

Research Article

In-Depth Computational Analysis of Natural and Artificial Carbon Fixation Pathways

Hannes Löwe  and Andreas Kremling

Systems Biotechnology, Technical University of Munich, Germany

Correspondence should be addressed to Hannes Löwe; h.loewe@tum.de

Received 6 April 2021; Accepted 2 August 2021; Published 1 September 2021

Copyright © 2021 Hannes Löwe and Andreas Kremling. Exclusive Licensee Nanjing Agricultural University. Distributed under a Creative Commons Attribution License (CC BY 4.0).

In the recent years, engineering new-to-nature CO₂- and C1-fixing metabolic pathways made a leap forward. New, artificial pathways promise higher yields and activity than natural ones like the Calvin-Benson-Bassham (CBB) cycle. The question remains how to best predict their *in vivo* performance and what actually makes one pathway “better” than another. In this context, we explore aerobic carbon fixation pathways by a computational approach and compare them based on their specific activity and yield on methanol, formate, and CO₂/H₂ considering the kinetics and thermodynamics of the reactions. Besides pathways found in nature or implemented in the laboratory, this included two completely new cycles with favorable features: the reductive citramalyl-CoA cycle and the 2-hydroxyglutarate-reverse tricarboxylic acid cycle. A comprehensive kinetic data set was collected for all enzymes of all pathways, and missing kinetic data were sampled with the Parameter Balancing algorithm. Kinetic and thermodynamic data were fed to the Enzyme Cost Minimization algorithm to check for respective inconsistencies and calculate pathway-specific activities. The specific activities of the reductive glycine pathway, the CETCH cycle, and the new reductive citramalyl-CoA cycle were predicted to match the best natural cycles with superior product-substrate yield. However, the CBB cycle performed better in terms of activity compared to the alternative pathways than previously thought. We make an argument that stoichiometric yield is likely not the most important design criterion of the CBB cycle. Still, alternative carbon fixation pathways were paretooptimal for specific activity and product-substrate yield in simulations with C1 substrates and CO₂/H₂ and therefore hold great potential for future applications in Industrial Biotechnology and Synthetic Biology.

1. Introduction

As of 2020, most of the earth’s surface is cultivated by mankind: around 50% of the habitable land is actually agricultural land with 37% of the remaining land being forests (with a decreasing trend) [1]. In order to produce more crop to feed a growing population or to fuel the bioeconomy, it is clear that either natural ecosystems have to be converted to agricultural land or agriculture has to be intensified, both with potentially tremendous downsides for nature. Crop plant areal productivity is finally limited by photosynthetic efficiency that is rather low: only around 2.4-3.7% of the incoming sunlight is converted to a usable product [2], and this is only during the growth season. The low efficiency has been attributed to the suboptimal efficiency of the reductive pentose phosphate cycle, better known as the Calvin-Benson-Bassham cycle (CBB cycle) which is the central carbon fixation pathway in plants, green algae, and cyanobacteria but

also in many chemotrophic and heterotrophic microorganisms. Apart from the cycle’s high ATP demand, its key enzyme, Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (RuBisCO), is a rather slow enzyme (turnover numbers of around 1-10 s⁻¹ [3]) and uses oxygen in a wasteful side reaction. The product of this reaction, phosphoglycolate, has to be recovered by the energetically wasteful photorespiratory reactions. In nature, some carboxylating enzymes can be found that have supposedly better characteristics than RuBisCO [4] which led to the question why evolution developed the dark reaction of photosynthesis the way it is. This has inspired researchers to explore whether the dark reaction of photosynthesis can be augmented to support higher productivity by mainly four mutually nonexclusive approaches: (i) enzyme engineering of RuBisCO to increase its activity [5]; (ii) implementation of CO₂ concentration mechanisms to reduce photorespiration [6, 7]; (iii) replacing C3 with C4 photosynthesis or with completely new carbon fixation

pathways, like the CETCH cycle [8]; and (iv) installing more efficient pathways for the recycling of phosphoglycolate during photorespiration [9–11]. A more detailed overview of the efforts to improve photosynthesis can be found in Bar-Even [12].

A further option to circumvent the limitations of the areal productivity of photosynthesis is to use alternative sources of energy and carbon, like the C1 compounds formic acid and methanol or a mixture of H_2 and CO_2 gas. These electron and carbon sources can be derived from electrolysis of water or from CO_2 which increases the areal productivity compared to crop-based primary production by an order of magnitude [13]. Alternatively, methanol or syngas can be produced from waste or biomass which are abundantly available and do not need additional area. For the assimilation of these compounds, nature also has evolved specialized microorganisms with dedicated pathways like the Serine cycle or the Ribulose Monophosphate (RuMP) cycle that are intrinsically more electron-efficient than the traditional CBB cycle [14]. For biotechnological purposes, it was tried to implant these pathways into *Escherichia coli* with mixed success in the beginning ([15]; H. [16, 17]), but finally, the RuMP cycle could be implemented in *E. coli* to support growth on methanol as the sole carbon and energy source (F. Y. H. [18]). Additionally, multiple optimized pathways for C1-assimilation have been proposed for bioproduction, and their basic working principle could be shown [19–21]. Among them, the reductive glycine pathway (rGlyP) was first proposed as an alternative pathway used for CO_2 fixation in acetogens [22] and later engineered in aerobic hosts by the group of Arren Bar-Even and others. This pathway shines as it is the potentially most efficient aerobic formate-dependent pathway and one of the few pathways to be integrated *in vivo* to support growth with formate as the sole carbon and energy source [23–25]. To acknowledge this important achievement, this work is dedicated to Arren Bar-Even, a visionary in pathway design and metabolic engineering, who passed away far too early in the zenith of his scientific achievements.

Throughout this article, we will call biochemical pathways simply “pathways,” which may be linear biochemical pathways or cycles, for simplification. When comparing different pathways, metabolic routes were usually thought to be better because of the following reasons: (i) a lower demand of ATP to form a product [4, 8, 20], (ii) key enzymes with a higher activity ([26], and [4, 12]), (iii) a higher mean thermodynamic driving force (i.e., free energy ΔG) of the reactions [21, 27, 28], or (iv) a higher affinity for CO_2 or HCO_3^- of the carboxylating enzymes ([26], and [4, 12]). All these criteria have limitations, however, and there hasn’t been a conclusive comparison to our knowledge that systematically took into account the enzyme kinetics apart from the forward rate constant. As more efficient pathways are supposed to have lower energy losses and thus usually operate closer to thermodynamic equilibrium, reversibility of reactions accounts for a higher demand for the respective enzymes. Additionally, kinetic undersaturation will limit a pathway’s activity with RuBisCO’s affinity for CO_2 being a prominent example. Noor et al. developed a method to inte-

grate these additional thermodynamic and kinetic constraints by means of the Enzyme Cost Minimization (ECM) algorithm [29]. This algorithm predicts optimal pathway activities if thermodynamic and kinetic data are available. Missing kinetic parameters can be estimated using the Parameter Balancing algorithm [30] to yield a complete and consistent set of parameters for each reaction.

In this work, we made an attempt to objectively compare natural occurring and artificial pathways considering their specific activity and the predicted yield on different industrially relevant electron and carbon sources with a computational approach. Only oxygen-tolerant pathways were considered since data for most of the anaerobic pathways are scarce and ATP-intensive metabolites cannot be produced without by-product formation [31, 32]. By recombining reactions of existing pathways and by adding reactions that were introduced in recent publications, we present 2 new artificial pathways: the reductive citramalyl-CoA cycle (rCCC) and the 2-hydroxyglutarate-reductive tricarboxylic acid cycle (2-HG-rTCA cycle). Both were designed taking into account common thermodynamic and kinetic bottlenecks of other pathways, resulting in a good trade-off between activity and product-substrate yield. Finally, we compare natural and artificial pathways using the ECM algorithm and draw conclusions about evolution, design principles, and future perspectives of these pathways.

2. Material and Methods

2.1. Pathway Selection and Construction of New Routes. The metabolic pathways that were explored in this study were chosen based on the following criteria:

- (i) The metabolic routes needed to include a reaction that uses CO_2 or a C1 compound (formate, methanol) as a substrate
- (ii) The enzymes constituting the pathways are active under aerobic conditions
- (iii) Sufficient kinetic data were available from the literature
- (iv) Additionally, more common routes that connect isolated pathways were added, like glycolysis or the TCA cycle

To design modified versions or combinations of pathways, we followed an intuitive, manual approach in contrast to the systematic methodology suggested in former studies [4, 33] as we sought new routes aside of the previously proposed ones. This was achieved by providing new reaction stoichiometries, oftentimes side reactions of known enzymes, but also by the theoretical dissection of pathways into functional modules that could be freely combined.

2.2. Data Collection. In order to quantify the pathway-specific activities of CO_2 - and C1-fixing pathways, a comprehensive data set was collected from different sources which was then used for further calculations. Thermodynamic and enzymatic data were extracted from the databases

KEGG [34], eQuilibrator [35], BRENDA [36] and UniProt [37]. The details are listed in Supplementary Table S1 specifying which kinetic data were extracted from each source. Some equilibrium constants of reactions that were critical for certain carbon fixation pathways were adjusted based on enzymatic data as described in Section 3 in the Supplementary Information. For most enzymes, kinetic data exists for enzymes from various organisms under varying conditions. We chose parameters of optimal enzymes (i.e., those that allow for the highest activity) to explore the potential of each pathway. These enzymes could be derived from any organisms in all domains of life. All primary sources were manually curated from original publications since parameters in BRENDA are often stored with wrong units or are only partially extracted. The complete workflow is illustrated in Figure 1. During data collection, kinetic parameter sets with more data for Michaelis constants and those with forward and backward rate constants were preferred over sets with less measured data. The most important parameters which are decisive for a good estimate are the rate constant of the direction the reaction is supposed to operate and the K_m of the primary substrate which is usually the carbon intermediate in the respective pathway.

All kinetic data were brought together in an SBtab model file and SBtab data file as specified in [29, 38]. The templates of the respective SBtab files were provided by the ECM pathway analysis platform on the eQuilibrator website (<http://equilibrator.weizmann.ac.il/pathway/>). An additional column for parameter uncertainties was added, and table headings were slightly modified to match the current implementation of the Metabolic Network Toolbox (<https://github.com/liebermeister/metabolic-network-toolbox>) in MATLAB that was used to process the data. The final SBtab files are included in the Supplementary Information (available here) and also at <https://github.com/hannesloewe/pathway-comparison-ECM>.

2.3. Integration of Quinone-Dependent and Proton-Translocating Reactions. During data collection, it had to be decided which enzymes are used for each reaction as there are cases in which a reaction with identical main substrate and product is catalyzed by different enzymes. The only difference in these cases might be the use of different cofactors or whether the reaction is coupled to ion translocation through the membrane. This was especially important for the electron transport chain. It is known that depending on the organism and the conditions, different variants of the electron transport chain occur [39]. Additionally, different quinones with varying redox potentials might be used. For the analysis of pathways, quinones were treated as cofactors like other electron carriers, e.g., NADH. In the Tricaric Acid cycle (TCA cycle), the direction of the succinate dehydrogenase complex is towards the oxidation of succinate which is usually accompanied by ubiquinone reduction due to the high redox potential of the ubiquinone (UQ)/ubiquinol (UQH₂) redox couple [40]. To poise the reaction in the reverse, reductive direction, bacteria, e.g., *E. coli* [40], usually use quinones with a lower redox potential like the menaqui-

none couple. Irrespective, whether ubiquinone, menaquinone, or even rhodoquinone is used, the oxidation of succinate was reported to be reversible [41–43]. The direction of the reaction is rather determined by the redox potential and ratio of the reduced and oxidized form of the quinone. The use of a soluble NADH-dependent fumarate reductase that is operating in the rTCA cycle of certain bacteria [44] was suggested for the implementation of new pathways by others as an alternative with a higher driving force due to the higher redox potential of NADH compared to quinones [4]. This enzyme was not considered here since the reaction is practically irreversible and hence energetically more wasteful. Instead, we used the data of the succinate dehydrogenase of *Bacillus subtilis* that uses menaquinone as it reversibly works under aerobic conditions and kinetic data were available for forward and backward reaction [42]. In *B. subtilis*, the oxidation of succinate is accompanied by the translocation of 2 protons from the periplasm to the cytosol. To account for the 2 protons that are pumped by the enzyme through the membrane, the equilibrium constant of this reaction was adjusted since the proton gradient was not included in the equilibrium constant taken from eQuilibrator. The change in the equilibrium constant was calculated according to the following formula, assuming a membrane potential $\Delta\psi$ of -150 mV:

$$K'_{\text{eq}} = K_{\text{eq}}^0 e^{zF\Delta\psi/RT}, \quad (1)$$

in which K'_{eq} is the adjusted equilibrium constant, K_{eq}^0 is the original equilibrium constant provided by eQuilibrator, F is the Faraday constant, z is the number of protons transferred from the cytosol to the periplasm, $\Delta\psi$ is the transmembrane potential, R is the general gas constant, and T is the temperature.

Depending on the direction of the proton translocation, the exponent of the exponential function in Equation (1) will have a positive or negative exponent (per definition, z is positive for transport from the cytosol to the periplasm). When the reaction is using succinate as the substrate and fumarate as the product, the equilibrium constant has to be higher than without considering the proton gradient (2 protons from the periplasm are transferred to the cytoplasm), and hence, the exponent is positive. Apart from the succinate dehydrogenase, only one other type reaction with quinones was implemented: acyl-CoA dehydrogenases that desaturate carbon-carbon single bonds, as found, e.g., in β -oxidation. These enzymes were assumed to channel the electrons to the electron transport chain at the level of ubiquinone which is an exergonic reaction under physiological conditions.

2.4. Kinetic and Thermodynamic Parameter Estimation by Parameter Balancing. The SBtab files with the collected data were used as the input for the Parameter Balancing algorithm [30] to estimate the missing kinetic parameters following essentially the approach used in a former study [29]. The theoretical background concerning this algorithm is explained in detail in the supplementary information in Section 4 and in the original publication. Each reaction was balanced alone

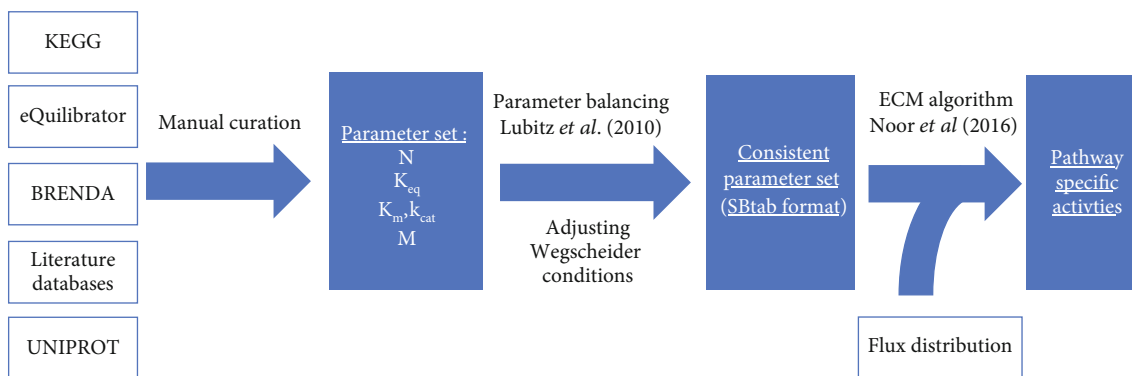


FIGURE 1: Schematic workflow that was applied to calculate the pathway-specific activities from the kinetic parameter set and the flux distributions of each pathway. Box with white-colored filling indicate input data(sources), blue boxes indicate processed data. N : stoichiometric matrix; K_{eq} : equilibrium constants; K_m : Michaelis constants; k_{cat} : forward rate constants; M : molecular weight.

without considering the other reactions in the network using the MATLAB implementation of Parameter Balancing. The equilibrium constants were not adjusted but kept fixed during balancing to ensure thermodynamic consistency across reaction networks. This was done by setting the option “fix_Keq_in_sampling” to 1 in the MATLAB code. After Parameter Balancing, the resulting parameters should fulfil the Haldane relationship which can be formulated as follows [29]:

$$K_{eq} = \frac{k_{cat}^+ \prod_i (K_i^P)^{m_i^P}}{k_{cat}^- \prod_i (K_i^S)^{m_i^S}}, \quad (2)$$

in which K_{eq} is the equilibrium constant, k_{cat} is the forward and backward rate constants, K_i^P is the Michaelis constant of product i , m_i^P is the stoichiometric factor of product i , K_i^S is the Michaelis constant of substrate i , and m_i^S is the stoichiometric factor of substrate i .

For some reactions, the algorithm did not produce consistent sets of parameters because the collected parameter values were contradictory with respect to the equilibrium constants of the reactions. This would typically happen if forward and backward kinetic constants were both available. In these cases, the uncertainties of the measured constants as well as the parameter prior uncertainties of kinetic rate constants and Michaelis constants were iteratively increased and balanced until the Haldane relationships were fulfilled. The Parameter Balancing algorithm did not give estimates only for parameter values but also for their uncertainties which were used later to estimate the error of optimization results. Both parameter values and their uncertainties were stored in data structures for further calculations. After this step, the data structure with the network kinetics contained a complete set of parameters for all reactions and all reactants. The corresponding MATLAB code can be found at <https://github.com/hannesloewe/pathway-comparison-ECM>.

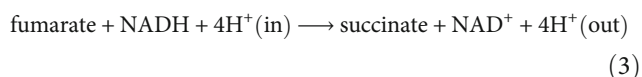
2.5. Determination of Pathway-Specific Activities. To calculate the pathway-specific activity of a selected pathway, the MATLAB implementation of the ECM algorithm was used

[29]. The MATLAB function for ECM needs four arguments: (i) a biochemical network containing the reaction stoichiometries, (ii) the kinetic parameters of the reactions, (iii) a flux distribution representing the selected pathway, and (iv) an option data structure which can be used to specify upper and lower bounds for metabolites and cost weights for enzyme concentrations as well as further options. These four arguments were provided as follows: The reaction stoichiometries (i) were specified in the SBtab model file that was assembled during data collection. The stoichiometries were loaded into a data structure with the Metabolic Network Toolbox. The data structure with the kinetic parameters (ii) were obtained by Parameter Balancing (for details, compare the respective section in the Material and Methods). For each pathway, a flux distribution (iii) was set up as a vector that reflects fluxes in the selected pathway. For the upper and lower bounds of metabolite concentrations (iv), the standard bounds provided by the ECM pathway analysis platform of eQuilibrator were taken and slightly modified with values from the literature to represent a realistic profile of metabolite concentrations (detailed values for metabolite bounds can be found in the Supplementary Information in Table S3). Depending on the analysis, CO_2 was either assumed to equal $10 \mu\text{M}$ which is in the order of air-saturated water or 1 mM to account for scenarios where CO_2 can be actively fed (e.g., in bioreactors). HCO_3^- which is used by many carboxylating enzymes instead of CO_2 was assumed to be approximately in equilibrium with CO_2 at a pH of 7.5. As enzyme cost weights (iv), the molecular weights of the enzymes were taken as it was assumed that bigger enzymes will be costlier to synthesize for the cells [29, 45]. The molecular weights of enzymes are stored in a table in the SBtab model file and were loaded into a data structure in MATLAB.

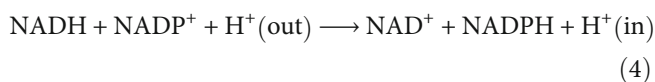
The ECM algorithm returns the minimal enzyme concentrations to support the given flux distribution and thereby offers a measure of the optimal pathway activity per milligram of total enzyme. In addition, the enzyme capacities, reversibility, and kinetic factors [29] (also compare Supplementary Equation S2) are returned by the algorithm. With these factors, it can be evaluated if a certain

reaction is thermodynamically or kinetically limited as described in the section on Parameter Balancing. Uncertainties of pathway-specific activities were estimated by a Monte Carlo method, where in every iteration, the input parameters (i.e., the kinetic parameters) were randomly varied according to their geometric standard deviations assuming a log-normal distribution. The final values that are presented in Results are the geometric mean values and the geometric standard deviations of 100 iterations.

2.6. Calculation of Projected ATP Costs and Product-to-Substrate Yields of Pathways. To compare the pathways in terms of efficiency, a common criterion for all pathways had to be found as they use different cofactors. This was done by calculating the maximal predicted yield of a desired product from a given substrate. To cover the ATP demand of pathways, it was assumed that it is regenerated by the respiratory chain from NADH to yield a total of 2.5 ATP during optimal aerobic respiration [46] which represents an optimistic estimate. In reactions that produce AMP, it was assumed that two equivalents of ATP are required for its regeneration to ATP. FADH₂ which is a cofactor of the succinate dehydrogenase complex is supposed to yield 1.5 ATP when it is fully oxidized with O₂ [46]; the same should be true for electrons derived from the oxidation of FADH₂-dependent acyl-CoA dehydrogenases that are found in β -oxidation. This results in a yield of 1.5 ATP per ubiquinol as ubiquinone is the direct acceptor of electrons from FADH₂. The oxidation of methanol by PQQ-dependent methanol dehydrogenases (MDH) was assumed to yield 1 ATP per methanol by the oxidation of reduced cytochrome *c*_L. The NADH:ubiquinone oxidoreductase complex is supposed to transport 4 protons from the cytosol to the periplasm, equaling around 1 ATP [47]. During fumarate respiration in *E. coli*, the proton translocating NADH dehydrogenase (type I) is used [48] which leads to the assumption that the maximal achievable stoichiometry would be



Assuming a H⁺/ATP ratio of the ATP synthase of 4, this results in the generation of 1 ATP. The difference in the costs of NADPH and NADH depends on the ratio of their reduced and oxidized form. NADPH is usually more reduced than NADH, i.e., the NADPH/NADP⁺ ratio is thought to be a lot higher than the NADH/NAD⁺ ratio in many bacteria [49]. The ratio of the two cofactors can be balanced by the membrane-bound transhydrogenase:



Using this stoichiometry, all needed NADPH could be regenerated from NADH and the cofactors could be poised to the optimal ratios. The regeneration of one NADPH from NADH thus comes at the cost of 0.25 ATP (H⁺/ATP ration of 4). It also has to be stated that different organisms use dif-

ferent respiratory chains [39, 50] and the stoichiometry of ATP synthase may also vary, e.g., in chloroplasts, the H⁺/ATP ratio is 4.67 rather than 4 [51] or even 3.3 [46]. Therefore, the projected ATP costs (and the product-substrate yields) represent rather optimistic estimates. This does not influence the overall comparison between pathways too much, however.

As electron sources, either methanol, formate, or H₂ was used to regenerate NADH in some of the following analyses. Methanol was oxidized via a NADH- or PQQ-dependent dehydrogenase and formaldehyde dehydrogenase. Formate was oxidized by the action of formate dehydrogenase and H₂ via an oxygen-tolerant hydrogenase. These electron-delivering reactions are summed up in Figure S10 in the Supplementary Information. To support formate-dependent pathways with H₂ as a substrate, the formate dehydrogenase was used in the reductive direction with NADPH (X. [52]) assuming that a NADPH-dependent version of the enzyme with equal kinetics can be found in nature or evolved in the laboratory.

3. Results

3.1. Overview of the Natural and Artificial Pathways Used in This Study. Nine original pathways were chosen from the literature and implemented by adding their stoichiometries and relevant kinetic and thermodynamic parameters to the SBTAB model file and SBTAB data file as described in Material and Methods. An overview of these pathways is given in Figure 2. We included three natural occurring, oxygen-insensitive CO₂-fixing pathways: the Calvin-Benson-Bassham cycle (CBB cycle), the 3-hydroxypropionate bicycle (3-HP bicycle), and the 3-hydroxypropionate/4-hydroxybutyrate cycle (3-HP/4-HB cycle). For the 3-HP/4-HB cycle, two different variants have been described [53], both of which were integrated. They were designated as crenarchaeal or thaumarchaeal, respectively. Another CO₂-fixing cycle, the reverse TCA (rTCA) operates in several organisms with an aerobic lifestyle in the genera *Nitrospira*, *Aquifex*, and among the Epsilonproteobacteria. Although the rTCA and the recently discovered roTCA variant [54] are both highly efficient, they were not included in this work because of insufficient kinetic data of the oxygen-tolerant 2-oxoglutarate:ferredoxin oxidoreductase and because it is still unclear how the necessary ferredoxin is generated under aerobic conditions. Additionally, two common C1-fixing pathways were added to the SBTAB model file: the Serine cycle that is known for its efficient assimilation of formate and the Ribulose Monophosphate cycle (RuMP cycle) that uses methanol as the main substrate. The RuMP cycle exists in various variants: we integrated the sedoheptulose-1,7-bisphosphatase variant with either 2-dehydro-3-deoxy-phosphogluconate or fructose-1,6-bisphosphat aldolase. Ribulose-5-phosphate was regenerated with transketolase and not with transaldolase. Furthermore, a NAD-dependent methanol dehydrogenase (MDH) as found, e.g., in the thermophilic *Bacillus methanolicus*, and a PQQ-dependent MDH as found in many mesophilic methanotrophs were used as options in pathway analysis.

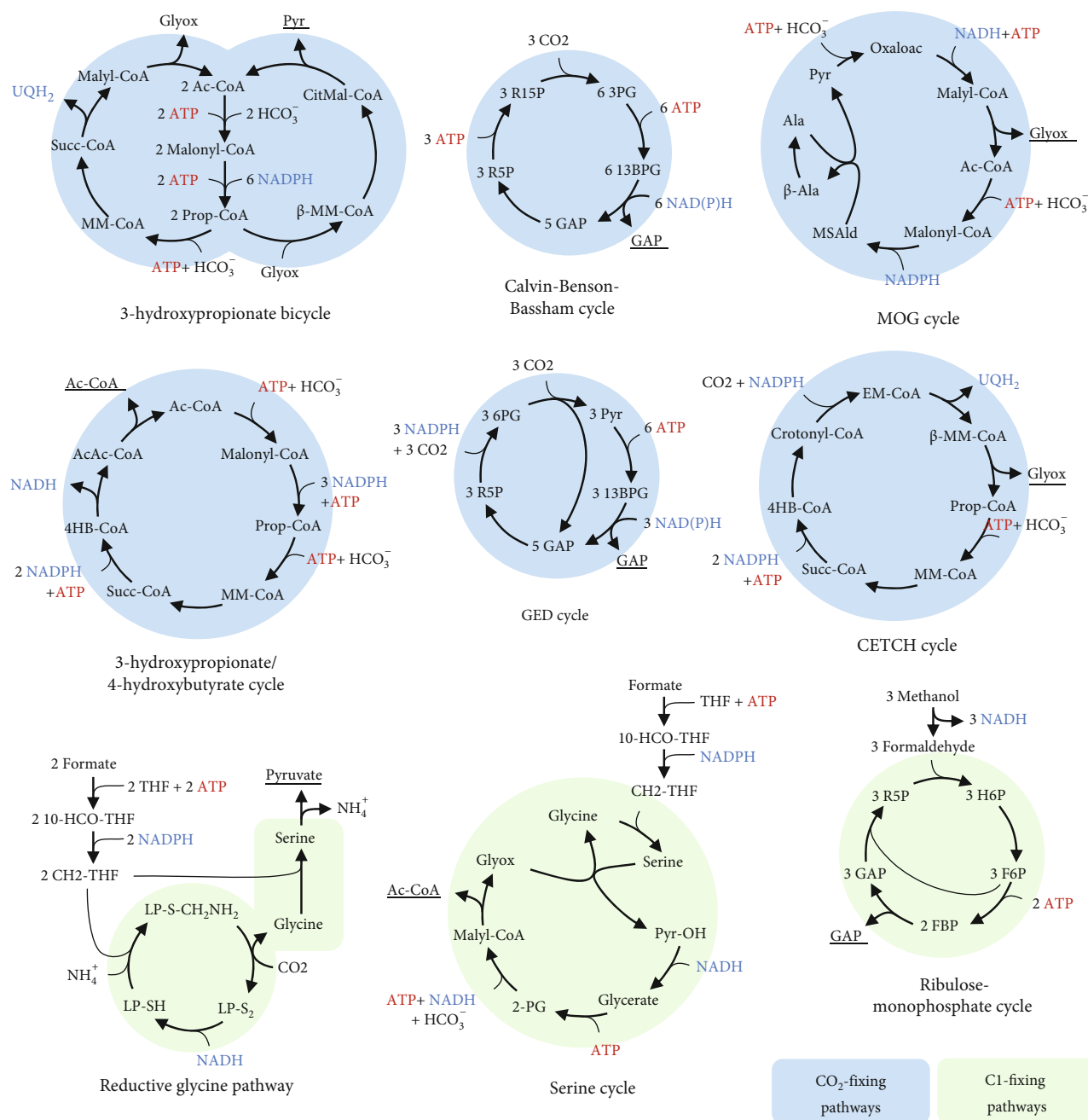


FIGURE 2: Overview of natural occurring and artificial CO₂- and C₁-fixing pathways that were integrated in the model. For simplification, reaction arrows can include multiple reactions and skip some metabolites. ADP, AMP, phosphate, water, and oxidized forms of electron carriers were also left out to improve clarity. Primary products of the pathways are underlined. Abbreviations: Ac-CoA: acetyl-CoA; Prop-CoA: propanoyl-CoA; β-MM-CoA: β-methylmalyl-CoA; CitMal-CoA: citramalyl-CoA; Pyr: pyruvate; MM-CoA: methylmalonyl-CoA; Succ-CoA: succinyl-CoA; Glyox: glyoxylate; 4HB-CoA: 4-hydroxybutyrate; AcAc-CoA: acetoacetyl-CoA; R15P: ribulose-1,5-bisphosphate; 3PG: 3-phosphoglycerate; 13BPG: 1,3-bisphosphoglycerate; GAP: glyceraldehyde-3-phosphate; R5P: ribulose-5-phosphate; H6P: hexulose-6-phosphate; MSAlD: malonate semialdehyde; Ala: alanine; β-Ala: β-alanine; Oxaloac: oxaloacetate; 6-PG: 6-phosphoglycerate; OH-Pyr: hydroxypyruvate; 2-PG: 2-phosphoglycerate; 10-CHO-THF: 10-formyltetrahydrofolate; CH₂-THF: 5,10-methylenetetrahydrofolate; LP-S₂: [glycine-cleavage complex H protein]-N₆-lipoyl-L-lysine; LP-S-CH₂NH₂: [glycine-cleavage complex H protein]-S-aminomethyl-N₆-dihydrolipoyl-L-lysine; LP-SH: [glycine-cleavage complex H protein]-dihydrolipoyl-L-lysine.

Apart from the presented natural pathways, four less common pathways that were described as potential candidates for efficient bioproduction were also included in the SBTab files: the CETCH cycle, the linear reductive glycine

pathway (rGlyP), the Gnd-Entner-Doudoroff cycle (GED cycle) [28], and the aminomutase variant of the malonyl-CoA-oxaloacetate-glyoxylate (MOG) pathways [4]. Although there is a number of possible carbon fixation pathways that

have been proposed or discovered in the past, the four chosen pathways have favorable characteristics for the analysis in this work. All pathways are supposed to work under aerobic conditions, and sufficient kinetic data are available. In addition, the CETCH cycle is one of the few artificial CO₂-fixing pathways that have been implemented experimentally [8]. The rGlyP was engineered *in vivo* to support growth on formate or methanol as the sole carbon and energy source and is also the potentially most efficient aerobic formatotrophic pathway. The GED cycle shows structural resemblance to the CBB cycle with the carboxylation of a C5 sugar and the subsequent cleavage into two C3 bodies, followed by regeneration of the C5 sugar. The key reaction of the cycle, the carboxylation of ribulose-5-phosphate which was characterized well, avoids photorespiration, and the activity of the whole pathway was already shown *in vivo*. The artificial MOG pathways were described as an optimal compromise between kinetics and stoichiometry [4], and therefore, one of its variants (using alanine aminomutase) was used as a benchmark for the predictions as a purely *in silico*-designed pathway. Exotic pathways that rely on high concentrations of formaldehyde [19–21] and pathways that are variations of existing pathways ([55]; H. [17]) were also not considered. The detailed pathway maps that were integrated here, as well as the corresponding metabolites and enzymes, the respective kinetic data, and their literature sources, can be found in the Supplementary Information in Supplementary Figures S1–10 and Table S2.

3.2. Design of New Pathways Related to the Reverse TCA Cycle. In addition to pathways that were described by other authors or found in nature, we also designed two completely new pathways that are insensitive to oxygen and feature a favorable stoichiometry: the reductive citramalyl-CoA cycle (rCCC) and the 2-hydroxyglutarate-reverse TCA cycle (2-HG-rTCA cycle) that are depicted in detail in Figure 3. Both pathways share a common structure with the reverse TCA cycle that is regarded as one of the most efficient pathways for CO₂ fixation [4, 21, 54] but is limited to a very specialized group of microorganisms. The rCCC is a composite of existing pathways: it combines one part of the reverse TCA cycle (from oxaloacetate to succinyl-CoA), one part of the CETCH cycle (from succinyl-CoA to mesaconyl-C1-CoA), and one part of the 3-HP bicycle (from mesaconyl-C1-CoA to pyruvate and acetyl-CoA). The cycle is closed by pyruvate carboxylase, a carboxylating enzyme with high activity that is supposed to be superior to RuBisCO ([4, 26], and [12]). The cycle only needs 3 ATP to produce 1 acetyl-CoA and is thus even more efficient than the thau-marchaeal 3-HP/4-HB cycle that was described as the most ATP-efficient aerobic CO₂-fixing pathway [53] and needs 4 ATP to produce 1 acetyl-CoA. The 2-HG-rTCA cycle, the second pathway that was designed, has an even higher overlap with the reverse TCA cycle (from ketoglutarate to succinyl-CoA in the reductive direction). Hence, the only reaction that is missing from the rTCA cycle is the ferredoxin-dependent 2-oxoglutarate synthase which is also the only oxygen-sensitive enzyme of the cycle. This carboxylating enzyme was replaced in the 2-HG-rTCA cycle by

an enzyme that extends the C4-carboxylic acid not by CO₂ but with formyl-CoA in the following fashion:



This reaction is catalyzed by a 2-hydroxyacyl-CoA lyase, an enzyme that has recently gained attention as an enzyme to assimilate C1 bodies [19, 56]. In their original work, Chou et al. use this enzyme to extend formaldehyde with formyl-CoA. Besides the high toxicity of formaldehyde, the enzyme also has a very high K_m for this substrate, making it less than ideal. Longer chain length aldehydes show lower K_m values which help to limit toxic aldehyde concentrations. We assumed the K_m value and turnover number for succinate semialdehyde to be the geometric mean of propanal and pentanal since succinate semialdehyde has 4 carbon atoms. For future applications, this is probably a conservative estimate since the enzyme was not yet engineered for the shorter chained substrates and is actually supposed to catalyze the reverse reaction. In fact, enzyme engineering was successfully used to reduce the K_m of formaldehyde of this enzyme [56]. So higher affinities and turnover numbers for succinic semialdehyde should be possible. In the 2-HG-rTCA cycle, 2-hydroxyglutaryl-CoA is then supposed to transfer its CoA group to an acceptor carboxylic acid. Here, it was assumed to be transferred to succinate. Considering the kinetics, it was estimated that similar kinetic parameters as determined for succinyl-CoA:malate CoA-transferase can be achieved. In a recent study, it could be demonstrated that highly active 2-hydroxyacyl-CoA transferases can be found in nature [57], so from a design perspective, this should not be a problematic reaction step. Next, 2-hydroxyglutarate is oxidized to 2-ketoglutarate. This thermodynamically difficult step (at least with NAD⁺ as the oxidizing agent) is catalyzed by the lactate-malate transhydrogenase that also accepts 2-hydroxyglutarate as a substrate, with a lower specificity and rate, however. Again, this is a conservative estimation as the enzyme has not yet been optimized for the substrates. From 2-ketoglutarate on, the pathway follows the traditional reverse TCA cycle.

The detailed pathway maps, as well as the corresponding enzymes, the respective kinetic data, and their sources can be found in the Supplementary Information in Figure S4 and Table S2. An overview on the overall stoichiometries of all pathways is listed in Tables S4 and S5 in the Supplementary Information.

3.3. Design of a New Phosphatase-Less Calvin-Benson-Bassham Cycle Variant. As a thought experiment to analyze and to understand the natural design of the CBB cycle, a hypothetically augmented version of the pathway was designed that uses 2 mol ATP less for the production of 1 mol C3 sugar. This is achieved by replacing the irreversible phosphatase reactions with phosphotransferase reactions in the following fashion:

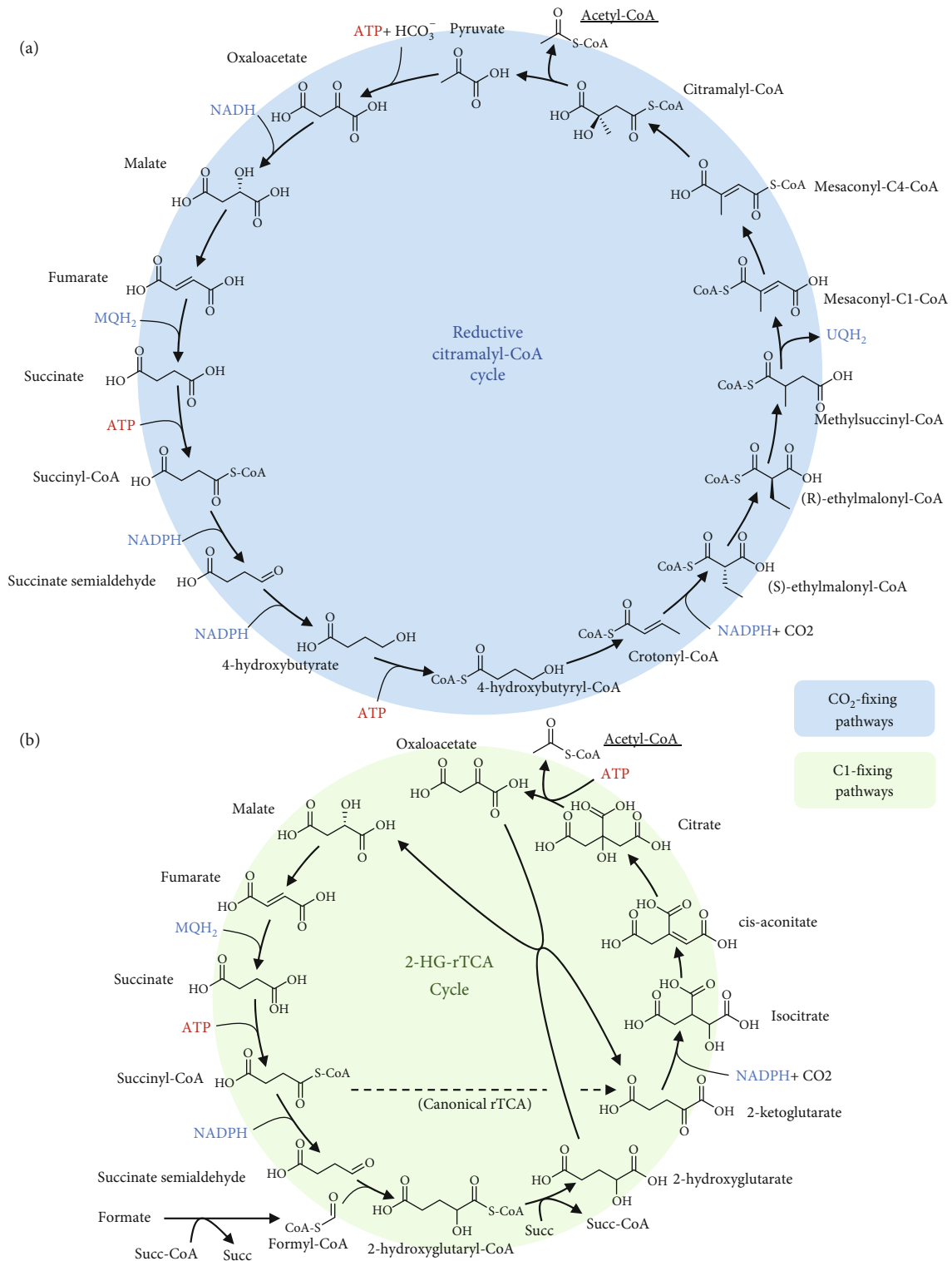
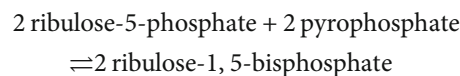
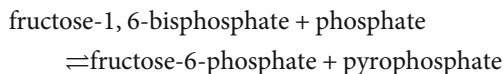
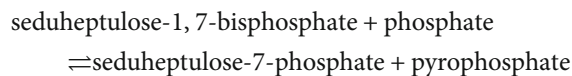


FIGURE 3: Overview of (a) the reductive citramalyl-CoA cycle (rCCC) and (b) the 2-hydroxyglutarate-reverse TCA cycle (2-HG-rTCA cycle) that were designed in this work. For simplification, reaction arrows can include multiple reactions and skip some metabolites. ADP, AMP, phosphate, water, and oxidized forms of electron carriers were also left out to improve clarity. Abbreviations: Succ-CoA: succinyl-CoA; Succ: succinate.



(6)

This net phosphate transfers are exergonic under standard conditions with $\Delta_r G'^{\circ} = -11 \pm 4 \text{ kJ/mol}$ for the transfer of phosphate from seduheptulose-1,7-bisphosphate to ribulose-5-phosphate and with $\Delta_r G'^{\circ} = -4 \pm 5 \text{ kJ/mol}$ for transfer of the phosphate from fructose-bisphosphate to ribulose-5-phosphate. While the phosphotransferase reaction with fructose-1,6-bisphosphate is known to operate in plants and even replaces the fructose-1,6-bisphosphate phosphatase in some organisms [58], the phosphorylation of ribulose-5-phosphate has not been observed to our best knowledge. A simplified pathway map of this imaginary cycle is illustrated in Figure 4, and a more detailed map specifying the enzymes for each reaction can be found in the Supplementary Information in Figure S5.

Enzymes were found that are able to catalyze the described phosphate transfer reactions via pyrophosphate [59] in a promiscuous fashion. Nevertheless, enzymes that can directly transfer the phosphate group from one sugar to the next might also appear in nature and thus might avoid possible regulatory problems associated with the use of pyrophosphate as a substrate.

3.4. Connecting Products of Pathways with Each Other in a Modular Fashion. The primary products of the pathways presented in the last sections are different for each pathway when looking at the illustrated stoichiometries (Figures 2–4):

- (i) The 3-HP/4-HB cycle, the rCCC, the 2-HG-rTCA cycle, and the Serine cycle produce acetyl-CoA
- (ii) The CBB cycle and RuMP cycle produce glyceraldehyde-3-phosphate
- (iii) The 3-HP bicycle and rGlyP produce pyruvate
- (iv) The CETCH cycle produces glyoxylate as the primary product

This is a challenge for the comparison of the pathways as it is unclear how the “value” of a produced metabolite relates to another. To solve this, pathways were added to the SBtab model file that connect the primary products with each other. These additional connecting pathways will be called “connecting modules” as they can modularly be combined with the original primary pathways. The connecting modules that were implemented are described in the following sections and summed up with their net stoichiometries in Table 1. Again, the respective reactions, their stoichiome-

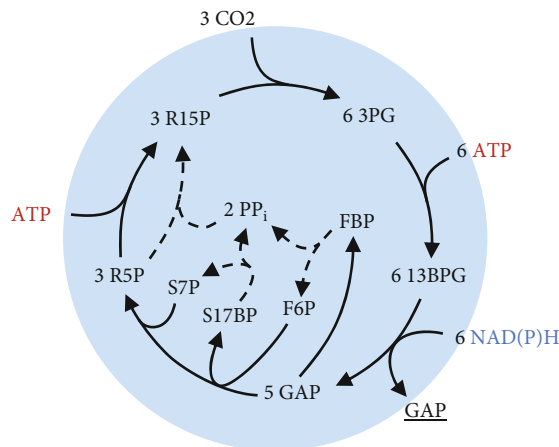
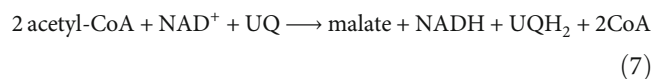


FIGURE 4: Overview of the phosphatase-less Calvin-Benson-Bassham cycle variant that was designed in this work. Dashed lines represent the phosphotransferase reactions that were added which replace the phosphatase reaction and transfer 2 phosphate groups to ribulose-5-phosphate (R5P). For simplification, reaction arrows can include multiple reactions and skip some metabolites. ADP, AMP, phosphate, water, and oxidized forms of electron carriers were also left out to improve clarity. Abbreviations: R15P: ribulose-1,5-bisphosphate; 3PG: 3-phosphoglycerate; 13BPG: 1,3-bisphosphoglycerate; GAP: glyceraldehyde-3-phosphate; FBP: fructose-1,6-bisphosphate; F6P: fructose-6-phosphate; S17BP: seduheptulose-1,7-bisphosphate; PP_i: pyrophosphate; S7P: seduheptulose-7-phosphate; R5P: ribulose-5-phosphate.

tries, and kinetic parameters were added to the SBtab model and data files. Some of these connections are obvious at first: pyruvate can be turned into glyceraldehyde-3-phosphate via gluconeogenesis, for instance. Glyceraldehyde-3-phosphate can inversely be converted back to pyruvate or oxaloacetate via the last steps of glycolysis. For other primary pathway products, the situation is more difficult, however: assimilation of acetyl-CoA to C4 dicarboxylic acids is done by many organisms with the glyoxylate cycle that follows stoichiometry:



where UQ is ubiquinone and UQH₂ is ubiquinol.

This pathway is oxidative (i.e., electrons are freed) which is contrary to the reductive nature of carbon fixation. Besides, for gluconeogenesis, two equivalents of CO₂ are produced per molecule of formed C₆ sugar using the glyoxylate cycle. This is why some bacteria use different pathways for the assimilation of acetyl-CoA that are reductive and even fix CO₂, like the ethylmalonyl-CoA pathway [60]. Archaea with the 3-HP/4-HB cycle can turn acetyl-CoA to succinyl-CoA with the 3-HP route that is an intrinsic part of their cycle [61]. In case of the rCCC, a part of the pathway can be used for the conversion of acetyl-CoA to pyruvate if 3 reactions from the ethylmalonyl-CoA pathway are added as showcased in Figure 5. This pathway module does not need ATP. Following the calculation of projected ATP costs

TABLE 1: Modules connecting pathways that were implemented in the SBtab model file.

Originating pathway ^e	Stoichiometry (main substrates underlined, main products bold)	$\Delta_r G'^0$ (kJ/mol)
<i>Glyoxylate-converting modules</i>		
β -Hydroxyaspartate cycle	$2 \text{ glyoxylate} + \text{NADH} \rightarrow \text{oxaloacetate} + \text{NAD}^+$	-71 ± 6
3-HP-bicycle (modified)	$\text{glyoxylate} + 3\text{NADPH} + 2\text{HCO}_3^- + \text{ATP} \rightarrow$ $\rightarrow \text{oxaloacetate} + 3\text{NADP}^+ + \text{ADP} + \text{P}_i$	-28 ± 13
Reverse glyoxylate shunt	$\text{Glyoxylate} + \text{ATP} + \text{MQH}_2 + \text{NADH} + \text{CoA} + 2\text{H}^+_{\text{in}} \rightarrow$ $\rightarrow \text{acetyl-CoA} + \text{ADP} + \text{P}_i + \text{MQ} + \text{NAD}^+ + 2\text{H}^+_{\text{out}}$	-15 ± 7^a
Serine cycle	$\text{Glyoxylate} + \text{formate} + \text{NADH} + \text{NADPH} + \text{HCO}_3^- + 2\text{ATP} \rightarrow$ $\rightarrow \text{oxaloacetate} + \text{NAD}^+ + \text{NADP}^+ + 2\text{ADP} + 2\text{P}_i$	-76 ± 8
Glyoxylate carboligase/glycerate dehydrogenase	$2 \text{ glyoxylate} + \text{NADH} \rightarrow \text{glycerate} + \text{NAD}^+ + \text{CO}_2$	-46 ± 8
<i>Acetyl-CoA-converting modules</i>		
Ethylmalonyl-CoA pathway (modified)	$2 \text{ acetyl-CoA} + 2\text{NADPH} + 0.5\text{NAD}^+ + \text{UQ} + \text{MQ} + \text{HCO}_3^- + \text{CO}_2 + 2\text{H}^+_{\text{out}} \rightarrow$ $\rightarrow 1.5$ $\text{oxaloacetate} + 2\text{NADP}^+ + 0.5\text{NADH} + \text{UQH}_2 + \text{MQH}_2 + 2\text{CoA} + 2\text{H}^+_{\text{in}}$	-65 ± 13^a
3-HP/4-HB cycle (modified) ^b	$\text{Acetyl-CoA} + 3\text{ATP} + 3\text{NADPH} + \text{NAD}^+ + \text{MQ} + 2\text{HCO}_3^- + 2\text{H}^+_{\text{out}} \rightarrow$ $\rightarrow \text{oxaloacetate} + 3\text{ADP} + 3\text{P}_i + 3\text{NADP}^+ + \text{NADH} + \text{MQH}_2 + \text{CoA} + 2\text{H}^+_{\text{in}}$	-101 ± 14^a
Reductive citramalyl/ethylmalonyl-CoA pathway	$\text{Acetyl-CoA} + \text{HCO}_3^- + 2\text{NADPH} + \text{UQ} \rightarrow$ $\rightarrow \text{pyruvate} + 2\text{NADP}^+ + \text{UQH}_2 + \text{CoA}$	-39 ± 9
<i>Pyruvate-converting modules^c</i>		
4-HB pathway ^b (+pyruvate carboxylase)	$\text{Pyruvate} + 2\text{CoA} + 3\text{ATP} + \text{HCO}_3^- + \text{MQH}_2 + 2\text{NADPH} + 2\text{H}^+_{\text{in}} \rightarrow$ $\rightarrow 2 \text{ acetyl-CoA} + 3\text{ADP} + 3\text{P}_i + \text{MQ} + 2\text{NADP}^+ + 2\text{H}^+_{\text{out}}$	-42 ± 10^a
MCG-like cycle ^d (+pyruvate carboxylase)	$\text{Pyruvate} + 3\text{ATP} + \text{HCO}_3^- + 2\text{CoA} + 3\text{NADH} \rightarrow$ $\rightarrow 2 \text{ acetyl-CoA} + 3\text{ADP} + 3\text{P}_i + 3\text{NAD}^+$	-115 ± 8
<i>General modules</i>		
Glycolysis (starting from GAP)	$\text{Glyceraldehyde-3-phosphate} + \text{ADP} + \text{NAD(P)}^+ + \text{HCO}_3^- \rightarrow$ $\rightarrow \text{oxaloacetate} + \text{NAD(P)H} + \text{ATP}$	-49 ± 6
Gluconeogenesis (starting from oxaloacetate)	$\text{Oxaloacetate} + 2\text{ATP} + \text{NADH} \rightarrow$ $\rightarrow \text{glyceraldehyde-3-phosphate} + 2\text{ADP} + \text{P}_i + \text{NAD}^+ + \text{CO}_2$	28 ± 6

^aThe free energy was calculated assuming a transmembrane potential of -150 mV. ^bThis module exists in two variants (like the pathway it is derived from), depending on whether an AMP- or ADP-producing 3-hydroxypropyl-CoA/4-hydroxybutyryl-CoA synthase is used or not. In the here presented form, it corresponds to the crenarchaeal form. Both variants were considered, however, for the following analyses. ^cIn addition to the modules listed here, single reactions such as the pyruvate dehydrogenase complex or the pyruvate carboxylase were also used to connect metabolites. ^dThis cycle has the same working principle as the malyl-CoA-glycerate cycle (MCG cycle) presented in Yu et al. [66]. Instead of turning 2 glyoxylate into glycerate which is then converted to oxaloacetate, this cycle directly uses the β -hydroxyaspartate cycle to turn 2 glyoxylate into oxaloacetate. A detailed pathway map can be found in Supplementary Figure S8B. ^eOriginal pathway in this context relates to the pathway in which the respective enzymes are usually found. Many of the connecting modules are subnetworks of the full, original pathways.

presented in Material and Methods, the net transfer of 2 electrons from ubiquinone to NADPH comes at a projected cost of 1.25 ATP, however. This is stoichiometrically on par with the most efficient modules introduced by Bar-Even et al. [4] in their seminal work almost a decade ago as part of the MOG pathways albeit using more reactions. In contrast to their approach that uses an aminomutase or acryloyl-CoA hydratase that is more or less oxygen sensitive and features low specific activity, the design presented in Figure 5 is assembled purely from oxygen-insensitive reactions that appear in actual carbon fixation pathways in nature.

For the conversion of glyoxylate to other metabolites, microorganisms employ different strategies: glyoxylate can be turned to glycine and then sequestered by the glycine

cleavage system to finally yield glycerate [62]; alternatively, 2 mol of glyoxylate can be converted to 1 mol of glycerate and CO_2 with tartronate-semialdehyde as an intermediate [4, 63] by tartronate-semialdehyde synthase and glycerate dehydrogenase. Recently, it was found that glyoxylate can also be sequestered to oxaloacetate by the β -hydroxyaspartate cycle (Schada [64]). Both glycerate and oxaloacetate can easily be converted to phosphoenolpyruvate at the cost of one ATP, but glycerate needs one ATP to react to oxaloacetate (via 2-phosphoglycerate). Here, the β -hydroxyaspartate cycle and the pathway using tartronate-semialdehyde synthase were both implemented in the SBtab model file. The reverse glyoxylate shunt introduced by Mainguet et al. [65] was also implemented as a module that turns glyoxylate to acetyl-CoA.

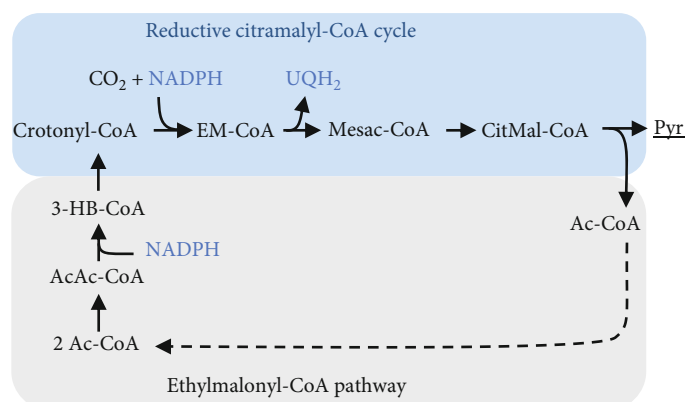


FIGURE 5: Design of a pathway to convert acetyl-CoA (Ac-CoA) to pyruvate (Pyr) using parts of the reductive citramalyl-CoA cycle (blue) and the ethylmalonyl-CoA pathway (grey). The dashed arrow denotes the possible recycling of acetyl-CoA for the first reaction of the ethylmalonyl-CoA pathway. Abbreviations: EM-CoA: ethylmalonyl-CoA; Mesac-CoA: mesaconyl-CoA; CitMal-CoA: citramalyl-CoA; Ac-CoA: acetyl-CoA; Pyr: pyruvate; AcAc-CoA: acetoacetyl-CoA; 3-HB-CoA: 3-hydroxybutyryl-CoA.

Another interesting case is the conversion of pyruvate or C4 dicarboxylic acids to acetyl-CoA. Plants use the pyruvate dehydrogenase complex, an oxidative decarboxylation reaction, which is in contrast to the reductive carbon fixing CBB cycle. To capitalize on this potential, it was proposed to rather split a C4 body into two C2 bodies, e.g., in the synthetic malyl-CoA-glycerate carbon fixation pathway (MCG pathway) (H. [66]). In this pathway, malate is split into acetyl-CoA and glyoxylate. Glyoxylate is then fixed to glycerate as described earlier. Finally, malate can be regenerated from glycerate. We implemented a variant of this pathway in the SBtab model file that uses the β -hydroxyaspartate cycle for glyoxylate assimilation instead which results in a cycle that needs 1 mol of ATP less than the original pathway per 1 mol of acetyl-CoA. A detailed map of this MCG-like pathway can be found in the Supplementary Information in Figure S8B. Nature also has invented a way to convert C4 dicarboxylic acids to 2 acetyl-CoA as part of the 3-HP/4-HB cycle (compare Figure 2), starting from succinyl-CoA and ending in two molecules of acetyl-CoA. Both pathway modules were added to the SBtab model file. Detailed pathway maps of the connecting modules with the respective enzymes can be found in Supplementary Figures S6-S10.

3.5. General Comparison of CO₂ Pathways at Atmospheric CO₂ Concentration. Many pathway designs implemented in this work bear the potential to be superior to the pathways that are widespread in nature—at least in certain scenarios, like biotechnological applications. But how can we actually predict whether or not these pathways are feasible or even better apart from their stoichiometry? To rule out that pathways are thermodynamically and kinetically limited or even infeasible, the ECM algorithm was used which is able to factor in the thermodynamics and kinetics of the reactions based on all parameters that were collected. After assembly of all pathways and connecting modules, the SBtab model file comprised 113 reactions with 104 metabolites and 615 enzymatic parameters (before Parameter Balancing). The data file was then processed by Parameter Balancing to yield

a complete and consistent parameter set for all reactions. Next, we evaluated the pathway-specific activities of all pathways with ECM towards the production of a certain metabolite.

First, only the CO₂-fixing pathways were compared for the production of glyceraldehyde-3-phosphate at atmospheric CO₂ partial pressure as shown in Figure 6 in order to reflect the conditions found in nature without an external CO₂ source like hydrothermal vents. The flux distributions that were assumed for each pathway are defined by the pathway stoichiometries combined with the connecting modules, as listed on the left side of Figure 6 and detailed in the Supplementary Information (available here). The source of reducing equivalents (mainly NAD(P)H) was not considered in the beginning to get a picture of the capacity and limitations of the core pathways without interference of redox metabolism. The pathway-specific activity of the CBB cycle variants was in the order of 0.25 $\mu\text{mol min}^{-1} \text{mg}^{-1}$ which is in line with former estimates [4]. The phosphatase-less CBB cycle (CBB(PTS)) exhibited a slightly lower activity compared to the canonical CBB cycle while the GED cycle only reached around 0.04 $\mu\text{mol min}^{-1} \text{mg}^{-1}$ which makes sense given the low CO₂ partial pressure. Most pathways were predicted to have pathway activities in the same order of magnitude; some, however, were considerably slower: the MOG cycle was predicted to have only marginal activity which is in sharp contrast to the former predictions [4]. Even though the kinetic data of an oxygen-tolerant aminomutase was taken using its optimal substrate (as data for alanine is sparse), the kinetics and thermodynamics of both the aminomutase and the aminotransferase reaction are so unfavorable that the cycle is severely limited by them. Methods of enzyme engineering might solve these shortcomings in the future, however. Comparing the CO₂-fixing pathways, the rCCC, the CETCH cycle, and the crenarchaeal 3-HB/4-HB cycle showed pathway-specific activities equal to the CBB cycle with 10% of photorespiration.

Glyceraldehyde-3-phosphate is the main product of the CBB cycle which is why this pathway was likely favored when choosing it for comparison. Therefore, the same

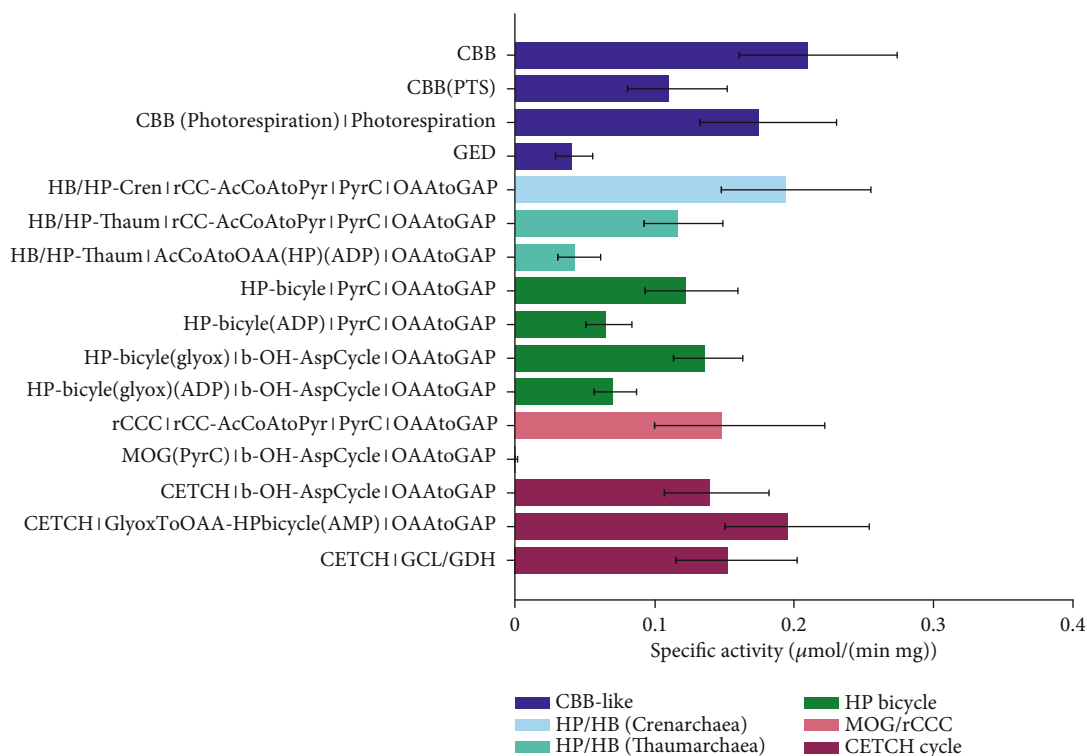


FIGURE 6: Comparison of the specific activity of natural and artificial carbon fixation pathways with the Enzyme Cost Minimization algorithm. The final product of choice was glyceraldehyde-3-phosphate in this case. The concentrations of CO_2 and HCO_3^- were assumed to not exceed $10 \mu\text{M}$ and $100 \mu\text{M}$, respectively, corresponding approximately to air saturation. Pathway-specific activities with standard deviations and the full name of the main pathways and the connecting modules to transform their primary product to glyceraldehyde-3-phosphate. Pathway abbreviations: CBB(PTS): phosphatase-free CBB cycle; HP-bicycle(glyox): glyoxylate-producing subcycle of the 3-HP bicycle; PyrC: pyruvate carboxylase; OAAtoGAP: gluconeogenesis; b-OH-AspCycle: β -hydroxyaspartate cycle; AcCoAtoOAA(HP): acetyl-CoA to oxaloacetate-converting module derived from the 3-HP/4-HB cycle using either ADP- or AMP-producing CoA-ester synthases; rCC-AcCoAtoPyr: acetyl-CoA to pyruvate-converting module derived from the rCCC; GlyoxToOAA: glyoxylate- to oxaloacetate-converting module based on the Serine cycle; GlyoxToGAP(GCL/GDH): glyoxylate conversion by glyoxylate carboligase and glycerate dehydrogenase. The overall stoichiometries of all pathways are listed in supplementary Table S5.

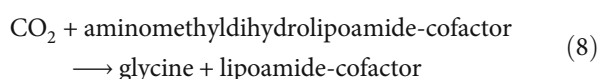
analysis was repeated using acetyl-CoA as the product of choice which is shown in Figure S12 in the Supplementary Information. As expected, the acetyl-CoA-producing pathways performed better for the production of acetyl-CoA in this case compared to the glyceraldehyde-3-phosphate-producing ones. The general trend of pathway-specific activities is not influenced by the choice of product, however: again, the rCCC and the CETCH cycle are on par with the CBB cycle in terms of activity. For all pathways that produce C3 or C4 bodies, different connecting modules were tested that were able to generate 2 acetyl-CoA molecules from one C3 or C4. Among them, the route via 4-hydroxybutyrate from the Thaumarchaeota showed the highest activity and yield. Since the CBB cycle is mainly limited by RuBisCO, lowering the CO_2 concentration has a dramatic impact on its pathway-specific activity which is less of a concern in alternative CO_2 -fixing pathways.

3.6. Evaluation of the Potential of Aerobic CO_2 - and C1-Fixing Pathways for Biotechnological Applications. In most biotechnological scenarios, bioreactors can be gasified with CO_2 , and thus, respective limitations and also photorespiration can be avoided to a certain degree which will likely have

a big influence on the resulting activities. Therefore, three different electron and carbon sources that are regarded as possible platform substrates for the future bioeconomy were evaluated as feedstock for the pathways: methanol, formate, and a mixture of CO_2 and H_2 . The concentrations of CO_2 and carbonate were increased by a factor of 100 to simulate active gassing with CO_2 .

To use methanol as an electron source, MDH was employed in this study. According to the calculations performed here, the RuMP cycle is heavily limited by the NAD-dependent MDH reaction which is thermodynamically and kinetically very unfavorable. Also, the forward rate constant at a temperature of 37°C was taken which is a lot lower than the temperature of its native host [67]. Apart from the unfavorable kinetics of the MDH, the RuMP cycle using NAD-dependent MDH was only thermodynamically feasible if the upper and lower bounds for NAD/NADH and NADP/NADPH were extended further by a factor of 10. It remains enigmatic how microorganisms like *Bacillus methanolicus* use this enzyme at a high specific rate, but the elevated temperature requirement might play a role here. Alternatively, the enzyme kinetics and maybe even the thermodynamics could be fundamentally different *in vivo* since,

for engineered *E. coli* with a functional RuMP cycle, it was speculated that formaldehyde was accumulating, and thus, the MDH was not the limiting factor (F. Y. H. [18]). The PQQ-dependent MDH catalyzes an exergonic reaction, and thus, the pathways with this enzyme are in general predicted to have a higher specific activity, but with a lower product yield as the electrons are directly channeled to the respiratory chain. Indeed, all pathways using the NAD-dependent MDH were limited by the methanol oxidation reaction rather than by the enzymes of the pathways themselves in our calculations. Therefore, only the results with the PQQ-dependent MDH are shown in Figure 7. The NAD-dependent specific activities and product-substrate yields can be found in the supplement in Figure S13, though. Considering the calculations with the PQQ-dependent MDH, the rGlyP (w/o SHMT) showed the highest specific activity and yield. The “(w/o SHMT)” label denoted that the serine hydroxymethyltransferase was not included when calculating the pathway-specific activity. As already noted by others [68], the kinetics of this enzyme seem to be estimated inaccurately by the usual assays which is why the enzyme accounted for a major fraction of enzyme demand in the respective pathways. The rGlyP (w/o SHMT) showed a pathway-specific activity around twice as high as the next best pathway, albeit with high uncertainty. This can be attributed to the small number of reactions steps and the favorable thermodynamics and kinetics of the rGlyP although it has to be pointed out that the kinetic parameters from the literature were mainly based on activity measurements with unnatural substrates. Kinetic data of the assembled glycine cleavage system, especially in the carboxylating direction, are sparse. Additionally, eEquilibrator had trouble estimating the free energy of the carboxylation reaction that is catalyzed by the P-protein ($\Delta_r G'^{\circ} = -17.3 \pm 15.9$ kJ/mol):



As this is one of the key reactions and we also did not find a value for the forward kinetic constant, this standard deviation has a crucial impact on the total estimation of pathway-specific activity. All in all, the estimation for the rGlyP are consequently not completely trustworthy which stresses how urgently kinetic data for this system are needed to make accurate predictions. Disregarding the rGlyP, the rCCC, the 2-HG-rTCA cycle, and the CETCH cycle were paretooptimal for both criteria. The 2-HG-rTCA cycle in conjunction with the acetyl-CoA conversion module from the rCCC has a very low ATP-requirement per product. Although it has a lower pathway activity, it might still be a promising route, considering that conservative assumptions were made about the kinetic parameters of the enzymes.

When formate was used as the substrate of choice, NADH was regenerated via formate dehydrogenase, a highly active enzyme in this reaction direction. Therefore, the pathway activity is in general less limited by the supply of reducing equivalents compared to methanol. The RuMP cycle

needs formaldehyde which is difficult to produce from formate because formaldehyde oxidation with NAD^+ is almost irreversible. Likewise, reduction of formyl-CoA to formaldehyde that has been proposed by others [19, 21] was strongly limited by the formaldehyde: NADP+ oxidoreductase (CoA-acetylating) (data not shown). The only route that was functional in our simulation was the production of formaldehyde from 5,10-methylenetetrahydrofolate by the combined action of L-serine formaldehyde-lyase (glycine-forming) and 5,10-methylenetetrahydrofolate:glycine hydroxymethyltransferase. The former enzyme was also proposed for an optimized methanol assimilation pathway recently [20]. This variant of the RuMP cycle is denoted as RuMP (formate/THF). Compared to the general analysis in the previous section, the pathways' activity profited from the higher concentration of $\text{CO}_2/\text{HCO}_3^-$ which can be seen in Figure 8. The RuMP cycle starting from formate showed little activity and performed worse than the 2-HG-rTCA cycle and the rGlyP as it was again limited by the reactions supplying the formaldehyde.

Looking at the single contributions of the reactions to the enzyme demand, the Enzyme Cost Minimization algorithm makes consistent predictions as can be seen in Figure 9. The formate-based pathway analysis was well suited to evaluate the new proposed pathways in this regard because they were not supposed to be limited by the electron supply (i.e., formate) in this case. For the CBB cycle, RuBisCO (identifier “R_ribulose_b_R00024_2”) is the enzyme with the highest demand in the cycle where most of the demand is due to the relatively low turnover number (represented by “Capacity”), and a small fraction is caused by undersaturation with CO_2 (represented by “Saturation” which is considerably higher when the CO_2 partial pressure is low). In comparison to the CBB cycle, the new rCCC and the 2-HG-rTCA cycle (Figure 9) do not have a single limiting enzyme, but the enzyme demand is more evenly distributed among the enzymes. For the rCCC, the enzymes with the highest demand are the pyruvate carboxylase (Q00074) and the 4-hydroxybutyrate-CoA ligase (ADP-forming) (Q00071). While the pyruvate carboxylase is limited by the low concentration of carbonate, the 4-hydroxybutyrate-CoA ligase is supposed to be slower and thermodynamically more challenging than the AMP-forming variant [53]. Considering the 2-HG-rTCA cycle, the enzyme with the highest demand is the succinyl-CoA:2-hydroxyglutarate CoA-transferase (identifier Q00010). As noted before, this enzyme activity is just a side reaction of the succinyl-CoA:(S)-malate CoA-transferase, and thus, with a more specific enzyme, the pathway activity will be higher. In the 2-HG-rTCA cycle, the isocitrate dehydrogenase (carboxylating) is severely limited by the low availability of CO_2 (K_m of 1.3 mM) and is therefore the second costliest enzyme in the cycle. Taken together, the estimates for the new pathways show a lot of potential for the new artificial pathways. The contributions of each reaction on the other pathways can be found in the Supplementary Information in Figures S17-S24.

For the biotechnological scenario with a mixture of CO_2 and H_2 , an oxygen-tolerant soluble hydrogenase (kinetic

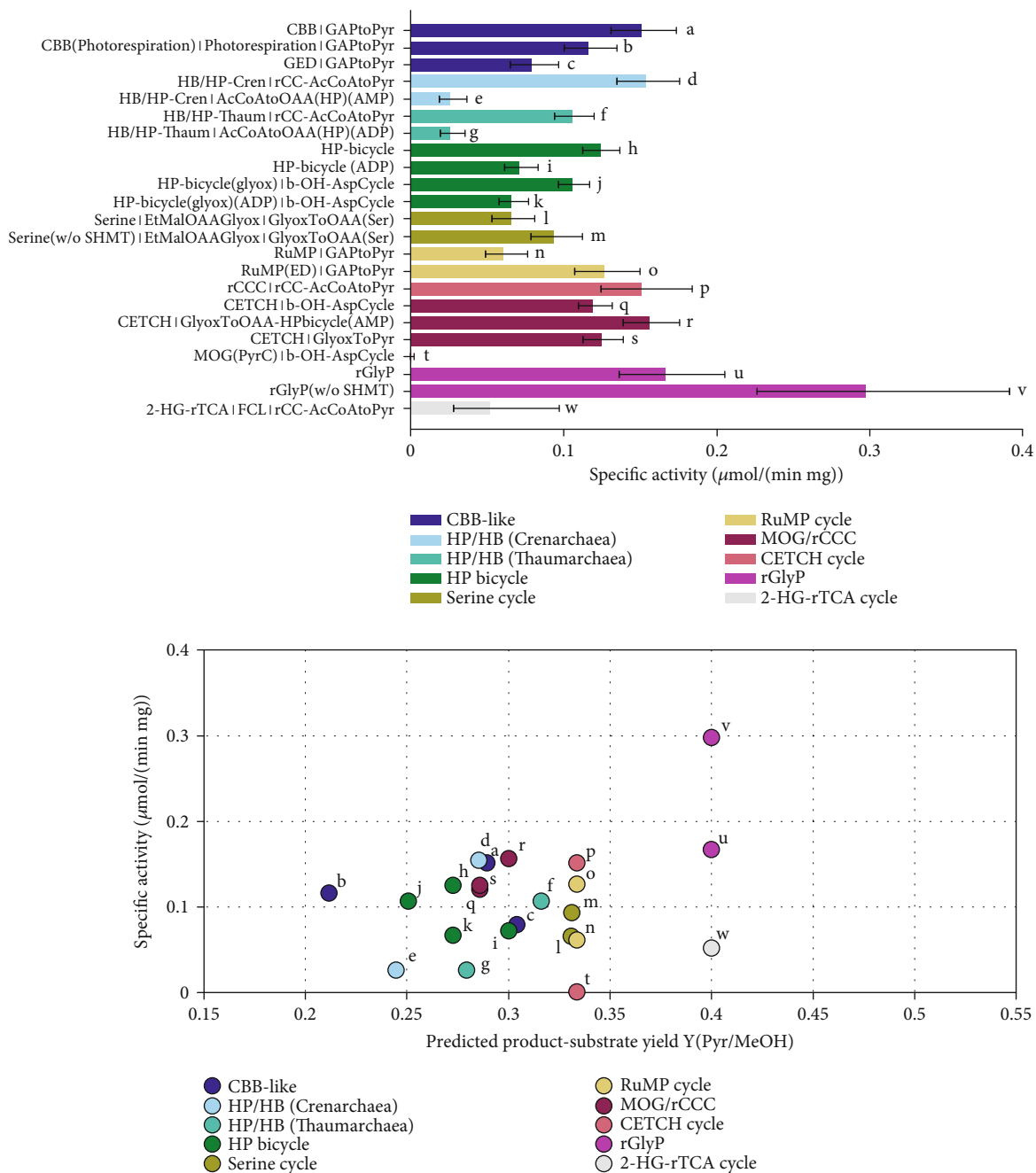


FIGURE 7: Comparison of carbon fixation pathways using methanol as a substrate for the production of pyruvate with the Enzyme Cost Minimization algorithm. “(w/o SHMT)” indicates that the cost of the serine hydroxymethyltransferase are not included in the respective pathway’s activity. The concentration of CO₂ was assumed to be 1 mM and the concentration of HCO₃⁻ to be 10 mM. (a) Pathway-specific activities with standard deviations and the full name of the main pathways and the connecting modules to transform their primary product to pyruvate. Pathway abbreviations: CBB(PTS): phosphatase-free CBB cycle; HP-bicycle(glyox): glyoxylate producing subcycle of the 3-HP bicycle; GAPtoPyr: glycolysis; FCL: formaldehyde:NADP⁺ oxidoreductase (formyl-CoA-forming); GlyoxToPyr: GCL/GDH route; b-OH-AspCycle: β-hydroxyaspartate cycle; AcCoAtoOAA(HP): acetyl-CoA to oxaloacetate-converting module derived from the 3-HP/4-HB cycle using either ADP- or AMP-producing CoA-ester synthases; rCC-AcCoAtoPyr: acetyl-CoA to pyruvate-converting module derived from the rCCC; GlyoxToOAA: glyoxylate- to oxaloacetate-converting module based on the Serine cycle. (b) Pathway-specific activities compared to product-substrate yield. The overall stoichiometries of all pathways and modules are listed in Table 1 and supplementary Table S4 and S5. The small letters label each pathway and correspond to the labels and the respective pathway combinations in (a).

data from *C. necator* H16) was assumed to regenerate NADH. In order to also include formate-dependent pathways, the formate dehydrogenase was run in the reductive

direction as described or proposed by others (X. [26, 52], and [12]). This is only possible at elevated CO₂ concentrations, however, as activity at atmospheric concentrations

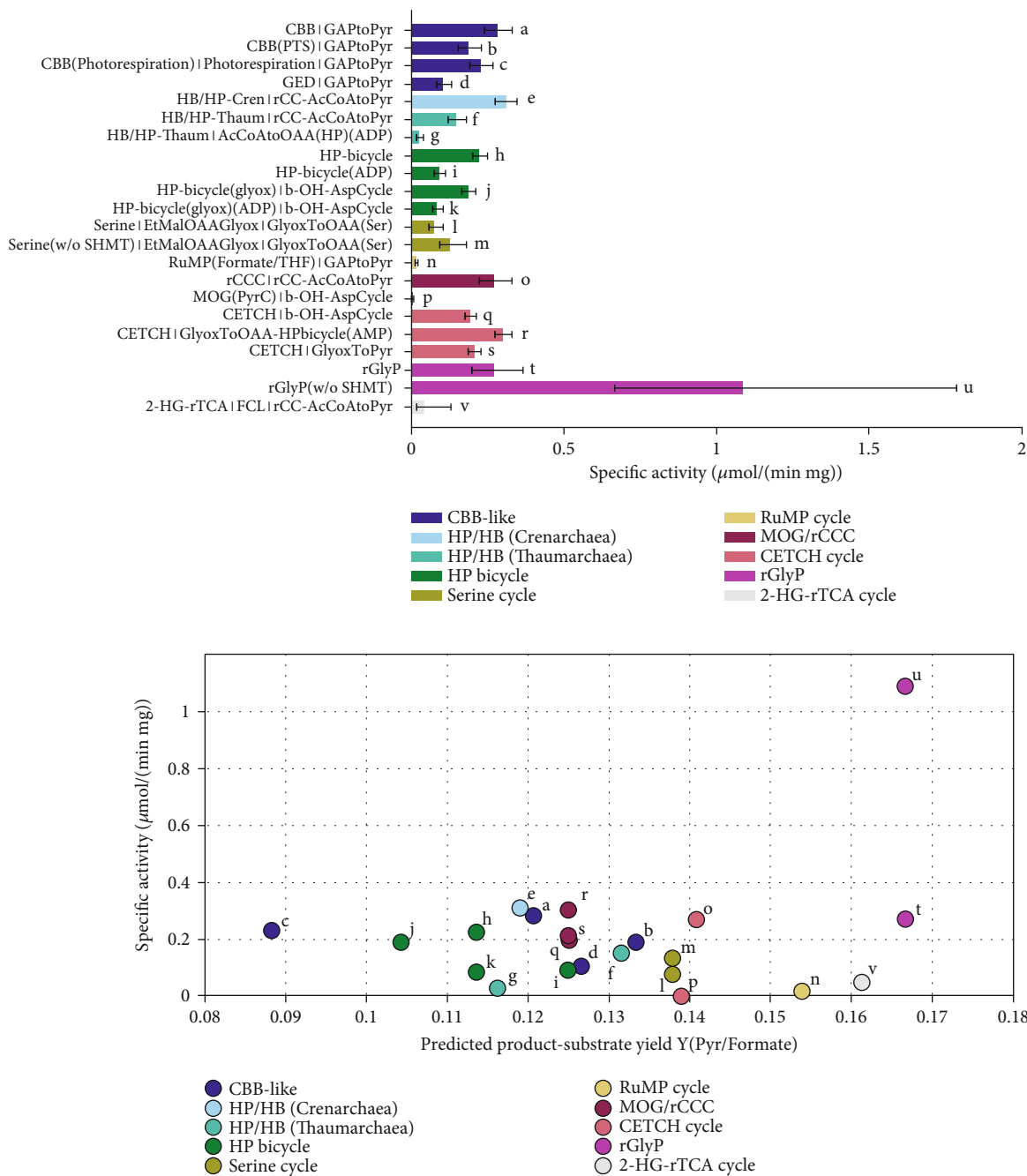


FIGURE 8: Comparison carbon fixation pathways using formate as a substrate for the production of pyruvate with the Enzyme Cost Minimization algorithm. “(w/o SHMT)” indicates that the cost of the serine hydroxymethyltransferase is not included in the respective pathway’s activity. The concentration of CO₂ was assumed to be 1 mM and the concentration of HCO₃⁻ to be 10 mM. (a) Pathway-specific activities with standard deviations and the full name of the main pathways and the connecting modules to transform their primary product to pyruvate. Pathway abbreviations: CBB(PTS): phosphatase-free CBB cycle; HP-bicycle(glyox): glyoxylate producing subcycle of the 3-HP bicycle; GAPtoPyr: glycolysis; FCL: formaldehyde:NADP⁺ oxidoreductase (formyl-CoA-forming); GlyoxToPyr: GCL/GDH route; b-OH-AspCycle: β-hydroxyaspartate cycle; AcCoAtoOAA(HP): acetyl-CoA- to oxaloacetate-converting module derived from the 3-HP/4-HB cycle using either ADP- or AMP-producing CoA-ester synthases; rCC-AcCoAtoPyr: acetyl-CoA- to pyruvate-converting module derived from the rCCC; GlyoxToOAA: glyoxylate to oxaloacetate converting module based on the Serine cycle. (b) Pathway-specific activities compared to product-substrate yield. The overall stoichiometries of all pathways and modules are listed in Table 1 and supplementary Table S4 and S5. The small letters label each pathway and correspond to the labels and the respective pathway combinations in (a).

showed only marginal activity when formate dehydrogenase was run in “reverse.” The kinetic data of the *C. necator* enzyme were taken as it exhibited reasonably high activity

in the reductive direction with NADH and is stable under aerobic conditions. It was assumed that a NADPH-dependent variant can be found in nature, engineered

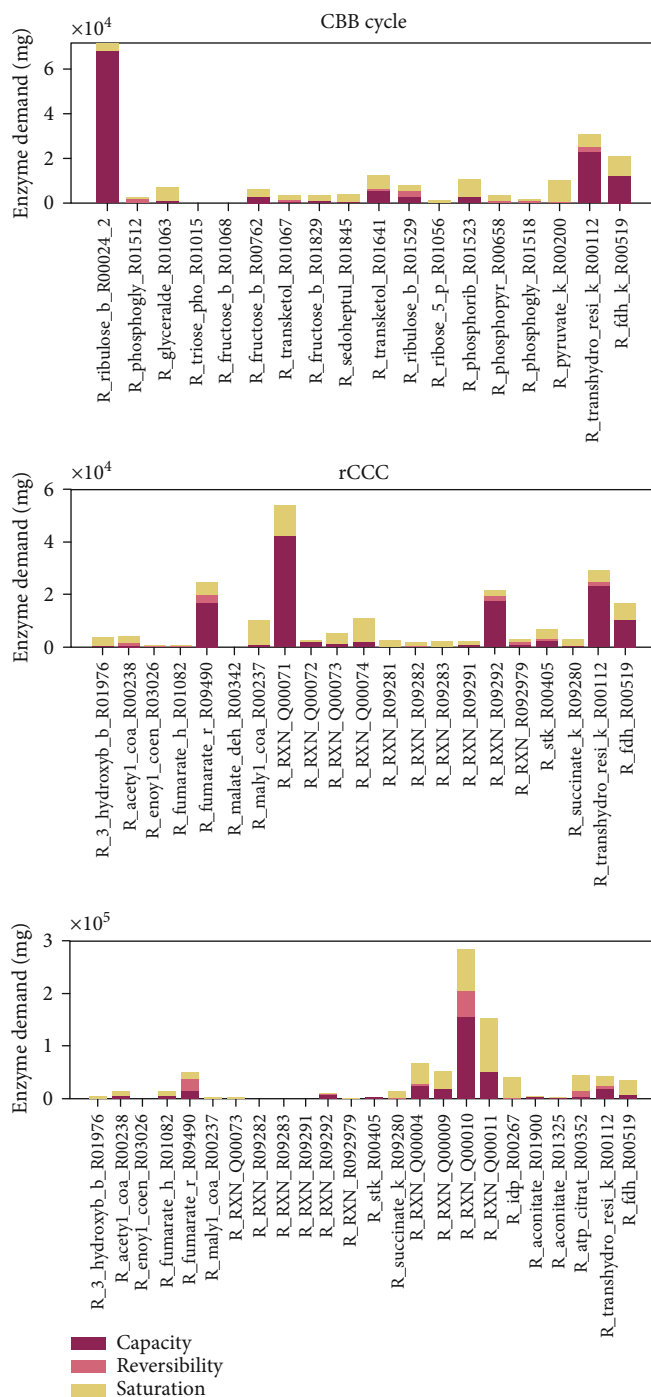


FIGURE 9: Enzyme demands to sustain a total pathway activity of 1 mmol product per second. Contribution of the capacity, the reversibility, and the saturation with substrates or products of each reaction to the demand of enzymes for the CBB cycle, the rCCC, and the 2-HG-rTCA cycle. The figure follows the wording of Noor et al. [29]. “Capacity”: demand of enzyme caused by a limitation by the catalytic rate constant; “Reversibility”: extra amount of enzyme needed because of a backward flux; “Saturation”: additional enzyme necessary because of undersaturation with a substrate and oversaturation with a product. The values present the optimized state as predicted by the ECM algorithm assuming a CO_2 concentration of $10 \mu\text{M}$ and HCO_3^- concentration of $100 \mu\text{M}$. Glyceraldehyde-3-phosphate was chosen as a product for the pathways. Reaction names correspond to their identifiers in the SBtab model file (Supplementary file Reactions_Composite22_model.tsv (available here)).

and/or evolved with similar characteristics, and that diaphorase activity can be neglected at high CO_2 concentrations. For the RuMP cycle, again the formate/THF variant was

evaluated. At high $\text{CO}_2/\text{HCO}_3^-$ and with NADPH as the reducing equivalent, the formate dehydrogenase could indeed reduce CO_2 to formate to power the formate-

dependent pathways according to our calculations as shown in Figure 10. The activity of these pathways was in general lower, however, than when formate was already present (Figure 8). The rGlyP still showed acceptable activity with the highest yield among all pathways making it a viable option even for H_2/CO_2 -based processes. Apart from that, the general trend of the performance of different pathways was unchanged compared to formate. Again, the 2-HG-rTCA cycle, rCCC, and CETCH cycle were pareto-optimal for both criteria when not considering the rGlyP. Given the high rate constant of the hydrogenase, the supply of reducing equivalents was not the limiting factor in this case in contrast to the methanol-based calculation.

All three scenarios presented in this section have also been performed with acetyl-CoA as the product of choice (compare Supplementary Figures S14-S16). Once more, the acetyl-CoA-generating pathways performed better than the production of pyruvate as in Section 3.5. In particular, the Serine cycle had a much improved activity, indicating that for this cycle, the ethylmalonyl-CoA pathway was the limiting part in the calculations for GAP production. However, the general trend of the pathways' performances did not change much.

4. Discussion

4.1. Strengths, Weakness, and Niches of Carbon Fixation Pathways. In the simulations described in the last sections, it could be seen that the orders of magnitude of both specific activity and product-substrate yield are similar for most pathways. Given the uncertainties of the enzymatic parameters that were used and the fact that enzyme kinetics were measured in vitro, there is no clear "optimal" pathway. It is rather likely that most pathways could have a niche in nature in which they are competitive which is in agreement with the sheer existence of many of them. The Serine cycle was very efficient and highly active for production of acetyl-CoA from formate, and the RuMP pathway was very efficient with methanol as a substrate which fits their natural roles. In the biotechnological scenarios described in the last section, the 3-HP/4-HP cycle and the 3-HP bicycle were predicted to be as active as the CBB cycle variants. Since these pathways occur in aquatic microorganisms and all carboxylation reactions are dependent on HCO_3^- , they can profit from the option to actively transport carbonate into the cell. For a similar concentration mechanism for CO_2 , dedicated (micro-) compartments are necessary. Additionally, carbonate concentrations at a pH greater than 7—as found in most aquatic habitats—are generally higher than the CO_2 concentrations at saturation, and therefore, the Michaelis constants of carbonate-fixing enzymes are not as restricted to low values.

The alternative and engineered carbon-fixing pathways match the activity of the CBB cycle but yield more product per substrate: the rCCC and the CETCH cycle outperform the CBB cycle in this regard. With formate as a substrate, the rGlyP was superior to the Serine cycle. Taken together, the rCCC, CETCH cycle, and rGlyP pathways have a high potential for the implementation in living organisms to cre-

ate chassis organisms for the fixation of CO_2 and C1 compounds in Industrial Biotechnology. The GED pathway and the phosphatase-less CBB cycle were shown to not fall short off the CBB cycle in specific activity and might be interesting alternatives if better enzymes can be found for the critical reactions. However, the aminomutase variant of the MOG cycle seemed to be infeasible judging from our calculations. All model substrates—formate, methanol, and H_2/CO_2 —showed good specific activities and hence are promising options for sustainable bioprocesses.

For future applications, the CO_2 -fixing pathways might also be implemented in crop plants which can be regarded as the Holy Grail of plant biotechnology and might boost crop productivity. To achieve this, the plant and chloroplast genome would have to be radically changed on a metabolic and regulatory level which, as of today, seems out of range of the methods of Synthetic Biology. First applications will consequently rather be found in microorganisms that can be manipulated more easily and cultivated under controlled conditions.

4.2. The Activity of the Calvin-Benson-Bassham Cycle Has Been Underestimated. Apart from the reductive glycine pathway which was difficult to predict because of a lack of kinetic data, most pathways actually have a pathway-specific activity in the same order of magnitude (compare Figures 6–8). This is in contrast to previous reports that concentrated on the weaknesses of the CBB cycle [4, 69–71]. So what is the reason that the activity of the CBB cycle was higher than expected? Judging from the results in this work, apart from RuBisCO, all enzymes in the CBB cycle are highly active and thermodynamics and kinetics are favorable. All other pathways were suffering either from kinetic or thermodynamic limitations or because of enzymes with equally low activity as RuBisCO. These enzymes should be better characterized in the future to exclude that they are indeed the limiting factor for the respective pathways. In Figures S17-S24 in the Supplementary Information, a detailed overview on pathway bottlenecks can be found. While the activity of the CBB cycle is higher than expected, this does not change its relatively high demand of ATP resulting in a lower product-substrate yield compared to other alternative pathways that were part of this work. The CBB cycle performed best when the glyceraldehyde-3-phosphate was chosen as the product of choice. In fact, the glyceraldehyde-3-phosphate as a product is a realistic case for comparison since plants rely on sucrose as a transport sugar. As sucrose is chemically relatively inert, it can be used in high concentrations in the conducting vascular cells of plants in which sucrose accumulates up to a concentration of one molar [72]. This "high-voltage" power transmission of the plant can hardly be realized with fats or acids. Therefore, the advantages that other carbon fixation pathways show for the production for nonsugar products might not be relevant for plants. In addition, the CBB cycle generates sugars that can be directly used as a precursor for different biomolecules, like ribose that is part DNA and RNA. It also has to be noted that only C3 photosynthesis was

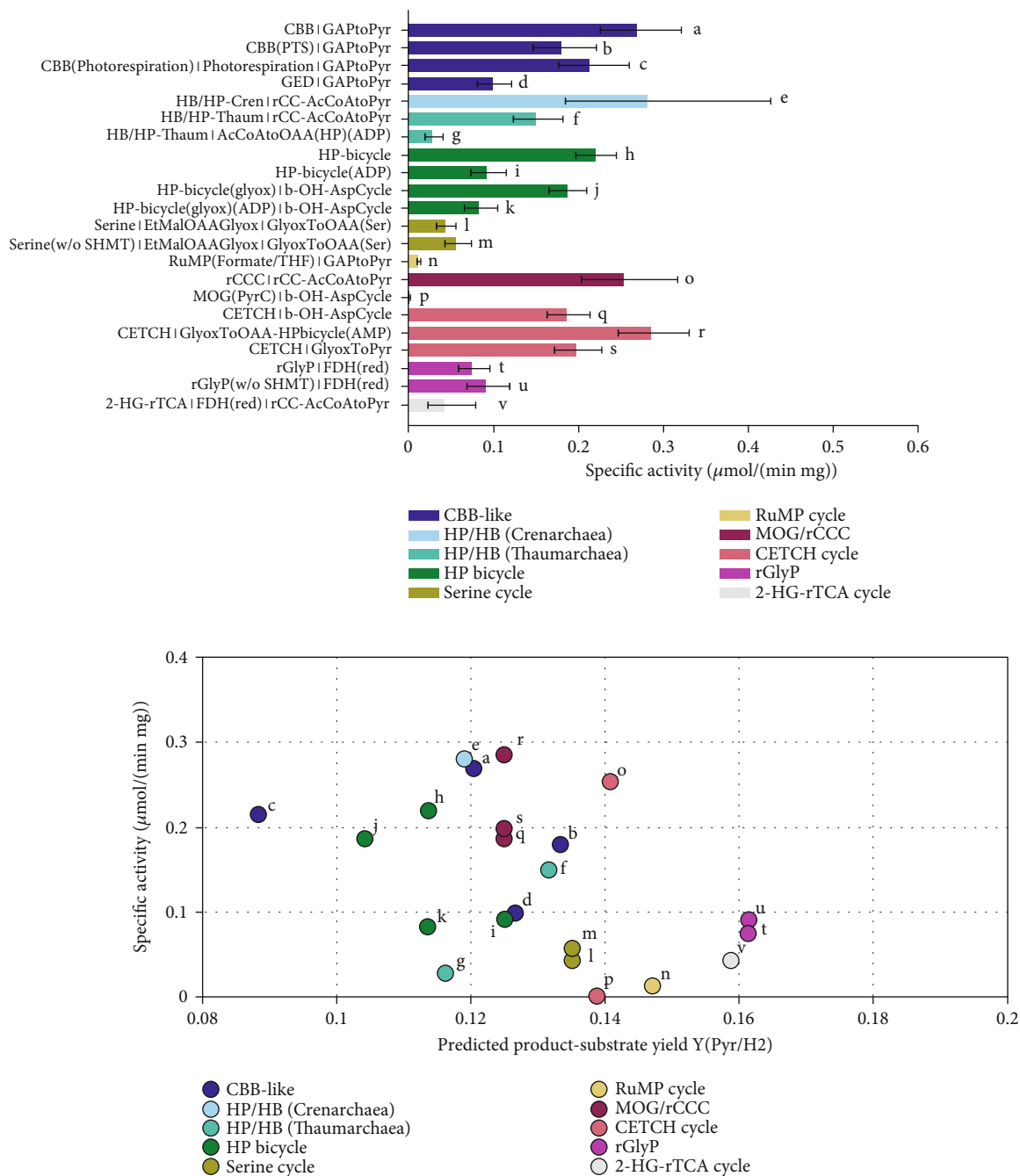
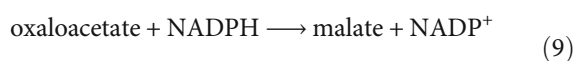
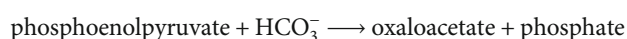
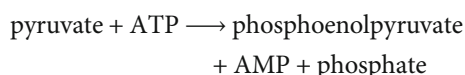


FIGURE 10: Comparison carbon fixation pathways using H_2/CO_2 as a substrate for the production of pyruvate with the Enzyme Cost Minimization algorithm. “(w/o SHMT)” indicates that the cost of the serine hydroxymethyltransferase is not included in the respective pathway’s activity. The concentration of CO_2 was assumed to be 1 mM and the concentration of HCO_3^- to be 10 mM. (a) Pathway-specific activities with standard deviations and the full name of the main pathways and the connecting modules to transform their primary product to pyruvate. Pathway abbreviations: CBB(PTS): phosphatase-free CBB cycle; HP-bicycle(glyox): glyoxylate-producing subcycle of the 3-HP bicycle; GAPtoPyr: glycolysis; FCL: formaldehyde:NADP⁺ oxidoreductase (formyl-CoA-forming); GlyoxToPyr: GCL/GDH route; b-OH-AspCycle: β -hydroxyaspartate cycle; AcCoAtoOAA(HP): acetyl-CoA- to oxaloacetate-converting module derived from the 3-HP/4-HB cycle using either ADP- or AMP-producing CoA-ester synthases; rCC-AcCoAtoPyr: acetyl-CoA- to pyruvate-converting module derived from the rCCC; GlyoxToOAA: glyoxylate- to oxaloacetate-converting module based on the Serine cycle. (b) Pathway-specific activities compared to product-substrate yield. The overall stoichiometries of all pathways and modules are listed in Table 1 and supplementary Table S4 and S5. The small letters label each pathway and correspond to the labels and the respective pathway combinations in (a).

included in the analysis which is strongly limited by low CO_2 concentration and low RuBisCO activity (Figure 9). Using CO_2 concentration mechanisms and RuBisCO variants of C4 plants with a higher K_m and k_{cat} , the pathway-specific activity would be significantly higher and thus probably surpass even the best designed pathways. Consequently, the CBB cycle is not as bad as commonly assumed, according to our calculations, and there is a good reason that most of the carbon fixation on this planet is facilitated by this pathway. After all, RuBisCO might not be as “bad” as previously thought as others have already noted [73].

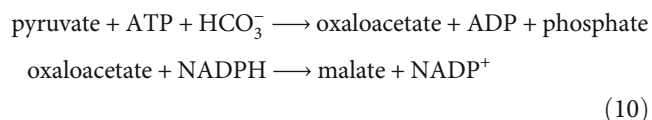
4.3. Plant-Like Photosynthesis Is Not Designed to Be as ATP-Efficient as Possible. Although the CBB cycle has a higher ATP demand for 1 mol of a specific product than other pathways, this difference is actually not very high from an energetic perspective: the free burning energy of glyceraldehyde-3-phosphate is around -1500 kJ/mol, for instance, while the energy of ATP hydrolysis is around -40 kJ/mol. So the energy that is saved by more efficient pathways might not be as important as other factors which will be discussed in the next section. In fact, there are indications that ATP efficiency is not a decisive factor for the design of carbon fixation in photosynthetic organisms. The phosphatase-less CBB cycle that we introduced in this work is a good example for this. With its modified reactions, the CBB cycle would save 2 ATP per formed C3 sugar. In addition, the pathway also did not lose much of its activity, although enzymes with promiscuous activity were taken that were not specialized to run this cycle. Why has nature not made this “easy” step in evolution to optimize the CBB cycle? While it might well be possible that there are actually organisms that use this modified cycle which just have not been found yet, there is no evidence so far that this variant is used in photosynthetic organisms. Not having the irreversible phosphatase reactions might disrupt the regulation and/or robustness of the cycle which might be more important.

A second example that ATP efficiency is not the main driver in the evolution of photosynthesis can be found in C4-plant metabolism. C4 plants use a CO_2 concentration mechanism that gives these plants a selective advantage in certain climates [74]. CO_2 concentration in the mesophyll cells can be achieved by the actions of PEP synthase, PEP carboxylase, and malate dehydrogenase (malic enzyme variant of C4 photosynthesis):



This set of reactions is quite exothermic in total with $\Delta_r G'^{\circ} = -75 \pm 7 \text{ kJ/mol}$. Malate is then shuttled to the bundle sheath cells where CO_2 is released by the malic enzyme and refixed by the CBB cycle. Although a total minimum

of 15 ATP instead of 9 ATP is necessary for the formation of one C3 sugar, C4 plants easily grow faster than C3 plants in most situations. Additionally, thinking of optimal design, the first steps in a CO_2 concentration mechanism could also be catalyzed by pyruvate carboxylase:



These reactions are also practically irreversible ($\Delta_r G'^{\circ} = -35 \pm 7 \text{ kJ/mol}$) with the difference that one ATP is saved. In total, using pyruvate carboxylase instead of the natural occurring enzymes would hence save 3 ATP per C3 sugar in C4 plants. Again, this variant has not been found in nature although all enzymes exist in plants. Therefore, it seems that plants prefer to have a slight edge in driving force or activity instead of being more efficient ATP-wise.

4.4. Unconsidered Factors in Pathway Comparison. Specific activity and yields of different metabolic pathways were the main criteria for comparison in this work. While those are important factors, they are not necessarily decisive. One unconsidered factor in this context are the vitamins: interestingly, cobalamin (vitamin B12) is necessary for all CO_2 -fixing pathways except for the CBB cycle and the rGlyP. Being the most structurally complex of all vitamins, B12 biosynthesis is a complex pathway that does not occur in plants and animals. In addition, it also contains cobalt. The costs associated to biosynthesis of B12 and other important cofactors were not taken into account in this work. The other vitamins are structurally more compact, less demanding in their biosynthesis, and independent of metal ions and are needed for many other vital reactions in the metabolism as well. In the future, this requirement could be addressed by adding B12 as a metabolite and its biosynthesis to the STab model file if adequate data are available. This will help to evaluate whether or not B12 requirement has a critical impact.

Other unconsidered factors in the analysis presented in this work are the controllability and robustness of the pathways. Previous works have shown that these factors are an essential part of the design principles in nature [75–77]. Allosteric control of enzymes was also not included in the reaction kinetics. More dedicated models and algorithms that combine pathway performance and criteria for the robustness could be established in the future. This way, good trade-offs between the two criteria could be predicted in artificial pathways or be explained in naturally occurring pathways. New efforts in the design of artificial allosteric control of enzymes [78, 79] could even make possible the prediction and design of a regulatory layer on top of the reactivity layer in the network to build pathway designs that are optimal for both criteria.

So far, metabolite toxicity has also not been addressed in this work. Toxic metabolites in the pathways include, for example, aldehydes like formaldehyde, glyoxylate, succinate semialdehyde, malonate and semialdehyde, but also

propionyl-CoA. To include this in the future, tighter constraints on these toxic compounds may be implemented, possibly based on concentrations that are actually measured within cells with a respective metabolism.

4.5. Implications on Future Design and Implementation of Pathways for Carbon Fixation. Although the design space for artificial CO₂-fixing pathways was thoroughly investigated in a former study [4], we could still find several new routes that have not yet been addressed. This is probably due to missing reactions in the metabolic networks the authors used which were collected from KEGG as the remaining reactions of the 3-hydroxypropionate bicycle were just described at the same time. Likewise, side reactions that were used to construct the pathways here (2-HG-rTCA cycle and phosphatase-less CBB cycle) are also not annotated. Using more complete, updated versions of reaction stoichiometries or even extending the network on all possible reactivities with tools of retrosynthesis like RetroPath [33, 80] will possibly open up many new possible routes that should be explored in the future.

Using ECM and other advanced tools for pathway analysis can predict how to best implement a metabolic pathway based on measured data and which ones are most promising. This relies heavily on the quality of the kinetic and thermodynamic data that are used. For some sensitive parameters of reactions, there were only single studies available in which the enzymes were characterized. The ECM algorithm can predict, however, which reactions will be likely limiting the pathways and thereby gives good suggestions for enzymes that need to be better characterized and/or engineered. This stresses how important enzymology is for the understanding and the design of biochemical pathways as has also been noted by Erb recently [81]. In fact, the predictions made in this work are only possible because of the thorough characterization of all the enzymes by others. In the future, omics data in addition to more enzyme kinetics should be integrated into the analysis to make use of *in vivo* data which might be very different from *in vitro* enzyme kinetics as has been pointed out for the ethylmalonyl-CoA pathway in this study. In this context, modelling approaches that are able to quantify uncertainties and identify missing data are ideal. This way, models and simulations can be iteratively improved to get closer and closer to the biological reality and a deeper understanding of metabolism and design principles of natural pathways.

Data Availability

The kinetic and thermodynamic data of all enzymes and reactions used to support the findings of this study are included within the supplementary information file "Reactions_Composite22_model.tsv" and in the supplementary information document. The MATLAB code that was used to calculate pathway activities is deposited at <https://github.com/HannesLoewe/pathway-comparison-ECM> and can be reused under the GNU General Public License.

Conflicts of Interest

We declare no competing interests.

Authors' Contributions

H. Löwe was responsible for the conceptualization, methodology, validation, visualization, investigation, writing of the original draft, and review and editing of the article. A. Kremling was responsible for the funding acquisition, project administration, supervision, and review and editing of the article.

Acknowledgments

This research was supported by the Deutsche Forschungsgemeinschaft (DFG), grant KR 2963/3-1. We would like to thank Wolfram Liebermeister for the fruitful discussion and the support with setting up the MATLAB implementations of the toolboxes that were used in this study. We also thank Franziska Kratzl for her suggestions to the manuscript.

Supplementary Materials

The kinetic and thermodynamic data of all enzymes and reaction used to support the findings of this study are included within the supplementary information file "Reactions_Composite22_model.tsv" and in the supplementary information document. (*Supplementary Materials*)

References

- [1] World Bank, and Food and Agriculture Organization of the UN, *Agricultural land (% of land area)*, 2016.
- [2] X. G. Zhu, S. P. Long, and D. R. Ort, "What is the maximum efficiency with which photosynthesis can convert solar energy into biomass?," *Current Opinion in Biotechnology*, vol. 19, no. 2, pp. 153–159, 2008.
- [3] A. Flamholz, N. Prywes, U. Moran et al., "Revisiting trade-offs between Rubisco kinetic parameters," *Biochemistry*, vol. 58, no. 31, pp. 3365–3376, 2019.
- [4] A. Bar-Even, E. Noor, N. E. Lewis, and R. Milo, "Design and analysis of synthetic carbon fixation pathways," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 19, pp. 8889–8894, 2010.
- [5] R. H. Wilson and S. M. Whitney, "Improving CO₂ fixation by enhancing Rubisco performance," in *Directed Enzyme Evolution: Advances and Applications*, pp. 101–126, Springer International Publishing, 2017.
- [6] A. I. Flamholz, E. Dugan, C. Blikstad et al., "Functional reconstitution of a bacterial CO₂ concentrating mechanism in *Escherichia coli*," *ELife*, vol. 9, no. October, pp. 1–57, 2020.
- [7] B. M. Long, W. Y. Hee, R. E. Sharwood et al., "Carboxysome encapsulation of the CO₂-fixing enzyme Rubisco in tobacco chloroplasts," *Nature Communications*, vol. 9, no. 1, 2018.
- [8] T. Schwander, L. S. von Borzyskowski, S. Burgener, N. S. Cortina, and T. J. Erb, "A synthetic pathway for the fixation of carbon dioxide in vitro," *Science*, vol. 354, no. 6314, pp. 900–904, 2016.

- [9] M. Scheffen, D. G. Marchal, T. Beneyton et al., "A new-to-nature carboxylation module to improve natural and synthetic CO₂ fixation," *Nature Catalysis*, vol. 4, no. 2, pp. 105–115, 2021.
- [10] P. F. South, A. P. Cavanagh, H. W. Liu, and D. R. Ort, "Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field," *Science*, vol. 363, no. 6422, 2019.
- [11] D. L. Trudeau, C. Edlich-Muth, J. Zarzycki et al., "Design and in vitro realization of carbon-conserving photorespiration," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 49, pp. E11455–E11464, 2018.
- [12] A. Bar-Even, "Daring metabolic designs for enhanced plant carbon fixation," *Plant Science*, vol. 273, no. August, pp. 71–83, 2018.
- [13] N. J. Claassens, D. Z. Sousa, V. A. P. Martins, D. Santos, W. M. De Vos, and J. Van Der Oost, "Harnessing the power of microbial autotrophy," *Nature Reviews Microbiology*, vol. 14, no. 11, pp. 692–706, 2016.
- [14] N. J. Claassens, C. A. R. Cotton, D. Kopljar, and A. Bar-Even, "Making quantitative sense of electromicrobial production," *Nature Catalysis*, vol. 2, no. 5, pp. 437–447, 2019.
- [15] P. Keller, E. Noor, F. Meyer et al., "Methanol-dependent *Escherichia coli* strains with a complete ribulose monophosphate cycle," *Nature Communications*, vol. 11, no. 1, pp. 5403–5403, 2020.
- [16] H. He, C. Edlich-Muth, S. N. Lindner, and A. Bar-Even, "Ribulose monophosphate shunt provides nearly all biomass and energy required for growth of *E. coli*," *ACS Synthetic Biology*, vol. 7, no. 6, pp. 1601–1611, 2018.
- [17] H. Yu and J. C. Liao, "A modified serine cycle in *Escherichia coli* converts methanol and CO₂ to two-carbon compounds," *Nature Communications*, vol. 9, no. 1, 2018.
- [18] F. Y. H. Chen, H. W. Jung, C. Y. Tsuei, and J. C. Liao, "Converting *Escherichia coli* to a synthetic methylotroph growing solely on methanol," *Cell*, vol. 182, no. 4, pp. 933–946.e14, 2020.
- [19] A. Chou, J. M. Clomburg, S. Qian, and R. Gonzalez, "2-Hydroxyacyl-CoA lyase catalyzes acyloin condensation for one-carbon bioconversion," *Nature Chemical Biology*, vol. 15, no. 9, pp. 900–906, 2019.
- [20] H. He, R. Höper, M. Dodenhöft, P. Marlière, and A. Bar-Even, "An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in *E. coli*," *Metabolic Engineering*, vol. 60, no. July, pp. 1–13, 2020.
- [21] J. B. Siegel, A. L. Smith, S. Poust et al., "Computational protein design enables a novel one-carbon assimilation pathway," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 12, pp. 3704–3709, 2015.
- [22] G. Fuchs, "CO₂ fixation in acetogenic bacteria: variations on a theme," *FEMS Microbiology Letters*, vol. 39, no. 3, pp. 181–213, 1986.
- [23] J. Bang and S. Y. Lee, "Assimilation of formic acid and CO₂ by engineered *Escherichia coli* equipped with reconstructed one-carbon assimilation pathways," *Proceedings of the National Academy of Sciences*, vol. 115, no. 40, pp. E9271–E9279, 2018.
- [24] N. J. Claassens, G. Bordanaba-Florit, C. A. Cotton et al., "Replacing the Calvin cycle with the reductive glycine pathway in *Cupriavidus necator*," *Metabolic Engineering*, vol. 62, no. -November, pp. 30–41, 2020.
- [25] S. Kim, S. N. Lindner, S. Aslan et al., "Growth of *E. coli* on formate and methanol via the reductive glycine pathway," *Nature Chemical Biology*, vol. 16, no. 5, pp. 538–545, 2020.
- [26] C. A. R. Cotton, C. Edlich-Muth, and A. Bar-Even, "Reinforcing carbon fixation: CO₂ reduction replacing and supporting carboxylation," *Current Opinion in Biotechnology*, vol. 49, pp. 49–56, 2018.
- [27] O. Hädicke, A. von Kamp, T. Aydogan, and S. Klamt, "OptMDFP pathway: identification of metabolic pathways with maximal thermodynamic driving force and its application for analyzing the endogenous CO₂ fixation potential of *Escherichia coli*," *PLoS Computational Biology*, vol. 14, no. 9, article e1006492, 2018.
- [28] A. Satanowski, B. Dronsella, E. Noor et al., "Awakening a latent carbon fixation cycle in *Escherichia coli*," *Nature Communications*, vol. 11, no. 1, 2020.
- [29] E. Noor, A. Flamholz, A. Bar-Even, D. Davidi, R. Milo, and W. Liebermeister, "The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization," *PLoS Computational Biology*, vol. 12, no. 11, article e1005167, 2016.
- [30] T. Lubitz, M. Schulz, E. Klipp, and W. Liebermeister, "Parameter balancing in kinetic models of cell metabolism," *Journal of Physical Chemistry B*, vol. 114, no. 49, pp. 16298–16303, 2010.
- [31] J. Bertsch and V. Müller, "Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria," *Biotechnology for Biofuels*, vol. 8, no. 1, pp. 1–12, 2015.
- [32] R. Takors, M. Kopf, J. Mampel et al., "Using gas mixtures of CO, CO₂ and H₂ as microbial substrates: the do's and don'ts of successful technology transfer from laboratory to production scale," *Microbial Biotechnology*, vol. 11, no. 4, pp. 606–625, 2018.
- [33] T. J. Erb, P. R. Jones, and A. Bar-Even, "Synthetic metabolism: metabolic engineering meets enzyme design," *Current Opinion in Chemical Biology*, vol. 37, pp. 56–62, 2017.
- [34] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [35] A. Flamholz, E. Noor, A. Bar-Even, and R. Milo, "Equilibrator—the biochemical thermodynamics calculator," *Nucleic Acids Research*, vol. 40, no. D1, pp. D770–D775, 2012.
- [36] L. Jeske, S. Placzek, I. Schomburg, A. Chang, and D. Schomburg, "BRENDA in 2019: a European ELIXIR core data resource," *Nucleic Acids Research*, vol. 47, no. D1, pp. D542–D549, 2019.
- [37] A. Bateman, "UniProt: a worldwide hub of protein knowledge," *Nucleic Acids Research*, vol. 47, no. D1, pp. D506–D515, 2019.
- [38] T. Lubitz, J. Hahn, F. T. Bergmann, E. Noor, E. Klipp, and W. Liebermeister, "SBtab: a flexible table format for data exchange in systems biology," *Bioinformatics*, vol. 32, no. 16, pp. 2559–2561, 2016.
- [39] V. B. Borisov, R. Murali, M. L. Verkhovskaya et al., "Aerobic respiratory chain of *Escherichia coli* is not allowed to work in fully uncoupled mode," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 42, pp. 17320–17324, 2011.
- [40] G. Cecchini, I. Schröder, R. P. Gunsalus, and E. Maklashina, "Succinate dehydrogenase and fumarate reductase from *Escherichia coli*," *Biochimica et Biophysica Acta-Bioenergetics*, vol. 1553, no. 1-2, pp. 140–157, 2002.

- [41] C. Sakai, E. Tomitsuka, H. Esumi, S. Harada, and K. Kita, "Mitochondrial fumarate reductase as a target of chemotherapy: from parasites to cancer cells," *Biochimica et Biophysica Acta-General Subjects*, vol. 1820, no. 5, pp. 643–651, 2012.
- [42] M. Schnorpfel, I. G. Janausch, S. Biel, A. Kröger, and G. Unden, "Generation of a proton potential by succinate dehydrogenase of *Bacillus subtilis* functioning as a fumarate reductase," *European Journal of Biochemistry*, vol. 268, no. 10, pp. 3069–3074, 2001.
- [43] E. Tomitsuka, K. Kita, and H. Esumi, "Regulation of succinate-ubiquinone reductase and fumarate reductase activities in human complex II by phosphorylation of its flavoprotein subunit," *Proceedings of the Japan Academy, Series B*, vol. 85, no. 7, pp. 258–265, 2009.
- [44] A. Miura, M. Kameya, H. Arai, M. Ishii, and Y. Igarashi, "A soluble NADH-dependent fumarate reductase in the reductive tricarboxylic acid cycle of hydrogenobacter thermophilus TK-6," *Journal of Bacteriology*, vol. 190, no. 21, pp. 7170–7177, 2008.
- [45] C. Kaleta, S. Schäuble, U. Rinas, and S. Schuster, "Metabolic costs of amino acid and protein production in *Escherichia coli*," *Biotechnology Journal*, vol. 8, no. 9, pp. 1105–1114, 2013.
- [46] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry*, W. H. Freeman, 2010, <https://books.google.de/books?id=u8yofUtTQqsC>.
- [47] T. Friedrich, "The NADH: ubiquinone oxidoreductase (complex I) from *Escherichia coli*," *Biochimica et Biophysica Acta-Bioenergetics*, vol. 1364, no. 2, pp. 134–146, 1998.
- [48] Q. H. Tran, J. Bongaerts, D. Vlad, and G. Unden, "Requirement for the proton-pumping NADH dehydrogenase I of *Escherichia coli* in respiration of NADH to fumarate and its bioenergetic implications," *European Journal of Biochemistry*, vol. 244, no. 1, pp. 155–160, 1997.
- [49] U. Sauer, F. Canonaco, S. Heri, A. Perrenoud, and E. Fischer, "The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*," *Journal of Biological Chemistry*, vol. 279, no. 8, pp. 6613–6619, 2004.
- [50] M. Shepherd and R. K. Poole, "Bacterial respiratory chains," in *Encyclopedia of Biophysics*, pp. 172–177, Springer, Berlin Heidelberg, 2013.
- [51] G. A. Davis and D. M. Kramer, "Optimization of ATP synthase c-rings for oxygenic photosynthesis," *Frontiers in Plant Science*, vol. 10, no. January, p. 30, 2020.
- [52] X. Yu, D. Niks, A. Mulchandani, and R. Hille, "Efficient reduction of CO₂ by the molybdenum-containing formate dehydrogenase from *Cupriavidus necator* (*Ralstonia eutropha*)," *Journal of Biological Chemistry*, vol. 292, no. 41, pp. 16872–16879, 2017.
- [53] M. Könneke, D. M. Schubert, P. C. Brown et al., "Ammonia-oxidizing Archaea use the most energy-efficient aerobic pathway for CO₂ fixation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 22, pp. 8239–8244, 2014.
- [54] A. Mall, J. Sobotta, C. Huber et al., "Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium," *Science*, vol. 359, no. 6375, pp. 563–567, 2018.
- [55] X. Yang, Q. Yuan, H. Luo et al., "Systematic design and in vitro validation of novel one-carbon assimilation pathways," *Metabolic Engineering*, vol. 56, no. December, pp. 142–153, 2019.
- [56] M. Nattermann, S. Burgener, P. Pfister et al., "Engineering a highly efficient carbolligase for synthetic one-carbon metabolism," *ACS Catalysis*, vol. 11, no. 9, pp. 5396–5404, 2021.
- [57] X. Zhang, Y. Mao, B. Wang et al., "Screening, expression, purification and characterization of CoA-transferases for lactoyl-CoA generation," *Journal of Industrial Microbiology and Biotechnology*, vol. 46, no. 7, pp. 899–909, 2019.
- [58] J. K. Zorz, J. A. Kozłowski, L. Y. Stein, M. Strous, and M. Kleiner, "Comparative proteomics of three species of ammonia-oxidizing bacteria," *Frontiers in Microbiology*, vol. 9, no. MAY, p. 938, 2018.
- [59] A. S. Reshetnikov, O. N. Rozova, V. N. Khmelenina et al., "Characterization of the pyrophosphate-dependent 6-phosphofructokinase from *Methylococcus Capsulatus bath*," *FEMS Microbiology Letters*, vol. 288, no. 2, pp. 202–210, 2008.
- [60] T. J. Erb, I. A. Berg, V. Brecht, M. Müller, G. Fuchs, and B. E. Alber, "Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 25, pp. 10631–10636, 2007.
- [61] S. Estelmann, M. Hügler, W. Eisenreich et al., "Labeling and enzyme studies of the central carbon metabolism in *Metallosphaera sedula*," *Journal of Bacteriology*, vol. 193, no. 5, pp. 1191–1200, 2011.
- [62] R. Douce, J. Bourguignon, M. Neuburger, and F. Rébeillé, "The glycine decarboxylase system: a fascinating complex," *Trends in Plant Science*, vol. 6, no. 4, pp. 167–176, 2001.
- [63] A. Grostern, C. M. Sales, W. Q. Zhuang, O. Erbilgin, and L. Alvarez-Cohana, "Glyoxylate metabolism is a key feature of the metabolic degradation of 1,4-dioxane by *Pseudonocardia dioxanivorans* strain CB1190," *Applied and Environmental Microbiology*, vol. 78, no. 9, pp. 3298–3308, 2012.
- [64] L. S. von Borzyskowski, F. Severi, K. Krüger et al., "Marine proteobacteria metabolize glycolate via the β -hydroxyaspartate cycle," *Nature*, vol. 575, no. 7783, pp. 500–504, 2019.
- [65] S. E. Mainguet, L. S. Gronenberg, S. S. Wong, and J. C. Liao, "A reverse glyoxylate shunt to build a non-native route from C4 to C2 in *Escherichia coli*," *Metabolic Engineering*, vol. 19, pp. 116–127, 2013.
- [66] H. Yu, X. Li, F. Duchoud, D. S. Chuang, and J. C. Liao, "Augmenting the Calvin-Benson-Bassham cycle by a synthetic malyl-CoA-glycerate carbon fixation pathway," *Nature Communications*, vol. 9, no. 1, 2018.
- [67] T. Brautaset, Ø. M. Jakobsen, K. D. Josefsen, M. C. Flickinger, and T. E. Ellingsen, "*Bacillus methanolicus*: a candidate for industrial production of amino acids from methanol at 50°C," *Applied Microbiology and Biotechnology*, vol. 74, no. 1, pp. 22–34, 2007.
- [68] H. Šmejkalová, T. J. Erb, and G. Fuchs, "Methanol assimilation in *Methylobacterium extorquens* AM1: demonstration of all enzymes and their regulation," *PLoS ONE*, vol. 5, no. 10, article e13001, 2010.
- [69] N. J. Claassens, "A warm welcome for alternative CO₂ fixation pathways in microbial biotechnology," *Microbial Biotechnology*, vol. 10, no. 1, pp. 31–34, 2017.
- [70] T. J. Erb and J. Zarzycki, "A short history of RubisCO: the rise and fall (?) of Nature's predominant CO₂ fixing enzyme," *Current Opinion in Biotechnology*, vol. 49, pp. 100–107, 2018.
- [71] G. M. M. Stoffel, D. A. Saez, H. DeMirci et al., "Four amino acids define the CO₂ binding pocket of enoyl-CoA

- carboxylases/reductases,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 28, pp. 13964–13969, 2019.
- [72] D. Geiger, “Plant sucrose transporters from a biophysical point of view,” *Molecular Plant*, vol. 4, no. 3, pp. 395–406, 2011.
- [73] C. Bathellier, G. Tcherkez, G. H. Lorimer, and G. D. Farquhar, “Rubisco is not really so bad,” *Plant, Cell & Environment*, vol. 41, no. 4, pp. 705–716, 2018.
- [74] R. R. L. Atkinson, E. J. Mockford, C. Bennett et al., “C4 photosynthesis boosts growth by altering physiology, allocation and size,” *Nature Plants*, vol. 2, no. 5, 2016.
- [75] L. Bromig, A. Kremling, and A. Marin-Sanguino, “Understanding biochemical design principles with ensembles of canonical non-linear models,” *PLOS ONE*, vol. 15, no. 4, article e0230599, 2020.
- [76] Y. Lee, J. G. Lafontaine Rivera, and J. C. Liao, “Ensemble modeling for robustness analysis in engineering non-native metabolic pathways,” *Metabolic Engineering*, vol. 25, pp. 63–71, 2014.
- [77] J. M. Whitacre, “Biological robustness: paradigms, mechanisms, and systems principles,” *Frontiers in Genetics*, vol. 3, 2012.
- [78] Z. Chen, S. Rappert, and A. P. Zeng, “Rational design of allosteric regulation of homoserine dehydrogenase by a non-natural inhibitor l-lysine,” *ACS Synthetic Biology*, vol. 4, no. 2, pp. 126–131, 2015.
- [79] P. J. Cross, T. M. Allison, R. C. J. Dobson, G. B. Jameson, and E. J. Parker, “Engineering allosteric control to an unregulated enzyme by transfer of a regulatory domain,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 6, pp. 2111–2116, 2013.
- [80] B. Delépine, T. Duigou, P. Carbonell, and J. L. Faulon, “Retro-Path2.0: a retrosynthesis workflow for metabolic engineers,” *Metabolic Engineering*, vol. 45, no. January, pp. 158–170, 2018.
- [81] T. J. Erb, “Back to the future: why we need enzymology to build a synthetic metabolism of the future,” *Beilstein Journal of Organic Chemistry*, vol. 15, pp. 551–557, 2019.