

# MEMO: A Method for Computing Metabolic Modules for Cell-Free Production Systems

Axel von Kamp and Steffen Klamt\*



Cite This: *ACS Synth. Biol.* 2020, 9, 556–566



Read Online

ACCESS |



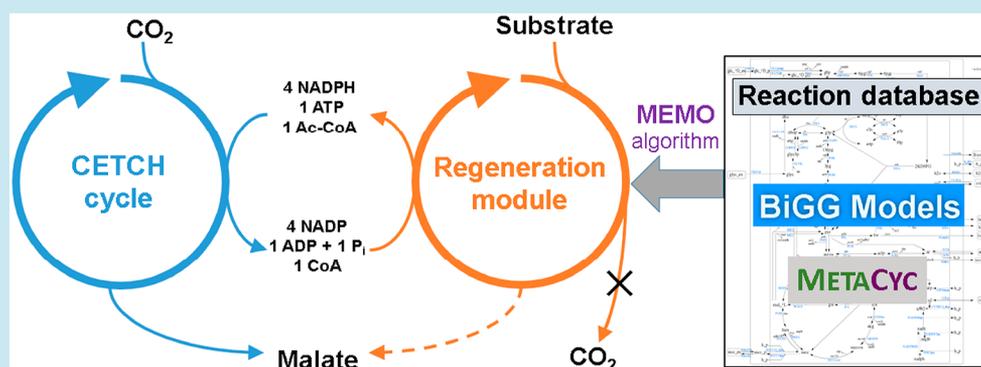
Metrics & More



Article Recommendations



Supporting Information



**ABSTRACT:** Cell-free bioproduction systems represent a promising alternative to classical microbial fermentation processes to synthesize value-added products from biological feedstocks. An essential step for establishing cell-free production systems is the identification of suitable metabolic modules with defined properties. Here we present MEMO, a novel computational approach to find smallest metabolic modules with specified stoichiometric and thermodynamic constraints supporting the design of cell-free systems in various regards. In particular, one key challenge for a sustained operation of cell-free systems is the regeneration of utilized cofactors (such as ATP and NAD(P)H). Given a production pathway with certain cofactor requirements, MEMO can be used to compute smallest regeneration modules that recover these cofactors with required stoichiometries. MEMO incorporates the stoichiometric and thermodynamic constraints in a single mixed-integer linear program, which can then be solved to find smallest suitable modules from a given reaction database. We illustrate the applicability of MEMO by calculating regeneration modules for the recently published synthetic CETCH cycle for *in vitro* carbon dioxide fixation. We demonstrate that MEMO is very flexible in taking into account the diverse constraints of the CETCH cycle (e.g., regeneration of 1 ATP, 4 NADPH and of 1 acetyl-group without net production of CO<sub>2</sub> and with permitted side production of malate) and is able to determine multiple solutions in reasonable time in two large reaction databases (MetaCyc and BiGG). The most promising regeneration modules found utilize glycerol as substrate and require only 8 enzymatic steps. It is also shown that some of these modules are robust against spontaneous loss of cofactors (e.g., oxidation of NAD(P)H or hydrolysis of ATP). Furthermore, we demonstrate that MEMO can also find cell-free production systems with integrated product synthesis and cofactor regeneration. Overall, MEMO provides a powerful method for finding metabolic modules and for designing cell-free production systems as one particular application.

**KEYWORDS:** synthetic biology, design of cell-free systems, constraint-based modeling, metabolic networks, cofactor regeneration, mixed-integer linear programming

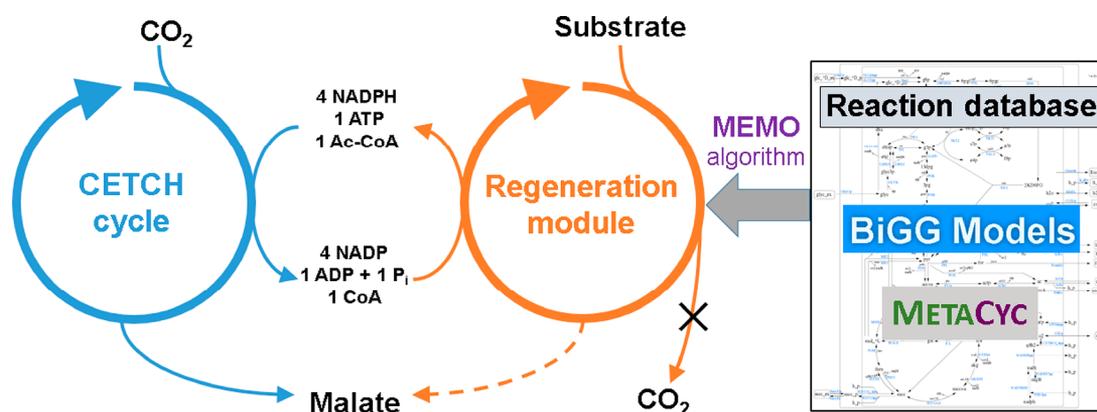
The biobased production of chemicals, materials, and fuels represents one important building block to overcome the dependency on fossil feedstocks. Apart from the replacement of petrochemical feedstocks with biorenewable resources, this includes the development and use of biochemical conversion systems to synthesize value-added products. Biochemical production processes often rely on microbial fermentation where specifically selected or/and genetically (re)designed micro-organisms convert a given substrate to the target product. The optimization of suitable cell factories with superior performance is a key topic of metabolic engineering.<sup>1</sup> This top-down approach has proven successful for biosynthesis

of a number of industrially relevant products.<sup>2</sup> An alternative—mainly bottom-up—approach is the use of cell-free production systems, where purified enzymes and selected metabolites are combined *in vitro* to facilitate bioconversion.<sup>3–6</sup> While cell-free

Received: October 24, 2019

Published: February 18, 2020



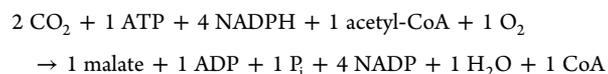


**Figure 1.** Application of the MEMO algorithm for finding suitable regeneration modules for the cell-free operation of the CETCH cycle.

systems initially consisted only of a single or a few enzymes (e.g., in the classical field of biocatalysis), the complexity and scope of cell-free bioproduction systems has grown rapidly in recent years. This includes the successful *in vitro* reconstruction of entire central metabolic pathways<sup>7,8</sup> and of cell-free production systems for a range of chemicals, for example, ethanol, isobutanol,<sup>9</sup> 1,3-propanediol,<sup>10</sup> and monoterpenes.<sup>8</sup> Moreover, cell-free systems allow the construction of completely new (synthetic) pathways such as the recently published CETCH cycle for carbon dioxide fixation which involves 17 reaction steps.<sup>11</sup> As detailed in the literature,<sup>4,5</sup> both cell-based and cell-free production processes have particular advantages and weaknesses. For example, while cellular systems provide a robust environment for enzyme operation and enable a cheap and integrated synthesis of the required pathway components, cell-free systems allow one to assemble completely new pathways, to precisely control system parameters, and to circumvent the intricate cellular (e.g., genetic) regulation often impeding higher yields and productivities of whole-cell systems. Furthermore, since biomass synthesis does not take place in cell-free systems, up to 100% conversion of the substrate into the desired product is possible and pathways involving toxic compounds can be used *in vitro* that would not be feasible *in vivo*. However, one particular challenge for cell-free systems is the balancing and regeneration of cofactors (mainly but not exclusively ADP/ATP and NAD(P)/NAD(P)H) consumed or produced by the *in vitro* metabolic pathway.<sup>4</sup> While cells can balance the net consumption or net production of those cofactors *via* multiple dedicated metabolic pathways, their imbalance in cell-free systems will quickly stop the operation of the *in vitro* system. Accordingly, different regeneration submodules have been used to replenish cofactor pools. For example, for regenerating ATP from ADP, Schwander *et al.*<sup>11</sup> added polyphosphate and a polyphosphate transferase to the carbon dioxide fixing CETCH cycle. A similar approach has been used for cell-free production of nucleotide sugars.<sup>12</sup> Although this kind of ATP regeneration system is simple and relatively cheap and worked sufficiently in these examples, it has the disadvantage of accumulation of inorganic phosphate which, for thermodynamic reasons, may slow down the ATP regeneration step and thus negatively affect the efficiency of the whole system in longer batch runs. Similar negative effects arise when using other compounds with high-energy phosphate bonds (such as PEP), which have frequently been used for cell-free protein synthesis.<sup>3</sup> Alternative ATP regeneration systems have therefore been

developed, for example, *via* generation of acetyl-phosphate from pyruvate<sup>13</sup> (possibly extend by several upstream glycolytic steps<sup>14</sup>), which is then used to produce ATP *via* acetate kinase. This approach circumvents net production of inorganic phosphate but leads in turn to the accumulation of acetate, which may again have adverse (e.g., inhibitory) effects on the efficiency of the ATP regeneration cycle or/and of the whole process when running over longer time periods. Regarding regeneration of NAD(P)H, Schwander *et al.*<sup>11</sup> used formate as an electron donor to replenish the NADPH pool *via* a formate dehydrogenase. While this one-step module is simple and cheap and was fully sufficient to demonstrate the CO<sub>2</sub>-fixing capability of the synthetic CETCH cycle, it would, in a realistic application, negatively affect the overall stoichiometry of CO<sub>2</sub> fixation because CO<sub>2</sub> is released as side product of the formate dehydrogenase reaction (however, if formate is produced electrochemically (from CO<sub>2</sub>), then its use would be carbon neutral).

These examples underline the necessity of a systematic approach to identify suitable regeneration modules for cell-free production systems. As a minimal requirement, a regeneration module must fulfill certain stoichiometric and possibly further specifications so that its coupling with the actual production module leads to an optimal balancing of all cofactors in the entire system. Figure 1 illustrates the situation for the CETCH cycle, which will serve as a running example in this study. As mentioned above, this synthetic cycle is a collection of enzyme-catalyzed reaction steps that assimilate CO<sub>2</sub> *in vitro* to form glyoxylate as primary product.<sup>11</sup> Since its first conception it has been improved in many iterations, and in this study we focus on version 5.0, which includes also reactions that combine glyoxylate with externally provided acetyl-CoA to produce the end product malate.<sup>11</sup> CETCH 5.0 comprises 17 enzyme-catalyzed reactions, thereof 14 core reactions, a catalase reaction decomposing hydrogen peroxide (produced as side product during the CETCH cycle) to water and oxygen, as well as the two aforementioned reactions for regenerating ATP (*via* polyphosphate and polyphosphate kinase) and NADPH (*via* formate and formate dehydrogenase). Leaving out the two reactions for cofactor regeneration the net stoichiometry of the CETCH cycle reads:



Assuming external provision of oxygen and carbon dioxide, the CETCH cycle thus requires for one iteration the provision of 1

ATP, 4 NADPH, and 1 acetyl-CoA (Figure 1). In the original work, ATP and NADPH were regenerated with the two described mechanisms, whereas acetyl-CoA was externally provided. However, since acetyl-CoA is very expensive and because its consumption leads to net accumulation of CoA and thus potentially to inhibitory effects over longer batch runs, we will herein consider the task to find a suitable regeneration module that replenishes not only the pools of ATP (from ADP and P<sub>i</sub>) and NADPH (from NADP) but also of acetyl-CoA (from CoA). For the special context of the CETCH cycle we demand in addition (1) that no carbon dioxide is released as this would compromise the ultimate goal of the CETCH cycle, namely carbon dioxide fixation, (2) that malate is the only allowed organic byproduct of the regeneration module (but its production is not mandatory), and (3) that the algorithm may choose from a given set of possible (typical) substrates (including glucose, glycerol, acetate, and others). Constraint (2) renders the problem significantly harder but ensures that malate is the only organic product of the CETCH cycle when being coupled with the regeneration module, which will simplify downstream processing.

Herein we present MEMO, a computational method to find smallest metabolic modules with specified stoichiometric and thermodynamic constraints. MEMO supports the design of cell-free systems in various regards, and the design of cofactor regeneration modules is one particular application on which we will focus herein. In this context, MEMO takes as input a universe of metabolic reactions and identifies smallest, thermodynamically feasible regeneration modules delivering cofactors with specified stoichiometric requirements, e.g., as posed for the CETCH cycle (Figure 1). MEMO is generically applicable and based on constraint-based modeling techniques. We demonstrate its applicability for the CETCH cycle example showing that even for this case with rather specific requirements metabolic regeneration modules can be found with as few as 8 reactions. As reaction universe we use modified versions of the BiGG<sup>15</sup> and the MetaCyc<sup>16</sup> databases (Figure 1), where reactions between different compartments (and within the membrane compartment) have been removed as their use would complicate the cell-free production system. Apart from the application to find regeneration modules, at the end we will demonstrate that MEMO can also be used to find entire cell-free systems with integrated production pathway and regeneration module ensuring balanced overall conversions.

## METHODS

**Mixed-Integer Linear Program for Calculating Metabolic Modules.** In the following, we describe MEMO, a method to compute, from a given reaction database, metabolic modules fulfilling certain specifications. The method is based on an extended version of OptMDFpathway,<sup>17</sup> which is a mixed-integer linear program (MILP) originally developed to identify pathways with highest thermodynamic driving force (according to the max-min driving force (MDF) definition given in<sup>18</sup>) within a given metabolic reaction network. Using OptMDFpathway as a starting point ensures that the computed metabolic modules are (1) balanced and (2) thermodynamically feasible with a (maximal) overall driving force (MDF) exceeding a specified threshold. However, here we need to extend OptMDFpathway in order to enforce the required stoichiometries of the metabolic modules to be identified and to ensure that we find the *minimal* module (in

terms of number of reactions) with these properties. Furthermore, while the original OptMDFpathway implementation was based on the standard Gibbs free energies  $\Delta_r G'^0$  of reactions, the extended formulation can alternatively use the Gibbs free energies of formation  $\Delta_f G'^0$  of the metabolites.

Given a metabolic network (in our application, this will be a reaction database), we consider the following equations of a classical flux balance analysis (FBA) problem on the reaction rates  $\mathbf{r}$ :

$$\mathbf{N}\mathbf{r} = 0 \quad (1)$$

$$\alpha_i \leq r_i \leq \beta_i \quad (2)$$

$$\mathbf{D}\mathbf{r} \leq \mathbf{d} \quad (3)$$

As usual, the matrix  $\mathbf{N}$  comprises the internal metabolites that are considered to be in steady state (eq 1). We assume that the network contains only irreversible reactions ( $\alpha_i \geq 0$ ), that is, reversible reactions have been split into forward and backward directions. Furthermore, for technical reasons we demand that the upper flux bounds  $\beta_i$  must be finite. Constraint eq 3 can be used to add other inequality constraints (like yield constraints) to specify desired properties. Importantly, eqs 2 and 3 will later also be used to incorporate desired stoichiometric requirements of the metabolic module (in the context of regeneration modules, these will be the required cofactor stoichiometries). The thermodynamic driving forces of the reactions are defined as their negative change of Gibbs free energy and collected in vector  $\mathbf{f}$ . If the standard Gibbs free energies changes of the reactions ( $\Delta_r G'_i$ ) are available and used as thermodynamic parameters then the driving force of reaction  $i$  is given by

$$f_i = -\Delta_r G'_i = -(\Delta_r G'_i{}^0 + RT \cdot \hat{\mathbf{N}}_{\bullet,i}^T \mathbf{x}) \quad (4)$$

$\hat{\mathbf{N}}_{\bullet,i}^T$  is the transposed  $i$ -th column (reaction) of the full stoichiometric matrix  $\hat{\mathbf{N}}$ .  $\hat{\mathbf{N}}$  extends the stoichiometric matrix of the internal metabolites ( $\mathbf{N}$ ) with external metabolites which need not be in steady state.  $R$  is the ideal gas constant,  $T$  the temperature used for determining the Gibbs free energies, and  $\mathbf{x}$  contains the logarithmized metabolite concentrations. As the exact concentration values are typically not known one demands that  $\mathbf{x}$  must comply with certain concentration ranges:

$$\ln(\mathbf{C}_{\min}) \leq \mathbf{x} \leq \ln(\mathbf{C}_{\max}) \quad (5)$$

Eq 4 can be alternatively written with the Gibbs free energies of formation of metabolites ( $\Delta_f G'$ ):

$$f_i = -\Delta_r G'_i = -\hat{\mathbf{N}}_{\bullet,i}^T \Delta_f G' \quad (6)$$

If the standard Gibbs free energy of formation of metabolite  $j$  ( $\Delta_f G'_j{}^0$ ) is known then

$$\Delta_f G'_j = \Delta_f G'_j{}^0 + RT \cdot x_j \quad (7a)$$

otherwise the value of  $\Delta_f G'_j$  is allowed to assume an arbitrary value within a specified range by

$$-L \leq \Delta_f G'_j \leq L \quad (7b)$$

where  $L$  is a sufficiently large constant (e.g., 10 000). Using constraint eq 6 together with eqs 7a or 7b ensures that no thermodynamically infeasible cycles will arise in the identified solutions for the rate vector  $\mathbf{r}$ . This even holds true for eq 7b where the  $\Delta_f G'_j$  is practically unspecified (this approach is similar to st-FBA as proposed by Noor<sup>19</sup>).

In a preprocessing step we determine the minimum ( $f_{i,\min}$ ) and maximum ( $f_{i,\max}$ ) values for each driving force  $f_i$ , which depend on the concentration ranges eq 5 and/or on the value of  $L$  in case constraints of type eq 7b are present.

Each reaction is associated with a binary variable  $z_i$  that must be 1 when there is a flux through this reaction. This is achieved by the constraints

$$r_i \leq z_i \beta_i \quad (8)$$

In order to ensure that the minimal driving force of all active reactions ( $r_i > 0$ ) is greater than a given value  $B > 0$  (the driving force of inactive reactions with zero flux is not taken into account) the following constraints are added to the optimization problem:

$$f_i + (1 - z_i) \cdot M_i \geq B \quad (9)$$

With  $K = \max(f_{i,\max})$ , we set  $M_i = K - f_{i,\min}$ , and with  $B \leq K$ , these constraints are always fulfilled for all reactions with  $r_i = z_i = 0$ . For all reactions with  $z_i = 1$  it must hold that  $f_i \geq B$ . Thus,  $B$  acts as a demanded lower bound on the MDF of the solution.

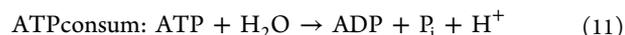
Finally, in order to minimize the number of active reactions in a given solution the objective function

$$\underset{\mathbf{x}, \mathbf{r}, \mathbf{z}}{\text{minimize}} \sum_i z_i \quad (10)$$

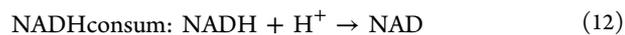
is imposed, subject to the relevant constraints in eqs 1–9. Its solution ( $\mathbf{x}$ ,  $\mathbf{r}$ ,  $\mathbf{z}$ ) will deliver a flux distribution  $\mathbf{r}$  where, given the calculated logarithmized concentration vector  $\mathbf{x}$ , each active reaction has a driving force of at least  $B$ . With the enforced minimization of the number of active reactions, the rate vector  $\mathbf{r}$  of the solution will represent an elementary flux vector (EFV<sup>20</sup>) (cf. also ref 17). This means that in networks, where a full enumeration of EFVs is computationally feasible, the optimal solution could be identified by screening the set of these vectors, and even a ranking of all possible minimal solutions could be generated. However, herein we will deal with large networks with thousands of reactions where this is not possible. Instead, the optimal solution of the mixed-integer linear programming (MILP) problem posed by eqs 1–10 (with  $\mathbf{x}$  and  $\mathbf{r}$  as continuous and  $\mathbf{z}$  as integer variables) will be identified with appropriate MILP solvers.

**Incorporation of Required Stoichiometries of Regeneration Modules.** Generally, when using MEMO to find certain metabolic modules, the stoichiometric input/output requirements (i.e., the net conversion) of the module must be specified appropriately. In particular, employing MEMO to identify suitable cofactor regeneration modules for a given cell-free production system, one needs to specify (a) a set of possible substrates that can be used, (b) certain stoichiometries of the cofactors to be regenerated, and (c) allowable byproducts of the regeneration module. The set of allowed substrates and byproducts can be easily specified by setting the respective upper bounds  $\beta_i$  of the substrate uptake and product excretion reactions either to  $\beta_i > 0$  (allowed substrate or byproduct) or to zero (not allowed). Compared to classical metabolic network models, which often allow synthesis of, e.g., certain fermentation products, typically only very few (or even no) byproducts will be allowed for regeneration modules. To finally consider the demanded net synthesis of cofactors such as ATP, NAD(P)H, or others, we need to include appropriate demand (pseudo)reactions. For example, the provision of ATP

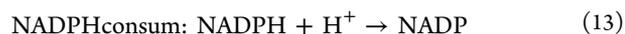
by the regeneration module (and its consumption in the cell-free production system) is modeled by including a reaction that consumes (hydrolyzes) ATP yielding ADP and orthophosphate:



Likewise, for the provision (regeneration module) and consumption (production module) of NADH and NADPH, we may include the reactions



and



Other cofactors or metabolites to be regenerated can be included as well. In certain applications, NAD(P)H might occur as byproducts of the production module, in this case the regeneration module would need to replenish the NAD(P) pool; hence, the reactions 12 or 13 would be specified in the reverse direction. Finally, the concrete requirements of the cofactor stoichiometries are configured by setting the respective bounds of these exchange reactions in eq 2. For example, a demanded stoichiometric ATP/NADH ratio of 2:1 would be enforced by setting  $\alpha_{\text{ATPconsum}} = \beta_{\text{ATPconsum}} = 2$  and  $\alpha_{\text{NADHconsum}} = \beta_{\text{NADHconsum}} = 1$  effectively fixing  $r_{\text{ATPconsum}} = 2$  and  $r_{\text{NADHconsum}} = 1$ . Sometimes the sum of the produced reduction equivalents must give a certain number (e.g.,  $r_{\text{NADHconsum}} + r_{\text{NADPHconsum}} = 4$ ); such constraints can be incorporated in eq 3.

**Reaction Databases (Universal Networks).** An essential input for MEMO is a metabolic network or, more generally, a reaction universe or database, within which the metabolic module can be identified. Due to the here intended use of MEMO to identify metabolic regeneration modules for cell-free production systems, we are free to compile enzymatic reaction steps from different organisms. On the other hand, we assume that the use of cellular compartments (e.g., mitochondria, chloroplasts etc.) within the cell-free production system is not possible or to be avoided as they would add an additional layer of complexity. We therefore do not consider reactions that involve metabolites from different compartments. Likewise, reactions involving quinone metabolites (which reside in membranes) are not taken into account. Of course, these constraints can be adapted or relaxed for certain applications.

In the following we describe the setup of two such universal network models, which were derived from the BiGG<sup>15</sup> and from the MetaCyc<sup>16</sup> database, respectively.

**BiGG Model.** The BiGG model is a fusion of the 85 models present in the BiGG database<sup>15</sup> (accessed February 2018). As a first step, the models were successively merged while at the same time technical reactions (exchanges, demands, sinks) were removed. For model merging the mergeTwoModels function of the COBRA toolbox<sup>21</sup> was used. This function removes duplicate reactions via the checkDuplicateRxn function, which we extended with a new option that recognizes duplicate reactions by their stoichiometry (up to a scalar factor) and also properly handles the reaction reversibility. Next, reactions that involve metabolites from different compartments were removed for the reasons described above. Although the metabolite IDs in the BiGG models are largely unique, some of them have different formulas in different compartments (mostly due to different protonation

states). Such IDs were semiautomatically disambiguated by choosing either a common formula or by making the IDs explicitly distinct. After this the mass balance for the elements C, H, O, N, S, P as well as the charge balance for all reactions were determined and the unbalanced reactions removed. Metabolites occurring in different compartments were mapped onto one ID and the compartment information was removed. Reaction duplicates that had been produced during this process were afterward merged with the extended checkDuplicateRxn function. Finally, reactions that involve any quinone metabolites (which reside in membranes) were deactivated for the calculations. The resulting adapted BiGG network model comprises 5863 metabolites and 8983 reactions. The  $\Delta_f G'^0$  values [kJ/mol] for the metabolites were retrieved, as far as available, from the eQuilibrator API<sup>22</sup> by applying a mapping from BiGG metabolite IDs to KEGG IDs. For reactions in which all involved metabolites have a known  $\Delta_f G'^0$ , reaction reversibility was restricted according to their minimal/maximal driving force. The reversibility of the remaining reactions was left as defined in the BiGG models. For the calculations, driving forces were set up according to eq 6 together with  $\Delta_f G'$  constraints of the form eq 7a or 7b (accordingly, eq 4 was left out).

**MetaCyc Model.** This model was constructed from the flat file distribution of MetaCyc<sup>16</sup> version 22.6, which contains the files compounds.dat (metabolites) and metabolic-reactions.xml, which were parsed for this purpose. As for the BiGG model, only those reactions were kept whose metabolites all come from the same compartment, and the compartment information for the metabolites was then removed. Duplicate reactions were removed, the mass balance for the elements C, H, O, N, S, P as well as the charge balance were determined, and the unbalanced reactions were removed. Furthermore, reactions that involve metabolites without known  $\Delta_f G'^0$  values were removed, and  $\Delta_r G'^0$  values (in MetaCyc given in [kcal/mol]) for the remaining reactions were calculated from the  $\Delta_f G'^0$ . For each reaction, its reversibility was restricted in accordance with its minimal/maximal driving force. Because in this model all reactions have a defined  $\Delta_r G'^0$  value, constraints of type eq 4 were used for the driving forces (accordingly, eqs 6, 7a, and 7b were left out). It should be noted that although the  $\Delta_r G'^0$  values in MetaCyc have been determined by a predecessor of the eQuilibrator method,<sup>22</sup> the actual values for the same reaction can be quite distinct between both. The adapted MetaCyc network model comprises 8964 metabolites and 11634 reactions.

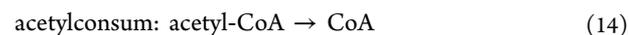
**Implementation and Availability.** Calculations were performed using dedicated functions from (version 2019.3) CellNet-Analyzer,<sup>23,24</sup> including the OptMDFpathway function, which was extended as described in the Methods section. A package with the MATLAB script files for the calculations and with the models (in COBRA format) is available from the following GitHub repository: <https://github.com/ARB-Lab/MEMO.git>. CPLEX 12.8 was used as MILP solver. The calculations were performed on a cluster node with two Intel Xeon 8-core processors and 192 GB RAM.

## RESULTS AND DISCUSSION

**Computing Regeneration Modules for the CETCH Cycle.** To demonstrate the applicability of our MEMO method we employed it, together with the two adapted BiGG and MetaCyc reaction databases (see Methods), to compute regeneration modules that perfectly complement the

CETCH cycle.<sup>11</sup> The precise stoichiometric requirements of the CETCH 5.0 cycle were already described in the introduction section and are illustrated in Figure 1. The following constraints were accordingly set for the MEMO optimization problem:

In one iteration, the CETCH cycle consumes 4 NADPH, 1 ATP, and 1 acetyl-CoA (Figure 1). Accordingly, the flux through the NADPH consumption reaction 13 was fixed to 4, and the NADH consumption reaction was fixed to zero flux. The consumption of acetyl-CoA was accounted for by introducing another artificial consumption reaction



and by fixing its flux value to 1.

When computing regeneration modules, it is important to precisely specify the allowed end products (apart from the cofactor requirements). An accumulation of undesired by-products may not only slow down the regeneration module (and possibly also the actual production system) but also compromise the economic viability of the whole process, for example, due to necessary separation of the main product from byproducts in downstream processing. The main product of the CETCH cycle is malate, and we therefore allowed a net synthesis of this metabolite in the regeneration module as well. Thus, a malate exchange reaction was introduced in the model but without fixing its flux (its synthesis in the regeneration module is thus permitted but not mandatory *per se*). Additionally, water and oxygen can be exchanged with the environment. Since the CETCH cycle aims to fix carbon dioxide, we did not permit net production of carbon dioxide as this would compromise the efficiency of the whole process. Furthermore, in the BiGG model, protons can also be exchanged while in the MetaCyc model charged metabolites are balanced with protons in their exchange reactions (e.g., when acetate is exchanged this occurs together with one proton). Because all non-exchange reactions are H- and charge-balanced any proton exchange flux in the BiGG model only reflects the difference between the dissociated protons of substrate and product.

We separately computed regeneration modules for 9 simple and cheap substrates yielding a set of distributed entry points in the central metabolism: acetate, glycerol, glucose, sorbitol, formate, xylose, lactate, methanol, and succinate. The lower bound  $B$  for the thermodynamic driving force (MDF) in eq 9 was set to 0.01 kJ/mol. The metabolite concentrations were allowed to vary between 100 mM and 1  $\mu$ M, except for CO<sub>2</sub> and bicarbonate for which the lower/upper bound was set to 0.1  $\mu$ M/1 mM. We aimed to calculate the 10 shortest regeneration modules for each substrate. Computation time for each single optimization was limited to 1 h, which, in several cases, was sufficient to reach the guaranteed optimum. Typically, the optimum becomes more difficult to prove with an increasing number of reactions in the smallest regeneration module. Nonetheless, even if the optimum was not proven within 1 h one can usually get a good solution. This can then be evaluated together with the best bound (minimum number of reactions required for any solution) to decide whether or not it may be useful to invest more computation time to try and search for a smaller solution. In a few cases the solver was neither able to prove the infeasibility of the problem nor able to find a feasible solution within 1 h. In such cases the optimization might be repeated with different solver settings

and/or the time limit be extended to see if a solution can be found.

The results of the computations are shown in Table 1 and in the Supporting Information. Within the given time frame,

**Table 1. Smallest CETCH Regeneration Modules (Number of Reactions) and Their Largest Thermodynamic Driving Force (MDF) Found for the Different Substrates in the MetaCyc and in the BiGG Network<sup>a</sup>**

substrate	MetaCyc		BiGG	
	number of reactions in smallest module	max MDF [kJ/mol]	number of reactions in smallest module	max MDF [kJ/mol]
glucose	14	12.08	15	2.96
sorbitol	12	8.83	12	1.47
glycerol	8	18.01	8	12.38
methanol	11	3.33	12	3.27
xylose	13	16.28	16	11.50
lactate	11	8.84	14	1.13
succinate	not found		not found	
acetate	not found		not found	
formate	not found (proven to be infeasible)		not found	

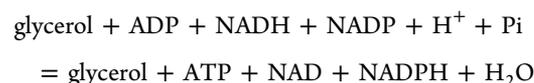
<sup>a</sup>Computation time limited to one hour per solution; at most 10 solutions per substrate were computed. The unit of the MDF values from the MetaCyc model were converted from [kcal/mol] to [kJ/mol].

CETCH regeneration modules were found for 6 out of the 9 substrates in both networks (glucose, glycerol, sorbitol, xylose, lactate, and methanol). All found solutions require between 8 and 16 reactions (excluding the exchange/consumption reactions), which is thus in all cases less than the 17 reaction steps of the entire CETCH cycle. Where both networks offered solutions, they are of the same size or shorter in MetaCyc. In contrast, no solutions were found in either network for the substrates acetate, succinate and formate. In the MetaCyc network, flux variability analysis (without thermodynamic constraints) already reveals that no solution is possible for formate.

The shortest regeneration modules were found with glycerol as substrate and require only 8 reactions (proven by the solver to be optimal) in both networks. One of the MetaCyc solutions of size 8 is shown in Figure 2. In this solution, glycerol is oxidized in two steps to glycerate which in total yields two NADPH. Glycerate is then phosphorylated to 2-phospho-glycerate (2-PG) *via* the backward direction of the phosphoglycerate phosphatase. Interestingly, this phosphatase has a  $\Delta_r G'^0$  of 6.7 kcal/mol and thus favors actually the kinase direction. 2-PG is then converted by the enolase to phosphoenol-pyruvate (PEP) where the pathway splits into two branches. In the first branch PEP is dephosphorylated to pyruvate from which acetyl-CoA, CO<sub>2</sub>, and NADH are produced. In the second branch PEP is carboxylated to oxaloacetate with concomitant ATP formation. Under consumption of NADH from the first branch oxaloacetate is then converted to malate.

For the BiGG network, one of the two solutions with 8 reactions (solution 1 in Supporting Information) is shown in Figure 3. Here, three molecules of glycerol (with concomitant NADPH production) are converted to dihydroxyacetone which is then phosphorylated. From dihydroxyacetone-phosphate (dhap) the pathway splits in two branches. In the

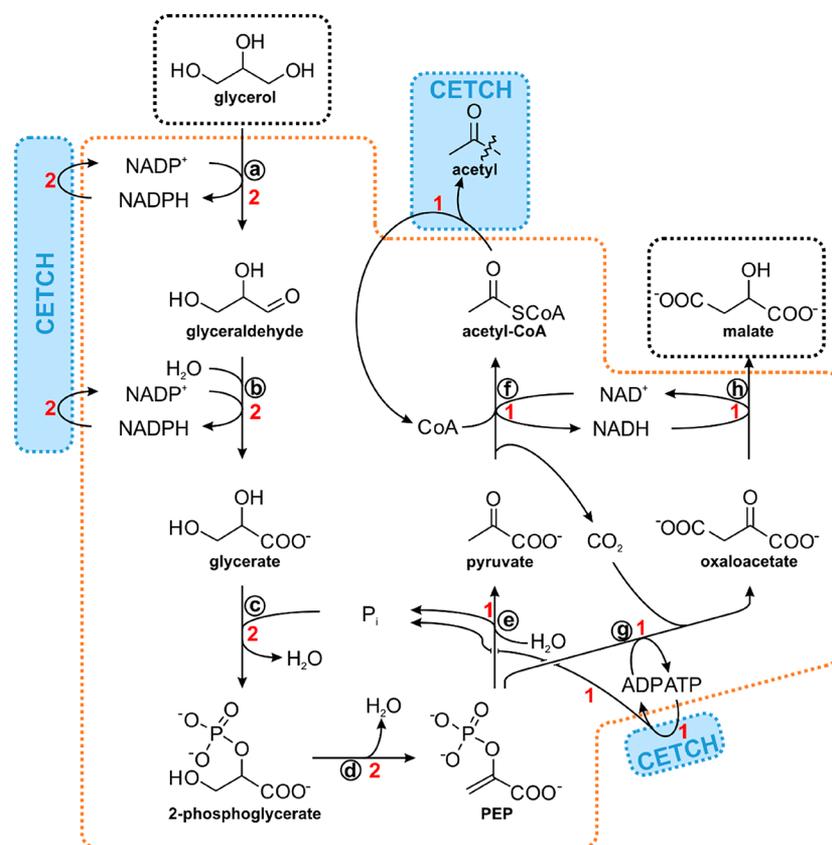
first branch two dhap are converted to methylglyoxal from which pyruvate and NADPH are produced. By two separate reactions, pyruvate is converted to acetyl-CoA and malate respectively. In the second branch from dhap, this compound is converted to glycerol-3-phosphate from which ATP and glycerol are produced. Therefore, this regeneration module contains a cyclic (sub)pathway through glycerol constituted by reactions GLYCD, r0242, r0202, and GLYK. However, this does not constitute a thermodynamically infeasible cycle because its net stoichiometry is not empty:



Although the  $\Delta_r G'^0$  of this partial overall reaction is 27.4 kJ/mol, its  $\Delta_r G'$  can indeed become negative under a suitable metabolite concentration profile, which is compatible with the concentration bounds used for the calculations here.

To further analyze the regeneration modules we calculated the elementary modes<sup>20</sup> within each module (*cf.* Supporting Information). This revealed that each of the two solutions described above is the only steady-state solution within its regeneration module. However, there are also regeneration modules that contain more than one elementary steady-state solution, and therefore can potentially realize multiple cofactor regeneration requirements. As an example, consider the second glycerol solution found in the MetaCyc model (Figure 4), where two elementary modes exist. The difference between the modes is that in the first mode the phosphorylation of (the two) glycerate molecules requires no ATP (reaction (c) in Figure 4), whereas ATP is consumed in the second mode (reaction (d) in Figure 4). The first mode thus has a 2:4:1 (ATP:NADPH:acetyl-CoA) ratio, while the second mode produces a 0:4:1 ratio. When both modes operate together in a 1:1 ratio the result is the desired 1:4:1 ratio, which was enforced by setting the respective fluxes of the cofactor exchange reactions when computing the modules. While the variability with two modes may, at a first glance, appear disadvantageous, it should be noted that when coupling production and regeneration module, operation of the regeneration module will push the net stoichiometry of the regeneration module toward what is consumed in the net by the production module. Moreover, regeneration modules consisting of multiple elementary modes have the advantage that they can, through variation of the flux ratio between the modes, compensate for potential fluctuations in the ATP level, caused, *e.g.*, by its spontaneous hydrolysis. Thereby, this regeneration module may act in a similar manner as the ATP rheostat described in ref 25.

Of all glycerol solutions, the discussed MetaCyc solution 2 in Figure 4 actually is one of those with the highest MDF for this substrate (18.01 kJ/mol, Table 1). The MetaCyc solution in Figure 2 has an MDF of only 3.22 kJ/mol, which is similar to that for the BiGG solution in Figure 3 (3.34 kJ/mol). Therefore, from a thermodynamic standpoint, MetaCyc solution 2 would be preferable. It has been suggested that a MDF of 3 kJ/mol should be sufficient to allow large net fluxes through all participating reactions.<sup>18</sup> Table 1 shows that not for all substrate/network combinations considered here this threshold can be reached with the solutions we calculated so far. However, such a criterion could be enforced during the calculations by setting the value of *B* in eq 8 to a desired minimum threshold.



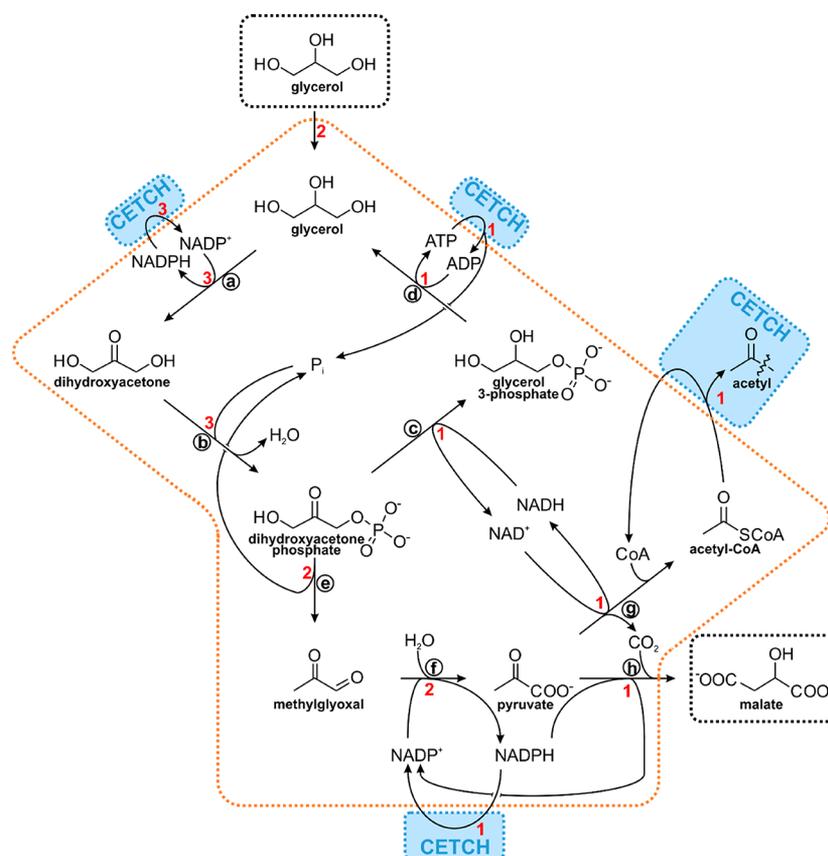
**Figure 2.** CETCH regeneration module of size 8 found with glycerol as substrate in the MetaCyc database (*cf.* MetaCyc solution 6 for glycerol in Supporting Information). Including the CETCH cycle, the overall reaction is  $2 \text{CO}_2 + 2 \text{glycerol} + \text{O}_2 \rightarrow 2 \text{malate} + 2 \text{H}_2\text{O} (+ 4 \text{ protons})$ . The reactions (MetaCyc identifiers) are (a) GLYCEROL-DEHYDROGENASE-NADP+-RXN, (b) RXN-15115, (c) PHOSPHOGLYCERATE-PHOSPHATASE-RXN, (d) 2PGADEHYDRAT-RXN, (e) PHOSPHOENOLPYRUVATE-PHOSPHATASE-RXN, (f) PYRUVDEH-RXN, (g) PEPCARBOXYKIN-RXN, (h) MALATE-DEH-RXN. The red numbers indicate the (relative) fluxes needed to deliver the cofactors in required stoichiometries to the CETCH cycle (indicated in blue).

Following glycerol, methanol is the substrate that allows for the next smallest regeneration modules with 11 reactions in MetaCyc and 12 reactions in BiGG. Exemplarily we discuss the first solution found in MetaCyc and BiGG. In both solutions, the first step after methanol uptake is its conversion to formaldehyde with the concomitant formation of NADH. After this, formaldehyde is bound to a tetrahydrofolate species as a methyl-group. From a part of the methylated tetrahydrofolate the methyl-group is released as glycine under the consumption of  $\text{CO}_2$ ,  $\text{NH}_4$  (which will be released again in later steps), and NADH. A part of this glycine combines with the remaining methylated tetrahydrofolate to produce serine. The remaining glycine is converted to glyoxylate and with acetyl-CoA further to malate, using different reactions in the two models. In the CETCH cycle, the fusion of glyoxylate and acetyl-CoA takes place as well,<sup>11</sup> and it is conceivable that the enzymes involved there could perform the same function if combined with the other reactions from the regeneration modules thereby reducing the number of additional regeneration enzymes needed. In both regeneration modules, serine is converted to pyruvate, which is the basis for acetyl-CoA formation *via* the pyruvate dehydrogenase. Also, both solutions include a NADH to NADPH transhydrogenase. Nonetheless, there are some differences: The MetaCyc solution requires one reaction less because ATP regeneration occurs together with malate formation and NADPH regeneration in the pyruvate dehydrogenase reaction, while in BiGG both regenerations

are coupled to an alpha-keto-glutarate/glutamate/glutamine conversion cycle. Furthermore, the BiGG solution requires 4 methanol (and yields 0.5 malate), while the MetaCyc solution uses 5 methanol and 1  $\text{CO}_2$  (and produces 1 malate). Both solutions have a relatively low MDF (3.33 kJ/mol for the MetaCyc and 0.54 kJ/mol for the BiGG solution).

Of the three carbohydrates glucose, xylose, and sorbitol, the last allows for the smallest regeneration modules with 12 reactions in both networks, making it the preferable substrate of these three. Despite the fact that an acetyl-group must be generated for the CETCH-cycle, acetate as substrate turned out not to be promising because in neither network a solution was found. Formate alone cannot be used for the regeneration tasks, probably because of its highly oxidized nature and the restriction that no net  $\text{CO}_2$  must be produced. For succinate no solutions were found either, despite the fact that its conversion to malate should quite easily render the required reduction equivalents. Lactate is then the only of the four acids for which regeneration modules were found. In MetaCyc, the smallest module contains only 11 reactions, while in BiGG it needs 14 reactions.

**Thermodynamics of the Combined Production (CETCH) and Regeneration Modules.** The MDF in Table 1 and discussed above are given for the stand-alone regeneration modules and can decrease when the modules are integrated with the CETCH cycle. However, it is straightforward to calculate the resulting MDF of the



**Figure 3.** CETCH regeneration module of size 8 found with glycerol as substrate in the BiGG database (*cf.* BiGG solution 1 for glycerol in [Supporting Information](#)). Including the CETCH cycle, the overall reaction is  $2 \text{ CO}_2 + 2 \text{ glycerol} + \text{O}_2 \rightarrow 2 \text{ malate} + 2 \text{ H}_2\text{O} + 4 \text{ protons}$ . The reactions (BiGG identifiers) are (a) GLYCDy, (b) r0242, (c) r0202, (d) GLYKm, (e) MGSA, (f) ALR, (g) PDHm, (h) ME\_x. The red numbers indicate the (relative) fluxes needed to deliver the cofactors in required stoichiometries to the CETCH cycle (indicated in blue).

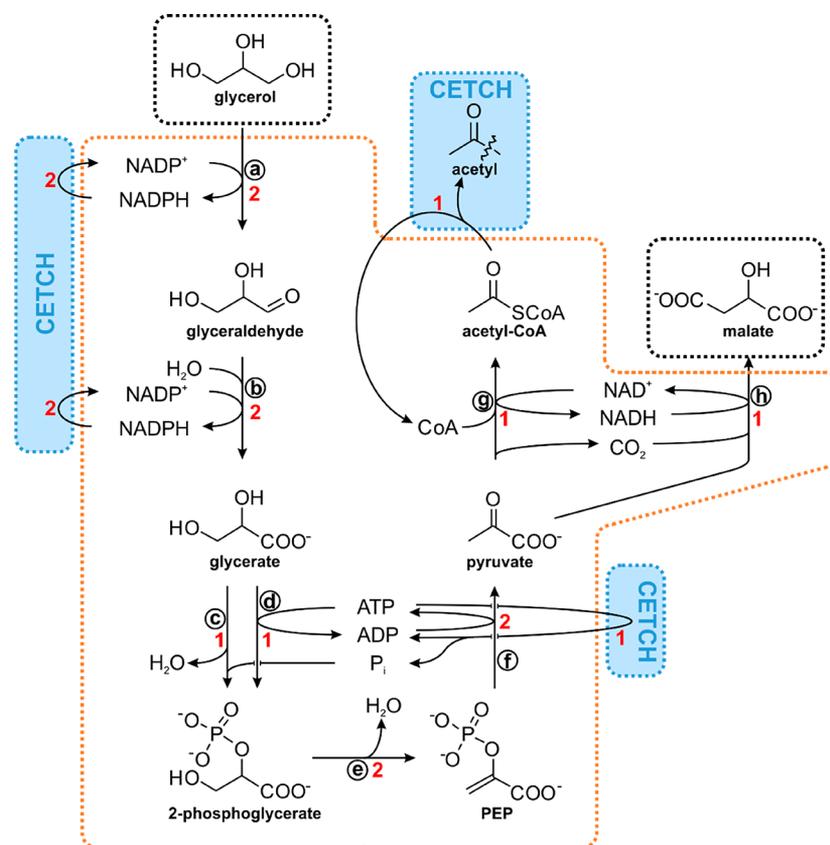
integrated system by taking the union of the regeneration module reactions (without the consumption pseudoreactions) and the CETCH reactions and recalculating the MDF. The results are shown in the [Supporting Information](#), and it turns out that the regeneration modules shown in [Figures 2–4](#) are still thermodynamically feasible when combined with the CETCH cycle, although with reduced MDF. However, several of the other solutions have an  $\text{MDF} < 0$  when combined with the CETCH cycle, which means that they would not be suitable regeneration modules in this context. It is possible to exclude such solutions during the MILP optimization directly by enforcing, *e.g.*, the CETCH reactions to be active with the required fluxes (using  $\alpha$  and  $\beta$  of [eq 2](#)) instead of the consumption pseudoreactions ([eqs 11–14](#)).

For the regeneration modules that are thermodynamically feasible when combined with the CETCH cycle, we also calculated the concentration ranges of the participating metabolites that are necessary to keep the  $\text{MDF} \geq 0.01$  (*cf.* [Supporting Information](#), only those metabolites are listed whose lower or upper concentration bound differs from the default concentration bound). Such results are potentially useful for assessing the overall practicality of the pathway.

**Regeneration Modules with Multiple Substrates and Alternative Cofactor Requirements.** So far, we only considered single inputs, but it is also possible to have multiple substrate uptakes open together to see whether this would allow for smaller regeneration modules. We therefore calculated 20 solutions in each network with all 9 substrate

uptakes open (*cf.* [Supporting Information](#)). It turns out that all these solutions require 8 reactions which means that, in this setting, multiple substrates do not make smaller regeneration modules possible. In fact, for the BiGG model, all 20 solutions use glycerol as single substrate only. All MetaCyc solutions also use glycerol, but there are some that co-utilize acetate or acetate and formate. Interestingly, neither acetate nor formate are usable as single substrate, but might be useful in conjunction with glycerol.

All the results above fulfill the requirements of the CETCH 5.0 cycle, but earlier CETCH versions had different cofactor requirements.<sup>11</sup> For instance, if the second reductive carboxylation step (acrylyl-CoA to methylmalonyl-CoA) is replaced by an ATP-dependent carboxylation step (propionyl-CoA to methylmalonyl-CoA) as in CETCH 1.0, then one iteration of this modified cycle requires 3 NADPH, 2 ATP and 1 AcCoA (instead of 4 NADPH, 1 ATP, and 1 AcCoA in CETCH 5.0). We repeated the single substrate calculations with these new requirements to assess the changes in minimal regeneration module sizes. The results (*cf.* [Supporting Information](#)) show that, depending on the network and the substrate, the minimal module sizes differ by at most one reaction. In the BiGG model, the minimal size increases slightly for glycerol and lactate while it remains unchanged for the other substrates. For the MetaCyc model, the size decreases for glucose, sorbitol, and methanol, but increases for xylose and lactate. Thus, whether trading one NADPH for



**Figure 4.** CETCH regeneration module of size 8 found with glycerol as substrate in the MetaCyc database (*cf.* MetaCyc solution 2 for glycerol in Supporting Information). Including the CETCH cycle, the overall reaction is  $2 \text{CO}_2 + 2 \text{glycerol} + \text{O}_2 \rightarrow 2 \text{malate} + 2 \text{H}_2\text{O} (+ 4 \text{ protons})$ . The reactions (MetaCyc identifiers) are (a) GLYCEROL-DEHYDROGENASE-NADP+-RXN, (b) RXN-15115, (c) PHOSPHOGLYCERATE-PHOSPHATASE-RXN, (d) GKI-RXN, (e) 2PGADEHYDRAT-RXN, (f) PEPDEPHOS-RXN, (g) PYRUVDEH-RXN, (h) RXN-19748. The red numbers indicate the (relative) fluxes needed to deliver the cofactors in required stoichiometries to the CETCH cycle (indicated in blue).

one ATP is favorable or not depends on the substrate used in the regeneration module.

**Computing Modules with Integrated Production and Regeneration.** So far, the focus of MEMO has been on regeneration modules for cofactors, which are often required for cell-free production system. However, MEMO is more general and can search for any metabolic module as long as the module specification can be formulated with the type of constraints used in MEMO. In fact, with MEMO it is also possible to design entire production systems with desired input-output stoichiometries where production and cofactor regeneration are integrated. Computing such integrated modules comes with the advantage that thermodynamic infeasibilities cannot arise when merging production and regeneration module (as seen for some regeneration modules computed for the CETCH cycle). As an example, we consider the production of the monoterpene limonene from glucose which has recently been achieved with a cell-free system.<sup>8</sup> This particular conversion comprises 23 reaction steps and produces one mol limonene per three mol glucose (Supporting Figure 1 in the original ref 8). It basically consists of two parts, the glycolysis which converts glucose to acetyl-CoA and cofactors and the mevalonate pathway that uses acetyl-CoA together with the cofactors to produce limonene. Because the glycolysis produces more reduction equivalents than needed for the mevalonate pathway the system has been designed to incorporate a purge valve that (i) removes excess reduction equivalents (*via* NoxE, an NADH oxidase) and (ii) produces

balanced amounts of NADH/NADPH (by a combined use of a NAD-specific and of a NADP-specific glyceraldehyde-3-phosphate dehydrogenase). We checked whether this system is in the solution space of the MetaCyc network by fixing the fluxes of the glycolysis, mevalonate pathway, and purge valve to the values required to the production of one limonene from three glucose (Supporting Information). This leads to a solution which is the same as in ref 8, which shows that this limonene production system can in principle be found by MEMO. To look for alternative production systems, we specified a minimal yield of 1/3 limonene per glucose as sole constraint and then calculated 10 smallest solutions (*cf.* Supporting Information). All the found solutions require 21 reactions and thus less than for the original system. However, the MDF of the original system is higher (1.03 kcal/mol = 4.31 kJ/mol) than the best MDF of the identified alternative solutions (0.75 kcal/mol = 3.14 kJ/mol). One of the computed solutions is similar to the original system<sup>8</sup> in that it uses many reactions from the glycolysis, but the others mix glycolytic reactions with those of the Entner–Doudoroff pathway. There also are variations as to how cofactor production and consumption are tied together, but in all cases excess reduction equivalent removal proceeds *via* NAD(P)H oxidases.

## CONCLUSIONS

In this work, we presented MEMO, a generic approach to identify smallest metabolic modules fulfilling specified stoichiometric and thermodynamic constraints. Although

MEMO has many possible applications, herein we focused on its use for finding regeneration modules for cell-free production systems. As exemplified with the CETCH cycle, MEMO is very flexible in taking into account diverse requirements (e.g., regeneration of two cofactors plus one acetyl-group with specific stoichiometry, net production of CO<sub>2</sub> not permitted, net synthesis of malate allowed, minimum threshold for MDF) and is able to enumerate multiple solutions in reasonable time also in very large universal networks. A method that may hold similar capabilities but was so far not used for the design of cell-free (regeneration) modules is optStoic.<sup>26</sup> For some predefined reactants and products, this two-step approach identifies in a first step overall conversions that maximize a given objective function (e.g., maximization of product yield). The calculations in this first step are solely based on stoichiometric balance and Gibbs energy changes of the reactants and products without taking into account the reaction steps between them. The actual metabolic pathway or network (generating the overall conversions found in the first step) is identified in the second step by searching for suitable combinations of reactions from a given reaction database. While the applicability of this approach has been proven in several case studies, for example, for finding alternative pathways in metabolic networks,<sup>26</sup> our approach is simpler and requires only one single optimization step to identify stoichiometrically and thermodynamically feasible solutions. In fact, overall conversions found in the first step of optStoic might not have a corresponding pathway in the reaction database used in the second step. Moreover, even if a stoichiometrically balanced pathway has been identified in the second step of optStoic, it may happen that it is thermodynamically infeasible within the defined concentration ranges of the intermediate metabolites (optStoic tests, in a preprocessing step, separately for each reaction whether it can proceed in a certain direction with the given metabolite concentration ranges, however, it does not ensure that a metabolite concentration vector within the specified concentration ranges exits where all reactions of the found pathway are active in the required direction). MEMO directly accounts for thermodynamic feasibility via the MDF approach and it even allows one to set lower thresholds for the MDF thus providing a flexible approach for an integrated search of metabolic modules.

The regeneration modules found for the CETCH cycle appear very interesting and, in contrast to the original system, they all avoid accumulation of side products and use of expensive compounds such as acetyl-CoA. The modules found with glycerol as substrate are especially promising and would require only 8 metabolic reactions (enzymes) which is even less than the half of the enzymes used so far in the CETCH cycle. The analysis of the glycerol solutions also revealed that the modules, although with identical net stoichiometry, may nevertheless have different properties. Apart from their thermodynamic driving force, this concerns in particular the phenomenon that the modules may either consist of a single elementary mode or of a combination of modes. In the first case, the demanded stoichiometry will be exactly fulfilled when the substrate is converted by the module. In the second case, the demanded stoichiometry is *feasible* with the respective module, however, alternative output stoichiometries would also be possible. When coupling production and regeneration module, operation of the production module will push the net stoichiometry of the regeneration module toward that of the

production module. However, regeneration modules consisting of multiple elementary modes have the advantage that they can balance the spontaneous loss (decay) of cofactors (e.g., oxidation of NAD(P)H or hydrolysis of ATP), which, in the long run, may lead to problems in modules with fixed stoichiometries (single modes). Such flexible regeneration modules have thus similar (desired) functionalities as the recently published ATP rheostat<sup>25</sup> or as purge valves.<sup>8,27</sup>

Although the application focus herein was on regeneration modules, MEMO is much more general and can be used to search for other types of cell-free (or even intracellular) metabolic modules as well. As one additional application for cell-free systems, we used the example of limonene synthesis to demonstrate that MEMO also supports the design of entire cell-free production systems with *integrated* production and regeneration: instead of fixing first the production module and finding then a suitable regeneration module, MEMO can also directly—in one step—identify the smallest integrated cell-free system with desired input–output stoichiometries. Overall, together with the adapted MetaCyc and BiGG master networks, our MEMO approach thus provides a powerful and flexible framework for designing cell-free production systems and can likewise be used for detecting or/and designing (intra)cellular metabolic modules with desired properties.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acssynbio.9b00434>.

Tables with the different computed regeneration modules for the CETCH cycle (separately for the MetaCyc and the BiGG reaction database) and with the computed (integrated) cell-free production systems for limonene (only MetaCyc) (XLS)

## ■ AUTHOR INFORMATION

### Corresponding Author

Steffen Klamt – Max Planck Institute for Dynamics of Complex Technical Systems, 39106 Magdeburg, Germany; [orcid.org/0000-0003-2563-7561](https://orcid.org/0000-0003-2563-7561); Email: [klamt@mpi-magdeburg.mpg.de](mailto:klamt@mpi-magdeburg.mpg.de)

### Author

Axel von Kamp – Max Planck Institute for Dynamics of Complex Technical Systems, 39106 Magdeburg, Germany

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acssynbio.9b00434>

### Author Contributions

SK conceived the study. AvK implemented the MEMO algorithm and performed the calculations. Both authors analyzed the results. Both authors wrote and approved the manuscript.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We acknowledge and thank Pavlos Bekiaris for providing the mapping between BiGG and KEGG metabolite identifiers and retrieving the  $\Delta_f G^{r0}$  for the BiGG model via the eEquilibrator API. We are grateful to Simon Boecker for helping in drawing

Figures 2–4. This work was financially supported by the German Ministry of Education and Research (FKZ: 031B0524B, 031L0104B).

## ■ ABBREVIATIONS

EFV, elementary flux vector; MDF, max-min driving force; MILP, mixed-integer linear program.

## ■ REFERENCES

- (1) Nielsen, J., and Keasling, J. D. (2016) Engineering Cellular Metabolism. *Cell* 164, 1185–1197.
- (2) Lee, S. Y., Kim, H. U., Chae, T. U., Cho, J. S., Kim, J. W., Shi, J. H., Kim, D. I., Ko, Y.-S., Jang, W. D., and Jang, Y.-S. (2019) A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* 2, 18–33.
- (3) Lim, H. J., and Kim, D.-M. (2019) Cell-free metabolic engineering: recent developments and future prospects. *Methods Protoc.* 2, 33.
- (4) Claassens, N. J., Burgener, S., Vögeli, B., Erb, T. J., and Bar-Even, A. (2019) A critical comparison of cellular and cell-free bioproduction systems. *Curr. Opin. Biotechnol.* 60, 221–229.
- (5) Dudley, Q. M., Karim, A. S., and Jewett, M. C. (2015) Cell-free metabolic engineering: biomanufacturing beyond the cell. *Biotechnol. J.* 10, 69–82.
- (6) Wilding, K. M., Schinn, S. M., Long, E. A., and Bundy, B. C. (2018) The emerging impact of cell-free chemical biosynthesis. *Curr. Opin. Biotechnol.* 53, 115–121.
- (7) Krutsakorn, B., Honda, K., Ye, X., Imagawa, T., Bei, X., Okano, K., and Ohtake, H. (2013) In vitro production of n-butanol from glucose. *Metab. Eng.* 20, 84–91.
- (8) Korman, T. P., Opgenorth, P. H., and Bowie, J. U. (2017) A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nat. Commun.* 8, 15526.
- (9) Guterl, J. K., Garbe, D., Carsten, J., Steffler, F., Sommer, B., Reiß, S., Philipp, A., Haack, M., Rühmann, B., Koltermann, A., Kettling, U., and Brück, T. (2012) Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. *ChemSusChem* 5, 2165–2172.
- (10) Rieckenberg, F., Ardao, I., Rujananon, R., and Zeng, A. P. (2014) Cell-free synthesis of 1,3-propanediol from glycerol with a high yield. *Eng. Life Sci.* 14, 380–386.
- (11) Schwander, T., Schada von Borzyskowski, L., Burgener, S., Cortina, N. S., and Erb, T. J. (2016) A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* 354, 900–904.
- (12) Mahour, R., Klapproth, J., Rexer, T., Schildbach, A., Klamt, S., Pietzsch, M., Rapp, E., and Reichl, U. (2018) Establishment of a five-enzyme cell-free cascade for the synthesis of uridine diphosphate N-acetylglucosamine. *J. Biotechnol.* 283, 120–129.
- (13) Kim, D. M., and Swartz, J. R. (1999) Prolonging cell-free protein synthesis with a novel ATP regeneration system. *Biotechnol. Bioeng.* 66, 180–188.
- (14) Kim, D. M., and Swartz, J. R. (2001) Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnol. Bioeng.* 74, 309–316.
- (15) King, Z. A., Lu, J. S., Dräger, A., Miller, P. C., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., and Lewis, N. E. (2016) BiGG Models: A platform for integrating, standardizing, and sharing genome-scale models. *Nucleic Acids Res.* 44, D515–D522.
- (16) Caspi, R., Billington, R. C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S., Subhraveti, P., and Karp, P. D. (2018) The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.* 46, D633–D639.
- (17) Hädicke, O., Von Kamp, A., Aydogan, T., and Klamt, S. (2018) OptMDF pathway: Identification of metabolic pathways with maximal thermodynamic driving force and its application for analyzing the endogenous CO<sub>2</sub> fixation potential of *Escherichia coli*. *PLoS Comput. Biol.* 14, No. e1006492.
- (18) Noor, E., Bar-Even, A., Flamholz, A., Reznik, E., Liebermeister, W., and Milo, R. (2014) Pathway Thermodynamics Highlights Kinetic Obstacles in Central Metabolism. *PLoS Comput. Biol.* 10, No. e1003483.
- (19) Noor, E. Removing both internal and unrealistic energy-generating cycles in flux balance analysis. *arXiv*, March 13, 2018, 1803.04999 [q-bio.MN]. <https://arxiv.org/abs/1803.04999> (accessed 2019/03/26).
- (20) Klamt, S., Regensburger, G., Gerstl, M. P., Jungreuthmayer, C., Schuster, S., Mahadevan, R., Zanghellini, J., and Müller, S. (2017) From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints. *PLoS Comput. Biol.* 13, No. e1005409.
- (21) Heirendt, L., et al. (2019) Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. *Nat. Protoc.* 14, 639–702.
- (22) Noor, E., Bar-Even, A., Flamholz, A., Lubling, Y., Davidi, D., and Milo, R. (2012) eQuilibrator - the biochemical thermodynamics calculator. *Nucleic Acids Res.* 40, D770–775.
- (23) von Kamp, A., Thiele, S., Hädicke, O., and Klamt, S. (2017) Use of CellNetAnalyzer in biotechnology and metabolic engineering. *J. Biotechnol.* 261, 221–228.
- (24) Klamt, S., Saez-Rodriguez, J., and Gilles, E. D. (2007) Structural and functional analysis of cellular networks with Cell Net Analyzer. *BMC Syst. Biol.* 1, 2.
- (25) Opgenorth, P. H., Korman, T. P., Iancu, L., and Bowie, J. U. (2017) A molecular rheostat maintains ATP levels to drive a synthetic biochemistry system. *Nat. Chem. Biol.* 13, 938–942.
- (26) Chowdhury, A., and Maranas, C. D. (2015) Designing overall stoichiometric conversions and intervening metabolic reactions. *Sci. Rep.* 5, 16009.
- (27) Opgenorth, P. H., Korman, T. P., and Bowie, J. U. (2014) A synthetic biochemistry molecular purge valve module that maintains redox balance. *Nat. Commun.* 5, 4113.