

Systems biology

Direct calculation of minimal cut sets involving a specific reaction knock-out

Luis Tobalina, Jon Pey and Francisco J. Planes*

CEIT and Tecnun, University of Navarra, San Sebastián 20018, Spain

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on September 23, 2015; revised on January 10, 2016; accepted on February 2, 2016

Abstract

Motivation: The concept of Minimal Cut Sets (MCSs) is used in metabolic network modeling to describe minimal groups of reactions or genes whose simultaneous deletion eliminates the capability of the network to perform a specific task. Previous work showed that MCSs were closely related to Elementary Flux Modes (EFMs) in a particular dual problem, opening up the possibility to use the tools developed for computing EFMs to compute MCSs. Until recently, however, there existed no method to compute an EFM with some specific characteristic, meaning that, in the case of MCSs, the only strategy to obtain them was to enumerate them using, for example, the standard K-shortest EFMs algorithm.

Results: In this work, we adapt the recently developed theory to compute EFMs satisfying several constraints to the calculation of MCSs involving a specific reaction knock-out. Importantly, we emphasize that not all the EFMs in the dual problem correspond to real MCSs, and propose a new formulation capable of correctly identifying the MCS wanted. Furthermore, this formulation brings interesting insights about the relationship between the primal and the dual problem of the MCS computation.

Availability and implementation: A Matlab-Cplex implementation of the proposed algorithm is available as a [supplementary material](#).

Contact: fplanes@ceit.es

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

The concept of Minimal Cut Sets was introduced in Klamt and Gilles (2004) and refined in Klamt (2006). MCSs were defined as a minimal set of reactions whose removal would render the functioning of a given objective reaction impossible. Their relationship with Elementary Flux Modes and the dualization of the main problem was mentioned in Klamt (2006) and formally exploited in Ballerstein *et al.* (2012). With this theoretical breakthrough, it became possible to calculate MCSs using algorithms designed to calculate EFMs. For example, the K-shortest EFM enumeration algorithm (de Figueiredo *et al.*, 2009) was used to enumerate MCSs in von Kamp and Klamt (2014).

Efficient calculation of EFMs has been an active area of research in the last years (de Figueiredo *et al.*, 2009; Kaleta *et al.*, 2009; von

Kamp and Schuster, 2006; Machado *et al.*, 2012; Pey and Planes, 2014; Pey *et al.*, 2015; Rezola *et al.*, 2011; Quek and Nielsen, 2014; Terzer and Stelling, 2008; Urbanczik and Wagner, 2005). Most of the algorithms have focused the attention on calculating as many EFMs as possible. However, one may only be interested in some EFMs that fulfill some constraints. While, in theory, a valid approach would be to calculate all the EFMs and then filter them according to the desired criteria, in practice it is currently impossible in large networks, as the number of EFMs grows exponentially with network size. It is this question that was addressed in a recent paper by Pey and Planes (2014), where a Mixed Integer Linear Programming (MILP) formulation was introduced to directly calculate EFMs satisfying a desired set of constraints. For example, an EFM containing several reactions of interest can be now directly calculated.

For the metabolic modeling community, it will be very valuable to directly calculate MCSs of certain characteristics. For instance, a MCS involving a specific reaction knock-out can be used to determine if it is possible to couple growth with the synthesis of a specific product (Klamt and Mahadevan, 2015). In fact, strain design for the optimization of chemical production, such as biofuels (Erdrich et al., 2014), is one of the main applications of MCSs. It is not necessary to solve a MCS problem to know if the production of a chemical is coupled to growth but, in case it is not, a MCS suggests the modifications necessary to make the coupling a reality. Another possible application of MCSs in the field of human health involves finding a complementary target to an already druggable reaction. Here, MCSs would be viewed as synthetic lethals for treating cancer or other diseases (Folger et al., 2011; Frezza et al., 2011; Kaelin, 2005; Suthers et al., 2009).

In this work, we adapt the formulation in Pey and Planes (2014) to achieve this goal. In the process, we stress that not all the EFMs in the dual problem correspond to correct MCSs of the original network, which implies that not all the algorithms designed for the calculation of EFMs can be directly applied to the calculation of MCSs without careful consideration. The modifications made to the algorithm in Pey and Planes (2014) are mathematically justified and bring interesting insights between the primal and the dual problem of the MCS computation.

2 Methods

A metabolic network of m metabolites and n reactions can be represented by an $m \times n$ stoichiometric matrix S , where each column represents a reaction with negative coefficients for the educts and positive coefficients for the products. The activity of the reactions is represented by the flux vector r . Under the steady-state assumption, the sum of fluxes that produce a compound are equal to the sum of fluxes that consume it (Eq. (1)). Irreversible reactions can only carry positive fluxes (Eq. (2)), while reversible reactions can carry negative and positive fluxes.

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

$$r_i \geq 0, \forall i \in Irrev \quad (2)$$

We would like to be able to perform a given task, in our case, to carry flux through a given reaction (usually, this reaction is the biomass reaction).

$$t^T \cdot r \geq r^* \quad (3)$$

where t is a vector of all zeros except for a 1 in the position of the reaction we want to carry flux through and r^* is the minimum amount of flux that such reaction should carry.

We are interested in a minimal group of reactions that, if blocked, would render this task impossible. To find those, we define the possible reaction knockout constraints we could impose (Eqs. (4) and (5)).

$$r_i = 0, \forall i \in Rev \quad (4)$$

$$r_i \leq 0, \forall i \in Irrev \quad (5)$$

Note that for the knock-out of irreversible reactions we only limit their upper bound (Eq. (5)), as their lower bound is already zero because of the irreversibility constraint (Eq. (2)).

As done in Ballerstein et al. (2012) and von Kamp and Klamt (2014), we formulate the dual problem of the infeasible primal problem defined by the previous constraints (Eqs. (1)–(5)). Note here that, as we modified the primal system with respect to Ballerstein

et al. (2012) and von Kamp and Klamt (2014), the arising dual system presented below is slightly different (see Supplementary Material I).

$$N \cdot \begin{pmatrix} u \\ vp \\ vn \\ w \end{pmatrix} = [S^T \quad I \quad -I \quad -t] \cdot \begin{pmatrix} u \\ vp \\ vn \\ w \end{pmatrix} = 0 \quad (6)$$

$$-r^* \cdot w \leq -c \quad (7)$$

$$vp \geq 0, vn \geq 0, w \geq 0, c > 0 \quad (8)$$

$$u \in R^m, vp \in R^n, vn \in R^n, w \in R \quad (9)$$

This dual problem can be viewed as a new stoichiometric matrix N with new flux variables u , vp , vn and w , with their respective reversibility constraints. When these variables take values different from zero, they indicate that their associated constraints in the primal problem are active. In particular, u variables are related to steady-state constraints, vp variables are related to constraints limiting the upper bound of a reaction, vn variables refer to constraints limiting the lower bound of a reaction and w variable is linked to Eq. (3). An MCS is an EFM in this dual problem that contains w and has minimal support in vp and vn , with the exception of vn variables related to the irreversibility constraints (Eq. (2)), which do not count for the minimal support (Ballerstein et al., 2012).

Constraint in Eq. (7) forces w to have a value different from zero, meaning that its associated constraint (Eq. (3)) must take part in the solution. This constraint is forcing flux through w , as c is a positive constant that rules out the trivial solution. If we want a MCS involving a specific reaction knock-out, we must also force flux through its related variable.

Recently, Pey and Planes (2014) formulated an optimization model to directly obtain EFMs fulfilling several biological constraints, such as carrying flux through a group of specific reactions. The key insight was to acknowledge in the problem formulation that an EFM has a single degree of freedom. When we have an EFM and the flux through one of its reactions is set, the rest of the fluxes of that EFM are automatically and univocally determined.

Certainly, if we add one reaction activation constraint to the steady-state constraint in a network composed of only irreversible reactions, the extreme points of the feasible region they define coincide with EFMs. In an n -dimensional space, extreme points lie on n linearly independent and binding constraints. They are mathematically represented as basic solutions, where variables are divided into basic and non-basic variables. The value of non-basic variables is set by their binding constraints (for example, a non-basic variable with a non-negativity binding constraint will be set to zero). The value of basic variables results from solving the system of linear equations that is obtained once the value of non-basic variables has been fixed.

An EFM can always be described with the steady-state constraints and flux inactivation constraints for those reactions that are not part of the EFM, collectively defining $n - 1$ linearly independent and binding constraints. Once we add one flux activation constraint to this system, all the fluxes are fully determined. Hence, if we want to include a second flux activation constraint and obtain EFMs as solutions, we need to ensure that the new constraint is redundant with respect to the binding constraints that describe the EFM. Redundancy is guaranteed by forcing that this second flux activation constraint can be written as a linear combination of the rest of the linearly independent and binding constraints. The tricky part here is

that we do not know *a priori* which reactions are active in the EFM and which ones are not. In [Pey and Planes \(2014\)](#), mixed-integer linear programming was used to select active reactions in the EFM and properly apply the linear combination requirement.

The direct application of the method described in [Pey and Planes \(2014\)](#) to our MCS problem would involve considering N as the initial network, and w and our knock-out related variable vp and vn the reactions we want to activate. Thus, we would make all the variables irreversible, splitting the u variables into two other irreversible variables and follow the implementation described in that work (see [Supplementary Material I](#) for the full formulation of the optimization problem obtained following this strategy). The solutions provided by this model constitute Elementary Flux Modes in the network N . However, not all the EFMs in N correspond to MCSs in the original network S ([Ballerstein et al., 2012](#)) and, more importantly, not all the EFMs that contain our two target variables correspond to MCSs in the original network S (see Section 3 for a toy example illustrating these events).

Following that approach, we obtain solutions that have minimal support in u , vp , vn and w , but MCSs are defined as solutions with minimal support in vp and vn variables related to knock-outs of reversible reactions, that include w . Hence, MCSs correspond to a subset of all the possible EFMs. Interestingly, we realize that MCSs are closer to the concept of Generating Flux Modes (GFMs) than EFMs. GFMs are elements of a convex basis and have minimal support with respect to the set of irreversible reactions ([Larhlami and Bockmayr, 2009](#); [Rezola et al., 2011](#)). In our case, irreversible reactions are vp , vn and w ; however, MCSs are minimal with respect to w , vp , and the subset of vn variables related to knock-outs of reversible reactions. This implies that not all the GFMs of the dual problem correspond to MCSs of the original network (see Section 3 for a toy example illustrating this event).

In order to limit the solutions of the dual problem to proper MCSs, we introduce two main modifications. The first one is to force the linear combination constraint for all the columns related to u variables, regardless of their value being different to zero or not. This means that we no longer need to split the u variables into two irreversible variables. As a bonus, these variables no longer need related binary z variables, which are now only associated with vp , vn and w variables. The second modification consists in treating the x variables, which represent the coefficients of the linear combination, as if they were the reaction flux variables of the original problem and constraining their reversibility accordingly. We present below the full mathematical model:

$$\text{minimize} \quad \sum_i zp_i + \sum_{i \in Rev} zn_i \quad (10)$$

subject to:

$$N \cdot \begin{pmatrix} u \\ vp \\ vn \\ w \end{pmatrix} = [S^T \quad I \quad -I \quad -t] \cdot \begin{pmatrix} u \\ vp \\ vn \\ w \end{pmatrix} = 0 \quad (11)$$

$$-r^* \cdot w \leq -c \quad (12)$$

$$(0 \quad d_p \quad d_n \quad 0) \cdot \begin{pmatrix} u \\ vp \\ vn \\ w \end{pmatrix} \geq b \quad (13)$$

$$\alpha \cdot \begin{pmatrix} zp \\ zn \\ zw \end{pmatrix} \leq \begin{pmatrix} vp \\ vn \\ w \end{pmatrix} \leq M \cdot \begin{pmatrix} zp \\ zn \\ zw \end{pmatrix} \quad (14)$$

$$zp_i + zn_i \leq 1, \forall i \in Rev \quad (15)$$

$$\begin{bmatrix} S & 0 \\ I & 0 \\ -I & 0 \\ -t^T & r^* \end{bmatrix} \cdot x = \begin{pmatrix} 0 \\ d_p + \epsilon_p - \delta_p \\ d_n + \epsilon_n - \delta_n \\ 0 + \epsilon_w - \delta_w \end{pmatrix} \quad (16)$$

$$M \cdot \begin{pmatrix} 1 - zp \\ 1 - zn \\ 1 - zw \end{pmatrix} \geq \begin{pmatrix} \epsilon_p + \delta_p \\ \epsilon_n + \delta_n \\ \epsilon_w + \delta_w \end{pmatrix} \quad (17)$$

$$\begin{pmatrix} vp \\ vn \\ w \end{pmatrix} \geq 0, \begin{pmatrix} \epsilon_p \\ \epsilon_n \\ \epsilon_w \end{pmatrix} \geq 0, \begin{pmatrix} \delta_p \\ \delta_n \\ \delta_w \end{pmatrix} \geq 0 \quad (18)$$

$$x_i \geq 0, \forall i \in Irrev \quad (19)$$

$$\begin{pmatrix} zp \\ zn \\ zw \end{pmatrix} \in \{0, 1\} \quad (20)$$

$$u \in R^m, vp \in R^n, vn \in R^n, w \in R \quad (21)$$

$$x \in R^{n+1} \quad (22)$$

$$\epsilon_p \in R^n, \epsilon_n \in R^n, \epsilon \in R \quad (23)$$

$$\delta_p \in R^n, \delta_n \in R^n, \delta \in R \quad (24)$$

$$c > 0, b > 0 \quad (25)$$

where α and M represent a sufficiently small and large constants, respectively, and d_p and d_n are vectors of all zeros except for a single 1 in the position related to the knock-out constraint that we want to activate. If the knock-out we want to enforce involves an irreversible reaction we will only set a 1 in d_p and leave d_n as a vector of all zeros. As in [Pey and Planes \(2014\)](#), variables ϵ_p , ϵ_n , ϵ_w , δ_p , δ_n and δ_w allow the linear combination constraint (Eq. (16)) to be applied only to active variables with the help of variables zp , zn and zw and their linking constraints (Eqs. (14) and (17)). The optimal solution of this optimization problem corresponds to a MCS that includes our desired reaction knock-out.

The formulation of the problem can be simplified (see [Supplementary Material I](#)), but we have written it here in a way that the similarities and differences with the formulation in [Pey and Planes \(2014\)](#) are easier to spot, aiming to facilitate the understanding of the model. To that end, we also provide in [Supplementary Material I](#) the resulting mathematical model for a particular toy example.

The reasons why these modifications work have a mathematical underpinning that we proceed to explain in the following paragraphs.

In contrast with [Pey and Planes \(2014\)](#), u variables are not split into two irreversible steps and, therefore, the linear combination

(Eq. (16)) must always apply to them. In particular, since the value 0 is not a bound for u variables and they can take any real value, they should always be considered as basic variables and thus, they must take part in the linear combination constraint. With this modification alone, the algorithm goes from calculating EFMs to obtaining GFMs satisfying several constraints (see [Supplementary Material I](#)). However, as noted above, this is not sufficient to obtain MCSs.

The justification for the second modification is somewhat more difficult to grasp. It has its roots in duality theory and it involves the correct manipulation of equality constraints in Eq. (11). Dual variables of equality constraints are always unrestricted sign variables, but dual variables of inequality constraints are always either non-negative or non-positive variables. In this context, and because we do not care about their value in the optimal solution, vn variables related to the irreversibility constraints are actually explicit excess variables in Eq. (11) and, in order to directly obtain MCSs, they can be removed without altering the solution space, transforming the equalities where they appear into *greater than or equal* inequalities (see [Supplementary Material I](#)). This action has consequences in the linear combination requirement introduced in [Pey and Planes \(2014\)](#), here Eq. (16), as now the linear combination coefficient variables x cannot be of unrestricted sign type by default and must be non-negative in these inequalities (see [Supplementary Material I](#)). In fact, those x variables are dual variables of the corresponding constraints and they must abide by duality theory. Eventually, we have that the x variables must obey the same irreversibility constraints that the r variables of the original problem do.

These modifications have a surprising consequence in the interpretability of the problem. First, x variables can be treated as the flux variables of the original problem and, second, we are forcing $S \cdot x = 0$, which resembles the steady-state constraint of the original model. This means that the values of the x variables in a valid solution of our direct MCS problem formulation can be interpreted as a valid flux distribution in the original network.

Finally, we can iteratively enumerate MCSs fulfilling our conditions by introducing a new constraint that eliminates previously obtained solutions ([Figueiredo et al., 2009](#); [von Kamp and Klamt, 2014](#)).

$$\sum_i z p_i^k \cdot z p_i + \sum_{i \in \text{Re } v} z n_i^k \cdot z n_i \leq \left(\sum_i z p_i^k + \sum_{i \in \text{Re } v} z n_i^k \right) - 1, \quad k = 1, \dots, K \quad (26)$$

3 Results

In this section, we first introduce a toy example where it can be seen that not every EFM in the dual network corresponds to a MCS in the primal network ([Ballerstein et al., 2012](#)) and check that our algorithm is capable of correctly identifying MCSs for each reaction. Then, we apply our algorithm to two networks of different sizes, namely the *E.coli* core metabolic network and the iAF1260 *E.coli* model, comparing the behaviour of our algorithm with that of the enumeration approach presented in [von Kamp and Klamt \(2014\)](#). The model was implemented in Matlab, using CPLEX as the underlying optimization software, and can be found in [Supplementary Material II](#). The computations were carried out on a 64 bit Intel Xeon E5-1620 v2 at 2.70 GHz (4 cores) and 16 GB of RAM.

3.1 Toy example

[Figure 1](#) shows a primal toy network and the dual network arising from its corresponding MCS problem. This network has only one irreversible reaction (r_2) and we will consider r_8 as the target reaction.

The network includes an irreversible reaction, r_9 , that is not related in any way to the target reaction r_8 , hence, there is no MCSs containing r_9 that blocks the activity of r_8 . However, there exists an EFM containing the input reaction for node r_9 in the dual network (e.g. $\{vp_9, vp_6, vn_5, u_B\}$), but that EFM does not contain our target related variable w .

Now we draw the attention to r_6 . This reaction participates in a MCS with r_2 and r_4 . It is easy to see that the dual network contains an EFM composed of the inputs to r_2 , r_4 and r_6 (vp_2 , vp_4 and vp_6 respectively), the arcs u_C , u_D and u_E and the target w . However, there is also another EFM that contains the input to node r_6 (vp_6) together with the input to node r_1 (vp_1), the output to node r_5 (vn_5), the arcs u_A , u_C , u_D and u_E and the target w . This EFM would correspond to a MCS composed of r_1 and r_6 , which is incorrect because r_1 is already a MCS on its own. It turns out that this second EFM involves less knock-out related dual variables than the former (2 versus 3, because the output of node r_5 , vn_5 , does not count), which implies that the direct application of the formulation in [Pey and Planes \(2014\)](#) to the dual network asking for the shortest EFM containing w and the input to r_6 would obtain this EFM first. This invalidates the direct use of their formulation for obtaining MCSs involving a specific reaction knock-out.

However, the new model introduced in this work correctly captures the MCS, as the incorrect MCS solution is infeasible for it. In particular, the values of the x variables (corresponding to the coefficients of the linear combination in Eq. (16)) in the case of the invalid EFM are all non-negative except for x_9 , which has a negative value that our new formulation forbids. We can interpret the values of the x variables as a flux distribution in the primal network, resulting in reactions r_6 , r_3 , r_7 and r_8 carrying flux in the forward direction and reaction r_9 carrying flux in the backward direction. Obviously, this flux distribution is not a valid one, because r_9 is irreversible. Our new model takes this into account and renders this solution infeasible.

Because of the way that MCSs are enumerated using k -shortest in [von Kamp and Klamt \(2014\)](#), from smallest to largest and taking into account only variables related to knock-out constraints, EFMs in the dual network that are not MCSs in the primal network are avoided. In the case of the network in [Figure 1](#), the EFM corresponding to the MCS that contains only r_1 is obtained before the invalid EFM that contains r_1 and r_6 . Hence, the solution enumeration constraint avoids the calculation of the invalid EFM, as it contains an already calculated MCS, r_1 .

3.2 Application to *E.coli* core metabolism

Here, we apply our algorithm to the *E.coli* core network available in the COBRA Toolbox ([Schellenberger et al., 2011](#)). This network contains 72 metabolites and 95 reactions. We conducted structural analysis of the reported network, so we did not take into account possible growth medium constraints (e.g. glucose and oxygen supply) or compound production requirements (e.g. ATP maintenance).

As with any other MILP, the time needed to find the optimal solution may widely vary depending on the chosen reaction. This disparity in problem difficulty can already be experienced in this simple network. Taking the biomass reaction as our target for the MCSs, we calculated one MCS for each reaction. We used the reduced formulation, presented in [Supplementary Material I](#), which deals with a lower number of variables than the one introduced in the Methods section. The median time to obtain a solution was 0.27 s, and only 3 reactions required more than 15 s to reach the optimal solution (the maximum was 380 s). Note here that, with other values for the Cplex parameters, we were able to obtain all these MCSs in less

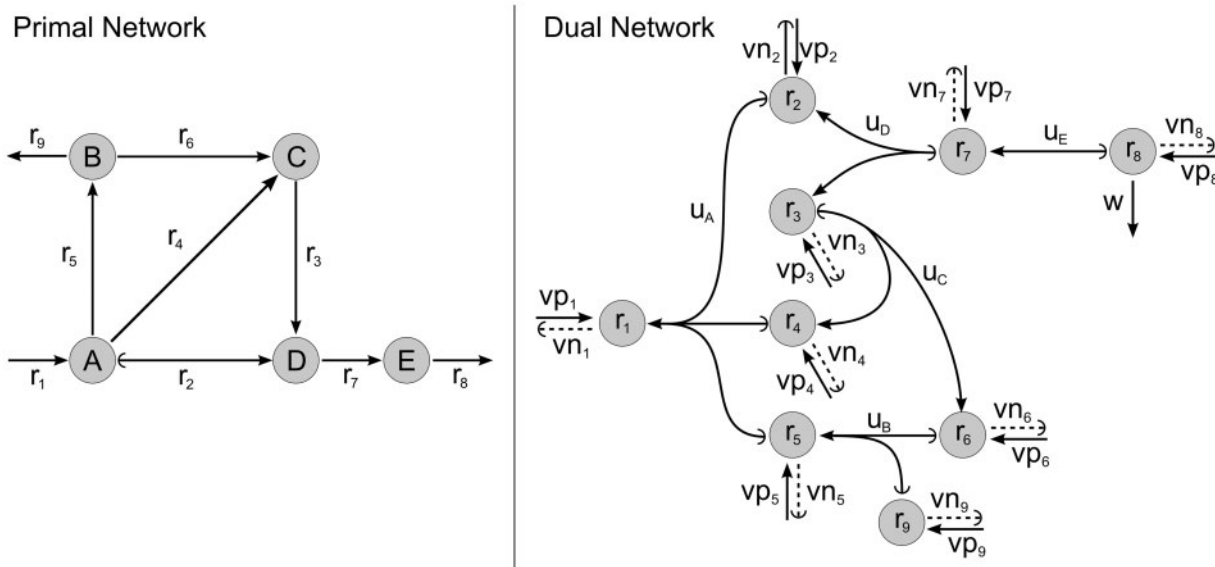


Fig. 1. Toy example. Example metabolic network illustrating an original network in the primal and the associated dual network of its MCS problem. The original network only has one reversible reaction, r_2 . The target reaction, for which we want to block its flux, is r_2 . The external input reactions to the reaction nodes in the dual network correspond to vp variables and the external output reactions to vn variables. MCSs of this network are: $\{r_1\}$, $\{r_2\}$, $\{r_2, r_3\}$, $\{r_2, r_4, r_5\}$ and $\{r_2, r_4, r_6\}$. Two dual EFMs of r_1 knock-out that include w are $\{vp_1, vn_9, u_A, u_B, u_C, u_D, u_E, w\}$ and $\{vp_1, vp_6, vn_5, u_A, u_C, u_D, u_E, w\}$. The former corresponds to a MCS, while the latter does not

than 60 s, but the current ones achieved better results in larger networks.

We found that 18 reactions where MCSs on their own, 36 participated in a 2 reaction MCS, 11 in a 3 reaction MCS, 7 in a 4 reaction MCS, 10 in a 5 reaction MCS, 2 in a 6 reaction MCS, and 3 in an 8 reaction MCS, while 8 reactions did not participate in any MCS.

We also enumerated MCSs up to size 10, i.e. involving a maximum of 10 reaction knock-outs, using the implementation of the enumeration approach provided in von Kamp and Klamt (2014), which applies a compression technique prior to the enumeration of the MCSs (the reduced network has 33 metabolites and 57 reactions). We obtained 240 186 MCSs in 270 s, 18 of size 1, 111 of size 2, 223 of size 3, 396 of size 4, 1302 of size 5, 3633 of size 6, 4370 of size 7, 14 305 of size 8, 62 046 of size 9 and 153 782 of size 10. In the compressed network, these MCSs are represented by 11 of size 1, 48 of size 2, 67 of size 3, 130 of size 4, 130 of size 5, 188 of size 6, 176 of size 7, 220 of size 8, 334 of size 9 and 411 of size 10. It is clear that the number of MCSs grows exponentially as the size of the network increases. This poses an issue to the performance of the enumeration approach in larger networks.

If one is interested in collecting many short MCSs, or if the network is small, the enumeration approach could be more appropriate than our method. However, if one is interested in the MCSs in which a particular reaction participates, our approach will be more suited. In particular, if our knock-out of interest only participates in high-order MCSs and the network we are dealing with has a considerable size, with the enumeration approach we would require an impractical amount of time to obtain it. If it does not participate in any MCS, we would need to enumerate all the possible MCSs to find it out. Our method clearly poses an advantage in these situations.

3.3 Application to *E.coli* iAF1260 metabolic network

The *E.coli* iAF1260 (Feist et al., 2007), available for download at the BiGG database (Schellenberger et al., 2010), contains 1668

metabolites and 2382 reactions. As before, structural analysis of the reported network was conducted. We used the implementation of the enumeration approach provided in von Kamp and Klamt (2014) to calculate all the MCSs up to size 4 (4572 MCSs in total), and calculated one MCS for each reaction with the reduced formulation of our method (2382 MCSs in total). These MCSs are all related to the biomass reaction. In this case, we only allowed our algorithm a maximum of 5 min to find each MCS.

The enumeration approach needed 13 h to calculate all the MCSs with a maximum of 4 reaction knock-outs, while our approach was capable of discovering MCSs involving up to 4 reaction knock-outs among the optimal solutions returned in less than 5 min. Out of the 1447 solutions returned under the 5 min time limit, there were 279 MCSs of size 1, 128 of size 2, 145 of size 3 and 56 of size 4. The mean time to obtain these solutions was 32.4 s. The remaining 839 reactions corresponded to blocked reactions and did not participate in any MCS (our algorithm discovered these cases in less than 1 s on average).

More importantly, solutions returned by the algorithm after hitting the 5 min time limit show that the size of the MCSs can go as far as to include 139 reactions. These solutions that are returned because the time limit is reached, however, are not guaranteed to be optimal and thus they may not be correct MCSs (although some of the reactions included in that solution will form a proper MCS) or a guarantee that the reaction has no associated MCS if no solution was returned. However, out of the 935 cases hitting the time limit, only 4 solutions were incorrect MCS (which contained a set of reactions that formed a proper MCS), 722 solutions were correct MCSs (although not necessarily optimal) and no solution was returned in the 209 remaining cases, which does not guarantee infeasibility. A histogram of the sizes of these MCSs is shown in Figure 2.

Going back to the 209 cases that contained no MCSs mentioned above, we found that 24 of them had some associated MCS (within the 5 min time limit) by changing some of the solver parameter settings. With a complicated MILP, such as the one proposed here, the

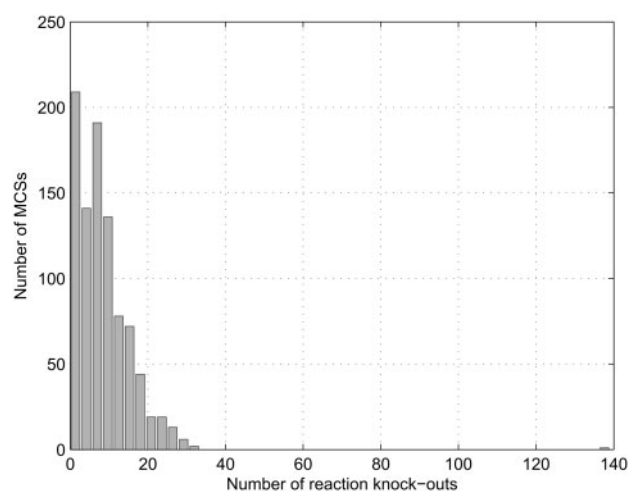


Fig. 2. MCS length histogram. Histogram of the number of reaction knock-outs participating in the MCSs returned by the direct MCS calculation algorithm when hitting the time limit on the *E.coli* iAF1260 metabolic network case study

tuning of the solver parameter values may have a big influence on the time needed to find a solution or guarantee infeasibility. The performance even changes depending on the reaction for which we are looking for a MCS. Hence, finding a set of parameters that works equally well in all cases may be impossible. Notwithstanding, the implementation we provide has shown an acceptable performance.

Overall, the application of our algorithm to the *E.coli* iAF1260 metabolic networks shows that it is applicable to the genome-scale setting.

4 Conclusion

In this work, we have introduced an optimization model to calculate MCSs involving a specific reaction knock-out. We have emphasized that not all the EFMs in the dual problem correspond to valid MCSs in the primal problem (Ballerstein et al., 2012). Also, we acknowledge that MCSs are closer to the definition of GFM of the dual problem than to EFMs, although not all the GFMs in the dual problem correspond to valid MCSs in the primal problem either. As a side effect, this work extends the formulation in Pey and Planes (2014) to the calculation of GFMs satisfying several constraints and deepens in the mathematical understanding of the model.

This formulation is important because it makes possible to find if each reaction can participate in a MCS or not. Without it, the only possible way would be to enumerate all the MCSs (using the method in von Kamp and Klamt, 2014, for example) until one incorporating the reaction of interest is found. This may become impractical, as the enumeration process is currently limited by memory and time constraints. Given that some reactions may participate in very high order MCSs, this drawback is evident. Our formulation can also be used to enumerate those MCSs that include a particular reaction knock-out.

Looking at the model and the solution a bit more in depth, we can realize that, whenever the model chooses to activate a particular reaction knock-out, the x variable associated to that reaction that has been knocked-out gets equal to zero. This is in line with interpreting the x variables as fluxes in the primal network: if the reaction has been knocked-out it cannot carry any flux. As an exception,

the x variable associated to the reaction whose knock-out we are interested in gets a fixed value different from zero. Thus, all the possible knock-outs of reactions that participate in the flux distribution (x value different from zero) are discarded, except for the one we were interested in. Then, the algorithm chooses the minimal number of knock-outs that make all the other remaining knock-out constraints redundant.

Regarding non-optimal solutions to the MCS problem formulation, if we stop the solution process before reaching optimality, the last feasible solution calculated will not necessarily be a MCS, although it will contain one. The MCS contained in that solution will not be necessarily optimal according to our objective function either. It may exist another MCS involving less reaction knock-outs than the MCS contained in the non-optimal solution.

Future work will include considering how to deal with more general reaction bounds as well as considering multiple reaction knock-out or gene knock-out constraints. The formulation can also be easily extended to consider the calculation of MCSs including several specific reaction knock-outs constraints.

We also consider the incorporation of experimental data through weights in the objective function, for example. With information about the probabilities of each reaction being really active in the actual organism, longer MCSs may become more interesting than shorter ones.

In summary, we believe that this work will boost new research in metabolic engineering around MCSs, including its applications in strain design and human health. Moreover, the insights brought by our formulation could lead to more efficient algorithms for the calculation of MCSs.

Funding

This work has been supported by the Basque Government [to L.T.] and the Minister of Economy and Competitiveness of Spain [BIO2013-48933].

Conflict of Interest: none declared.

References

- Ballerstein, K. et al. (2012) Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics*, **28**, 381–387.
- Erdrich, P. et al. (2014) Cyanobacterial biofuels: new insights and strain design strategies revealed by computational modeling. *Microb. Cell Fact.*, **13**, 128.
- Feist, A.M. et al. (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.
- de Figueiredo, L.F. et al. (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics*, **25**, 3158–3165.
- Folger, O. et al. (2011) Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.*, **7**, 501.
- Frezza, C. et al. (2011) Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature*, **477**, 225–228.
- Kaelin, W.G. (2005) The concept of synthetic lethality in the context of anti-cancer therapy. *Nat. Rev. Cancer*, **5**, 689–698.
- Kaleta, C. et al. (2009) EFMEvolver: Computing elementary flux modes in genome-scale metabolic networks. In: *Lecture Notes in Informatics Proceedings*, pp. 179–189.
- von Kamp, A. and Klamt, S. (2014) Enumeration of smallest intervention strategies in genome-scale metabolic networks. *PLoS Comput. Biol.*, **10**, e1003378.
- von Kamp, A. and Schuster, S. (2006) Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, **22**, 1930–1931.
- Klamt, S. (2006) Generalized concept of minimal cut sets in biochemical networks. *Biosystems*, **83**, 233–247.

- Klamt, S. and Gilles, E.D. (2004) Minimal cut sets in biochemical reaction networks. *Bioinformatics*, **20**, 226–234.
- Klamt, S. and Mahadevan, R. (2015) On the feasibility of growth-coupled product synthesis in microbial strains. *Metab. Eng.*, **30**, 166–178.
- Larhlami, A. and Bockmayr, A. (2009) A new constraint-based description of the steady-state flux cone of metabolic networks. *Discrete Appl. Math.*, **157**, 2257–2266.
- Machado, D. *et al.* (2012) Random sampling of elementary flux modes in large-scale metabolic networks. *Bioinformatics*, **28**, i515–i521.
- Pey, J. *et al.* (2015) TreeEFM: calculating elementary flux modes using linear optimization in a tree-based algorithm. *Bioinformatics*, **31**, 897–904.
- Pey, J. and Planes, F.J. (2014) Direct calculation of elementary flux modes satisfying several biological constraints in genome-scale metabolic networks. *Bioinformatics*, **30**, 2197–2203.
- Quek, L.E. and Nielsen, L.K. (2014) A depth-first search algorithm to compute elementary flux modes by linear programming. *BMC Syst. Biol.*, **8**, 94.
- Rezola, A. *et al.* (2011) Exploring metabolic pathways in genome-scale networks via generating flux modes. *Bioinformatics*, **27**, 534–540.
- Schellenberger, J. *et al.* (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinf.*, **11**, 213.
- Schellenberger, J. *et al.* (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protocols*, **6**, 1290–1307.
- Suthers, P.F. *et al.* (2009) Genome-scale gene/reaction essentiality and synthetic lethality analysis. *Mol. Syst. Biol.*, **5**, 301.
- Terzer, M. and Stelling, J. (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, **24**, 2229–2235.
- Urbanczik, R. and Wagner, C. (2005) An improved algorithm for stoichiometric network analysis: theory and applications. *Bioinformatics*, **21**, 1203–1210.