Journal of Synthetic Computability and Biology (JSCB)

https://utilitasmathematicapublishing.com
Vol: 1, ISSUE: 1, 2025
Pg No: 1-12

MEMO: A Technique for Calculating Metabolic Modules in Production Systems Without Cells

Dieter, Noah, Friedrich

University of Greifswald, Greifswald, Mecklenburg-Vorpommern, Germany

ABSTRACT:

A viable substitute for traditional microbial fermentation methods for creating value-added products from biological feedstocks is cell-free bioproduction systems. Finding appropriate metabolic modules with certain characteristics is a crucial first step in creating cell-free production systems. Here, we introduce MEMO, a brand-new computational method that supports the design of cell-free systems in a number of ways by identifying the smallest metabolic modules with predetermined stoichiometric and thermodynamic restrictions. Specifically, the regeneration of used cofactors (such ATP and NAD(P)H) is a major obstacle to the long-term operation of cell-free systems. The smallest regeneration modules that recover these cofactors with the necessary stoichiometries can be calculated using MEMO given a production pathway with specific cofactor requirements. In order to identify the fewest appropriate modules from a given reaction database, MEMO combines the stoichiometric and thermodynamic restrictions into a single mixed-integer linear program. By computing regeneration modules for the recently released synthetic CETCH cycle for in vitro carbon dioxide fixation, we demonstrate the usefulness of MEMO. We show that MEMO can determine multiple solutions in a reasonable amount of time in two large reaction databases (MetaCyc and BiGG) and is very flexible in considering the various constraints of the CETCH cycle (e.g., regeneration of 1 ATP, 4 NADPH, and of 1 acetyl-group without net production of CO2 and with permitted side production of malate). Glycerol is used as the substrate in the most promising regeneration modules, which only need eight enzymatic steps. Additionally, it is demonstrated that certain of these modules are resistant to the oxidation of NAD(P)H and the hydrolysis of ATP, two examples of spontaneous cofactor loss. Additionally, we show that MEMO can identify cell-free production systems that incorporate cofactor regeneration and product synthesis. All things considered, MEMO offers a strong technique for identifying metabolic modules and, as one specific use, for creating cell-free production systems.

KEYWORDS: Synthetic biology, Design of cell-free systems, Constraint-based modeling, Metabolic networks, Cofactor regeneration, Mixed-integer linear programming

INTRODUCTION

One crucial step in reducing reliance on fossil fuels is the manufacturing of chemicals, minerals, and fuels using biobased processes. This covers the creation and application of biochemical conversion systems to create value-added products in addition to substituting biorenewable resources for petrochemical feedstocks. Microbial fermentation, in which specially chosen or genetically (re)designed microorganisms transform a given substrate into the desired output, is a common component in biochemical production processes. One of the main areas of metabolic engineering is the optimization of appropriate cell factories with exceptional performance.One This top-down strategy has worked well for biosynthesis.

of several goods with industrial relevance.2. The use of cell-free production systems, which combine purified enzymes and specific metabolites in vitro to promote bioconversion, is an alternative—primarily bottom-up—approach.3–6 The complexity and reach of cell-free bioproduction systems have increased dramatically in recent years, since at first they only included one or a small number of enzymes (for example, in the traditional field of biocatalysis). In vitro reconstruction of whole major metabolic pathways7,8 and cell-free manufacturing systems for various compounds, such as ethanol, isobutanol9,1,3-propanediol10, and monoterpenes, are examples of this.8. Furthermore, cell-free systems enable the development of entirely new (synthetic) pathways, as the recently described 17-step CETCH cycle for fixing carbon dioxide.11. Both cell-based and cell-free production methods have unique benefits and drawbacks, as the literature 4,5 explains. Cell-free systems, for instance, allow one to precisely control system parameters, assemble entirely new pathways, and get around complex cellular (e.g., genetic) regulation that frequently prevents higher yields and productivities of whole-cell systems. In contrast, cellular systems offer a robust environment for enzyme operation and enable a cheap and integrated synthesis of the necessary pathway components. Furthermore, methods using hazardous substances that would not be practical in vivo can be used in vitro, and up to 100% conversion of the substrate into the target product is achievable because biomass synthesis does not occur in cell-free systems. The balancing and regeneration of cofactors (mostly but not only ADP/ATP and NAD(P)/NAD(P)H) consumed or created by the in vitro metabolic pathway, however, present a unique challenge for cell-free systems.4. Although cells can use a variety of specialized metabolic pathways to balance the net synthesis or net consumption of certain cofactors, in cell-free systems, their imbalance will rapidly halt the in vitro system's function. In order to restore cofactor pools, various regeneration submodules have been employed. For instance, Schwander et al.11 introduced polyphosphate and a polyphosphate transferase to the carbon dioxide fixing CETCH cycle in order to regenerate ATP from ADP. Nucleotide sugars have been produced cell-free using a similar method.Twelve Despite being straightforward, reasonably priced, and effective in these examples, this type of ATP regeneration system has the drawback of accumulating inorganic phosphate, which can slow down the ATP regeneration step for thermodynamic reasons and thereby reduce the system's overall efficiency in longer batch runs. Other substances with high-energy phosphate bonds (like PEP), which have been widely employed for cell-free protein synthesis, have comparable adverse consequences.3. Consequently, several ATP regeneration techniques have been produced, for instance, by generating acetyl-phosphate from pyruvate13 (which may be extended by many upstream glycolytic steps14), which is subsequently utilized by acetate kinase to generate ATP. This method avoids the net creation of inorganic phosphate, but it causes acetate to build up, which can also have negative (e.g., inhibitory) consequences on the ATP regeneration cycle's or the process's overall efficiency when it is used for extended periods of time. In order to restore the NADPH pool through a formate dehydrogenase, Schwander et al.11 employed formate as an electron donor in relation to NAD(P)H regeneration. A realistic application would have a negative impact on the overall stoichiometry of CO_2 fixation because CO_2 is released as a byproduct of the formate dehydrogenase reaction, even though this one-step module is straightforward, inexpensive, and fully sufficient to demonstrate the CO_2-fixing capability of the synthetic CETCH cycle. If formate is produced electrochemically (from CO_2), on the other hand, its use would be carbon neutral.

These illustrations highlight the need for a methodical approach to find appropriate regeneration modules for manufacturing systems that do not require cells. A regeneration module must, at the very least, meet particular stoichiometric and maybe additional requirements in order for its coupling with the production module to result in the best possible balance of all cofactors in the system. The CETCH cycle, which will be used as a running example in this work, is depicted in Figure 1. As previously stated, this synthetic cycle consists of a series of enzyme-catalyzed reaction stages that absorb CO2 in vitro, with the principal result being glyoxylate.11. Since its inception, it has undergone numerous improvements; in this investigation, we concentrate on version 5.0, which incorporates reactions that mix glyoxylate with externally supplied acetyl-CoA to yield malate as the final product.11. In addition to the two previously mentioned reactions for regenerating ATP (via polyphosphate and polyphosphate kinase) and NADPH (via formate and formate dehydrogenase), CETCH 5.0 includes 17 enzyme-catalyzed reactions, including 14 core reactions, a catalase reaction that breaks down hydrogen peroxide (a byproduct of the CETCH cycle) to water and oxygen, and more.

For one loop of the CETCH cycle, one ATP, four NADPH, and one acetyl-CoA must be provided, assuming that oxygen and carbon dioxide are provided externally (Figure 1). Acetyl-CoA was supplied externally in the original work, whereas ATP and NADPH were renewed using the two procedures mentioned. The task of finding a suitable regeneration module that replenishes not only the pools of ATP (from ADP and Pi) and NADPH (from NADP), but also of acetyl-CoA (from CoA) will be considered here due to the high cost of acetyl-CoA and the fact that its consumption results in a net accumulation of CoA, which may have inhibitory effects over longer batch runs. In addition, we require that (1) no carbon dioxide be released because this would undermine the CETCH cycle's ultimate objective, which is carbon dioxide fixation; (2) that malate be the only organic byproduct permitted by the regeneration module (though its production is not required); and (3) that the algorithm be able to select from a predetermined set of potential (typical) substrates, such as glucose, glycerol, acetate, and others. When combined with the regeneration module, constraint (2) makes the task much more difficult but guarantees that malate is the only organic product of the CETCH cycle, which will make downstream processing easier.

In this article, we introduce MEMO, a computational technique for locating the fewest metabolic modules under predetermined thermodynamic and stoichiometric restrictions. The design of cofactor regeneration modules is one specific application on which we will concentrate in this article. MEMO facilitates the design of cell-free systems in a number of ways. A universe of metabolic events is fed into MEMO in this situation, which then finds the smallest, thermodynamically possible regeneration modules that give cofactors with predetermined stoichiometric requirements, such as those for the CETCH cycle (Figure 1). MEMO is built on constraint-based modeling techniques and can be used in a variety of situations. We illustrate its applicability using the example of the CETCH cycle, demonstrating that metabolic regeneration modules can be discovered with as little as eight processes, even in this case with very strict requirements. We use modified versions of the BiGG15 and MetaCyc16 databases (Figure 1) as reaction universes. Reactions within the membrane compartment and between various compartments have been eliminated since their use would make the cell-free manufacturing system more difficult. Finally, we will show that MEMO may be used to locate whole cell-free systems with an integrated production pathway and regeneration module that ensure balanced overall conversions, in addition to the use for finding regeneration modules.

METHODS

A Mixed-Integer Linear Program for Metabolic Module Calculation. In this article, we outline MEMO, a technique for calculating metabolic modules that meet particular requirements from a given response database. The technique is based on identifying the pathways within a particular metabolic reaction network that have the highest thermodynamic driving force (as defined by the max-min driving force (MDF) definition in 18). The calculated metabolic modules are (1) balanced and (2) thermodynamically viable with a (maximal) overall driving force (MDF) above a predetermined threshold when OptMDFpathway is used as the starting point. To ensure that we identify the smallest module (in terms of the number of reactions) with these characteristics, we must expand OptMDFpathway in order to impose the necessary stoichiometries of the metabolic modules to be identified. Additionally, the extended formulation can employ the Gibbs free energies of production Δf G′0 of the metabolites instead of the usual Gibbs free energies ΔrG′0 of reactions, which were the basis for the original OptMDFpathway implementation.

Databases for reactions (universal networks). A metabolic network, or more broadly, a reaction universe or database, that contains the metabolic module is a necessary input for MEMO. We are free to gather enzymatic reaction steps from many organisms because MEMO is intended to be used here to find metabolic regeneration modules for cell-free manufacturing systems. However, since they would introduce another level of complexity, we assume that using cellular compartments (such as mitochondria, chloroplasts, etc.) in the cell-free production system is either impossible or should be avoided. As a result, reactions involving metabolites from various compartments are not taken into account. Similarly, processes that involve quinone metabolites—which are found in membranes—are not considered. Of course, for some applications, these restrictions can be modified or loosened.

The setup of two such universal network models, which were taken from the MetaCyc16 database and the BiGG15 database, respectively, is described below.

The BiGG Model. The 85 models in the BiGG database15 (accessed February 2018) are combined to create the BiGG model. Initially, the models were integrated one after the other while removing technical reactions (exchanges, demands, sinks). The COBRA toolbox21's mergeTwoModels function was utilized for model merging. This code eliminates duplicate reactions using the checkDuplicateRxn function, which we expanded with a new feature that correctly manages the reversibility of the reaction and identifies duplicate reactions based on their stoichiometry (up to a scalar factor). Reactions involving metabolites from various compartments were then eliminated for the previously mentioned reasons. Some of the metabolite IDs in the BiGG models have various formulas in different compartments, despite the fact that they are mainly unique (primarily due to differing protonation states). By using a common formula or by expressly differentiating the IDs, these IDs were semiautomatically disambiguated. Following this, the charge balance for every reaction and the mass balance for the elements C, H, O, N, S, and P were ascertained, and the imbalanced reactions were eliminated. After removing the compartment information, metabolites that occurred in various compartments were mapped onto a single ID. The enhanced checkDuplicateRxn function was then combined with reaction duplicates that had been created throughout this procedure. Lastly, for the purposes of the computations, all processes involving quinone metabolites—which are found in membranes—were deactivated. There are 8983 reactions and 5863 metabolites in the final modified BiGG network model. Using a mapping from BiGG metabolite IDs to KEGG IDs, the metabolites' Δf G′0 values [kJ/mol] were obtained, as much as possible, using the eQuilibrator API22. Reaction reversibility was limited based on the minimal/maximal driving force of each implicated metabolite in reactions where their Δf G′0 is known. The BiGG models' definition of reversibility was applied to the remaining reactions. Driving forces and Δf G′ limitations of the type eq 7a or 7b were set up for the computations using eq 6 (eq 4 was omitted correspondingly).

The MetaCyc Model. The flat file distribution of MetaCyc16 version 22.6, which includes the files compounds.dat (metabolites) and metabolic-reac-tions.xml, was used to build this model. These files were processed for this purpose. Regarding the BiGG model, only reactions with metabolites originating from the same compartment were retained; the metabolite compartment information was then eliminated. Unbalanced reactions were eliminated, duplicate reactions were eliminated, and the charge balance and mass balance for the elements C, H, O, N, S, and P were ascertained. Moreover, ΔrG′0 values (in MetaCyc reported in [kcal/mol]) for the remaining reactions were computed using the Δf G′0 after reactions involving metabolites without known Δf G′0 values were eliminated. Reversibility was limited for each reaction based on its minimum and maximum driving forces. Constraints of type eq 4 were utilized for the driving forces because all reactions in this model had a definite ΔrG′0 value (eqs 6, 7a, and 7b were left out correspondingly). It should be mentioned that even while the eQuilibrator method's predecessor established the ΔrG′0 values in MetaCyc,22 the actual results for the same reaction can differ significantly between the two methods. There are 8964 metabolites and 11634 reactions in the modified MetaCyc network model.

Availability and Implementation. Dedicated functions from CellNet-Analyzer (version 2019.3), including the OptMDFpathway function, which was expanded as detailed in the Methods section, were used to carry out the calculations. The following GitHub repository offers a package containing the models (in COBRA format) and the MATLAB script files needed for the computations: MEMO.git (https://github.com/ARB-Lab). The MILP solver used was CPLEX 12.8. A cluster node equipped with two Intel Xeon 8-core CPUs and 192 GB of RAM was used to carry out the computations.

RESULTS AND DISCUSSION

Calculating CETCH Cycle Regeneration Modules. We utilized our MEMO approach in conjunction with the two modified BiGG and MetaCyc reaction databases (refer to Methods) to compute regeneration modules that precisely match the

CETCH cycle 11. Figure 1 depicts the exact stoichiometric requirements of the CETCH 5.0 cycle, which were previously discussed in the introductory section. As a result, the MEMO optimization issue was subject to the following constraints:

Four NADPH, one ATP, and one acetyl-CoA are consumed by the CETCH cycle in a single round (Figure 1). As a result, the NADH consumption reaction was set to zero flux and the flux through the NADPH consumption reaction 13 was set to 4. By adding another artificial consumption reaction, acetylconsum: acetyl-CoA → CoA (14) and setting its ℂux value to 1, the consumption of acetyl-CoA was taken into consideration.

It is crucial to properly define the permitted end products (aside from the cofactor needs) while calculating regeneration modules. Because the main product must be separated from byproducts in downstream processing, for example, an accumulation of undesirable byproducts may not only slow down the regeneration module (and possibly the actual production system) but also jeopardize the process's overall economic viability. We permitted a net synthesis of malate, the primary product of the CETCH cycle, in the regeneration module as well. Since its synthesis in the regeneration module is therefore allowed but not required per se, a malate exchange reaction was added to the model without fixing its flux. It is also possible to exchange oxygen and water with the environment. Since fixing carbon dioxide is the goal of the CETCH cycle, we did not allow net carbon dioxide production because doing so would reduce the process' overall efficiency. Moreover, in the MetaCyc model, charged metabolites are balanced with protons in their exchange reactions (for example, when acetate is exchanged, this happens along with one proton), whereas in the BiGG model, protons can also be swapped. Any proton exchange flux in the BiGG model solely reflects the difference between the dissociated protons of the substrate and product because all non-exchange reactions are H- and charge-balanced.

In order to determine a set of dispersed entrance points in the central metabolism, we independently calculated regeneration modules for nine straightforward and inexpensive substrates: acetate, glycerol, glucose, sorbitol, formate, xylose, lactate, methanol, and succinate. For the thermodynamic driving force (MDF) in equation 9, the lower bound B was established at 0.01 kJ/mol. With the exception of CO2 and bicarbonate, for which the lower/upper bound was fixed at 0.1 µM/1 mM, the metabolite concentrations were permitted to fluctuate between 100 mM and 1 µM. Our goal was to determine each substrate's ten shortest regeneration modules. Each optimization's computation time was restricted to one hour, which in multiple instances was enough to get the desired result. Generally, the more reactions in the smallest regeneration module, the harder it is to prove the optimal. However, one can typically obtain a good solution even if the optimal answer was not demonstrated in one hour. To determine whether it could be beneficial to devote extra computational effort to attempting to find a smaller solution, this can then be compared with the best bound, which is the minimal number of reactions needed for any given solution. In a few instances, the solver failed to provide a workable solution in less than an hour and failed to demonstrate that the problem was not practicable. To test if a solution can be discovered in these situations, the optimization may be performed with various solver parameters and/or the time restriction extended.

Table 1 and the Supporting Information display the computations' outcomes. Six of the nine substrates in both networks (glucose, glycerol, sorbitol, xylose, lactate, and methanol) had CETCH regeneration modules within the allotted period. With the exception of the exchange/consumption reactions, all discovered solutions call for eight to sixteen reactions, which is always less than the total of the CETCH cycle's seventeen reaction stages. In MetaCyc, the solutions provided by both networks are either the same size or less. On the other hand, neither network contained any solutions for the substrates formate, succinate, or acetate. Without thermodynamic restrictions, flux variability study in the MetaCyc network already shows that formate cannot be solved.

In both networks, the shortest regeneration modules were discovered using glycerol as the substrate, requiring only eight reactions—which the solver determined to be ideal. Figure 2 displays one of the size 8 MetaCyc solutions. Glycerol undergoes two stages of oxidation in this solution to produce glycerate, which produces two NADPH in total. The phosphoglycerate phosphatase then phosphorylates the glycerate in reverse to produce 2-phospho-glycerate (2-PG). Remarkably, this phosphatase actually favors the kinase direction due to its ΔrG’0 of 6.7 kcal/mol. After the enolase converts 2-PG to phosphoenol-pyruvate (PEP), the route divides into two branches. Acetyl-CoA, CO2, and NADH are generated in the first branch when PEP is dephosphorylated to pyruvate.

PEP undergoes carboxylation to oxaloacetate in the second branch, which also results in the production of ATP. The conversion of oxaloacetate to malate occurs when NADH from the first branch is consumed.

Figure 3 displays one of the two solutions with eight reactions for the BiGG network (solution 1 in Supporting Information). In this case, three glycerol molecules are transformed into dihydroxyacetone, which is phosphorylated, along with the creation of NADPH. Beginning with dihydroxyacetone-phosphate (dhap), the route divides into two branches. In the

In the first branch, two dhap are transformed into methyglyoxal, which yields NADPH and pyruvate. Pyruvate undergoes two distinct processes to produce malate and acetyl-CoA, respectively. This molecule is transformed into glycerol-3-phosphate in the second branch from dhap, which yields glycerol and ATP. Consequently, a cyclic (sub)pathway through glycerol made up of the reactions GLYCD, r0242, r0202, and GLYK is present in this regeneration module. However, because its net stoichiometry is not empty, this cycle is not thermodynamically impossible:

ATP + NAD + NADPH + H2O = glycerol + ADP + NADH + NADP + H+ + Pi

This partial overall reaction's ΔrG’0 is 27.4 kJ/mol, but under the right metabolite concentration profile, its ΔrG’ can actually become negative, which is consistent with the concentration limitations employed in these computations.

We determined the elementary modes20 in each regeneration module to conduct a more thorough analysis (see Supporting Information). This demonstrated that the regeneration module of each of the two solutions mentioned above contains just one steady-state solution. Regeneration modules that comprise numerous elementary steady-state solutions, however, may be able to meet the needs of multiple cofactor regeneration. Take the MetaCyc model (Figure 4), which has two elementary modes, and the second glycerol solution as an example. The two modalities differ in that ATP is used in the second mode (reaction (d) in Figure 4) whereas it is not needed in the first mode (reaction (c) in Figure 4) for the phosphorylation of (the two) glycerate molecules. As a result, the ratio of ATP:NADPH:acetyl-CoA is 2:4:1 in the first mode and 0:4:1 in the second. By adjusting the respective fluxes of the cofactor exchange reactions when computing the modules, the desired 1:4:1 ratio is achieved when both modes operate together in a 1:1 ratio. It should be noted that when the production and regeneration modules are coupled, the operation of the production module will push the net stoichiometry of the regeneration module toward what is consumed in the net by the production module, even though the variability with two modes may initially seem disadvantageous. Further, regeneration modules with several elementary modes offer the benefit of compensating for possible changes in the ATP level, such as those brought on by its spontaneous hydrolysis, by altering the flux ratio between the modes. Thus, this regeneration module might function similarly to the ATP rheostat mentioned in reference 25.

For this substrate, the MetaCyc solution 2 mentioned in Figure 4 has one of the highest MDFs of any glycerol solution (18.01 kJ/mol, Table 1). Similar to the BiGG solution in Figure 3 (3.34 kJ/mol), the MetaCyc solution in Figure 2 has an MDF of just 3.22 kJ/mol. Thus, MetaCyc solution 2 would be better from a thermodynamic perspective. Large net fluxes via all participating processes should be possible with an MDF of 3 kJ/mol, according to some theories.18 Table 1 demonstrates that not all substrate/network combinations taken into consideration here can be attained by the solutions we have computed thus far. However, by setting the value of B in equation 8 to a desired minimum threshold, such a requirement might be imposed during the computations.

With 11 reactions in MetaCyc and 12 in BiGG, methanol is the substrate that enables the next smallest regeneration modules after glycerol. As an example, we go over the initial solution discovered in MetaCyc and BiGG. Methanol absorption in both solutions is followed immediately by its conversion to formaldehyde and the subsequent production of NADH. Formaldehyde then forms a methyl-group bond with a tetrahydrofolate species. When CO2, NH4 (which will be released again in subsequent steps), and NADH are consumed, the methyl-group from a portion of the methylated tetrahydrofolate is released as glycine. Serine is created when a portion of this glycine reacts with the leftover methylated tetrahydrofolate. Different processes are used in the two models to convert the residual glycine to glyoxylate and then to malate with acetyl-CoA. The fusion of glyoxylate and acetyl-CoA also occurs in the CETCH cycle11, and it is possible that the enzymes involved there could serve the same purpose if paired with the other regeneration module reactions, thereby lowering the requirement for additional regeneration enzymes. The pyruvate dehydrogenase in both regeneration modules uses the pyruvate that is produced from serine to generate acetyl-CoA. Additionally, a NADH to NADPH transhydrogenase is present in both solutions. However, some distinctions exist: Because ATP regeneration happens in tandem with malate creation and NADPH regeneration in the pyruvate dehydrogenase reaction, whereas both regenerations occur in BiGG, the MetaCyc solution requires one fewer reaction.

Alpha-ketoglutarate/glutamate/glutamine conversion cycles are connected to them. Additionally, the MetaCyc solution utilizes 5 methanol and 1 CO2 (and yields 1 malate) but the BiGG solution needs 4 methanol (and yields 0.5 malate). The MDF of both solutions is comparatively modest, at 0.54 kJ/mol for the BiGG solution and 3.33 kJ/mol for the MetaCyc.

Among the three carbohydrates—glucose, xylose, and sorbitol—the latter is the preferred substrate because it permits the smallest regeneration modules, with 12 reactions in both networks. Acetate as a substrate did not prove to be promising because no solution was found in either network, even though the CETCH-cycle requires the generation of an acetyl-group. Due to its highly oxidized nature and the need that no net CO2 be created, formate by itself is not suitable for the regeneration duties. Even though succinate's conversion to malate should easily yield the necessary reduction equivalents, no solutions were found for it either. The only one of the four acids for which regeneration modules were discovered is lactate. The simplest module in MetaCyc only has 11 reactions, whereas BiGG requires 14.

Thermodynamics of the Regeneration and Combined Production Modules (CETCH).

When the modules are integrated with the CETCH cycle, the MDF, which is provided for the standalone regeneration modules in Table 1 and previously described, may drop. The resulting MDF of the integrated system may be easily calculated, though, by recalculating the MDF using the union of the CETCH reactions and the regeneration module reactions (without the consumption pseudoreactions). It turns out that the regeneration modules in Figures 2–4 are still thermodynamically possible when integrated with the CETCH cycle, albeit with less MDF. The results are displayed in the Supporting Information. Nevertheless, when paired with the CETCH cycle, a number of the alternative solutions have an MDF < 0, indicating that they would not be appropriate regeneration modules in this situation. By requiring, for example, that the CETCH reactions be active with the necessary fluxes (using α and β of eq 2) rather than the consumption pseudoreactions (eqs 11−14), such solutions can be immediately excluded during the MILP optimization.

We also determined the concentration ranges of the participating metabolites required to maintain the MDF ≥ 0.01 for the regeneration modules that are thermodynamically feasible when coupled with the CETCH cycle (see Supporting Information; only metabolites whose lower or upper concentration bound deviates from the default concentration bound are listed). These findings could be helpful in determining the pathway's overall viability.

Modules for regeneration that require different cofactors and several substrates. Although we have only looked at single inputs thus far, it is potentially feasible to have several substrate uptakes open simultaneously to determine if this would enable smaller regeneration modules. Thus, using all nine substrates, we computed 20 solutions in each network.

(See Supporting Information) Uptakes are open. It turns out that each of these solutions calls for eight reactions, indicating that using more than one substrate does not enable smaller regeneration modules in this situation. Actually, glycerol is used as a single substrate in all 20 solutions for the BiGG model. While glycerol is used in all MetaCyc solutions, some also employ acetate or acetate and formate. Curiously, formate and acetate cannot be used as a single substrate, although they may work well together with glycerol.

Although previous CETCH versions had differing cofactor criteria, all of the aforementioned results meet the requirements of the CETCH 5.0 cycle.11. For example, 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) if the second reductive carboxylation step (acrylyl-CoA to methylmalonyl-CoA) is substituted with an ATP-dependent carboxylation step (propionyl-CoA to methylmalonyl-CoA) as in CETCH 1.0. To evaluate the changes in minimal regeneration module sizes, we recalculated the single substrate computations using these new specifications. The findings (see Supporting Information) indicate that the minimal module sizes vary by no more than one reaction, depending on the substrate and network. The minimal size in the BiGG model stays constant for the other substrates but significantly increases for glycerol and lactate. According to the MetaCyc model, xylose and lactate cause the size to increase while glucose, sorbitol, and methanol cause it to shrink. Therefore, the substrate utilized in the regeneration module determines whether or not it is advantageous to exchange one NADPH for one ATP.

Modules for computing that combine production and regeneration. Up to now, MEMO has concentrated on cofactor regeneration modules, which are frequently necessary for cell-free manufacturing systems. As long as the module definition can be created using the same kind of constraints that MEMO uses, MEMO is more universal and can look for any metabolic module. Actually, MEMO may be used to build whole production systems that include cofactor regeneration and manufacture, with the necessary input-output stoichiometries. The benefit of computing such integrated modules is that, unlike other regeneration modules calculated for the CETCH cycle, thermodynamic infeasibilities cannot occur when combining production and regeneration modules. We use the recent achievement of producing the monoterpene limonene from glucose using a cell-free system as an example.8. One mol of limonene is produced for every three mol of glucose in this specific conversion, which consists of 23 chemical steps (Supporting Figure 1 in the original ref 8). Glycolysis, which turns glucose into acetyl-CoA and cofactors, and the mevalonate pathway, which uses acetyl-CoA and cofactors to create limonene, are its two main components. A purge valve has been incorporated into the system because glycolysis generates more reduction equivalents than the mevalonate pathway requires. This valve (i) eliminates surplus reduction equivalents (using the NADH oxidase NoxE) and (ii) produces balanced levels of NADH/NADPH (by using a glyceraldehyde-3-phosphate dehydrogenase that is specific to both NAAD and NADP). By setting the purge valve, glycolysis, and mevalonate pathway fluxes to the values necessary to produce one limonene from three glucose, we were able to determine whether this system is in the MetaCyc network's solution space (Supporting Information). This results in the same solution as in ref 8, demonstrating that MEMO can theoretically locate this limonene producing mechanism. We determined 10 smallest solutions and set a minimum yield of 1/3 limonene per glucose as the only constraint in order to search for alternate production methods (see Supporting Information). The number of reactions needed for all of the discovered solutions is 21, which is fewer than for the original system. Nevertheless, the optimal MDF of the found alternative solutions (0.75 kcal/mol = 3.14 kJ/mol) is lower than the MDF of the original system (1.03 kcal/mol = 4.31 kJ/mol). While some of the calculated solutions combine glycolytic and Entner-Doudoroff pathway reactions, one is comparable to the original system8 in that it involves a lot of glycolysis processes. Although the relationship between cofactor production and consumption varies, excess reduction equivalent elimination always occurs through NAD(P)H oxidases.

CONCLUSION

In this study, we introduced MEMO, a general method for locating the smallest metabolic modules that satisfy predetermined thermodynamic and stoichiometric requirements. We concentrated on using MEMO to identify regeneration modules for cell-free manufacturing systems, despite the fact that it has a wide range of potential uses. MEMO is very flexible in considering a variety of requirements, such as the regeneration of two cofactors plus one acetyl-group with a specific stoichiometry, the prohibition of net production of CO2, the allowance of net synthesis of malate, and the minimum threshold for MDF. It can also enumerate multiple solutions in a reasonable amount of time in very large universal networks. This is demonstrated by the CETCH cycle. OptStoic is a technique that hasn't been applied to the construction of cell-free (regeneration) modules yet but might have comparable potential.26 This two-step method finds overall conversions that optimize a specified objective function (such as the maximum product yield) for a few predetermined reactants and products in the first phase. This first step's computations ignore the reaction steps between the reactants and products and are only based on stoichiometric balance and the Gibbs energy changes of the reactants and products. The second stage looks for appropriate reaction combinations from a given reaction database to determine the actual metabolic route or network that is responsible for producing the total conversions discovered in the first step. Our method is more straightforward and only needs one optimization step to uncover stoichiometrically and thermodynamically feasible solutions, even though the relevance of this methodology has been demonstrated in other case studies, such as for identifying alternate paths in metabolic networks26. Actually, the reaction database used in the second step of optStoic may not contain a pathway that corresponds to the overall conversions discovered in the first stage. Furthermore, even if a stoichiometrically balanced pathway is found in the second step of optStoic, it might not be thermodynamically feasible within the specified concentration ranges of the intermediate metabolites (optStoic tests, separately for each reaction, in a preprocessing step, whether it can proceed in a certain direction with the given metabolite concentration ranges, but 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 offers a flexible method for an integrated search of metabolic modules by directly accounting for thermodynamic feasibility using the MDF approach and even enabling the setting of lower limits for the MDF.

The regeneration modules discovered for the CETCH cycle seem highly intriguing, and unlike the original system, none of them require costly substances like acetyl-CoA or accumulate side products. Particularly encouraging are the modules that employ glycerol as a substrate; they would only need eight metabolic reactions (enzymes), which is even fewer than half of the enzymes that have been used thus far in the CETCH cycle. The modules may have various properties even though they have the same net stoichiometry, according to the examination of the glycerol solutions. This relates specifically to the phenomena that the modules may consist of a single elementary mode or a collection of modes, aside from their thermodynamic driving force. In the first scenario, the module's conversion of the substrate will precisely meet the required stoichiometry. Alternative output stoichiometries would also be available in the second scenario, even though the required stoichiometry is achievable with the relevant module. When the production and regeneration modules are coupled, the production module's activity will cause the regeneration module's net stoichiometry to move closer to the module for manufacturing. The ability to balance the spontaneous loss (decay) of cofactors (such as the oxidation of NAD(P)H or the hydrolysis of ATP) is an advantage of regeneration modules with multiple elementary modes. Over time, this could cause issues in modules with fixed stoichiometries (single modes). Therefore, these flexible regeneration modules have similar (desired) functions to purge valves or the recently described ATP rheostat25.8, 27 MEMO is far more versatile and can be used to look for many kinds of cell-free (or even intracellular) metabolic modules, even if the application focus here was on regeneration modules. We used the synthesis of limonene as an example to show that MEMO can also be used to design whole cell-free production systems with integrated production and regeneration. This is because MEMO can identify the smallest integrated cell-free system with desired input-output stoichiometries in a single step, rather than first fixing the production module and then finding a suitable regeneration module. All things considered, in conjunction with the modified MetaCyc and BiGG master networks, our MEMO approach offers a strong and adaptable framework for creating cell-free production systems. It can also be utilized to identify or create (intra)cellular metabolic modules with specific characteristics.

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 metabolicengineering: 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., Reiße, 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. *Chem- SusChem 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) OptMDFpathway: Identification of metabolic pathways with maximal thermodynamic driving force and its application for analyzing the endogenous CO2 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.