in *E. coli* showed that expression of the heterologous C1-module from *M. extorquens* and overexpression of the native GCS were sufficient to supply ~11% of biomass for *E. coli*, whereas overexpression of the native GlyA was not yet needed for this selection [98].

3.2 *Selections for Full Biomass Assimilation and Full Formatotrophy via rGlyP*

The next challenges after implementing these three modules until serine biosynthesis were to generate all biomass from formate via conversion of serine to pyruvate by serine deaminase (SdaA), and after that to make all energy from formate through the formate dehydrogenase (FDH or "energy module"). To only select for biomass production via the rGlyP, and not yet for energy production from formate, a selection scheme was developed to separate energy metabolism from carbon assimilation (Fig. 4g). This separation was realized in *E. coli* by designing a selection strain in which the gene encoding the isocitrate lyase enzyme (AceA) was deleted on top of the previous open-ended serine+glycine+C1-auxotroph (Δ*aceA* Δ*kbl* Δ*ltaE* Δ*serA*) [27]. The isocitrate lyase knockout in *E. coli* blocks the so-called glyoxylate shunt. The glyoxylate shunt is essential for growth on acetate as a carbon substrate. By blocking this shunt acetate can still be "burned" in the TCA cycle to generate energy, and be used to produce the precursor acetyl-CoA (~18% of biomass), but all other essential precursors of biomass (~82%, including, e.g., pyruvate and metabolites in glycolysis and TCA cycle) cannot be generated from acetate anymore.

This selection scheme was chosen to show operation of the full rGlyP. However, it was suspected that the burden of using several plasmids for expressing all needed four modules could impair formatotrophic growth. Hence, all pathway genes that required overexpression were expressed from the genome by integrating a synthetic operon of previously confirmed heterologous *M. extorquens* genes with strong constitutive promoters and introducing strong constitutive promoters on the genome to overexpress native operons. This strategy successfully allowed growth via the rGlyP, in which almost all biomass (except from molecules coming from acetyl-CoA) were generated from formate and CO_2, while all energy was supplied by acetate.

Next, to test growth using formate both as carbon and energy source, a heterologous FDH gene from *Pseudomonas sp*.101 was integrated into the genome of the previous strain. It was observed that this strain could directly grow on formate (and CO_2) as sole carbon and energy source, albeit at a relatively slow growth rate of ~70 h doubling time [27]. Still, this demonstrated for the first time formatotrophic growth via a rational implementation of the rGlyP in *E. coli*, a strain which was called K4. Another recent study followed a similar approach leading to the same result of full formatotrophic growth of *E. coli* by a rational approach [29]. However, this study optimized the expression of some genes on low copy number plasmids and performed some additional mutations that the authors expected to contribute to achieve full formatotrophic growth. This included the deletion of GcvR, a repressor of the GCS and the regulator PpsR, the deletion of which could stimulate gluconeogenic flux from pyruvate upwards, as pyruvate is the key product of the rGlyP. Furthermore, a heterologous NADPH-dependent FDH was included in addition to a heterologous NADH-dependent variant, and PurT was deleted, an enzyme which could counteract FTL activity. Despite these additional rational steps, this approach failed to beat the performance of the previously published K4 strain and led to a maximum growth rate of ~157 h doubling time, and only a few doublings were shown in this work [29] .

To further improve the growth rate apparently other changes were required. Hence the previously mentioned rationally engineered K4 strain was subjected to ALE by several serial transfers on formate minimal medium. This approach improved the formatotrophic growth rate and yield of the resulting strain K4e successfully, to reach a doubling time of ~8 h and a yield of 2.3 gCDW/mol formate, approaching yields of natural formatotrophs [27]. Key mutations identified in this strain were the increased expression of heterologous FDH, which was sensible as the non-metal dependent *Pseudomonas* FDH was already expected to be a bottleneck based on simple theoretical calculation [23]. In addition, a mutation was found to upregulate the expression of native membrane-found transhydrogenase (PntAB), the latter probably reflecting a limitation in NADPH in the non-evolved strain, as the rGlyP requires NADPH. Another interesting improvement of the evolved K4e strain is its higher tolerance to formate. Whereas the initial K4 strain already had decreased growth at >30 mM formate, the evolved strain could grow up to 100 mM formate, and in the presence of bicarbonate (probably due to pH buffering) even up to 300 mM. Growth on higher formate concentrations is beneficial for bioprocesses. However, given the relatively high toxicity of formate some kind of fed-batch or continuous process will be required to achieve sufficiently high biomass concentrations and product titers on formate.

The only other study so far that showed full growth via an engineered rGlyP was done in *C. necator* [28]. Here, the expression levels of the C1-module and GCS-module were first optimized in a glycine+C1-auxotroph. Then, growth of *C. necator* on glycine was tested, which after short-term evolution was established, demonstrating a good potential for *C. necator* to perform glycine assimilation. However, transcriptomics and knockouts of this glycine growth phenotype revealed *C. necator* employed wasteful glycine oxidation to glyoxylate and glyoxylate

assimilation via the glycerate pathway. Still, it was attempted to integrate the C1 + GCS modules with this native glycine assimilation capacity. This was tested in a strain that was knocked out for the CBB cycle, which is the native formatotrophic pathway of *C. necator*, by removing both copies of the Rubisco genes. This already led to formatotrophic growth, but only at low growth rate (~60 h doubling time) and low yield, likely caused by the native inefficient glycine assimilation via the glycine oxidase route. After overexpression of native enzymes for the more efficient serine deaminase route, as well as short-term ALE, this led to relatively fast formatotrophic growth (~12 h doubling time) and reasonably high yield (2.6 gCDW/mol formate), albeit still slightly lower than yield of formatotrophic growth via the CBB cycle (2.9 gCDW/mol formate (Table 2). One of the interesting mutations found after ALE in *C. necator* led to the inactivation of the CBB cycle regulator gene (CbbR), which probably lowered the expression of remaining CBB cycle enzymes, which were not required in this strain, thus reducing burden. In addition, a mutation was found in the strong constitutive promoter integrated in the genome to regulate the GCS, which was shown to further upregulate its genomic expression. Interestingly, in an earlier *C. necator* rGlyP strain harboring a plasmid-based GCS expression a mutation lowering GCS expression was found after short-term ALE. This underlines that fine-tuning GCS expression to a sweet spot between limiting flux and expression burden is important for efficient rGlyP operation and can be achieved by ALE.

3.3 Other Selection Schemes and Considerations for Future Engineering

For the engineering of rGlyP with an alternative module for glycine assimilation rather than the canonical serine deaminase variant, other selection schemes can be used. For several of the glycine conversion variant modules, P-glycerate will be a key intermediate. A possible selection scheme for the selection of such pathways could be a strain in which glycolysis is disrupted below P-glycerate by an enolase knockout (Δ*eno*) (Fig. 4h). In this case modules of a rGlyP variant making P-glycerate could supply upper metabolism (~28% of biomass) and the rest of metabolism could, for example, be fed with pyruvate as helper substrate. This selection scheme is very similar to a scheme used to evolve the full functionality of an engineered CBB cycle in *E. coli* [99]. More selection schemes can be envisioned for different variants and in different hosts. Suitable selection schemes will also be partly host-dependent. To be able to create specific auxotrophs in a host several conditions need to be met. For example, all genes towards the auxotroph metabolite need to be known, the knockout of the genes should be technically and biologically possible and the host should be able to transport the auxotrophic metabolites inside the cell from the medium. This may lead to the selection of specific selections schemes when engineering the rGlyP in a different host. In

general, based on the experience of engineering the rGlyP in a few host, one can start with already somewhat "high selection" strains covering multiple modules (e.g., C1-module and GCS module in the glycine +C1 auxotroph) at once. If this directly works one will be able to proceed faster. In case no growth can be achieved, one needs to fall back on lower level selection strains and first optimize an individual module.

3.4 Pathway Module Confirmation by ^{13}C-Labelling Studies

Even though the growth of a selection strain is a strong indication for the functionality of the rGlyP, the confirmation of module or pathway functionality can be obtained more conclusively by ^{13}C-labelling studies. In such a study, labelled substrates, in the case of rGlyP typically ^{13}C-formate or $^{13}CO_2$, are fed to the strain operating a module or the full pathway. Then, a convenient way to prove functionality is the analysis of steady-state labelling patterns in the proteinogenic amino acids. Proteinogenic amino acids are abundant and can be easily hydrolyzed from culture samples and analyzed by LC-MS. For more details on related protocols, we refer to Wenk et al. [94]. Distinct labelling patterns can determine if, for example, methylene-THF (via the labelling of the C1-derived carbon in methionine and histidine), glycine, serine or pyruvate (via labelling of derived amino acids such as alanine and valine) are indeed synthesized as expected by the rGlyP modules (see also Table 2). Full operation of the rGlyP should lead to full labelling of all proteinogenic amino acids when grown on ^{13}C-formate plus $^{13}CO_2$. Growing strains with the full rGlyP on either ^{13}C-formate or $^{13}CO_2$ can be potentially used to prove the operation of certain variants of the pathway. In this way the use of the glycine reductase route could be demonstrated in the natural rGlyP auxotroph *D. desulfuricans* [30]. However, neither ^{13}C-formate nor $^{13}CO_2$ labelling can differentiate between the other glycine to biomass conversion pathways. In addition, labelling with either 13-formate or $^{13}CO_2$ during growth via the full rGlyP can be used to determine how high the "wasteful" TCA cycle flux is. Labelling of proteinogenic amino acids derived from TCA cycle metabolites can be used to determine if fluxes to these metabolites primarily comes from anaplerosis or from full TCA cycling. This method was shown for *E. coli* and *C. necator* with an engineered rGlyP, for both it was found that the anaplerotic flux is higher than the "wasteful" TCA cycling [27, 28].

3.5 Non-growth Coupled Engineering Efforts

Some studies have expressed heterologous enzymes for the rGlyP without following a growth-coupled approach. In some of the *E. coli* studies on engineering the rGlyP, modules were tested without coupling them (fully) to growth [36, 100]. Even though

this led to partially functional modules, it is harder to assess module functionality in these studies. And once the modules were confirmed without selection, it was realized that a serine auxotroph strain could be beneficial to force more flux to serine and even downstream to pyruvate [100].

In some other hosts, initial steps were taken to implement the rGlyP, such as the overexpression of the GCS enzymes in the anaerobic glycerol fermenter *Clostridium pasteurianum*, with the goal to create mixotrophic fermentation of glycerol and formate [69]. However, the overexpression of these genes was not coupled to growth via a selection scheme, and high expression burden of the GCS overexpression was observed. This suggests optimization of expression levels of the GCS genes is needed for which a growth-coupled selection could be helpful and allow for expression optimization via evolution, as also demonstrated in *C. necator*. In this study still some physiological impacts of the GCS were observed in the form of increased formate consumption in the presence of glucose, but this work could not yet prove high flux through rGlyP modules.

In another study FTL was overexpressed to establish modules of the rGlyP as a pathway for formate assimilation to support carbon fixation by the CBB cycle in the cyanobacterium *Synechocystis* sp. PCC6803. In this non-growth coupled approach some physiological effects were observed that could potentially be linked to flux through the rGlyP [101]. However, labelling and transcriptomics revealed that the overexpression of FTL led to unanticipated changes in the glycine, serine, C1 and nitrogen metabolism. Also here, a growth-coupled approach could potentially further force flux through the modules towards establishing the rGlyP in this host in the future.

A seemingly more successful establishment of the full rGlyP was achieved in the acetogen *E. limosum* [42]. Here, the GCS from the rGlyP-operating acetogen *C. drakei* was heterologously introduced on a plasmid. This was done without any modular or growth-coupling based approaches. Direct introduction of the GCS led to improved growth and acetate biosynthesis on H_2/CO_2. However, it must be noted that the rGlyP here did not lead to full autotrophic growth, as the pathway operates in concert with the native rAcP. Also, no labelling data on the engineered pathway activity were provided.

4 Comparing the rGlyP to Other C1-Pathways

Apart from the rGlyP there are several other pathways for assimilation of C1-substrates and CO_2 in nature, as well as hundreds of theoretical pathways that have been proposed and analyzed in silico [19, 23, 81, 103–105]. Only for a subset of these pathways, experimental studies have attempted to implement them in vitro or in vivo [52, 104, 106]. For even fewer pathways, full functional operation has been realized so far in engineered hosts. Apart from the rGlyP, successful implementation was recently shown for the RuMP cycle for synthetic methylotrophy in *E. coli* and for the CBB cycle for synthetic formatotrophy in *E. coli* and

methylotrophy in *Pichia pastoris* [88, 107]. In this section, we provide a comparison of the rGlyP versus other key pathways for the assimilation of formate, methanol and/or CO_2 that are present in nature, have been engineered or have been proposed as promising pathways by in silico analyses.

4.1 Assessing the rGlyP for Formatotrophic Growth

Some natural and many more theoretical formate-assimilating pathways have been described in literature [23, 33] (Table 3). Briefly, these can be distinguished into two groups based on their formate utilization strategy. To the first cluster belong those pathways which oxidize formate to CO_2, exploiting the extracted electrons as reducing power for CO_2 fixation routes. Examples of such pathways are the naturally occurring CBB cycle, reductive tricarboxylic acid cycle, and synthetic CO_2 fixations pathways such as the CETCH cycle and GED (6-phosphogluconate dehydrogenase (Gnd) – Entner–Doudoroff) cycle (Table 3). To the second group belong those metabolic routes which directly assimilate formate. The latter category includes the rGlyP and other major natural formate-assimilation pathways, the rAcP and Serine cycle, which all use FTL for the direct assimilation and activation of formate.

A key parameter to compare different pathways is their theoretical yield, which is mostly determined by the ATP investments (and in some cases also by "wasteful" reducing equivalent investments) required to produce a certain metabolite (e.g., pyruvate) or biomass. Another key parameter to compare pathways is their kinetic performance, as the overall pathway kinetics determine potential growth rates, productivities, and enzyme burden. Due to limited, trustworthy estimates for the kinetic parameters that accurately reflect in vivo conditions, this performance is harder to quantify and compare. Still, several studies have compared different formate natural and synthetic assimilation pathways [19, 23, 33] including a recent detailed assessment taking into account thermodynamic driving force and kinetic parameters [81]. The general picture that emerges is that the rGlyP outcompetes most other formate-assimilation pathways – as well as CO_2 fixation pathways – in terms of theoretical yields, due to the lower ATP consumption for pyruvate and biomass biosynthesis. The only pathways that can rival the rGlyP in terms of theoretical yield are the rAcP and rTCA, which have a lower ATP consumption than most variants of the rGlyP (Table 3). However, the rAcP and rTCA usually require anaerobic conditions due to oxygen-sensitive enzymes [108], and relatedly have ATP-limitations and a limited product spectrum. When comparing the rGlyP further to other oxygen-tolerant formate-pathways, the ATP efficiency is higher, leading to small yield advantages over most pathways [19, 81]. Also in terms of kinetics based on available parameter estimates, the rGlyP is suggested to be among the best performing pathways.

Another advantage of the rGlyP is its architecture as a linear pathway. We note here that some studies refer to the rGlyP as a circular pathway, since the co-factor THF is used as an acceptor at the start of the pathway and released later in the GCS

Table 3 Comparison of the rGlyP to other natural, engineered, and proposed pathways for C1-assimilation and CO_2 fixation

Pathway name (variant name)	Naturally present, (partly) engineered or only theoretically proposed per C1-substrate [refs.]	ATP costs/ pyruvate	Kinetic performance based on [81]	Oxygen tolerant	CO_2 concentration dependence	Key product	Topology	Overlap central metabolism
rGlyP (serine deaminase)	Engineered: formate, methanol [27–29] Theoretical: CO_2, CO	2	+ (+/− for CO_2)	Yes	High	Pyruvate	Linear	Low
rGlyP (glycine reductase[a])	Natural: CO_2, formate [30] Theoretical: CO, methanol	1	?	No	High	Acetyl-coA	Linear	Low
rAcP	Natural: CO_2, CO, formate, methanol	<1	?	No	High	Acetyl-coA	Linear	Low
CBB cycle	Natural: CO_2, CO, formate, methanol engineered: CO_2, formate, methanol [107, 116]	7	+	Yes	Low	Glyceraldehyde-3-phosphate	Cyclic	High
rTCA cycle	Natural: CO_2 [117, 118] Theoretical: formate, CO methanol	1–2	?	No[b]	High	Acetyl-coA	Cyclic	High
RuMP (FBA-TAL variant) cycle	Natural: methanol Engineered: methanol [88] Theoretical: CO, formate	−1 to 5[c]	+/−	Yes	None	Glyceraldehyde-3-phosphate	Cyclic	High
Serine cycle (natural)	Natural: formate, methanol Theoretical: CO_2, CO	6[d]	+/−	Yes	Low	Acetyl-coA	Cyclic	High
Serine cycle (modified serine cycle)	Partially engineered: formate, methanol Theoretical: CO_2, CO	8[d]	?	Yes	Low	Acetyl-coA	Cyclic	High
Serine cycle (homoserine cycle)	Partially engineered: methanol Theoretical: CO_2, CO, formate	4[d]	?	Yes	None	Acetyl-coA	Cyclic	Low
Formolase pathway (dihydroxy-acetone)	Partly engineered: formate [106] Theoretical: CO_2, CO, methanol	-1 – 5[c]	-(?)	Yes	None	Dihydroxy-acetone-phosphate	Linear	Low

(continued)

Table 3 (continued)

Pathway name (variant name)	Naturally present, (partly) engineered or only theoretically proposed per C1-substrate [refs.]	ATP costs/ pyruvate	Kinetic performance based on [81]	Oxygen tolerant	CO_2 concentration dependence	Key product	Topology	Overlap central metabolism
Formolase pathway (SACA)	Partly engineered: methanol [111] Theoretical: CO_2, CO, formate	0–8[c,d]	-(?)	Yes	None	Acetyl-coA	Linear	Low
CETCH cycle (7.0)	Partly engineered (in vitro): CO_2 [114] Theoretical: formate, methanol, CO	6	+	Yes	Low	Glyoxylate	Cyclic	Low
GED cycle	Partially engineered: CO_2 [115] Theoretical: formate, methanol, CO	6	?	Yes	High	Pyruvate	Cyclic	High
MOG cycle (PyrC-alanine)	Theoretical [103]: CO_2, CO, formate, methanol	6	–	Yes	Low	Glyoxylate	Cyclic	High

For each pathway it is indicated if it is known to exist in nature, and for which C1-substrates full growth is observed. Engineered pathways demonstrated full functionality, while supporting growth on the indicated C1-source without any additional carbon source (other than CO_2). For partly engineered pathways, experimental evidence on functioning on most of the pathway enzymes was demonstrated in vivo and/or in vitro, but full growth via the pathway has not yet been realized. Theoretical means that those pathways can function for the indicated substrates on paper or in silico but there is no experimental evidence towards this yet. Note that pathways which are indicated to support grow on methanol, formate or CO still may require additional CO_2, whereas the indication CO_2 means that the pathway can sustain growth on CO_2 as sole carbon source (in the presence of an energy source such as H_2). For several pathways, growth on formate, methanol and CO requires the full oxidation of these reduced C1-donors and re-assimilation of CO_2. The rough indication of kinetic pathway performance (−, +/− or +) is based on estimates of overall pathway specific activities reported by Löwe et al. [81] as calculated using the Enzyme Cost Minimization (ECM) framework [119]. For some pathways, no estimates were or could be made due to exclusion in their analysis of anaerobic pathways and some other engineered pathways; these are marked with "?". Pathways which include the notoriously slow engineered formolase enzyme were estimated to be in the lowest category of kinetic performance (−). If a pathway is not tolerant to oxygen, some oxygen-sensitive (generally ferredoxin-dependent) enzymes are present. CO_2 concentration dependence indicates if the pathway contains carboxylating enzymes, which can function at low CO_2 concentrations (around ambient CO_2: 0.04% in air), or require higher, elevated CO_2 concentrations. Some C1-assimlation pathways do not have carboxylation steps and hence require no CO_2 for pathway operation. They key pathway product is a key molecule coming out of the pathway that is available for further biosynthesis. Pathway abbreviations: CETCH cycle: crotonyl–coenzyme A (CoA)/ethylmalonyl-CoA/hydroxybutyryl-CoA cycle; GED Cycle: 6-phosphogluconate dehydrogenase (Gnd)–Entner–Doudoroff cycle: MOG cycle (PyrC-alanine variant): malonyl-CoA oxaloacetate glyoxylate cycle (pyruvate carboxylase – alanine bypass variant); rAcP: reductive acetyl-CoA pathway; rGlyP: reductive glycine pathway; rTCA cycle: reductive tricarboxylic acid cycle; RuMP cycle (FBA-TA variant): ribulose monophosphate cycle (fructose bisphosphatase/transaldolase variant); SACA pathway: synthetic acetyl-CoA pathway

[a]The ATP consumption value reported here is for the glycine reductase variant with phospho-acetyltransferase, whereas naturally demonstrated variant is less ATP efficient, using acetyl-CoA synthetase

[b]Despite the presence of oxygen-sensitive enzymes in the rTCA cycle some occurrences of the pathway in aerobic are known [108]

[c]The ATP consumption per pyruvate of these formaldehyde-assimilation pathways can be low (even negative) when methanol is used as C1-substrate. The use of formate (including CO_2 reduction to formate) will lead to the higher ATP-cost value, due to the cost of 2 ATP/formate for formyl-CoA synthetase, which can next be reduced to formaldehyde

[d]For ATP consumption, aerobic assimilation of acetyl-CoA is assumed to proceed via the glyoxylate shunt, which leads to the overall reaction of 2 acetyl-CoA = pyruvate + CO_2 + CoA

reaction to be re-used again ("THF cycle") [38, 100]. Similarly the H-protein is recycled during the GCS reaction. However, we nonetheless consider the pathway to be linear since only co-factors rather than intermediates of the pathway are recycled. By contrast, in typical circular pathways such as the Serine cycle and CBB cycle, carbon intermediates must be (partially) regenerated to serve as acceptor molecules.

The rGlyP can furthermore be considered advantageous as it has relatively limited overlap with central metabolism. Although glycine, C1 and THF metabolism are key nodes in most metabolic networks, most other C1 assimilation pathways have more significant overlap with central carbon and energy metabolism. For instance, the overlap of the RuMP cycle and CBB Cycle with glycolysis and the pentose phosphate pathway is likely to have a more drastic impact on metabolic fluxes when they are engineered into a novel host. However, recent breakthroughs in successful engineering in *E. coli* of full circular pathways with high overlap with central metabolism (RuMP and CBB cycle) demonstrate that this challenge can be overcome [88, 107]. Still, the engineering of these pathways consistently required ALE to achieve growth, whereas initial growth on the rGlyP could be realized in *E. coli* and *C. necator* without ALE, supporting the relatively easier implementation of the latter pathway.

A small disadvantage of the rGlyP for biotechnological applications could be its dependence on elevated CO_2 concentrations, which requires CO_2 feeding into the bioreactor. Some natural or synthetic CO_2 fixation and formate-assimilation pathways can also function in ambient CO_2 (Table 3), which could relieve this requirement. In addition, some proposed formate-assimilation pathways do not require CO_2 fixation at all, including the RuMP cycle (which allows growth on formate only if formate to formaldehyde reduction via formyl-CoA is realized), Formolase pathways and Homoserine cycle. However, the Formolase pathways, Homoserine cycle and formate to formaldehyde reduction have been only partly demonstrated so far [106, 109–111]. Overall, the need for high CO_2 is also not a major issue, as the essential oxidation of a large part of the formate for energy production will likely release a surplus of CO_2 in the bioreactor.

Therefore, the rGlyP seems highly promising for aerobic bioproduction. However, so far natural organisms running the CBB cycle or Serine cycle still outperform engineered *E. coli* and *C. necator* with the rGlyP in terms of growth rates and yields on formate [19], demonstrating the need to further engineer and evolve towards superior rGlyP platform strains for formate-based bioproduction.

4.2 Assessing the rGlyP for Growth on Methanol

As discussed above, the rGlyP can also serve as a pathway to support growth on methanol. However, unlike the relatively clear superiority of the rGlyP for formatotrophy, this is not such a clear case for growth on methanol. Several aerobic, natural methanol assimilation pathways exist, i.e. several RuMP cycle variants, the Serine cycle and yeast-specific Xylulose-Monophosphate pathway

[19]. Furthermore, the anaerobic rAcP and CO_2-fixing aerobic CBB cycle support growth on methanol in nature [19, 112]. In addition to these natural pathways, some synthetic pathways for methanol assimilation have been proposed, including some variants to the natural Serine cycle, i.e. the "modified Serine cycle" and the Homoserine cycle, which were proposed and partly engineered to circumvent some toxicity and other issues of the natural Serine cycle [52, 110]. Among these pathways the theoretical biomass yield is predicted to be the highest for (variants of) the RuMP cycle (Table 3) [10, 19]. However, as methanol is a highly reduced substrate, the theoretical yields of some highly oxidized products can suffer from this. To supply enough carbon when methanol is the sole carbon substrate, some methanol must be wastefully oxidized. Hence, methanol assimilation pathways that allow for the co-fixation of CO_2, and which are still highly ATP-efficient, can result in higher yields for specific products. In that case the rGlyP is a very promising candidate [19, 81]. For example, for the production of relatively oxidized products such as lactate the theoretical yield of the rGlyP is higher than that of the RuMP cycle. Regarding kinetic performance, it is suggested that the rGlyP may be among the top-performing methylotrophic pathways and it may rival the RuMP cycle on that aspect [81]. In conclusion, the rGlyP is also interesting to consider for methylotrophy, especially for more oxidized products.

4.3 Assessing the rGlyP for Autotrophic Growth

As observed already in nature, the rGlyP can also serve as an autotrophic CO_2 fixation pathway. This can also be useful in biotechnological applications, especially for the growth on H_2 and CO_2. By now, seven CO_2 fixation pathways are known in nature including the rGlyP [30, 108]. As there are several carboxylase enzymes available in nature, as well as possible engineered carboxylases, a vast space of potential CO_2 fixation pathways can be imagined [103, 113]. For some of these synthetic pathways, initial steps in their engineering have been made, such as the in vitro realization of the CETCH cycle [104, 114] and the in vivo partial implementation of the GED cycle [115]. In terms of ATP efficiency, the rGlyP is again among the most efficient CO_2 fixation pathways, only rivalled by the rAcP and variants of the rTCA cycle. However, as stated above, both of these are anaerobic pathways, even though some (micro)aerobic occurrences of the rTCA cycle are known [108].

Despite the yield potential of the rGlyP, the required carboxylation by reverse operation of FDH could form a thermodynamic and kinetic bottleneck [76, 81]. This leads to a strict requirement for elevated CO_2, as well as potentially adverse kinetics of the rGlyP, when compared, for example, to the CBB cycle and CETCH cycle [81]. So overall, the potential of the rGlyP as an autotrophic pathway may be more limited than for assimilation of the reduced C1-substrates methanol and formate. However, this is highly related to the capacity to reverse the FDH reaction, and for example alternative co-factors (such as NADPH and ferredoxin), may partly

overcome this bottleneck as also some natural acetogens can grow relatively efficiently via the rAcP on H_2/CO_2 involving CO_2 reduction via FDH. So, the implementation of CO_2 reduction by FDH in combination with the rGlyP deserves further attention.

In conclusion, in this section we reviewed the rGlyP in relation to other potential C1-assimilation pathways. This field is rapidly evolving, and we expect the potential pathway list to expand further in the future. Nevertheless, overall we can conclude that several features such as i) low amount of ATP required, ii) little overlap with enzymes of the central metabolism, and iii) a linear pathway structure make the rGlyP an optimal synthetic route to engineer within microbial *chassis* for C1-assimilation within the context of industrial biotechnology.

5 Outlook on Future Directions and Applications of the Reductive Glycine Pathway

In this chapter we provided an overview of the rGlyP variants, the versatility of the rGlyP to support growth via multiple C1-feedstocks, approaches to engineer the rGlyP and the current state in several organisms, as well as a comparison with other C1-assimilation pathways, showcasing the high potential of the rGlyP for bioproduction. Finally, we provide a future outlook on where we expect challenges and opportunities to arise for the implementation of the rGlyP in the coming years.

5.1 Optimizing Inefficiencies in the rGlyP and the Central Metabolic Network

Based on the thermodynamics and experiences in engineering the rGlyP so far, potential bottlenecks in the pathway include the operation of the GCS, NADPH supply, and the FDH energy module.

So far, the engineered rGlyPs in *E. coli* and *C. necator* rely on the reverse activity of their native glycine cleavage systems (GCS) [27, 28]. As this enzyme system naturally functions in the direction of glycine cleavage, the enzyme's properties for glycine biosynthesis might not be optimal. Supporting this point, for the optimization of the C1-module, the use of enzymes which were evolved to operate in the formate assimilation direction from *M. extorquens* (in the Serine Sycle) rather than the native *E. coli* FolD enzyme improved the pathways' performance in *E. coli* [98]. Following a similar logic, it may be beneficial to replace the native GCS with homologous candidates, which naturally operate in the glycine synthesis direction (e.g., from *D. desulfuricans* [30]). Another approach is to optimize the GCS by enzyme engineering, where in vitro data already suggest possible protein

engineering targets [120], which is discussed in more detail in another chapter in this book chapter by Ren et al.

In addition to these approaches tackling the kinetics of the GCS, the thermodynamic properties of glycine biosynthesis should be positively influenced by changing the co-factor specificity of the GCS from NADH to NADPH. In contrast to the $NADH/NAD^+$ redox pair, the $NADPH/NADP^+$ ratio is kept high in the cell to serve as a driver for reductive reactions. If NADPH can be used to shift the thermodynamic equilibrium of the GCS reaction towards glycine synthesis, the necessary CO_2 concentrations could potentially be reduced. The described co-factor change of the GCS might be achievable, especially since the change of the co-factor preference to $NADP^+$ of the *E. coli* dihydrolipoyl dehydrogenase (the L-protein of the GCS) was described already [121] and the $NADP^+$ preference of this enzyme was shown to support NADPH generation in vivo [122].

The next logical step is the use of an $NADP^+$-dependent formate dehydrogenase, which was already attempted by *Bang* et al. [29] but without a clear comparison to a strain harboring only an NAD^+-dependent variant. NADPH is the driver of methenyl-THF reduction by methylene-THF dehydrogenase and the redox currency for biosynthesis of many biomass components of the cell. However, its synthesis from NADH catalyzed by the membrane-bound transhydrogenase comes with the cost of a proton translocation driven by the proton motive force. Also the implementation of faster metal-dependent FDHs may be important to consider, as metal-independent variants, like the one applied now in *E. coli,* likely form a kinetic bottleneck and demand a relatively high expression burden [23].

Outside of the pathway itself and its redox co-factor needs, likely improvements throughout the cellular metabolic network are needed to improve the yields and rates of engineered rGlyP-utilizing strains. This could include, for example, the optimization of gluconeogenic fluxes, as carbon enters the central metabolic network at the level of the lower-glycolytic metabolite pyruvate. Furthermore, the TCA cycle may need to be rebalanced in certain hosts, as the primary role of this cycle to provide NADH in aerobic conditions will be taken over by FDH or MDH generating NADH from formate or methanol. However, in the engineered formatotrophic *E. coli* and *C. necator* strains it was concluded from ^{13}C-labelling patterns that there remains only a relatively low TCA cycle flux.

5.2 Engineering of the rGlyP in More Organisms

The complete rGlyP has so far been completely established in two biotechnologically important bacteria, namely *E. coli* and *C. necator*. While the former has been extensively engineered to make a plethora of compounds [123], the demonstrated product portfolio for *C. necator* is less broad but steadily expanding [124]. Compared to these more established hosts, the organisms found to naturally harbor the rGlyP (*D. desulfuricans* and presumably *C. drakei*) are less well studied in regard to their accessible product spectrum, yields and other performance parameters relevant for bioproduction. Also, they so far lack crucial genetic tools, thus currently

representing suboptimal candidates for bioproduction. In order to extend the efficient conversion of methanol or formate via the rGlyP to a maximal diversity of possible products, the obvious next step therefore is to bring this pathway to other biotechnologically important microorganisms, ideally chosen according to a number of rational criteria.

We propose several features which make a microbial host particularly suitable for implementation of the rGlyP. First, arguably the most important criterion is the organism's ability to generate the desired (class of) products, natively or via previously engineered routes. Secondly, organisms with general, beneficial features for bioproduction (e.g., high tolerance to solvents or salts) may be preferred. Next, implementation of the rGlyP as a heterologous pathway can only be realized in the host if sufficient genetic tools are available to perform genomic editing and express relevant enzymes.

In addition, sufficient genomic and particularly metabolic knowledge is required, especially to enable a stepwise, modular implementation via growth-coupled approaches as described above. Such an implementation of the rGlyP seems particularly feasible if individual enzymes or segments of the pathway already operate in the given host. For instance, the known ability of a host to grow on glycine, glyoxylate or serine as sole carbon sources via one of the routes discussed above (see Sect. 2.2 and Fig. 2) would already provide a functional option for funneling the rGlyP product glycine into central metabolism. Similarly, operation of a fast (metal-dependent) formate dehydrogenase and/or methanol dehydrogenase in (genetically accessible) methylotrophs would make such organisms prime candidates for methanol and formate utilization via the rGlyP, which could outperform their native assimilation routes [16]. Compared to heterotrophic hosts, an organism natively capable of methylotrophic or chemolithoautotrophic growth may be better predisposed for utilization of one-carbon compounds as the sole source of energy and reducing power for driving the rGlyP.

Beyond these criteria, it is economically beneficial if the microbe natively achieves fast growth on minimal media (a likely prerequisite for high production rates) and exhibits a sufficiently high tolerance to formate (and methanol in case this is the desired C1-feedstock), thus enabling higher product titers in the case of batch cultures. In addition, the ability to use minimal (defined) media is highly desired for growth-dependent selection and optimization of rGlyP modules.

From an engineering point of view, we argue that the rGlyP is remarkably compatible with a wide variety of hosts. Its linear and modular structure and the ubiquitous presence of key enzymes in many organisms are expected to facilitate comparably fast establishment of the rGlyP. Furthermore, the lessons learned from previous successful implementations will likely aid the engineering in additional hosts. Introducing the rGlyP to more microorganisms and striving towards a larger variety of products will help to increase the impact and success of formate and methanol bio-economies.

Beyond interesting prokaryotic hosts, eukaryotes such as the yeasts *S. cerevisiae* in which implementation is already ongoing, *Yarrowia lipolytica* and methylotrophic *Pichia pastoris* (*Komagataella phaffii*) represent prime targets for

implementation of the rGlyP given their widespread industrial use and specific metabolic capabilities. In the first type of yeast species implementation is already ongoing [43]. More ambitious approaches could aim to engineer the pathway into higher eukaryotes as well. For example, the rGlyP could be envisioned as a sustainable C1-based production route to synthetic meat analogues in mammalian cell lines, or to increase agricultural yields by supporting carbon fixation in crop plants [125].

5.3 Developing Bioproduction Strains

The linear design of the rGlyP stabilizes it to perturbations brought about by dragging metabolites out for bioproduction. Circular pathways for the assimilation of C1-molecules (e.g., Serine, RuMP and CBB cycles) depend on the regeneration of acceptor metabolites (glycine, ribulose-5-phosphate and ribulose 1,5-bishosphate, respectively), which can also be consumed by biosynthetic pathways, and hence require delicate balancing of fluxes [126]. In contrast, the rGlyP only requires the regeneration of the co-factor THF and the reduced lipoyl moiety of the H-protein, which are typically not consumed for biosynthesis and relatively easily regenerated.

Pyruvate, which is the key output molecule of many rGlyP variants, is also a key precursor for many value-added products. For some products, which are, for example, produced from P-glycerate, from intermediates of upper glycolysis, or from oxaloacetate, alternative variants of the rGlyP may be advantageous. For anaerobic production also the glycine reductase variants and direct conversion of acetyl-CoA to products may be interesting to consider, although this may face some energetic challenges.

So far, the use of the rGlyP for generation of a bioproduct has yet to be shown. After optimization of growth and biomass yield performance of engineered rGlyP strains, a crucial step is to introduce heterologous production pathways for products of interest. Then some new challenges may emerge, such as the robustness of production strains and product yields, rates, and titers. To improve these production parameters this will required typical metabolic engineering workflows, in which potentially also some growth-coupled, modular optimization strategies could be employed for some products which can be coupled to growth. In addition to realizing bioproduction strains, an important challenge is to integrate those strains in bioprocesses that are well-connected to chemical processes producing these feedstocks sustainably from renewable electricity, water, and CO_2, as discussed elsewhere [10]. An efficient integration of the best of chemistry and biology, including the rGlyP, can provide truly sustainable solutions for chemical production.

Acknowledgements The authors are grateful to their mentor Arren Bar-Even. He was supposed to act as an editor and contribute a chapter for this book, which he could sadly not do anymore due to his unexpected, early demise in September 2020. He inspired us and many others to work on the reductive glycine pathway and the C1-bio-economy. In the spirit of Arren we want to further carry his brilliant ideas and work in this research area. NJC acknowledges support from the VENI grant

awarded to him by the Dutch Science Organization (NWO) (VI.Veni.192.156). SY is supported by the NWO-Gravitation Project BaSyC (024.003.019). EO and BD are supported by the German Ministry of Education and Research (BMBF) grant Transformate (033RC023G-). AS, VB, VR, SW, and SL acknowledge funding by the Max Planck Society.

Declaration of Interests The authors have no competing interest to declare.

References

1. Masson-Delmotte V, Zhai P, Pirani A et al (2021) IPCC, 2021: summary for policymakers. In: Climate change 2021: the physical science basis. Contribution of Working Group I to the sixth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
2. Kätelhön A, Meys R, Deutz S et al (2019) Climate change mitigation potential of carbon capture and utilization in the chemical industry. Proc Natl Acad Sci U S A 166:11187–11194. https://doi.org/10.1073/pnas.1821029116
3. Hepburn C, Adlen E, Beddington J et al (2019) The technological and economic prospects for CO_2 utilization and removal. Nature 575:87–97. https://doi.org/10.1038/s41586-019-1681-6
4. Artz J, Mu TE, Thenert K et al (2018) Sustainable conversion of carbon dioxide: an integrated review of catalysis and life cycle assessment. Chem Rev 118:434–504. https://doi.org/10.1021/acs.chemrev.7b00435
5. Blankenship RE, Tiede DM, Barber J et al (2011) Comparing photosynthetic and photovoltaic efficiencies and recognizing the potential for improvement. Science 332:805–809. https://doi.org/10.1126/science.1200165
6. Claassens NJ, Sousa DZ, dos Santos VAPM et al (2016) Harnessing the power of microbial autotrophy. Nat Rev Microbiol 14:692–706. https://doi.org/10.1038/nrmicro.2016.130
7. Leger D, Matassa S, Noor E et al (2021) Photovoltaic-driven microbial protein production can use land and sunlight more efficiently than conventional crops. Proc Natl Acad Sci U S A 118. https://doi.org/10.1073/pnas.2015025118
8. Grim RG, Huang Z, Guarnieri MT et al (2019) Transforming the carbon economy: challenges and opportunities in the convergence of low-cost electricity and reductive CO_2 utilization. Energ Environ Sci 13:472–494. https://doi.org/10.1039/C9EE02410G
9. Satanowski A, Bar-Even A (2020) A one-carbon path for fixing CO_2. EMBO Rep 21:e50273. https://doi.org/10.15252/embr.202050273
10. Claassens NJ, Cotton CAR, 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
11. De LP, Hahn C, Higgins D et al (2019) What would it take for renewably powered electrosynthesis to displace petrochemical processes? Science 364:eaav3506. https://doi.org/10.1126/science.aav3506
12. Clomburg JM, Crumbley AM, Gonzalez R (2017) Industrial biomanufacturing: the future of chemical production. Science 355. https://doi.org/10.1126/science.aag0804
13. Becker J, Wittmann C (2015) Advanced biotechnology: metabolically engineered cells for the bio-based production of chemicals and fuels, materials, and health-care products. Angew Chem Int Ed 54:3328–3350. https://doi.org/10.1002/anie.201409033
14. Li H, Opgenorth PH, Wernick DG et al (2012) Integrated electromicrobial conversion of CO_2 to higher alcohols. Science 335:1596–1596. https://doi.org/10.1126/science.1217643
15. Sydow A, Krieg T, Ulber R, Holtmann D (2017) Growth medium and electrolyte – how to combine the different requirements on the reaction solution in bioelectrochemical systems using *Cupriavidus necator*. Eng Life Sci:781–791. https://doi.org/10.1002/elsc.201600252

16. Stöckl M, Harms S, Dinges I et al (2020) From CO_2 to bioplastic – coupling the electrochemical CO_2 reduction with a microbial product generation by drop-in electrolysis. ChemSusChem 13:4086–4093. https://doi.org/10.1002/cssc.202001235
17. Yishai O, Lindner SN, Gonzalez de la Cruz J et al (2016) The formate bio-economy. Curr Opin Chem Biol 35:1–9. https://doi.org/10.1016/j.cbpa.2016.07.005
18. Smith WA, Burdyny T, Vermaas DA, Geerlings H (2019) Pathways to industrial-scale fuel out of thin air from CO_2 electrolysis. Joule 3:1822–1834. https://doi.org/10.1016/j.joule.2019.07.009
19. Cotton CAR, Claassens NJ, Benito-Vaquerizo 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
20. Bertsch J, Müller V (2015) Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria. Biotechnol Biofuels 8. https://doi.org/10.1186/s13068-015-0393-x
21. Whitaker WB, Sandoval NR, Bennett RK et al (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
22. Heux S, Brautaset T, Vorholt JA et al (2018) Synthetic methylotrophy: past, present, and future. In: Kalyuzhnaya MG, Xing X-H (eds) Methane biocatalysis: paving the way to sustainability. Springer, pp 133–1151
23. Bar-Even A, Noor E, Flamholz A, Milo R (2013) Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. Biochim Biophys Acta Bioenerg 1827:1039–1047. https://doi.org/10.1016/j.bbabio.2012.10.013
24. Waber LJ, Wood HG (1979) Mechanism of acetate synthesis from CO_2 by *Clostridium acidiurici*. J Bacteriol 140:468–478
25. Durrre P, Andreesen JR (1982) Pathway of carbon dioxide reduction to acetate without a net energy requirement in *Clostridium purinolyticum*. FEMS Microbiol Lett 15:51–56
26. Schneeberger A, Frings J, Schink B (1999) Net synthesis of acetate from CO_2 by *Eubacterium acidaminophilum* through the glycine reductase pathway. FEMS Microbiol Lett 177:2–7
27. Kim S, Lindner SN, Aslan S et al (2020) Growth of *E. coli* on formate and methanol via the reductive glycine pathway. Nat Chem Biol 16:538–545. https://doi.org/10.1038/s41589-020-0473-5
28. Claassens NJ, Bordanaba-Florit G, Cotton CAR et al (2020) Replacing the Calvin cycle with the reductive glycine pathway in *Cupriavidus necator*. Metab Eng 62:30–41. https://doi.org/10.1101/2020.03.11.987487
29. Bang J, Hwang CH, Ahn JH et al (2020) Escherichia coli is engineered to grow on CO2 and formic acid. Nat Microbiol 5:1459–1463. https://doi.org/10.1038/s41564-020-00793-9
30. Sánchez-Andrea I, Guedes IAIA, Hornung B et al (2020) The reductive glycine pathway allows autotrophic growth of *Desulfovibrio desulfuricans*. Nat Commun 11:doi.org/10.1038/s41467-020-18906-7 |
31. Claassens NJ (2021) Reductive glycine pathway: a versatile route for one-carbon biotech. Trends Biotechnol 39:327–329. https://doi.org/10.1016/j.tibtech.2021.02.005
32. Mejillano MR, Jahansouz H, Matsunaga TO et al (1989) Formation and utilization of formyl phosphate by N10-formyltetrahydrofolate synthetase: evidence for formyl phosphate as an intermediate in the reaction. Biochemistry 28:5136–5145. https://doi.org/10.1021/bi00438a034
33. 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
34. Christensen KE, MacKenzie RE (2008) Chapter 14 mitochondrial methylenetetrahydrofolate dehydrogenase, methenyltetrahydrofolate cyclohydrolase, and formyltetrahydrofolate synthetases. Vitam Horm 79:393–410. https://doi.org/10.1016/S0083-6729(08)00414-7
35. Dev IK, Harvey RJ (1978) A complex of N5,N10-methylentetrahydrofolate dehydrogenase and N5,N10-methenyltetrahydrofolate cyclohydrolase in *Escherichia coli*. Purification subunit

structure, and allosteric inhibition by N10-formyltetrahydrofolate. J Biol Chem 253:4245–4253. https://doi.org/10.1016/s0021-9258(17)34711-7
36. Tashiro Y, Hirano S, Matson MM et al (2018) Electrical-biological hybrid system for CO2 reduction. Metab Eng 47:211–218. https://doi.org/10.1016/j.ymben.2018.03.015
37. Kikuchi G, Motokawa Y, Yoshida T, Hiraga K (2008) Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. Proc Jpn Acad Ser B 84:246–263. https://doi.org/10.2183/pjab/84.246
38. Hong Y, Ren J, Zhang X et al (2020) Quantitative analysis of glycine related metabolic pathways for one-carbon synthetic biology. Curr Opin Biotechnol 64:70–78. https://doi.org/10.1016/j.copbio.2019.10.001
39. Tibbetts AS, Appling DR (2010) Compartmentalization of mammalian folate-mediated one-carbon metabolism. Annu Rev Nutr 30:57–81. https://doi.org/10.1146/annurev.nutr.012809.104810
40. Hassan F, Syed A, Ahmad B (2014) Folate: metabolism, gene, polymorphisms and the associated diseases. Gene 533:11–20. https://doi.org/10.1016/j.gene.2013.09.063
41. Flamholz A, Noor E, Bar-Even A, Milo R (2012) EQuilibrator – the biochemical thermodynamics calculator. Nucleic Acids Res 40:D770–D775. https://doi.org/10.1093/nar/gkr874
42. Song Y, Lee JS, Shin J et al (2020) Functional cooperation of the glycine synthasereductase and Wood-Ljungdahl pathways for autotrophic growth of *Clostridium drakei*. Proc Natl Acad Sci U S A 117:7516–7523. https://doi.org/10.1073/pnas.1912289117
43. Gonzalez J, Cruz D, Machens F et al (2019) Core catalysis of the reductive glycine pathway demonstrated in yeast. ACS Synth Biol 8:911–917. https://doi.org/10.1021/acssynbio.8b00464
44. Steiert PS, Stauffer LT, Stauffer GV (1990) The lpd gene product functions as the L protein in the *Escherichia coli* glycine cleavage enzyme system. J Bacteriol 172:6142–6144. https://doi.org/10.1128/jb.172.10.6142-6144.1990
45. Hartwich K, Poehlein A, Daniel R (2012) The purine-utilizing bacterium *Clostridium acidurici* 9a: a genome-guided metabolic reconsideration. PLoS One 7. https://doi.org/10.1371/journal.pone.0051662
46. Zhang X, Li M, Xu Y et al (2019) Quantitative study of H protein lipoylation of the glycine cleavage system and a strategy to increase its activity by co-expression of LplA. J Biol Eng 13. https://doi.org/10.1186/s13036-019-0164-5
47. Bar-Even A, Milo R, Noor E, Yishai O (2018) Use of the reductive glycine pathway for generating formatotrophic and autotrophic microorganisms. US Patent 10155933B2
48. Ren J, Zhou L, Wang C et al (2018) An unnatural pathway for efficient 5 – aminolevulinic acid biosynthesis with glycine from glyoxylate based on retrobiosynthetic design. ACS Synth Biol 7:2750–2757. https://doi.org/10.1021/acssynbio.8b00354
49. Claassens NJ, Scarinci G, Fischer A et al (2020) Phosphoglycolate salvage in a chemolithoautotroph using the Calvin cycle. Proc Natl Acad Sci U S A 117:22452–22461. https://doi.org/10.1073/pnas.2012288117
50. Zhao Z, Liu H (2008) A quantum mechanical/molecular mechanical study on the catalysis of the pyridoxal 5′-phosphate-dependent enzyme L-serine dehydratase. J Phys Chem B 112:13091–13100. https://doi.org/10.1021/jp802262m
51. Mehta PK, Hale TI, Christen P (1993) Aminotransferases: demonstration of homology and division into evolutionary subgroups. Eur J Biochem 214:549–561. https://doi.org/10.1111/j.1432-1033.1993.tb17953.x
52. Yu H, Liao JC (2018) A modified serine cycle in *Escherichia coli* coverts methanol and CO2 to two-carbon compounds. Nat Commun 9. https://doi.org/10.1038/s41467-018-06496-4
53. Kleiner D (1985) Bacterial ammonium transport. FEMS Microbiol Rev 32:87–100
54. Kim J, Copley SD (2012) Inhibitory cross-talk upon introduction of a new metabolic pathway into an existing metabolic network. Proc Natl Acad Sci U S A 109:16766–16767. https://doi.org/10.1073/pnas.1208509109

55. Ferguson SJ (2010) ATP synthase: from sequence to ring size to the P/O ratio. Proc Natl Acad Sci U S A 107:16755–16756. https://doi.org/10.1073/pnas.1012260107
56. Hinkle PC (2005) P/O ratios of mitochondrial oxidative phosphorylation. Biochim Biophys Acta Bioenerg 1706:1–11. https://doi.org/10.1016/j.bbabio.2004.09.004
57. Nakamura Y, Tolbert NE (1983) Serine:glyoxylate, alanine:glyoxylate, and glutamate: glyoxylate aminotransferase reactions in peroxisomes from spinach leaves. J Biol Chem 258:7631–7638. https://doi.org/10.1016/s0021-9258(18)32225-7
58. Job V, Marcone GL, Pilone MS, Pollegioni L (2002) Glycine oxidase from *Bacillus subtilis*. J Biol Chem 277:6985–6993. https://doi.org/10.1074/jbc.M111095200
59. Von Borzyskowski LS, Severi F, Krüger K et al (2019) Marine Proteobacteria metabolize glycolate via the β-hydroxyaspartate cycle. Nature 575:500–504. https://doi.org/10.1038/s41586-019-1748-4
60. Kornberg HL, Morris JG (1965) The utilization of glycollate by *Micrococcus Denitrificans*: the beta-hydroxyaspartate pathway. Biochem J 95:577–586
61. Claassens NJ, Sánchez-Andrea I, Sousa DZ, Bar-Even A (2018) Towards sustainable feedstocks: a guide to electron donors for microbial carbon fixation. Curr Opin Biotechnol 50:195–205. https://doi.org/10.1016/j.copbio.2018.01.019
62. Figueroa IA, Barnum TP, Somasekhar PY et al (2017) Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO_2 fixation pathway. Proc Natl Acad Sci 115:E92–E101. https://doi.org/10.1073/pnas.1715549114
63. Maia LB, Moura JJG, Moura I (2015) Molybdenum and tungsten – dependent formate dehydrogenases. J Biol Inorg Chem 20:287–309. https://doi.org/10.1007/s00775-014-1218-2
64. Tishkov VI, Popov VO (2006) Protein engineering of formate dehydrogenase. Biomol Eng 23:89–110. https://doi.org/10.1016/j.bioeng.2006.02.003
65. Sauer U, Canonaco F, Heri S et al (2004) The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. J Biol Chem 279:6613–6619. https://doi.org/10.1074/jbc.M311657200
66. Liliana Calzadiaz-Ramirez CC-T, Stoffel GMM, Lindner SN et al (2020) In vivo selection for formate dehydrogenases with high efficiency and specificity toward $NADP^+$. ACS Catal 10:7512–7525. https://doi.org/10.1021/acscatal.0c01487
67. Jiang H, Chen Q, Pan J et al (2020) Rational engineering of formate dehydrogenase substrate/cofactor affinity for better performance in NADPH regeneration. Appl Biochem Biotechnol 192:530–543. https://doi.org/10.1007/s12010-020-03317-7
68. Rousset M, Liebgott P (2014) Engineering hydrogenases for H2 production: bolts and goals. In: Zannoni D, De Philippis R (eds) Microbial BioEnergy: hydrogen production, Advances in photosynthesis and respiration. Springer, Dordrecht, pp 43–77. https://doi.org/10.1007/978-94-017-8554-9
69. Hong Y, Arbter P, Wang W et al (2020) Introduction of glycine synthase enables uptake of exogenous formate and strongly impacts the metabolism in clostridium pasteurianum. Biotechnol Bioeng 1–15. https://doi.org/10.1002/bit.27658
70. Buckel W, Thauer RK (2013) Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. Biochim Biophys Acta Bioenerg 1827:94–113. https://doi.org/10.1016/j.bbabio.2012.07.002
71. Schuchmann K, Müller V (2014) Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. Nat Rev Microbiol 12:809–821. https://doi.org/10.1038/nrmicro3365
72. Liew FM, Martin ME, Tappel RC et al (2016) Gas fermentation-A flexible platform for commercial scale production of low-carbon-fuels and chemicals from waste and renewable feedstocks. Front Microbiol 7. https://doi.org/10.3389/fmicb.2016.00694
73. Fuchs G (1986) CO_2 fixation in acetogenic bacteria: variations on a theme. FEMS Microbiol Lett 39:181–213. https://doi.org/10.1016/0378-1097(86)90446-5

74. Braakman R, Smith E (2012) The emergence and early evolution of biological carbon-fixation. PLoS Comput Biol 8. https://doi.org/10.1371/journal.pcbi.1002455
75. Cotton CAR, Edlich-Muth C, Bar-Even A (2018) Reinforcing carbon fixation: CO_2 reduction replacing and supporting carboxylation. Curr Opin Biotechnol 49:49–56. https://doi.org/10.1016/j.copbio.2017.07.014
76. Calzadiaz-Ramirez L, Meyer AS (2022) Formate dehydrogenases for CO2 utilization. Curr Opin Biotechnol 73:95–100. https://doi.org/10.1016/j.copbio.2021.07.011
77. Schuchmann K, Mueller V (2013) Direct and reversible hydrogenation of CO2 to formate by a bacterial carbon dioxide reductase. Science 342:1382–1386. https://doi.org/10.1126/science.1244758
78. Yu X, Niks D, Ge X et al (2019) Synthesis of formate from CO_2 gas catalyzed by an O_2-tolerant NAD-dependent formate dehydrogenase and glucose dehydrogenase. Biochemistry 58:1861–1868. https://doi.org/10.1021/acs.biochem.8b01301
79. Maia LB, Fonseca L, Moura I, Moura JJG (2016) Reduction of carbon dioxide by a molybdenum-containing formate dehydrogenase: a kinetic and mechanistic study. J Am Chem Soc 138:8834–8846. https://doi.org/10.1021/jacs.6b03941
80. Leo F, Schwarz FM, Schuchmann K, Müller V (2021) Capture of carbon dioxide and hydrogen by engineered *Escherichia coli*: hydrogen-dependent CO_2 reduction to formate. Appl Microbiol Biotechnol. https://doi.org/10.1007/s00253-021-11463-z
81. Löwe H, Kremling A (2021) In-depth computational analysis of natural and artificial carbon fixation pathways. BioDesign Res. https://doi.org/10.34133/2021/9898316
82. Huang J, Yu Z, Groom J et al (2019) Rare earth element alcohol dehydrogenases widely occur among globally distributed, numerically abundant and environmentally important microbes. ISME J 13:2005–2017. https://doi.org/10.1038/s41396-019-0414-z
83. Chistoserdova L (2011) Modularity of methylotrophy, revisited. Environ Microbiol 13:2603–2622. https://doi.org/10.1111/j.1462-2920.2011.02464.x
84. Yurimoto H, Oku M, Sakai Y (2011) Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. Int J Microbiol 2011. https://doi.org/10.1155/2011/101298
85. Krog A, Heggeset TMB, Mueller JEN et al (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. https://doi.org/10.1371/journal.pone.0059188
86. Meyer F, Keller P, Hartl J et al (2018) Methanol-essential growth of *Escherichia coli*. Nat Commun 9. https://doi.org/10.1038/s41467-018-03937-y
87. Wu TY, Chen CT, Liu JTJ et al (2016) Characterization and evolution of an activator-independent methanol dehydrogenase from *Cupriavidus necator* N-1. Appl Microbiol Biotechnol 100:4969–4983. https://doi.org/10.1007/s00253-016-7320-3
88. Chen FY, Jung H, Tsuei C, Liao JC (2020) Converting *Escherichia coli* to a synthetic methylotroph growing solely on methanol. Cell 182:1–14. https://doi.org/10.1016/j.cell.2020.07.010
89. Keller P, Noor E, Meyer F et al (2020) Methanol-dependent Escherichia coli strains with a complete ribulose monophosphate cycle. Nat Commun 11. https://doi.org/10.1038/s41467-020-19235-5
90. Smith TJ, Nichol T (2018) Engineering soluble methane monooxygenase for biocatalysis. In: Kalyuzhnaya MG, Xing X-H (eds) Methane biocatalysis: paving the way to sustainability. Springer, pp 153–168
91. Kim HJ, Huh J, Kwon YW et al (2019) Biological conversion of methane to methanol through genetic reassembly of native catalytic domains. Nat Catal 2:342–353. https://doi.org/10.1038/s41929-019-0255-1
92. Bennett RK, Dzvova N, Dillon M et al (2021) Expression of soluble methane monooxygenase in *Escherichia coli* enables methane conversion. bioRxiv. https://doi.org/10.1101/2021.08.05.455234

93. Claassens NJ, He H, Bar-even A (2019) Synthetic methanol and formate assimilation via modular engineering and selection. Curr Issues Mol Biol 33:237–248. https://doi.org/10.21775/9781912530045.14
94. Wenk S, Yishai O, Lindner SN (2018) An engineering approach for rewiring microbial metabolism. In: Methods in enzymology1st edn. Elsevier, pp 1–39
95. Sandberg TE, Salazar MJ, Weng LL et al (2019) The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. Metab Eng 56:1–16. https://doi.org/10.1016/j.ymben.2019.08.004
96. Neidhardt FC, Ingraham JL, Schaechter M (1990) Building blocks needed to produce 1g of *E. coli* protoplasm. In: Physiology of the bacterial cell: a molecular approach, pp 134–143
97. Yishai O, Goldbach L, Tenenboim H et al (2017) Engineered assimilation of exogenous and endogenous formate in *Escherichia coli*. ACS Synth Biol 6:1722–1731. https://doi.org/10.1021/acssynbio.7b00086
98. Yishai O, Bouzon M, Döring V, Bar-Even A (2018) In vivo assimilation of one-carbon via a synthetic reductive glycine pathway in *Escherichia coli*. ACS Synth Biol 7:2023–2028. https://doi.org/10.1021/acssynbio.8b00131
99. Antonovsky N, Gleizer S, Noor E et al (2016) Sugar synthesis from CO_2 in *Escherichia coli*. Cell 166:1–11. https://doi.org/10.1016/j.cell.2016.05.064
100. Bang J, Yup S, Lee SY (2018) Assimilation of formic acid and CO2 by engineered *Escherichia coli* equipped with reconstructed one-carbon assimilation pathways. Proc Natl Acad Sci 115:E9271–E9279. https://doi.org/10.1073/pnas.1810386115
101. Song S, Timm S, Lindner SN et al (2020) Expression of formate-tetrahydrofolate ligase did not improve growth but interferes with nitrogen and carbon metabolism of *Synechocystis* sp. PCC 6803. Front Microbiol 11. https://doi.org/10.3389/fmicb.2020.01650
102. Döring V, Darii E, Yishai O et al (2018) Implementation of a reductive route of one-carbon assimilation in *Escherichia coli* through directed evolution. ACS Synth Biol 7:2029–2036. https://doi.org/10.1021/acssynbio.8b00167
103. Bar-Even A, Noor E, Lewis NE, Milo R (2010) Design and analysis of synthetic carbon fixation pathways. Proc Natl Acad Sci U S A 107:8889–8894. https://doi.org/10.1073/pnas.0907176107
104. Schwander T, Schada von Borzyskowski L, Burgener S et al (2016) A synthetic pathway for the fixation of carbon dioxide *in vitro*. Science 354:900–904. https://doi.org/10.1126/science.aah5237
105. Yang X, Yuan Q, Luo H et al (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
106. Siegel JB, Smith AL, Poust S et al (2015) Computational protein design enables a novel one-carbon assimilation pathway. Proc Natl Acad Sci 112:3704–3709. https://doi.org/10.1073/pnas.1500545112
107. Gleizer S, Ben-Nissan R, Bar-on YM et al (2019) Conversion of *Escherichia coli* to generate all biomass carbon from CO_2. Cell 179:1255–1263. https://doi.org/10.1016/j.cell.2019.11.009
108. Berg IA (2011) Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. Appl Environ Microbiol 77:1925–1936. https://doi.org/10.1128/AEM.02473-10
109. Wang J, Anderson K, Wang J et al (2021) Enzyme engineering and *in vivo* testing of a formate-reduction pathway. bioRxiv. https://doi.org/10.1101/2021.02.15.431286
110. He H, Höper R, Dodenhöft M et al (2020) An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in *E. coli*. Metab Eng 60:1–13. https://doi.org/10.1016/j.ymben.2020.03.002
111. Lu X, Liu Y, Yang Y et al (2019) Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design. Nat Commun 10. https://doi.org/10.1038/s41467-019-09095-z

112. Müller V (2019) New horizons in acetogenic conversion of one-carbon substrates and biological hydrogen storage. Trends Biotechnol 37:1344–1354. https://doi.org/10.1016/j.tibtech.2019.05.008
113. Erb TJ (2011) Carboxylases in natural and synthetic microbial pathways. Appl Environ Microbiol 77:8466–8477. https://doi.org/10.1128/AEM.05702-11
114. Miller TE, Beneyton T, Schwander T et al (2020) Light-powered CO_2 fixation in a chloroplast mimic with natural and synthetic parts. Science 368:649–654. https://doi.org/10.1126/science.aaz6802
115. Satanowski A, Dronsella B, Noor E et al (2020) Awakening a latent carbon fixation cycle in *Escherichia coli*. Nat Commun 11. https://doi.org/10.1101/2020.05.18.102244
116. Gassler T, Sauer M, Gasser B et al (2020) The industrial yeast Pichia pastoris is converted from a heterotroph into an autotroph capable of growth on CO2. Nat Biotechnol 38:210–216. https://doi.org/10.1038/s41587-019-0363-0
117. Mall A, Sobotta J, Huber C et al (2018) Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. Science 359:563–567. https://doi.org/10.1126/science.aao2410
118. Nunoura T, Chikaraishi Y, Izaki R et al (2018) A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. Science 563:559–563. https://doi.org/10.1126/science.aao3407
119. Noor E, Flamholz A, Bar-Even A et al (2016) The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. PLoS Comput Biol. https://doi.org/10.1371/journal.pcbi.1005167
120. Zhang H, Li Y, Nie J et al (2020) Structure-based dynamic analysis of the glycine cleavage system suggests key residues for control of a key reaction step. Commun Biol 3. https://doi.org/10.1038/s42003-020-01401-6
121. Bocanegra JA, Scrutton NS, Perham RN (1993) Creation of an NADP-dependent pyruvate dehydrogenase multienzyme complex by protein engineering. Biochemistry 32:2737–2740. https://doi.org/10.1021/bi00062a001
122. Lindner SN, Calzadiaz Ramirez L, Krü JL et al (2018) NADPH-Auxotrophic *E. coli*: a sensor strain for testing *in vivo* regeneration of NADPH. ACS Synth Biol 7:2742–2749. https://doi.org/10.1021/acssynbio.8b00313
123. Pontrelli S, Chiu T, Lan EI, et al (2018) *Escherichia coli* as a host for metabolic engineering. Metab Eng 50:16–46. https://doi.org/10.1016/j.ymben.2018.04.008
124. Panich J, Fong B, Singer SW (2021) Metabolic engineering of *Cupriavidus necator* H16 for sustainable biofuels from CO2. Trends Biotechnol 39:412–424. https://doi.org/10.1016/j.tibtech.2021.01.001
125. Bar-Even A (2018) Daring metabolic designs for enhanced plant carbon fixation. Plant Sci 1–13. https://doi.org/10.1016/j.plantsci.2017.12.007
126. Barenholz U, Davidi D, Reznik E et al (2017) Design principles of autocatalytic cycles constrain enzyme kinetics and force low substrate saturation at flux branch points. Elife 6:1–32. https://doi.org/10.7554/eLife.20667.001