

# Development of a Framework for Metabolic Pathway Analysis-Driven Strain Optimization Methods

Vitor Vieira<sup>1</sup> · Paulo Maia<sup>2</sup> · Isabel Rocha<sup>1</sup> · Miguel Rocha<sup>1</sup> 

Received: 21 July 2016 / Revised: 1 February 2017 / Accepted: 6 February 2017 / Published online: 25 February 2017  
© International Association of Scientists in the Interdisciplinary Areas and Springer-Verlag Berlin Heidelberg 2017

**Abstract** Genome-scale metabolic models (GSMMs) have become important assets for rational design of compound overproduction using microbial cell factories. Most computational strain optimization methods (CSOM) using GSMMs, while useful in metabolic engineering, rely on the definition of questionable cell objectives, leading to some bias. Metabolic pathway analysis approaches do not require an objective function. Though their use brings immediate advantages, it has mostly been restricted to small scale models due to computational demands. Additionally, their complex parameterization and lack of intuitive tools pose an important challenge towards making these widely available to the community. Recently, MCSEnumerator has extended the scale of these methods, namely regarding enumeration of minimal cut sets, now able to handle GSMMs. This work proposes a tool implementing this method as a Java library and a plugin within the OptFlux metabolic engineering platform providing a friendly user interface. A standard enumeration problem and pipeline applicable to GSMMs is proposed, making use by the community simpler. To highlight the potential of these approaches, we devised a case study for overproduction of succinate, providing a phenotype analysis of a selected strategy and

comparing robustness with a selected solution from a bi-level CSOM.

**Keywords** Constraint-based modeling · Metabolic engineering · Pathway analysis · Minimal cut sets · Strain optimization

## Abbreviations

### List of Enzymes

ACoA	Acetyl-CoA hydrolase
ATPS	atp synthase
G6PDH	Glucose 6-phosphate dehydrogenase
GAPD	Glyceraldehyde 3-phosphate dehydrogenase
NADH2	NADH dehydrogenase
PGK	Phosphoglycerate kinase
PYRDC	Pyruvate decarboxylase
SUCD2	Succinate dehydrogenase iron-sulfur protein
SUCD3	Succinate dehydrogenase cytochrome b
SUCD	Succinate dehydrogenase complex

### List of Metabolites

$\alpha - kg$	Alpha-ketoglutarate
<i>atp</i>	Adenosine triphosphate
<i>dhap</i>	Dihydroxyacetone phosphate
<i>gaba</i>	Gamma-aminobutyrate
<i>nadh</i>	Nicotinamide adenine dinucleotide (reduced)
<i>nadh<sup>-</sup></i>	Nicotinamide adenine dinucleotide phosphate (reduced)
<i>nadh<sup>+</sup></i>	Nicotinamide adenine dinucleotide phosphate (oxidized)

**Electronic supplementary material** The online version of this article (doi:10.1007/s12539-017-0218-7) contains supplementary material, which is available to authorized users.

✉ Miguel Rocha  
mrocha@di.uminho.pt

<sup>1</sup> Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

<sup>2</sup> SilicoLife Lda., Rua do Canastreiro, 15, 4715-387 Braga, Portugal

## 1 Introduction

In silico modelling of cell metabolism has become an important tool for understanding and manipulating biochemical processes in living organisms. With recent advances in sequencing techniques, there has been an increase in the availability of sequenced whole-genomes. From these, a genome-scale metabolic model (GSMM) can be created using knowledge from annotations of metabolic genes. Such models can be used to predict cell behaviour and design intervention strategies to achieve certain goals. Metabolic engineering (ME), defined as the manipulation of cells with the purpose of achieving strains with desired phenotypes, benefits greatly from these approaches as it allows a rational design process as opposed to in vivo trial-and-error methods [1].

There are two main approaches for metabolic modeling, namely constraint-based models (CBMs) and kinetic models. Kinetic models attempt to represent metabolite concentration over time, requiring enzyme rate laws that can accurately represent the dynamic behavior of the system. While such models tend to be more accurate, there is a lack of knowledge of enzyme kinetics and associated parameters. CBMs, on the other hand, are more suited as GSMMs, adding assumptions regarding time, metabolite balance and enzyme kinetics to overcome the lack of knowledge of enzyme kinetics.

Phenotype prediction methods using CBMs are mostly based on flux balance analysis (FBA) first presented over two decades ago [2]. This approach has been applied and validated in a number of studies over the years [3] and attempts to find a flux distribution that is valid and satisfies a given cellular objective. This is normally defined as the maximization of an artificial flux representing cell growth. However, the validity of this objective is often disputed when modelling more complex or mutant cells, leading to the creation of alternative methods such as the minimization of metabolic adjustment (MOMA) [4] and regulatory on/off minimization (ROOM) [5]. These methods attempt to find a flux distribution in mutant cells that is similar to the wild-type strain.

Strain optimization approaches based on the principles set by the FBA began to emerge. The goal with these approaches is to find suitable modifications to a given model that lead to desired phenotypes (such as compound production). The focus of this work concerns the usage of reaction deletions to overproduce valuable end compounds. In silico approaches for this purpose, computational strain optimization methods (CSOMs), can be split into two major branches.

Constraint-based (CB) CSOMs are based upon the principles of FBA and employ a bi-level optimization approach. The outer optimization problem usually attempts to find an

intervention strategy to the ME problem, while being subject to an inner optimization problem, which is normally a phenotype prediction method that optimizes towards the cellular objective (normally growth). OptKnock [6] was the first CB CSOM to employ this generic framework and many methods were developed based on it. These optimization problems can also be solved with meta-heuristics, such as evolutionary algorithms (EA), as it is the case with OptGene [7] that first marked their usage in CSOMs.

Pathway analysis (PA) CSOMs, on the other hand, attempt to search for intervention strategies with methods that consider the entire solution space, describing all functional units in the metabolic model rather than considering a fixed objective. As such, these methods provide an unbiased perspective on the theoretical limits of the whole model [8]. Most of these approaches rely on the concept of elementary flux modes (EFMs), representing the smallest sets of reactions representing cell functions in the model and minimal cut sets (MCSs) which correspond to sets of reactions blocking undesired EFMs. MCSs can be exploited by CSOMs to provide ME strategies with optimal size.

An early example of a CSOM employing EFM analysis is the FluxDesign algorithm [9], that determines correlations between the cellular objective and the ME objective in a given set of EFMs and identifies up- and down-regulation targets based on this information. The CASOP framework [10] attempts to determine the relevance of each reaction to the ME objective through analysis of metabolic pathways described by selected EFMs. The end result is a ranking of reactions based on their potential for manipulation. The SMET approach attempts to find rate-limiting reactions and suitable intervention strategies with the purpose of increasing the productivity of a given cell for a ME objective using knowledge from EFMs [11].

Enumeration of MCSs usually requires pre-calculated EFMs, with early approaches being based mostly on the Berge algorithm for graphs. With this approach, candidate MCSs are checked for their ability to disrupt the undesired flux space and their minimality. If these candidates comply with these conditions, they are added as MCSs [12, 13].

These methods attempt to eliminate the optimality bias that FBA-based approaches usually suffer from. However, the enumeration of all EFMs is needed to represent all of the capabilities of the cell and provide accurate results. This is a very computationally expensive process and mostly infeasible with GSMMs. Despite this, an MCS enumeration method, MCSEnumerator, was recently proposed. It employs a MILP approach and allows partial enumeration of MCSs without requiring complete EFM enumeration. This method was validated with a GSMM of *Escherichia coli*, and was able to enumerate synthetic lethals up to five reactions, as well as several design strategies for relevant end-products [14].

In this scenario, given the potential of MCS enumeration to guarantee robust production, regardless of the phenotype prediction method used, this work pursues the development of an open-source software tool capable of handling relevant tasks in the MCS enumeration problem and applying those to strain optimization. Therefore, the main scientific/are:

- To implement a library compiling the necessary routines for enumeration of MCSs in CBMs.
- To integrate MCS enumeration tasks in the OptFlux metabolic engineering platform, providing novel tools for strain optimization.
- To provide a simple and intuitive user-interface for the implemented routines.
- To describe predicted phenotypes using MCS-derived design strategies.
- To compare MCSs with solutions from other approaches.

## 2 Methods

### 2.1 Constraint-Based Models and Pathway Analysis

Constraint-based models of metabolism typically contain  $m$  intracellular metabolites and  $n$  reactions acting on them. Reactions can represent pools of external metabolites to simulate uptake or production of certain compounds. This model can be represented by a  $m \times n$  matrix  $S$ , containing stoichiometric coefficients. Since CBMs assume metabolite concentrations are stable over time, this model leads to a system of linear equations:

$$S \cdot v = 0 \quad (1)$$

with  $v$  as a vector containing fluxes (or rates) for each reaction. Furthermore, thermodynamic or rate limits can be imposed upon the reactions, added as constraints in the following format:

$$\alpha \leq v \leq \beta \quad (2)$$

with  $\alpha$  and  $\beta$  being, respectively, the vectors containing lower and upper limits for the fluxes in  $v$ . An irreversible reaction  $j$  must have a lower limit  $\alpha_j = 0$ . The system defined by Eqs. 1 and 2 can be represented as a convex polyhedron hereby referred to as  $P$ , containing all feasible solutions to this system. Using this framework, the concept of elementary mode (EM) can be defined as representing a minimal functional unit within the model. Any elementary mode  $e$  equates to a flux distribution obeying three key properties [15]:

1. A flux distribution in  $e$  must comply with Eq. 1.

2. Irreversible reactions must carry flux only through a single direction in any EM. These are specified in Eq. 2.
3. Considering  $\text{supp}(e)$  as the reactions carrying flux in  $e$ , no subset of  $\text{supp}(e)$  can yield a flux distribution obeying Eqs. 1 and 2.

These approaches fully describe the polyhedron  $P$ , since any point contained within it is a linear combination of EMs. Desired flux distributions for the metabolic model can be found within non-null combinations of EMs within a desired set of flux vectors  $D$ .

Likewise, consider a set of undesired flux vectors  $T$ . Any set of reactions  $C$  that, when removed, can block the EMs contained in  $T$  is considered a cut set for  $T$ . If no subset of  $C$  can disable the EMs in  $T$  when blocked,  $C$  is considered a minimal cut set (MCS) [16]. However, the EMs in  $D$  are not guaranteed to remain active. If  $C$ , which is a MCS for  $T$ , does not block the EMs contained in  $D$ , it is considered a constrained MCS (cMCS) [17].

The definition of a ME problem for compound overproduction usually aims at coupling production fluxes with growth rate, represented as an artificial reaction converting metabolites required for growth into biomass. This feature is defined as growth-coupled product synthesis (GCPS) [18] and can be classified in two ways depending on the degree of coupling. Assuming  $Y^{B/S}$  as the biomass per substrate yield,  $Y^{P/S}$  as the product per substrate yield,  $p_{\min}$  as the minimum desired product yield and  $b_{\min}$  as the minimum desired biomass yield:

- Weak coupling occurs when  $Y^{P/S} > p_{\min}$  is always true if  $Y^{B/S} > b_{\min}$ .
- Strong coupling occurs when  $Y^{P/S} > p_{\min}$  is true for all considered flux vectors.

### 2.2 Enumerating Minimal Cut Sets

MCS enumeration methods have usually relied upon exhaustively enumerating the full set of EMs and applying combinatorial algorithms using that knowledge. As previously stated, this is mostly infeasible for models at the genome-scale due to the heavy computational demand this task poses, since the number of EMs scales with the model's complexity and size.

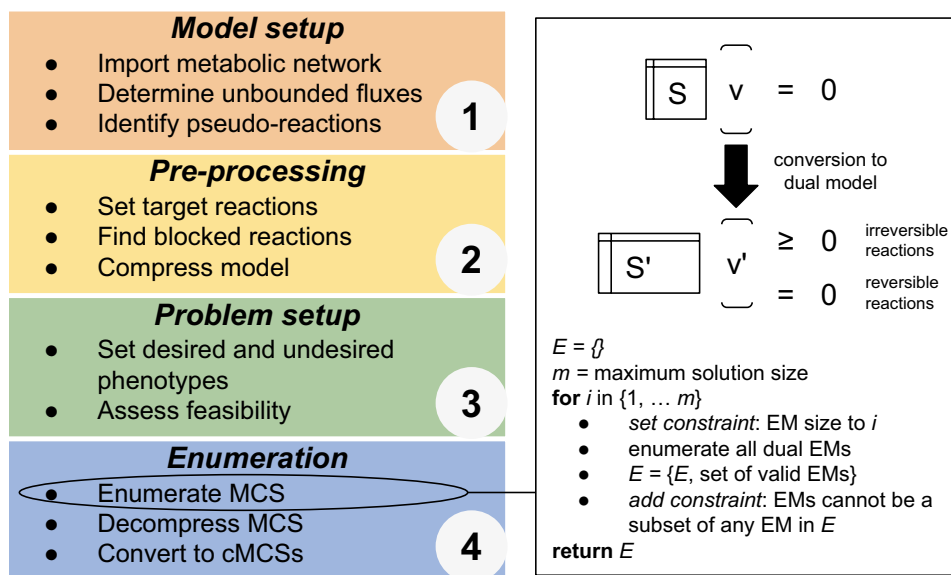
A recent approach, MCSEnumerator, circumvents this issue and allows enumeration of MCSs and cMCSs in GSMMs. Case studies using an *E. coli* GSMM resulted in compound overproduction strategies not greater than seven knockouts for various case studies, including serine, fumarate and ethanol production [14]. The algorithm employs a mixed-integer linear programming (MILP) problem that combines two approaches from earlier work. The first is a

mathematical formulation in which a dual-system is generated so that EMs calculated in it correspond to the primal system in which it was based [19]. The second approach concerns the *K*-shortest algorithm for partial EM enumeration, allowing enumeration of the smallest EMs using a simple MILP formulation [20]. By combining these two approaches, one can find the smallest EMs in the dual system and thus, the smallest MCSs for the original model.

A generic pipeline based on the original publication was assembled, containing all the required steps for any given enumeration problem, as represented on the left panel of Fig. 1. Model compression (pre-processing phase) and MCS decomposition are optional steps, but largely improve computation times. The MILP problem and basic algorithm for MCS enumeration are represented on the right panel of the previous figure.

### 3 Development

There are two key end-results regarding the software developed for this work. First, a *Java* library containing the MCS enumeration pipeline was implemented using the MCSEnumerator approach. The second end-result is a plugin for the OptFlux platform, developed to provide a user interface for the *Java* library.



**Fig. 1** Left representation of the pipeline used in this work. Step (1) concerns model setup which consists mainly on determining reaction reversibility and pseudo-reactions that will not be a part of any solution; Step (2) aims at reducing the size of the problem, mostly through removal of blocked reactions and network compression by lumping correlated reactions. Step (3) assembles the enumeration

### 3.1 Enumeration Library

MCSEnumerator was made available as a part of the CellNetAnalyzer platform currently available for MATLAB. One of the aims of this work was the development of a standalone library containing the necessary resources to solve MCS enumeration problems using the MCSEnumerator algorithm [14]. This library was built using the *Java* programming language to allow greater compatibility with most platforms and enable the use of advanced tools for the development of a graphical user interface (GUI). This software currently relies on the IBM ILOG CPLEX Optimization Studio to solve the MILP problem described in [14].

This library contains three newly developed packages:

1. *Enumeration* Contains methods needed to implement the MCSEnumerator MILP formulation, given a suitable problem.
2. *Metabolic* Provides a framework upon which CB models can be defined, as well as constraints for typical linear programming problems such as yield or capacity constraints on the reactions.
3. *Utilities* Includes methods that execute the entire pipeline given a set of parameters for the optimization. Functions to run the algorithms in a command-line environment are also provided, capable of reading parameters contained within text files.

problem and validates it, so that in Step (4) the proper formulation is built and solved. MCSs that are not feasible in the desired space are discarded, leaving only cMCSs. Right brief overview of the MILP formulation.  $S'$  and  $v'$  are derived from the dual model formulation in [19] which already includes the undesired phenotypes

The libraries provide routines capable of executing these tasks in a command-line environment using only a metabolic model in Systems Biology Markup Language (SBML) format and a file containing the parameters of the MCS enumeration problem, which may include undesired or desired limits for fluxes or yields, exclusion of target reactions, among other constraints. The source code for this library as well as resources to run the case studies featured in this article can be found on a public repository at <https://github.com/MEWorkbench/mcslibrary>.

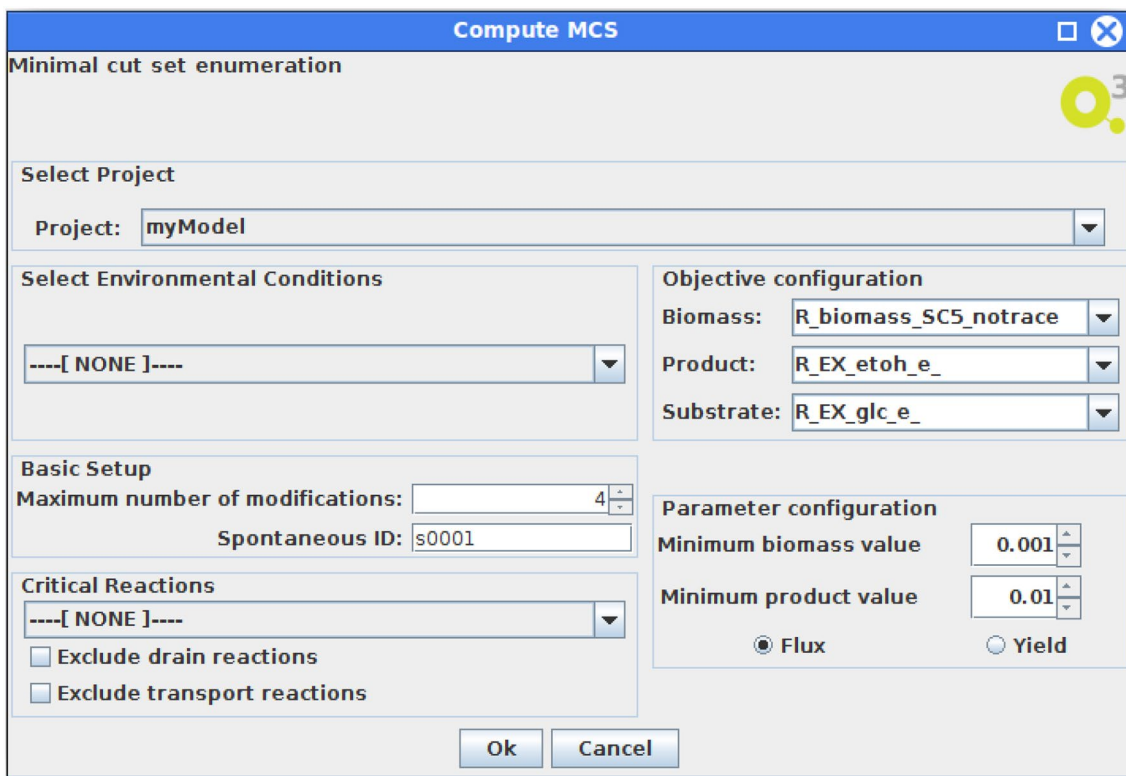
### 3.2 OptFlux Plugin

This work also aimed at providing a simple and clear user interface (GUI) as a plugin within OptFlux [21]. Currently, this framework includes important tools used in CB approaches, including phenotype simulation, analysis methods, and strain optimization algorithms developed in-house (OptGene [7] and derivatives [22]).

As far as this work is concerned, OptFlux provides the necessary methods to read and write metabolic models, serving as inputs for our algorithms.

The developed plugin provides a simple GUI (represented on Fig. 2 for the MCSEnumerator approach requiring minimal user input and providing a useful abstraction for the concepts discussed in the previous section. The user is only required to specify the maximum number of knockouts, which reactions correspond to biomass, product synthesis and substrate uptake, the desired thresholds for production and growth, and whether the production threshold is a yield or a rate constraint. Additionally, environmental conditions can be added and knockout targets can be discarded from the search either by supplying a list of critical reactions or a gene ID corresponding to spontaneous reactions; should the model represent those as being associated with a placeholder pseudo-gene?

The solutions are displayed using OptFlux's GUIs, using the format of previously available optimization algorithms. So, these solutions can be processed and simulated afterwards using other tools from OptFlux. OptFlux can be downloaded from <http://www.optflux.org/>. This plugin is available in the software's plugin repository as "Minimal cut set enumeration plugin for OptFlux", and is already included in the base installation from version 3.3 onwards.



**Fig. 2** Graphical interface provided by the plugin to formulate an enumeration problem

## 4 Results

### 4.1 Library Validation

The set of case studies was defined with the aim of ensuring that the outputs provided by the developed software match the ones from MCS Enumerator's original implementation. As such, the iAF1260 *E. coli* GSMM [23] was used with different enumeration problems for which the cMCSs were previously determined in the original publication [14]. The results from all case studies were accurately replicated and are highlighted in the Table 1.

### 4.2 Plugin Operation

This section shows in more detail the plugin's mode of operation using one of the case studies described above (ethanol anaerobic production in *E. coli* using glucose as carbon source). To run this case study:

1. Start a new OptFlux project using the **New project wizard** option on the **File** tab, click on **OptFlux model repository** and select the iAF1260 *E. coli* model. Assume default options in the process.
2. Create an environmental condition using the **New...** menu, option **Create...** and then click **Environmental condition**. Add a constraint for reaction R\_EX\_glc\_e with lower bound as  $-20$  and upper bound as  $999,999$  (definition of glucose uptake rate), and another for R\_EX\_o2\_e\_ with  $0$  as lower bound and  $999,999$  as upper bound (definition of anaerobic conditions).
3. Access the **Optimization** tab, and click on **Minimal cut sets**.
  - (a) Select the environmental condition that was created in the previous step.
  - (b) Allow at most 3 modifications and set the spontaneous ID for s0001.
  - (c) Set the objective configuration as follows:
    - Biomass as R\_Ec\_biomass\_core\_59p81M
    - Substrate as R\_EX\_glc\_e\_
    - Product as R\_EX\_etoh\_e\_
  - (d) Set the biomass value to 0.1, choose yield and then set the minimum product value to 0.2.

This example allows determination of 3 reaction deletions, guaranteeing a minimum ethanol per glucose yield of 0.2 and ensuring any solution determined maintains this production with a minimum growth rate of  $0.1 \text{ h}^{-1}$ . The results can be browsed and sorted and also saved to disk as a text file. Specific solutions (deletion sets) can be saved to the clipboard, and simulated or analyzed through other OptFlux tools.

## 5 Case Study

To prove the usefulness of the integrated framework, a case study is presented involving overproduction of succinate in the yeast *Saccharomyces cerevisiae* in a glucose-rich, aerobic medium. Design strategies for this task were enumerated using the MCS Enumerator implementation featured in this work and the Strength Pareto Evolutionary Algorithm (SPEA2) as CSOMs. For a closer comparison between both methods, the same target reactions were used. This was achieved through a list of reactions forbidden from being knocked-out, which was determined prior to this work. This list represents a set of reactions that compromise cell growth if knocked-out, and is specific to the environmental conditions that we attempt to replicate. Both sets of strategies are compared through analysis of the predicted phenotypes of a selected solution from both sets.

**Table 1** Overview of the validation case studies

Objective	Scenario	#MCS/#cMCS	Computation time (h)	Maximum size
Synthetic lethals	–	1018/–	17	4
Anaerobic ethanol production	$\text{Glc} \leq 10 \text{ Y} \geq 1.4$	185,302/8342	7.5	7
	$\text{Glc} \leq 10 \text{ Y} \geq 1.8$	153,338/1987	9.1	7
	$\text{Glc} \leq 18.5 \text{ Y} \geq 1.4$	156,477/8819	12.7	7
	$\text{Glc} \leq 18.5 \text{ Y} \geq 1.8$	138,675/4618	2	7
Fumarate production	$\text{Y} \geq 0.5 \text{ Glc} \leq 20$	17338/30	12.4	7
Serine production		18,449/140	1	6

Y represents product/substrate yield and Glc represents glucose uptake ( $\text{mmol gDW}^{-1} \text{ h}^{-1}$ ). Note that aerobic conditions were allowed only for fumarate and serine production. Computation times were determined using 12 processor cores and 30 GB of RAM

## 5.1 Problem Formulation

As there are two strategy groups featured in this section, we hereby name MCSEnumerator solutions as cMCSs (constrained minimal cut sets), while SPEA2 solutions are termed EA (evolutionary algorithm).

*Model A* modified version of the *iMM904* model reconstruction of *S. cerevisiae* [24] was chosen for this work. This model was chosen as it was already subject to previous work, being more consistent when predicting experimentally determined phenotypes [25].

*SPEA2* The SPEA2 algorithm [26] chosen to represent CB methods. It is a multi-objective optimization algorithm incorporating concepts initially used in OptGene [7]. Similarly to other bi-level heuristic frameworks, the inner problem pertains the assessment of the quality of the solution. However, unlike other algorithms, it can assume multiple objectives. In this case, the objective is the maximization of two objective functions, namely biomass and product flux. Parsimonious FBA (pFBA) [27] is used to obtain both these values.

Since this is an evolutionary algorithm, a pool of solutions is evolved until the algorithm stops. A cap of 100,000 objective function evaluations was applied as a stopping criterion and the algorithm was run 10 times. Solutions could not exceed 20 knockouts. A total of 331 distinct solutions were found.

*MCSEnumerator* Parameters used to build the enumeration problem for use with our implementation are described in Table 2. 1451 constrained minimal cut sets were successfully determined using this parameterization.

## 5.2 Phenotype Analysis

This section will focus on the changes on predicted phenotypes after applying constrained minimal cut sets to the model. An individual solution was selected (Solution 1 in Table 3) so that its phenotype could be further explored. The criteria for its selection was size, as it is beneficial for

**Table 3** Knockout strategies considered in the phenotype analysis

Abbreviation	Knockouts
WCS1	{SUCD2; SUCD3; G6PHD }
Solution 1	WCS1 $\cup$ {ACOAH; PYRDC ;PGK}

in vivo implementation purposes that the strategies contain the least number of knockouts. The goal is to find whether cMCSs can be based on other solutions from CB methods.

The selected solution highlighted on Table 3 (Solution 1) was decomposed and its predicted pFBA phenotype was determined. All possible combinations of knockouts using this solution were determined and simulated with the aim of determining the smallest set of knockouts allowing minimum product flux above a threshold of  $10^{-4}$  mmol gDW<sup>-1</sup> h<sup>-1</sup> at any minimum biomass fraction. Within these subsets, a solution was identified (WCS1 on Table 3) as appearing in the set of EA solutions.

### 5.2.1 Weakly Coupled Succinate Production

With two solutions from both sets, first we attempt to describe the phenotype of the smaller solution from the EA set by applying two different maximum biomass restrictions (90 and 100%) along with the knockouts and predicting the corresponding phenotypes.

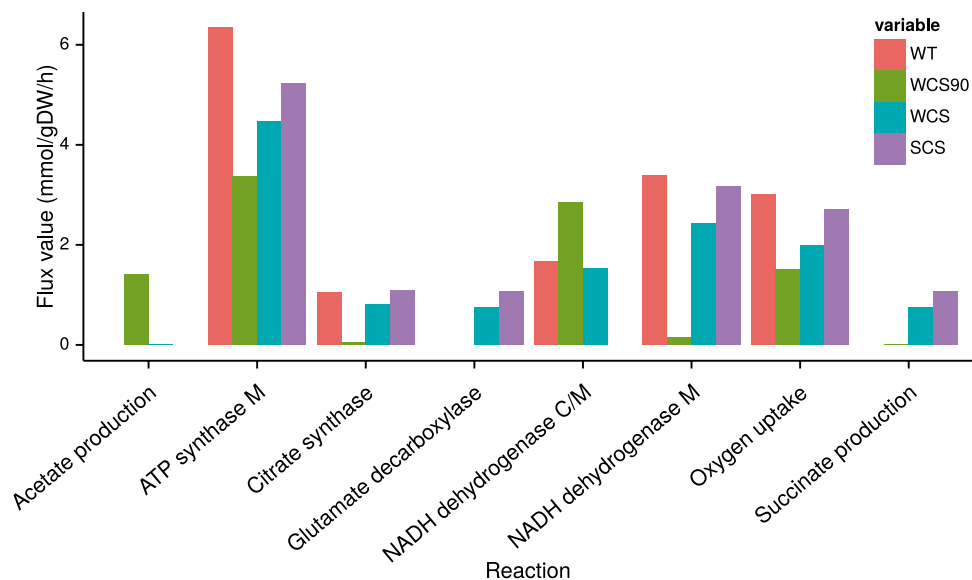
*Sub-optimal biomass* At 90% of the maximum biomass, there is a marked increase in acetate production from pyruvate, yielding reducing power (*nadh*) and protons over the wild type. In Fig. 3, the increased fluxes in cytosolic pathways, as well as the reduced flux through citrate synthase, show a metabolic shift from cell respiration to acetate production. As a result, there is little carbon for succinate production. Low flux values in mitochondrial *NADH2* can be explained by the almost absent flux in tricarboxylic acid

**Table 2** Parameters and settings used in the enumeration problem to achieve succinate-producing solutions

Parameter	Value
Undesired phenotypes	Glc $\geq -1.15$ mmol gDW <sup>-1</sup> h <sup>-1</sup> Succ/Glc $\geq -0.0001$
Desired phenotypes	Biomass $\geq 0.0001$ mmol gDW <sup>-1</sup> h <sup>-1</sup> Glc $\geq -1.15$ mmol gDW <sup>-1</sup> h <sup>-1</sup> Succ/Glc $\leq -0.0001$
Maximum solution size	9
Excluded reactions	1177
Forced uncompressed reactions	Drains, biomass and ATPM

Glc and Succ represent, respectively, glucose and succinate exchange from the external medium to the cell. Biomass is the growth pseudo-reaction. ATPM is the maintenance ATP flux

**Fig. 3** Comparison between different knockout scenarios based on Solution 1. C stands for cytosolic enzyme, while M stands for mitochondrial enzyme. C/M represents enzymes whose activity occurs across membranes. The scenarios included in this analysis are: *WT* for wild type, *WCS90* for the WCS knockout set under a 90% maximum biomass restriction, *WCS* for the WCS knockout set without restrictions and *SCS* for the Solution 1 knockout without restrictions



(TCA) cycle, with *nadh* being mostly produced in the cytosol as a result of acetate production. Even so, *atp* is still produced in the mitochondria since the proton gradient is maintained by cytosolic *NADH2* activity.

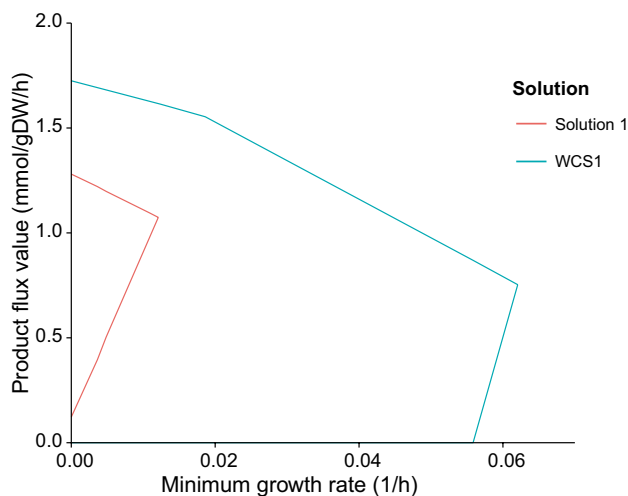
**Optimal biomass** Once biomass restrictions are lifted, there is less free carbon in the cell and pyruvate is no longer converted to acetate. In terms of *atp* production, it is still lower than wild type (WT), but is higher than the previous scenario. There is less *nadh* being produced outside the mitochondria since there is no fermentative process occurring. Oxaloacetate is produced outside the mitochondria (by pyruvate carboxylase) and enters it through the oxaloacetate-malate shuttle with *nadph* production. Ultimately, it is converted to  $\alpha$ -Kg with production of exported to the cytosol through exchange with malate. Outside the mitochondria, it is converted to glutamate through catalysis by glutamate dehydrogenase (*nadp*<sup>+</sup> forming). With the WCS1 at optimal biomass, TCA cycle flux is necessary for succinate production to occur through this pathway.

### 5.2.2 Strongly Coupled Succinate Production

With the WCS1 applied to the model, to achieve strong growth coupling, there must be a set of knockouts that can block pathways to which carbon can be diverted away from succinate production. Comparing with the WCS1 scenario at 90% biomass, acetate production from pyruvate is one of these pathways. Two of the exclusive knockouts in Solution 1 are reactions involved in acetate and ethanol fermentative synthesis (ACOAH and PYRDC). However, blocking fermentation isn't enough for strong coupling. The strongly coupled phenotype can only occur with the addition of a final knockout, PGK or GAPD (correlated reactions). The

differences in growth coupling are represented in Fig. 4, while an overview of the overall phenotype predicted to occur after the knockouts in Solution 1 is depicted in Fig. 5. This difference in growth-coupling repeats itself when comparing any two solutions from both sets.

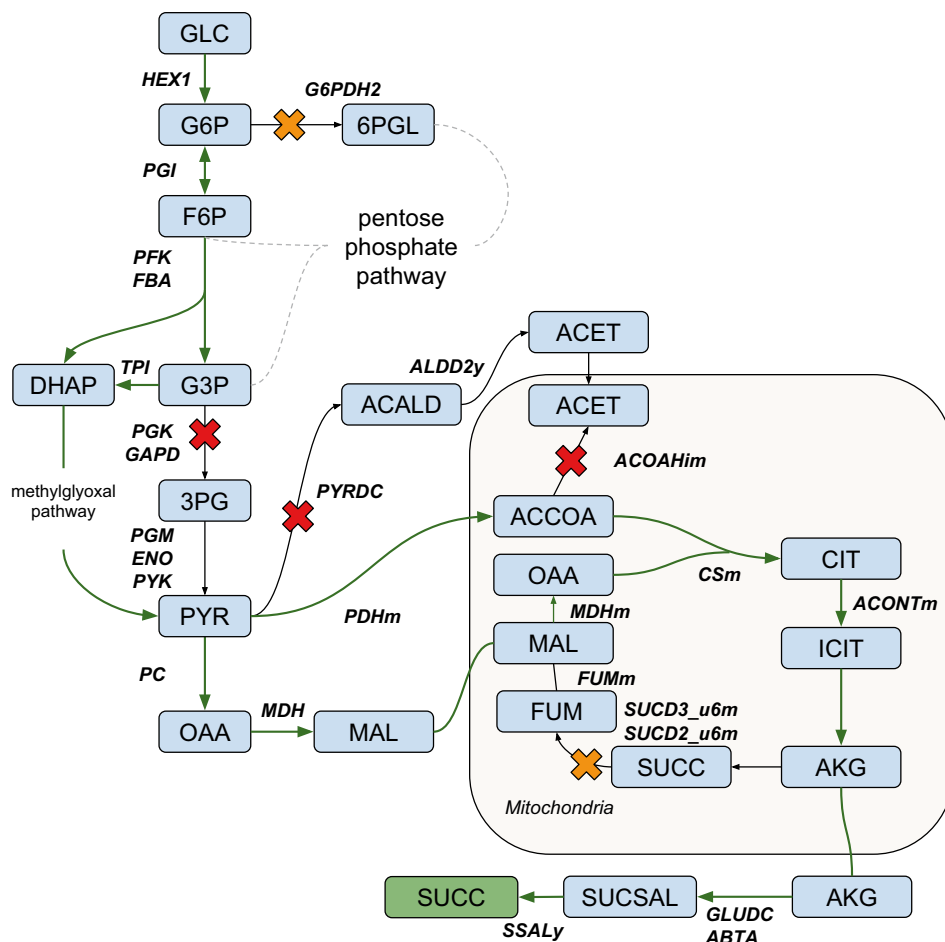
Comparing with the WCS1 phenotype at the same biomass restrictions, it is predicted that glycolysis is disrupted due to the knockout at the payoff phase. Glucose is thus converted into two *dhap* molecules at the end of the preparatory phase. The methylglyoxal pathway is now essential for pyruvate formation since it converts *dhap* to lactate (over a series of reactions), and this lactate is converted to pyruvate by lactate dehydrogenase.



**Fig. 4** Chart depicting the production envelope for the WCS and Solution 1 knockout simulations. The lines show the minimum and maximum boundaries for succinate excretion as a function of the lower bound imposed on the biomass flux



**Fig. 5** Overview of the pFBA flux distribution predicted for Solution 1. The boxes represent reactions responsible for a given conversion (closest arrow) and their flux value, with green arrows representing active fluxes on Solution 1. Crosses represent knockouts, with orange crosses corresponding to WCS1 knockouts and red ones representing knockouts exclusive to Solution 1 that are not present in the WCS1 solution



Shifting carbon through this pathway only produces 1 *nadh* per *dhap* molecule, with no *atp* production. Since there is no *atp* produced from cytosolic reactions, it must be produced inside the mitochondria by *atps*. *acoah* and *pyrdc* knockouts are likely needed so that there isn't any viable steady state fulfilling the desired phenotypes other than one carrying flux through the *tca* cycle.

Regarding the respiratory chain, there is a decrease in cytochrome oxidoreductase and an increase in cytochrome oxidase flux. Oxygen uptake and cytochrome oxidase activity is increased in this phenotype. Concurrently, there is also an increase in *NADH2* flux inside the mitochondria, which is inactive in the cytosol. There is a greater amount of carbon going in the malate–oxaloacetate shuttle and thus arriving at the *gaba* shunt, which also increases the succinate yield. Full coupling may occur since carbon is forced in the TCA cycle to allow cell growth, as *atp* is no longer produced in the cytosol. *atp* production occurs but succinate production must also occur since carbon in the cycle cannot be converted into fumarate through succinate.

## 6 Conclusions and Further Work

The availability of PA-based strain design methods is scarce when considering GSMs. The new library proposed in this work presents a useful resource for the metabolic engineering community, allowing for the enumeration of MCSs, in a way that is fit for most problems with generic CB models, while also allowing flexibility regarding problem setup.

The proposed OptFlux plugin facilitates an abstraction from complex concepts surrounding cMCS enumeration, improving ease of use and extending the already wide variety of optimization algorithms within OptFlux, maintaining a coherent overall computational interface. Also, the provided software is all made available to the community as open source allowing for third party contributions in the future.

As far as utility is concerned, it is demonstrated that predicted production robustness in MCS approaches is superior despite carrying the cost of lower maximum cell growth when compared with alternative CB methods. This

elicits the need for more research into the in vivo feasibility of strongly coupled designs, which will ultimately establish the viability of minimal cut set-based approaches altogether.

Despite all efforts, the library is still dependent on a commercial solver, but that provides a free academic license. Computation times for larger sets of deletions, even when using a state-of-the-art optimizer, can be time consuming for some enumeration problems, and others still remain out of reach. Heuristic methods or alternative formulations may help in achieving solutions for larger sizes and this will be a line of future work.

**Acknowledgements** The authors thank the project “DeYeastLibrary—Designer yeast strain library optimized for metabolic engineering applications”, Ref. ERA-IB-2/0003/2013, funded by national funds through “Fundação para a Ciência e Tecnologia / Ministério da Ciência, Tecnologia e Ensino Superior”.

## References

- Stephanopoulos G (1999) Metabolic fluxes and metabolic engineering. *Metab Eng* 1(1):1–11
- Varma A, Palsson BO, Arbor A, Varma A (1994) Stoichiometric flux balance models quantitatively predict. *Appl Environ Microbiol* 60(10):3724–3731
- Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nat Biotechnol* 28(3):245–248
- Segrè D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* 99(23):15112–15117
- Shlomi T, Berkman O, Ruppin E (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci USA* 102(21):7695–7700
- Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 84(6):647–657
- Patil KR, Rocha I, Forster J, Nielsen J (2005) Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinform* 6:308
- Schilling CH, Schuster S, Palsson BO, Heinrich R (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol Prog* 15(3):296–303
- Melzer G, Esfandabadi M, Franco-Lara E, Wittmann C (2009) Flux design: in silico design of cell factories based on correlation of pathway fluxes to desired properties. *BMC Syst Biol* 3(1):120
- Hädicke O, Klamt S (2010) CASOP: a computational approach for strain optimization aiming at high productivity. *J Biotechnol* 147(2):88–101
- Flowers D, Thompson RA, Birdwell D, Wang T, Trinh CT (2013) SMET: systematic multiple enzyme targeting—a method to rationally design optimal strains for target chemical overproduction. *Biotechnol J* 8(5):605–618
- Haus U-U, Klamt S, Stephen T (2008) Computing knock-out strategies in metabolic networks. *J Comput Biology: J Comput Mol Cell Biol* 15(3):259–268
- Jungreuthmayer C, Beurton-Aimar M, Zanghellini J (2013) Fast computation of minimal cut sets in metabolic networks with a Berge algorithm that utilizes binary bit pattern trees. *IEEE/ACM Trans Comput Biol Bioinform* 10(5):1329–1333
- von Kamp A, Klamt S (2014) Enumeration of smallest intervention strategies in genome-scale metabolic networks. *PLoS Comput Biol* 10(1):e1003378
- Schuster S, Hilgetag C (1994) On elementary flux modes in biochemical reaction systems at steady state. *J Biol Syst* 02(02):165–182
- Klamt S, Gilles ED (2004) Minimal cut sets in biochemical reaction networks. *Bioinformatics* 20(2):226–234
- Hädicke O, Klamt S (2011) Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metab Eng* 13(2):204–213
- Klamt S, Mahadevan R (2015) On the feasibility of growth-coupled product synthesis in microbial strains. *Metab Eng* 30:166–178
- Ballerstein K, von Kamp A, Klamt S, Haus U-U (2012) Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics* 28(3):381–387
- de Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S, Planes FJ (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics* 25(23):3158–3165
- Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC, Rocha M (2010) OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Syst Biol* 4(1):45
- Rocha M, Maia P, Mendes R, Pinto JP, Ferreira EC, Nielsen J, Patil K, Rocha I (2008) Natural computation meta-heuristics for the in silico optimization of microbial strains. *BMC Bioinform* 9(1):499
- Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 3(121):1–18
- Mo ML, Palsson BØ, Herrgård MJ (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* 3(1):37
- Pereira R, Nielsen J, Rocha I (2016) Improving the flux distributions simulated with genome-scale metabolic models of *Saccharomyces cerevisiae*. *Metab Eng Commun* 3:153–163
- Maia P, Rocha I, Ferreira EC, Rocha M (2008) Evaluating evolutionary multiobjective algorithms for the in silico optimization of mutant strains. In: 8th IEEE international conference on bioinformatics and bioEngineering, BIBE 2008
- Carreira R, Evangelista P, Maia P, Vilaça P, Pont M, Tomb J-F, Rocha I, Rocha M (2014) CBFA: phenotype prediction integrating metabolic models with constraints derived from experimental data. *BMC Syst Biol* 8(1):123