Metabolic network reductions

Mojtaba Tefagh[1*] Stephen P. Boyd[1]

March 13, 2019

Abstract

Genome-scale metabolic networks are exceptionally huge and even efficient algorithms can take a while to run because of the sheer size of the problem instances. To address this problem, metabolic network reductions can substantially reduce the overwhelming size of the problem instances at hand. We begin by formulating some reasonable axioms defining what it means for a metabolic network reduction to be "canonical" which conceptually enforces reversibility without loss of any information on the feasible flux distributions. Then, we start to search for an efficient way to deduce some of the attributes of the original network from the reduced one in order to improve the performance. As the next step, we will demonstrate how to reduce a metabolic network repeatedly until no more reductions are possible. In the end, we sum up by pointing out some of the biological implications of this study apart from the computational aspects discussed earlier.

Keywords: systems biology; metabolic network analysis; canonical metabolic network reduction; quantitative flux coupling analysis; QFCA.

Author summary

Metabolic networks appear at first sight to be nothing more than an enormous body of reactions. The dynamics of each reaction obey the same fundamental laws and a metabolic network as a whole is the melange of its reactions. The oversight in this kind of reductionist thinking is that although the behavior of a metabolic network is determined by the states of its reactions in theory, nevertheless it cannot be inferred directly from them in practice. Apart from the infeasibility of this viewpoint, metabolic pathways are what explain the biological functions of the organism and thus also what we are frequently concerned about at the system level.

Canonical metabolic network reductions decrease the number of reactions substantially despite leaving the metabolic pathways intact. In other words, the reduced metabolic networks are smaller in size while retaining the same metabolic pathways. The possibility of such operations is rooted in the fact that the total degrees of freedom of a metabolic network in the steady-state conditions are significantly lower than the number of its reactions because of some emergent redundancies. Strangely enough, these redundancies turn out to be very well-studied in the literature.

1 Introduction

Two decades ago, the first genome-scale metabolic network reconstruction of the cellular metabolism of an organism was published [1] shortly after the first genome was sequenced. From that time on, the ever-increasing advances in the high-throughput omics technologies have allowed for the more and more comprehensive reconstructions of exponentially growing sizes [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. These reconstructions have numerous applications in contextualization of high-throughput data, guidance of metabolic engineering, directing hypothesis-driven discovery, interrogation of multi-species relationships, and network property discovery [20]. However, the vast amount of data for

[1]Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA, USA
* Corresponding author:
E-mail: mtefagh@stanford.edu (MT)

some organisms can be a two-edged sword which makes many essential tasks in the metabolic network analysis computationally intractable.

To overcome the demands of systems biology, even while they are outpacing Moore's law [21], faster computational techniques are needed to enable the current methods to scale up to match the progress of high-throughput data generation in a prospective manner. As a natural solution, reducing the size of genome-scale metabolic networks has always been exploited to the advantage of performance for predetermined computational tasks [22, 23, 24, 25, 26]. However, all of these mentioned studies assume that a set of protected reactions which must be retained in the reduced metabolic network is given in advance and the other reactions are dispensable. In this way, these metabolic network reductions are not agnostic to the downstream analysis and lose information on some of the reactions which are irrelevant to the task at hand.

In the field of *flux coupling analysis* (FCA), from the very beginning *Flux Coupling Finder* (FCF) [27] considers aggregating all the isozymes and removing the blocked reactions. More recently, *fast flux coupling calculator* (F2C2) [28] considers merging the enzyme subsets [29] too. *FluxAnalyzer* [30] also detects conservation relations as a preprocessing step. *MONGOOSE* [31] employs a similar loss-free network reduction to convert the input model into a canonical form.

Although the practical implications of these findings affirm the plausibility of their general-purpose reductions in a common sense approach, to the best of our knowledge the formal definition of a metabolic network reduction, in general, is not clear so far. The aforementioned studies provide strong support for the idea that such general-purpose reductions preserve some interesting attributes of metabolic networks, *e.g.*, the flux coupling features. In this article, we specify a broad class of metabolic network reductions with respect to which some generally desired properties of metabolic networks are invariant.

Outline. In §2, we briefly overview the FCA framework. In §3, we define and study single and multiple sequential metabolic network reductions, respectively. Afterward in §4, we argue that the reduced metabolic networks are not only interesting from the computational perspective, but also they are biologically interpretable. Finally in §5, we conclude by highlighting the major contributions of this work.

2 Background

We specify a metabolic network by an ordered quadruple $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathcal{I})$ where $\mathcal{M} = \{M_i\}_{i=1}^m$ denotes the set of metabolites of size m, $\mathcal{R} = \{R_i\}_{i=1}^n$ denotes the set of reactions of size n, S denotes the $m \times n$ stoichiometric matrix, and $\mathcal{I} \subseteq \mathcal{R}$ denotes the set of irreversible reactions. Since it is often the case that $n \gg m$, we consider n as the size of $\mathcal{N}$ too.

In the constraint-based analysis of metabolic networks, the constraints $Sv = 0$ and $v_{\mathcal{I}} \geq 0$ are imposed on the metabolic network by the steady-state conditions and the definition of irreversible reactions, respectively. By a slight abuse of notation, $v_{\mathcal{I}} \geq 0$ means $v_i \geq 0$ for all the indices i for which $R_i \in \mathcal{I}$. We denote the *steady-state flux cone* [32] by

$$\mathcal{C} = \{v \in \mathbf{R}^n \mid Sv = 0,\ v_{\mathcal{I}} \geq 0\},$$

The feasible flux distributions of $\mathcal{N}$ are then defined to be the members of $\mathcal{C}$.

We call $R_k \in \mathcal{R}$ a blocked reaction if $v_k = 0$ for all the feasible flux distributions $v \in \mathcal{C}$. To the end of this paper, whenever we assume that all the blocked reactions are removed, by this, we also assume that if a reversible reaction R_k is blocked in only one direction, then this blocked direction is removed too meaning that R_k is included in $\mathcal{I}$ subsequently. In §3, we will review these trivial ways of reducing a metabolic network in more details.

In order to derive even more metabolic network reductions, we should exploit other redundancies of the steady-state flux cone analogous to the case of blocked reactions where the rate of a blocked reaction is always zero irrespective of what the other flux coefficients are equal to. A somewhat less obvious situation is when the rate of one reaction is unambiguously determined by another one as is the case in the following definitions of FCA.

Definition 1 ([27])**.** For an arbitrary pair of unblocked reactions $R_i, R_j \in \mathcal{R}$:

Directional coupling: R_i is *directionally coupled* to R_j, denoted by $R_i \longrightarrow R_j$, if for all feasible flux distributions $v_i \neq 0$ implies $v_j \neq 0$.

Partial coupling: R_i is *partially coupled* to R_j, denoted by $R_i \longleftrightarrow R_j$, if both $R_i \longrightarrow R_j$ and vice versa $R_j \longrightarrow R_i$ hold.

Full coupling: R_i is *fully coupled* to R_j, denoted by $R_i \Longleftrightarrow R_j$, if there exists a *full coupling equation* (FCE)

$$v_i = cv_j \qquad c \neq 0, \tag{1}$$

which holds for all $v \in \mathcal{C}$.

In this definition, full coupling is defined by means of a flux coupling equation, but directional and partial couplings are defined qualitatively. We can apply the following proposition to get the equivalent quantitative definitions of directional and partial couplings by introducing two other flux coupling equations in analogy to FCE. *Quantitative flux coupling analysis* (QFCA) [33] redefines the original flux coupling relations of FCA by the existence of flux coupling equations.

Proposition 1 ([33])**.** *Suppose that $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathcal{I})$ has no irreversible blocked reactions. Let R_j be an arbitrary unblocked reaction, and $\mathcal{D}_j \subseteq \mathcal{I}$ denote the set of all the irreversible reactions which are directionally coupled to R_j excluding itself. Then, $\mathcal{D}_j \neq \emptyset$ if and only if there exists $c_d > 0$ for each $R_d \in \mathcal{D}_j$, such that the following* directional coupling equation *(DCE)*

$$v_j = \sum_{d:R_d \in \mathcal{D}_j} c_d v_d, \tag{2}$$

holds for all $v \in \mathcal{C}$. Moreover, for any unblocked $R_i \notin \mathcal{I}$, we have $R_i \longrightarrow R_j$ if and only if there exists an extended directional coupling equation *(EDCE)*

$$v_j = \sum_{d:R_d \in \mathcal{D}_j} c'_d v_d + c'_i v_i \qquad c'_i \neq 0,$$

which holds for all $v \in \mathcal{C}$.

Prior to QFCA, the general attitude towards directional coupling was that it is a one-way relation which is, of course, true considering only one pair of reactions. However QFCA, for the first time, revealed that this relation becomes two-sided if we consider the set of directionally coupled reactions to a single reaction which are sufficient to infer the flux coefficient of it. This redundancy motivates generalizing the previously known application of FCE in the metabolic network reductions to DCE.

In the sequel, we will formally show that metabolic network reductions are intrinsically related to QFCA. The intuition is that redundancy in the steady-state flux cone creates flux coupling relations, and metabolic network reductions reduce redundancy by compressing this information. Identifying and eliminating statistical redundancy is called lossless compression in information theory [34]. We will prove results of the same flavor to show that QFCA induces lossless reductions and any other metabolic network reduction beyond that is a lossy compression.

We continue by the following general mathematical definition, and subsequently we use this terminology to formalize the notion of a minimal feasible flux distribution. The support of a flux distribution $v \in \mathbf{R}^n$ is denoted by $\mathrm{supp}(v)$, and is defined to be the set of those reactions in $\mathcal{R}$ which are active in this flux distribution, *i.e.*,

$$\mathrm{supp}(v) = \{R_i \in \mathcal{R} \mid v_i \neq 0\}.$$

Definition 2 ([35])**.** We call a nonzero feasible flux distribution $0 \neq v \in \mathcal{C}$ an *elementary mode* (EM), if its support is minimal, or equivalently, if there does not exist any other nonzero feasible flux distribution $0 \neq u \in \mathcal{C}$ such that $\mathrm{supp}(u) \subset \mathrm{supp}(v)$.

Given a finite number of vectors $v_1, v_2, \ldots, v_d$, a *conic combination* of these vectors is any weighted sum $\theta_1 v_1 + \theta_2 v_2 + \cdots + \theta_d v_d$ with nonnegative weights $\theta_1, \theta_2, \ldots, \theta_d \geq 0$. The set of all such conic combinations is called the polyhedral convex cone generated by $v_1, v_2, \ldots, v_d$ [36].

One of the main properties of EM is that the set of all EMs in a metabolic network generates its steady-state flux cone. Another way of expressing this is to say that any arbitrary feasible flux distribution can be written as a conic combination of EMs. With QFCA in mind, this property implies that to establish a flux coupling relation instead of going through all the feasible flux distributions, it suffices to validate

the corresponding flux coupling equation only for the EMs. Accordingly, reducing the size of a metabolic network so long as preserving the information on its EMs and in turn, the flux coupling relations leads to a significant speed-up during QFCA.

As we will see in the next section, the relationship between metabolic network reductions and QFCA does not end here and is surprisingly reciprocal. The uncovered connection between these two concepts sheds light on the mutual nature of both in relation to the core concept of EM. Besides presenting an axiomatic characterization to rigorously examine this connection, we will also derive metabolic network reductions inspired by QFCA which, though intuitive and straightforward, are nevertheless general enough to reduce any arbitrary metabolic network into a minimal irreducible one.

3 Methods

As an illustrative example of the method, consider the following toy metabolic network in Fig 1. We name it $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathcal{I})$, where $\mathcal{M} = \{M_1, M_2, M_3\}$, $\mathcal{R} = \{R_1, R_2, R_3, R_4, R_5\}$, $\mathcal{I} = \mathcal{R}$, and

$$S = \begin{bmatrix} +1 & -1 & 0 & +2 & 0 \\ 0 & +1 & -1 & 0 & 0 \\ 0 & 0 & 0 & +1 & -1 \end{bmatrix}.$$

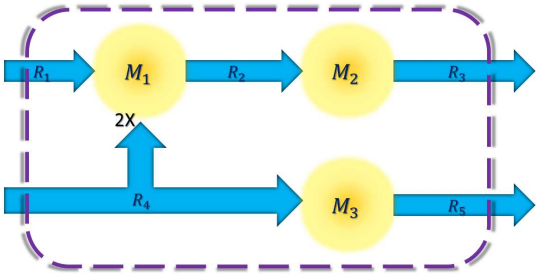

Figure 1: the original metabolic network

Note that, $R_2 \iff R_3$ by the corresponding FCE $v_2 = v_3$ for all $v \in \mathcal{C}$, which we will represent by $\lambda_1^T v = 0$ where

$$\lambda_1 = \begin{bmatrix} 0 \\ -1 \\ +1 \\ 0 \\ 0 \end{bmatrix}.$$

Therefore, for all $v \in \mathcal{C}$

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_3 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix}.$$

Let ϕ_1 denote the matrix on the right hand side. For all $v \in \mathbf{R}^5$,

$$Sv = S\phi_1 \begin{bmatrix} v_1 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = S \begin{bmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} +1 & -1 & +2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & +1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix}.$$

As a result, if we define

$$\tilde{S} = \begin{bmatrix} +1 & -1 & +2 & 0 \\ 0 & 0 & +1 & -1 \end{bmatrix},$$

then we have that $v \in \mathcal{C}$ if and only if

$$\tilde{S} \begin{bmatrix} v_1 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = 0,$$

and $v_i \geq 0$ for $i = 1, 3, 4, 5$.

Intuitively, if we incorporate R_2 into R_3 and remove M_2, then we arrive at the reduced metabolic network in Fig 2, namely $\tilde{\mathcal{N}} = (\tilde{\mathcal{M}}, \tilde{\mathcal{R}}, \tilde{S}, \tilde{\mathcal{I}})$, where $\tilde{\mathcal{M}} = \{\tilde{M}_1, \tilde{M}_3\}$, $\tilde{\mathcal{R}} = \{\tilde{R}_1, \tilde{R}_3, \tilde{R}_4, \tilde{R}_5\}$, $\tilde{\mathcal{I}} = \tilde{\mathcal{R}}$, and $S = \tilde{S}$.

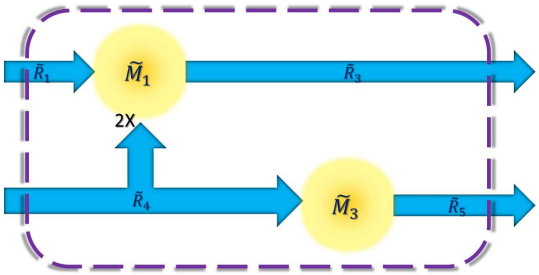

Figure 2: the reduced metabolic network

Similar to the previous case, $R_1, R_4 \longrightarrow R_2$ by the corresponding DCE $v_2 = v_1 + 2v_4$ for all $v \in \mathcal{C}$, which we will represent by $\lambda_2^T v = 0$ where

$$\lambda_2 = \begin{bmatrix} +1 \\ -1 \\ 0 \\ +2 \\ 0 \end{bmatrix}.$$

Similarly, we define ϕ_2 such that

$$\phi_2 \begin{bmatrix} v_1 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} +1 & 0 & 0 & 0 \\ +1 & 0 & +2 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_1 + 2v_4 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = v.$$

Therefore,

$$Sv = S\phi_2 \begin{bmatrix} v_1 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ +1 & -1 & +2 & 0 \\ 0 & 0 & +1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix}.$$

Comparing $\tilde{S}$ with $S\phi_2$ they are equal if we remove the row corresponding to M_1. We will come back to this point for a more careful inspection in §4, however for now, removing an all zeros row from the stoichiometric matrix does not affect the steady-state flux cone, hence, we can safely do it.

This time, if we incorporate R_2 into R_1 and R_4 and remove M_1, then we arrive at the reduced metabolic network in Fig 3, which is the same $\tilde{\mathcal{N}}$ except that $\tilde{M}_1$ is replaced by $\tilde{M}_2$ which again, does not affect the steady-state flux cone.

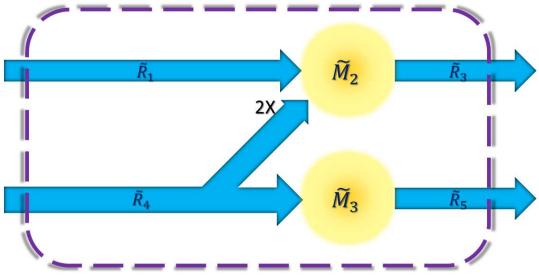

Figure 3: a DCE-induced reduction

Having said that, recall the correspondence between the original and reduced reactions in both cases. For the FCE-induced reduction we had,

$$\begin{array}{l} \tilde{R}_1 \xrightarrow{r_1} \{R_1\} \\ \tilde{R}_3 \xrightarrow{r_1} \{R_2, R_3\} \\ \tilde{R}_4 \xrightarrow{r_1} \{R_4\} \\ \tilde{R}_5 \xrightarrow{r_1} \{R_5\}, \end{array}$$

but for the DCE-induced reduction we had,

$$\begin{array}{l} \tilde{R}_1 \xrightarrow{r_2} \{R_1, R_2\} \\ \tilde{R}_3 \xrightarrow{r_2} \{R_3\} \\ \tilde{R}_4 \xrightarrow{r_2} \{R_2, R_4\} \\ \tilde{R}_5 \xrightarrow{r_2} \{R_5\}, \end{array}$$

where we call r_i the reduction map. Note that r_i is uniquely determined by the flux coupling relations in both cases.

Let e_k denote the kth unit vector of the standard basis of $\mathbf{R}^5$. Comparing ϕ_i with the following matrix

$$I + e_2\lambda_i^T = I + \begin{bmatrix} 0 \\ \lambda_i^T \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

they are equal for both $i = 1, 2$ if we remove the all zeros column corresponding to R_2. Hence, ϕ_i is uniquely determined by the flux coupling equations in both cases.

From now on, suppose λ is an arbitrary flux coupling equation for R_2, *i.e.*,

$$\lambda^T v = c_1v_1 - v_2 + c_3v_3 + c_4v_4 + c_5v_5 = 0.$$

For an index i, let X^i and x^i denote the ith column of the matrix X and the ith entry of the vector x. Furthermore, let $X^{(i)}$ and $x^{(i)}$ denote the result of removing X^i and x^i from X and x, respectively. By this notation, we define

$$\phi = I^{(2)} + e_2(\lambda^{(2)})^T.$$

Like before, ϕ maps the reduced $v^{(2)}$ to the original v.

$$\phi v^{(2)} = \phi \begin{bmatrix} v_1 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} v_1 \\ c_1v_1 + c_3v_3 + c_4v_4 + c_5v_5 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = v.$$

Moreover,

$$S(I+e_2\lambda^T)v = S\Big(I+\begin{bmatrix} 0 \\ \lambda^T \\ 0 \\ 0 \\ 0 \end{bmatrix}\Big)v = Sv + \begin{bmatrix} 0 \\ \lambda^T v \\ 0 \\ 0 \\ 0 \end{bmatrix} = 0$$

for all $v \in \mathcal{C}$. However, the second column of $I + e_2\lambda^T$ is zero so we can remove it together with the second entry of v to arrive at the equation $S\phi v^{(2)} = 0$. Hence, ϕ also provides the reduced stoichiometric matrix by setting $\tilde{S} = S\phi$.

Furthermore, the reversibility of the reduced reactions is determined by the simple rule that $\tilde{R}_k$ is irreversible if and only if $r(\tilde{R}_k)$ contains at least one irreversible reaction, and the reduction map r is uniquely determined by which flux coupling coefficients are nonzero in λ.

The last three paragraphs can be generalized from this toy metabolic network with five reactions to the most general case by following the same steps. In summary, we remove the blocked reactions, then merge the coupled reactions by adding the corresponding columns of the stoichiometric matrix and making the merged reaction irreversible if any of the original ones were irreversible. In the Appendix, we prove that all these QFCA reductions preserve the set of EMs. Moreover, we have abstracted the properties of these reductions used in the proof to show that this result holds for not only the QFCA reductions, but also for a broad class of reductions which we call canonical metabolic network reductions.

EM analysis, while still limited by the size of the metabolic networks it can handle, is a powerful metabolic pathway analysis tool for a wide variety of application such as characterizing cellular metabolism and reprogramming microbial metabolic pathways [37, 38, 39]. This potential encourages conceptualizing the notion of a canonical metabolic network reduction which respects EMs as has been proposed by [30] long ago. In order to realize this idea, we will propose three axioms for characterizing a canonical metabolic network reduction.

Let $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathcal{I})$ be an arbitrary metabolic network. We say that the metabolic network $\tilde{\mathcal{N}} = (\tilde{\mathcal{M}}, \tilde{\mathcal{R}}, \tilde{S}, \tilde{\mathcal{I}})$ is a reduction of $\mathcal{N}$ if

1. there exists a surjection $\phi : \tilde{\mathcal{C}} \to \mathcal{C}$, where $\tilde{\mathcal{C}}$ and $\mathcal{C}$ are the corresponding steady-state flux cones;
2. there exists a reduction map $r : \tilde{\mathcal{R}} \to \mathcal{P}(\mathcal{R})$ such that for any $\tilde{R}_i \in \tilde{\mathcal{R}}$ we have

$$r(\tilde{R}_i) \nsubseteq \bigcup_{k \neq i} r(\tilde{R}_k), \tag{3}$$

   where $\mathcal{P}(\mathcal{R})$ denotes the power set (the set of all the subsets) of $\mathcal{R}$;
3. and the following diagram commutes

$$\begin{array}{ccc} \tilde{\mathcal{C}} & \xrightarrow{\text{supp}} & \mathcal{P}(\tilde{\mathcal{R}}) \\ {\scriptstyle\phi}\downarrow & & \downarrow{\scriptstyle\tilde{r}} \\ \mathcal{C} & \xrightarrow{\text{supp}} & \mathcal{P}(\mathcal{R}) \end{array}$$

   where $\tilde{r} : \mathcal{P}(\tilde{\mathcal{R}}) \to \mathcal{P}(\mathcal{R})$ is defined by

$$\tilde{r}(\{\tilde{R}_i\}_{i \in I}) = \bigcup_{i \in I} r(\tilde{R}_i).$$

We have dedicated an entire Appendix to investigate the properties of such canonical metabolic network reduction. However, to continue reading the rest of the paper, we only need the following corollary of the reduction theorem which is proved in the Appendix.

Corollary 3.0.1. *By setting $i = j$ in the reduction theorem, any reaction in $r(\tilde{R}_i) \setminus \bigcup_{k \neq i} r(\tilde{R}_k)$ is directionally coupled to any reaction in $r(\tilde{R}_i)$.*

The first axiom enforces that each feasible flux distribution in either the reduced or the original metabolic network corresponds to at least one feasible flux distribution in the other one. The second axiom introduces the mapping r between reactions, which induces the injection $\tilde{r}$ between pathways essential for the third axiom to enforce that having minimal support is an invariant property under ϕ. This, in turn, implies that the EMs of these two metabolic networks are in correspondence with each other.

The only suspect property in here which seems like a strong assertion is (3) in the second axiom. Recall that by Corollary 3.0.1, all the reactions in $r(\tilde{R}_i) \setminus \bigcup_{k \neq i} r(\tilde{R}_k)$ turn out to be fully or partially coupled to each other, and hence, are intimately related. Consequently, instead of some abstract $\tilde{R}_i$ in an unreal metabolic network, we can think of each $\tilde{R}_i$ as a class of fully or partially coupled reactions in the current metabolic network which is very similar to the notion of enzyme subsets [29]. Whereas, if for some $\tilde{R}_i \in \tilde{\mathcal{R}}$ we had $r(\tilde{R}_i) \subseteq \bigcup_{k \neq i} r(\tilde{R}_k)$, then no reaction in $\mathcal{R}$ was uniquely associated to $\tilde{R}_i$. To sum up, the main reason behind this assertion is to ensure that reactions in $\tilde{\mathcal{R}}$ are meaningful from the biological point of view.

We give the example of coupled metabolites to further demonstrate this point. *Metabolite concentration coupling analysis* (MCCA) [40] is the same as FCA except it is applied to the left null space of the stoichiometric matrix [41]. We refer the interested reader to the results of [40] for an extensive discussion of the biological relevance of the coupled metabolites. Note that the subsets $r(\tilde{R}_i) \setminus \bigcup_{k \neq i} r(\tilde{R}_k)$ are exactly the same metabolite subsets in this instance. Furthermore, the reduced metabolic network preserves the reactions among these metabolically meaningful pools in addition to all the MCCA results derived from FCA. These reactions restrain the minimal conserved pools (the analogue of elementary modes) to be in one-to-one correspondence with the original set for *minimal conserved pool identification* (MCPI) [40].

Let $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathcal{I})$ be an arbitrary metabolic network, and $\tilde{\mathcal{N}}_1 = (\tilde{\mathcal{M}}_1, \tilde{\mathcal{R}}_1, \tilde{S}_1, \tilde{\mathcal{I}}_1)$ be a reduction of $\mathcal{N}$ by ϕ_1 and r_1. Suppose that we apply several metabolic network reductions to $\mathcal{N}$ successively. For instance, suppose that we reduce $\tilde{\mathcal{N}}_1$ once more to get $\tilde{\mathcal{N}}_2 = (\tilde{\mathcal{M}}_2, \tilde{\mathcal{R}}_2, \tilde{S}_2, \tilde{\mathcal{I}}_2)$ by a different metabolic network reduction, namely the one corresponding to ϕ_2 and r_2. We will show that the composition of consecutive reductions is again a reduction itself.

1. $\phi_1 \circ \phi_2 : \tilde{\mathcal{C}}_2 \to \mathcal{C}$ is a surjection because the composition of surjective functions is surjective;
2. $\tilde{r}_1 \circ r_2 : \tilde{\mathcal{R}}_2 \to \mathcal{P}(\mathcal{R})$ is a legitimate reduction map because for any $\tilde{R}_i \in \tilde{\mathcal{R}}_2$ we have

$$\exists \tilde{R}_j \in r_2(\tilde{R}_i) \setminus \bigcup_{k \neq i} r_2(\tilde{R}_k) \Rightarrow \exists R_t \in r_1(\tilde{R}_j) \setminus \bigcup_{k \neq j} r_1(\tilde{R}_k) \Rightarrow R_t \in \tilde{r}_1 \circ r_2(\tilde{R}_i) \setminus \bigcup_{k \neq i} \tilde{r}_1 \circ r_2(\tilde{R}_k);$$

3. and the following diagram commutes

$$\begin{tikzcd}
\tilde{\mathcal{C}}_2 \arrow[r, "\mathrm{supp}"] \arrow[d, "\phi_2"'] & \mathcal{P}(\tilde{\mathcal{R}}_2) \arrow[d, "\tilde{r}_2"] \\
\tilde{\mathcal{C}}_1 \arrow[r, "\mathrm{supp}"] \arrow[d, "\phi_1"'] & \mathcal{P}(\tilde{\mathcal{R}}_1) \arrow[d, "\tilde{r}_1"] \\
\mathcal{C} \arrow[r, "\mathrm{supp}"] & \mathcal{P}(\mathcal{R})
\end{tikzcd}$$

because for any $\tilde{v} \in \tilde{\mathcal{C}}_2$

$$\mathrm{supp}(\phi_1 \circ \phi_2(\tilde{v})) = \tilde{r}_1(\mathrm{supp}(\phi_2(\tilde{v}))) = \tilde{r}_1 \circ \tilde{r}_2(\mathrm{supp}(\tilde{v})).$$

Putting it all together, $\tilde{\mathcal{N}}_2$ is a reduction of $\mathcal{N}$ by the surjection $\phi_1 \circ \phi_2$ and the reduction map $\tilde{r}_1 \circ r_2$.

Up to now, we have covered how to reduce the metabolic network by the following three stages.

- First, we eliminate all the blocked reactions.
- Second, we merge all the fully coupled reactions.

- Third, we remove the eligible reactions by the DCE-induced reductions.

Suppose that after conducting all these three stages, the last metabolic network is $\tilde{\mathcal{N}} = (\tilde{\mathcal{M}}, \tilde{\mathcal{R}}, \tilde{S}, \tilde{\mathcal{I}})$ of size $\tilde{n}$, thus we obtain the chain of reductions

$$\mathcal{N} \xleftarrow{\phi_1, r_1} \tilde{\mathcal{N}}_1 \xleftarrow{\phi_2, r_2} \cdots \xleftarrow{\phi_{n-\tilde{n}}, r_{n-\tilde{n}}} \tilde{\mathcal{N}}_{n-\tilde{n}},$$

where $\tilde{\mathcal{N}}_{n-\tilde{n}} = \tilde{\mathcal{N}}$, since each reduction reduces exactly one reaction at a time. Throughout this process, we consider the aggregated metabolic network reduction resulting from combining the previously applied ones up to some point in time as a single reduction, *i.e.*,

$$\mathcal{N} \xleftarrow{\phi^i, r^i} \tilde{\mathcal{N}}_i,$$

where $\phi^i = \phi_1 \circ \cdots \circ \phi_i$ and $r^i = \tilde{r}_1 \circ \cdots \circ \tilde{r}_{i-1} \circ r_i$.

We claim that each ϕ^i is a one-to-one correspondence when restricted to the EMs. In the Appendix, we prove that $\tilde{v} \in \tilde{\mathcal{C}}_i$ is an EM if and only if $\phi^i(\tilde{v}) \in \mathcal{C}$ is an EM. Furthermore, ϕ^i is surjective by definition. It only remains to show that ϕ^i is injective too. One can easily verify the injectivity of the single metabolic network reductions ϕ_t of any of these three kinds separately. Hence, it is also proved for ϕ^i which is the composition of i of them.

As a consequence, the overall reduction generated by compositing all the preceding three stages is a metabolic network reduction which provides a one-to-one correspondence between the EMs of $\mathcal{N}$ and $\tilde{\mathcal{N}}$. This can also be shown independently, given the straightforward form of this specific reduction which we characterize next.

For the sake of simplicity, assume that we arrange $\mathcal{R}$ in an ordering such that first come the blocked reactions removed in the first stage, $\mathcal{B} = \{R_1, \ldots, R_b\}$. After $\mathcal{B}$ come the reactions which persist to the end, and by this we mean the final set $\tilde{\mathcal{R}}$ which is a subset of $\mathcal{R}$, because $\tilde{\mathcal{R}}_t \subseteq \tilde{\mathcal{R}}_{t-1}$ for each of these three types of reductions individually. Moreover, assume that $\tilde{\mathcal{R}} = \{\tilde{R}_i\}_{i=1}^{\tilde{n}}$ is sorted in such a way that $\tilde{R}_i$ is fully coupled to a_i other reactions in $\mathcal{R}$ where $a_1 \geq a_2 \geq \cdots \geq a_f > a_{f+1} = \cdots = a_{\tilde{n}} = 0$ (the descending order of the number of fully coupled reactions in $\mathcal{R}$). At last come the reactions which were removed in stages two and three, respectively.

By these simplifying assumptions, if P is the permutation matrix which permutes the columns of S to the ordering we have assumed, then we can give the following explicit formulas for $\tilde{S}$, $\phi^{n-\tilde{n}}$, and $r^{n-\tilde{n}}$ which is $\tilde{S} = SPA$, $\phi^{n-\tilde{n}}(\tilde{v}) = PA\tilde{v}$, and $r^{n-\tilde{n}}(\tilde{R}_i)$ is equal to the index set of the nonzero entries of the ith column of A with the following structure (see Fig 4).

As depicted in Fig 4, the blue block has b zero rows corresponding to the blocked reactions which always have zero flux coefficients. The identity matrix corresponds to $\tilde{\mathcal{R}}$ which in turn is associated with a subset of $\mathcal{R}$. The red block corresponds to the fully coupled reactions removed in stage two, and its nonzero entries are the associated FCE coefficients, *i.e.*, c in (1). In the end, each row of the nonnegative matrix C corresponds to a reaction removed in stage three, and its positive entries are the associated DCE coefficients, *i.e.*, c_d in (2).

Ultimately by Corollary 3.0.1, because of the fact that $\tilde{\mathcal{N}}$ has no directionally coupled pair of reactions, otherwise we could have continued stage three, the size of $\tilde{\mathcal{N}}$ cannot be reduced any further. Even more than that, not only $\tilde{\mathcal{N}}$ is minimal in the sense that it cannot be reduced any further without violating one of the principles we have assumed for the canonical metabolic network reductions, but also it has the minimum size, and in this sense, this is the most compact form of an equivalent steady-state flux cone with identical EMs.

For the proof of this statement, we show that no two distinct reactions in $\tilde{\mathcal{R}}$ can be reduced to the same reaction in any possible metabolic network reduction, hence $\tilde{n}$ is the minimum size. Let $\tilde{\mathcal{N}}' = (\tilde{\mathcal{M}}', \tilde{\mathcal{R}}', \tilde{S}', \tilde{\mathcal{I}}')$ be an arbitrary reduction of $\mathcal{N}$ by ϕ' and r'. By Corollary 3.0.1, if $\tilde{R}_p, \tilde{R}_q \in \tilde{\mathcal{R}} \subseteq \mathcal{R}$ are reduced to the same reaction $\tilde{R}'_t \in \tilde{\mathcal{R}}'$, then there exists $R_s \in r'(\tilde{R}'_t) \setminus \bigcup_{k \neq t} r'(\tilde{R}'_k)$ which is directionally coupled to both $\tilde{R}_p$ and $\tilde{R}_q$, simultaneously. However, $R_s \iff \tilde{R}_p$ if $R_s \longrightarrow \tilde{R}_p$, otherwise $\tilde{R}_p$ would have been removed in the third stage. Therefore $\tilde{R}_p \iff R_s \iff \tilde{R}_q$, which cannot hold unless $p = q$ since stage two guarantees that no two distinct reactions in $\tilde{\mathcal{R}}$ are fully coupled to each other.

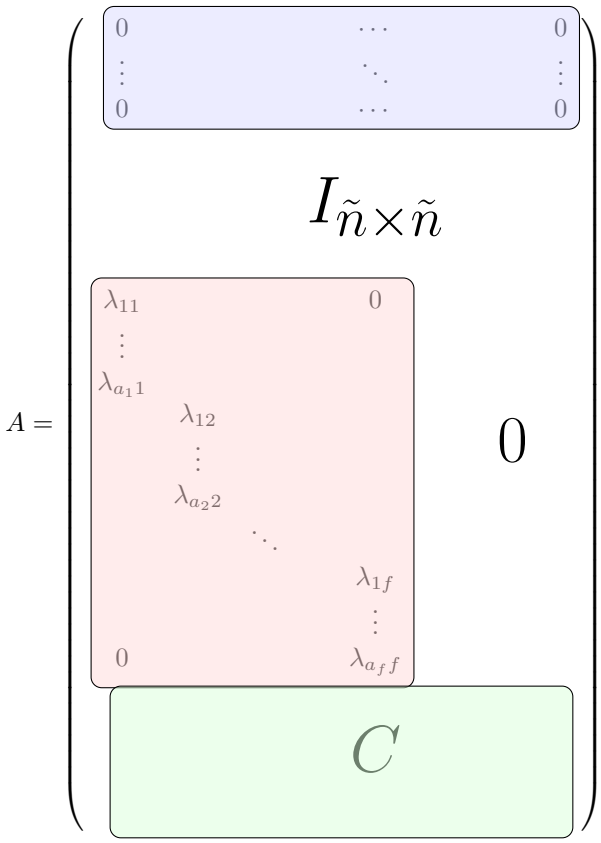

Figure 4: sparsity pattern of A

4 Discussion and results

QFCA[1] is implemented in the COBRA toolbox v3.0 [42], publicly available for academic and research purposes. This MATLAB© package is also the first implementation of the reduction algorithm presented here. It paves the way for the development of scalable solutions for various computational tasks. This brings us to a suggestion for the future research to work out the details of integrating this universal preprocessing (or postprocessing as in *NetworkReducer* [25]) step with the compatible pipelines of interest.

The reduced stoichiometric matrix introduced in §3 has a great potential to speed-up the downstream analyses which only rely on EMs, in particular, QFCA. The smaller the reduced $\tilde{S}$, the faster any downstream analysis on this proxy stoichiometric matrix. For a running example, iMM1415 (*Mus Musculus*, 1415 genes) [43] has 3726 reactions initially, and MONGOOSE reduces it into 1625 reactions while QFCA achieves a record 1171 reactions in the same way as we described in §3.

Even though $\tilde{S}$ has m rows in theory, it turns out that for almost every real-world metabolic network, many of its rows are identically zero and therefore, for all practical purposes they may be removed safely. For instance, consider the metabolic network in Fig 1 and suppose that we remove R_2 by the corresponding DCE-induced reduction. Although, M_1 is still there after conducting the metabolic network reduction according to the procedure of §3, it can be removed to arrive at the metabolic network in Fig 3 since it is no longer involved in any reactions.

Moreover, we remove the linearly dependent rows of $\tilde{S}$ as described under the title of the detection of conservation relations by [30]. In practice, $\tilde{S}$ may be much smaller than S both regarding the number of rows and columns. Going back to our running example, the number of metabolites can be reduced from 2775 to 335 by QFCA while the corresponding number for MONGOOSE is 643.

We should mention that in spite of the fact that the reduced $\tilde{S}$ has a smaller size than the original S,

[1] https://mtefagh.github.io/qfca/

it might be denser for some instances. As a result, if some algorithm is already exploiting the sparsity of S intricately, perhaps it does not gain much from this preprocessing step. This is a common theme in numerical analysis that sometimes sparse computation can make up for not including off-the-shelf preprocessing steps reducing the size of sparse matrices. Nonetheless, in our running example, the number of nonzero elements of S and $\tilde{S}$ are 14053 and 14385, respectively. These numbers are promising as they indicate that the reduced stoichiometric matrix is nearly as sparse as the original one.

To showcase canonical metabolic network reductions on a large-scale reference database, we have reduced the BiGG universal model [44] of 13249 metabolites, 24311 reactions, and 95774 nonzero stoichiometric coefficients into a metabolic network with 1278 metabolites, 10255 reactions, and 56457 nonzero stoichiometric coefficients. Furthermore, we have reconstructed consistent metabolic networks for randomly sampled core sets of varying sizes, from zero up to the number of unblocked reactions, by applying the algorithm SWIFTCORE [45] to both of these models. The runtime on the reduced network is 29% of the runtime on the original network averaged over 100 iterations (see Fig 5). Both results for each set of core reactions are double checked and are always flux consistent in both cases.

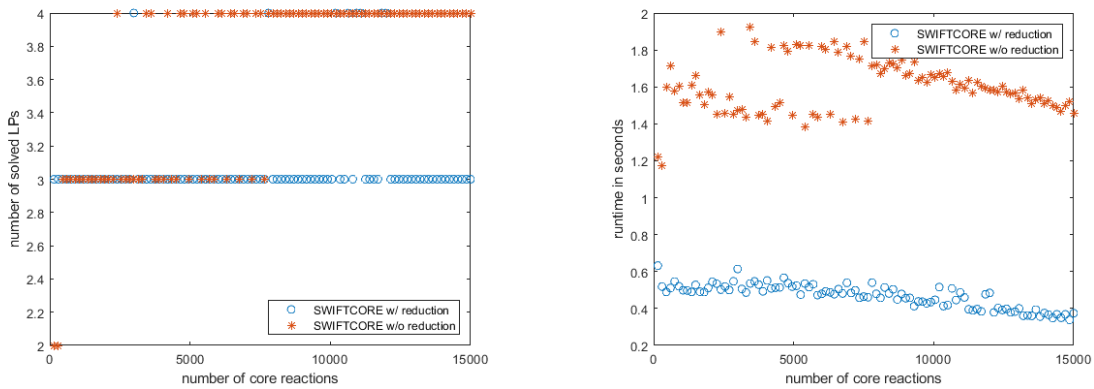

Figure 5: SWIFTCORE runs more than 3× faster on the reduced BiGG universal model

Apart from the computational benefits of canonical metabolic network reductions, the fact that the redundancies removed by the axiomatic framework coincide with the well-known biological QFCA relations is worth special attention. Recall that the reactions reduced during the suggested process are the reactions which the rest of the remaining reactions are coupled to. A recent study [46] has shown that these are the essential reactions enriching the vital metabolic processes of the cell. In addition, they are more conserved and of older evolutionary age. These reactions are essential in a wide range of conditions, and their associated genes are more expressed [47]. All these observations are compatible with the fact that canonical metabolic network reductions only keep the reactions which discriminate different elementary modes by their distinct activity patterns. For example, the first exchange reactions in pathways are merged into the intermediate reactions which differentiate those pathways from each other which is in agreement with the findings of [46].

5 Conclusions

Metabolic networks contain many essential reactions which are active regardless of the others. On the other hand, there are reactions which are specific to a few elementary modes and whose activity patterns determine the active pathway. Metabolic networks can be reduced by integrating the former reactions into the latter ones without mixing the elementary mode.

We have derived the minimum reduced metabolic networks by QFCA. Afterward, we have also proved a converse that any further reduction loses some information about the support of active pathways in analogy to the concept of lossy compression in information theory.

References

[1] Edwards JS, Palsson BØ. Systems properties of the Haemophilus influenzaeRd metabolic genotype. Journal of Biological Chemistry. 1999;274(25):17410–17416.

[2] Reed JL, Vo TD, Schilling CH, Palsson BØ. An expanded genome-scale model of Escherichia coli K-12 (i JR904 GSM/GPR). Genome biology. 2003;4(9):R54.

[3] Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. Genome research. 2003;13(2):244–253.

[4] Duarte NC, Herrgård MJ, Palsson BØ. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. Genome research. 2004;14(7):1298–1309.

[5] Feist AM, Scholten JC, Palsson BØ, Brockman FJ, Ideker T. Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri. Molecular systems biology. 2006;2(1).

[6] Jamshidi N, Palsson BØ. Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ 661 and proposing alternative drug targets. BMC systems biology. 2007;1(1):26.

[7] Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Molecular Systems Biology. 2007;3(1):121.

[8] Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proceedings of the National Academy of Sciences. 2007;104(6):1777–1782.

[9] Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, et al. The Edinburgh human metabolic network reconstruction and its functional analysis. Molecular systems biology. 2007;3(1):135.

[10] Herrgård MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, et al. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nature Biotechnology. 2008;26(10):1155–1160.

[11] Poolman MG, Miguet L, Sweetlove LJ, Fell DA. A genome-scale metabolic model of Arabidopsis and some of its properties. Plant physiology. 2009;151(3):1570–1581.

[12] de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. Plant physiology. 2010;152(2):579–589.

[13] Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive genome-scale reconstruction of Escherichia coli metabolism2011. Molecular systems biology. 2011;7(1):535.

[14] Mintz-Oron S, Meir S, Malitsky S, Ruppin E, Aharoni A, Shlomi T. Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. Proceedings of the National Academy of Sciences. 2012;109(1):339–344.

[15] Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, et al. EcoCyc: fusing model organism databases with systems biology. Nucleic acids research. 2012;41(D1):D605–D612.

[16] Aung HW, Henry SA, Walker LP. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. Industrial biotechnology. 2013;9(4):215–228.

[17] Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, et al. A community-driven global reconstruction of human metabolism. Nature biotechnology. 2013;31(5):419–425.

[18] Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hanscho M, et al. Recon 2.2: from reconstruction to model of human metabolism. Metabolomics. 2016 Jun;12(7):109. Available from: https://doi.org/10.1007/s11306-016-1051-4.

[19] Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. Nature biotechnology. 2018;36(3):272.

[20] Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. Molecular Systems Biology. 2009;5(1). Available from: http://msb.embopress.org/content/5/1/320.

[21] Berger B, Peng J, Singh M. Computational solutions for omics data. Nature Reviews Genetics. 2013;14(5):333–346.

[22] Burgard AP, Vaidyaraman S, Maranas CD. Minimal reaction sets for Escherichia coli metabolism under different growth requirements and uptake environments. Biotechnology progress. 2001;17(5):791–797.

[23] Jonnalagadda S, Srinivasan R. An efficient graph theory based method to identify every minimal reaction set in a metabolic network. BMC systems biology. 2014;8(1):28.

[24] Vlassis N, Pacheco MP, Sauter T. Fast reconstruction of compact context-specific metabolic network models. PLoS computational biology. 2014;10(1):e1003424.

[25] Erdrich P, Steuer R, Klamt S. An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. BMC systems biology. 2015;9(1):48.

[26] Röhl A, Bockmayr A. A mixed-integer linear programming approach to the reduction of genome-scale metabolic networks. BMC bioinformatics. 2017;18(1):2.

[27] Burgard AP, Nikolaev EV, Schilling CH, Maranas CD. Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Research. 2004;14(2):301–312.

[28] Larhlimi A, David L, Selbig J, Bockmayr A. F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. BMC Bioinformatics. 2012;13(1):57.

[29] Pfeiffer T, Snchez-Valdenebro I, Nu±o J, Montero F, Schuster S. METATOOL: for studying metabolic networks. Bioinformatics (Oxford, England). 1999;15(3):251–257.

[30] Gagneur J, Klamt S. Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinformatics. 2004 Nov;5(1):175. Available from: https://doi.org/10.1186/1471-2105-5-175.

[31] Chindelevitch L, Trigg J, Regev A, Berger B. An exact arithmetic toolbox for a consistent and reproducible structural analysis of metabolic network models. Nature communications. 2014;5:4893.

[32] Schilling CH, Schuster S, Palsson BØ, Heinrich R. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnology Progress. 1999;15(3):296–303.

[33] Tefagh M, Boyd SP. Quantitative flux coupling analysis. Journal of Mathematical Biology. 2018 Dec;Available from: https://doi.org/10.1007/s00285-018-1316-9.

[34] Shannon CE. A mathematical theory of communication. The Bell System Technical Journal. 1948 July;27(3):379–423.

[35] Schuster S, Hilgetag C. On elementary flux modes in biochemical reaction systems at steady state. Journal of Biological Systems. 1994;2(02):165–182.

[36] Boyd S, Vandenberghe L. Convex Optimization. Cambridge University Press; 2004.

[37] Trinh CT, Wlaschin A, Srienc F. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Applied microbiology and biotechnology. 2009;81(5):813.

[38] Trinh CT, Thompson RA. Elementary mode analysis: A useful metabolic pathway analysis tool for reprograming microbial metabolic pathways. In: Reprogramming microbial metabolic pathways. Springer; 2012. p. 21–42.

[39] Zanghellini J, Ruckerbauer DE, Hanscho M, Jungreuthmayer C. Elementary flux modes in a nutshell: properties, calculation and applications. Biotechnology journal. 2013;8(9):1009–1016.

[40] Nikolaev EV, Burgard AP, Maranas CD. Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions. Biophysical journal. 2005;88(1):37–49.

[41] Famili I, Palsson BØ. The Convex Basis of the Left Null Space of the Stoichiometric Matrix Leads to the Definition of Metabolically Meaningful Pools. Biophysical Journal. 2003;85(1):16 – 26. Available from: `http://www.sciencedirect.com/science/article/pii/S0006349503744506`.

[42] Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. arXiv preprint arXiv:171004038. 2017;.

[43] Sigurdsson MI, Jamshidi N, Steingrimsson E, Thiele I, Palsson BØ. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. BMC Systems Biology. 2010 Oct;4(1):140. Available from: `https://doi.org/10.1186/1752-0509-4-140`.

[44] King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic acids research. 2015;44(D1):D515–D522.

[45] Tefagh M, Boyd SP. SWIFTCORE: a tool for the context-specific reconstruction of genome-scale metabolic networks. In review;Available from: `https://mtefagh.github.io/swiftcore/`.

[46] Hosseini Z, Marashi SA. Hierarchical organization of fluxes in Escherichia coli metabolic network: Using flux coupling analysis for understanding the physiological properties of metabolic genes. Gene. 2015;561(2):199–208.

[47] Notebaart RA, Kensche PR, Huynen MA, Dutilh BE. Asymmetric relationships between proteins shape genome evolution. Genome Biology. 2009 Feb;10(2):R19. Available from: `https://doi.org/10.1186/gb-2009-10-2-r19`.

Appendix

Let $\phi^{-1}(v)$ denote the preimage of $v \in \mathcal{C}$ under ϕ. The term $\mathrm{supp}(\phi^{-1}(v))$ is well-defined as for any $\tilde{u}, \tilde{v} \in \phi^{-1}(v)$, we have

$$\tilde{r}(\mathrm{supp}(\tilde{u})) = \mathrm{supp}(\phi(\tilde{u})) = \mathrm{supp}(v) = \mathrm{supp}(\phi(\tilde{v})) = \tilde{r}(\mathrm{supp}(\tilde{v})).$$

Therefore, $\mathrm{supp}(\tilde{u}) = \mathrm{supp}(\tilde{v})$ since $\tilde{r}$ is injective from (3).

Another observation is that for any $\tilde{u}, \tilde{v} \in \tilde{\mathcal{C}}$

$$\mathrm{supp}(\tilde{u}) \subset \mathrm{supp}(\tilde{v}) \Leftrightarrow \tilde{r}(\mathrm{supp}(\tilde{u})) \subset \tilde{r}(\mathrm{supp}(\tilde{v})) \Leftrightarrow \mathrm{supp}(\phi(\tilde{u})) \subset \mathrm{supp}(\phi(\tilde{v})),$$

which is deduced from the second and third properties of the metabolic network reductions, respectively.

We claim that if $\tilde{v} \in \tilde{\mathcal{C}}$ is an EM, then $\phi(\tilde{v})$ is also an EM. The proof is by contradiction. Let us suppose that there exists $u \in \mathcal{C}$ such that:

$$\mathrm{supp}(0) \subset \mathrm{supp}(u) \subset \mathrm{supp}(\phi(\tilde{v})) \Rightarrow \mathrm{supp}(\phi^{-1}(0)) \subset \mathrm{supp}(\phi^{-1}(u)) \subset \mathrm{supp}(\tilde{v})$$

Since $\tilde{v} \in \tilde{\mathcal{C}}$ is an EM, for any $\tilde{u} \in \tilde{\mathcal{C}}$ if $\mathrm{supp}(\tilde{u}) \subset \mathrm{supp}(\tilde{v})$, then $\mathrm{supp}(\tilde{u}) = \emptyset$. However, $\mathrm{supp}(\phi^{-1}(0)) = \mathrm{supp}(\phi^{-1}(u)) = \emptyset$ is in contradiction with $\mathrm{supp}(\phi^{-1}(0)) \subset \mathrm{supp}(\phi^{-1}(u))$ which proves the desired result.

For the converse, we claim that if $v \in \mathcal{C}$ is an EM, then any arbitrary $\tilde{v} \in \phi^{-1}(v)$ is an EM too. Assume to the contrary that for a fixed $\tilde{v} \in \phi^{-1}(v)$ there exists $\tilde{u} \in \tilde{\mathcal{C}}$ such that:

$$\text{supp}(0) \subset \text{supp}(\tilde{u}) \subset \text{supp}(\tilde{v}) \Rightarrow \text{supp}(\phi(0)) \subset \text{supp}(\phi(\tilde{u})) \subset \text{supp}(\phi(\tilde{v})) = \text{supp}(v)$$

Since $v \in \mathcal{C}$ is an EM, for any $u \in \mathcal{C}$ if $\text{supp}(u) \subset \text{supp}(v)$, then $u = 0$. However, $\phi(0) = \phi(\tilde{u}) = 0$ is in contradiction with $\text{supp}(\phi(0)) \subset \text{supp}(\phi(\tilde{u}))$ which completes the proof.

Altogether, for any arbitrary metabolic network reduction $\tilde{v} \in \tilde{\mathcal{C}}$ is an EM if and only if $\phi(\tilde{v}) \in \mathcal{C}$ is an EM. As mentioned earlier, it is possible to infer flux coupling relations from EMs, solely. This evidence suggests that one should be able to deduce some of the flux coupling relations of the original and reduced networks from one another. The following theorem demonstrates one of the many possible ways to do so.

Theorem 5.1 (The reduction theorem)**.** *Suppose that $\tilde{\mathcal{N}} = (\tilde{\mathcal{M}}, \tilde{\mathcal{R}}, \tilde{S}, \tilde{\mathcal{I}})$ is a metabolic network reduction of $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathcal{I})$ by the surjection $\phi : \tilde{\mathcal{C}} \to \mathcal{C}$ and the reduction map $r : \tilde{\mathcal{R}} \to \mathcal{P}(\mathcal{R})$. For each $\tilde{R}_i, \tilde{R}_j \in \tilde{\mathcal{R}}$ such that $\tilde{R}_i \longrightarrow \tilde{R}_j$, any reaction in $r(\tilde{R}_i) \setminus \bigcup_{k \neq i} r(\tilde{R}_k)$ is directionally coupled to any reaction in $r(\tilde{R}_j)$. Conversely, if there exists a reaction in $r(\tilde{R}_i)$ which is directionally coupled to some reaction in $r(\tilde{R}_j) \setminus \bigcup_{k \neq j} r(\tilde{R}_k)$, then $\tilde{R}_i \longrightarrow \tilde{R}_j$.*

Proof. In order to show the forward direction of the theorem, assume that $R_1 \in r(\tilde{R}_i) \setminus \bigcup_{k \neq i} r(\tilde{R}_k)$ and $R_2 \in r(\tilde{R}_j)$. If $v_1 \neq 0$ for some arbitrary $v \in \mathcal{C}$, then we have

$$v_1 \neq 0 \Rightarrow R_1 \in \text{supp}(v) \Rightarrow R_1 \in \tilde{r}(\text{supp}(\phi^{-1}(v))) \Rightarrow R_1 \in \bigcup_{\tilde{R}_k \in \text{supp}(\phi^{-1}(v))} r(\tilde{R}_k).$$

From the assumption that $R_1 \notin \bigcup_{k \neq i} r(\tilde{R}_k)$, we have

$$\tilde{R}_i \in \text{supp}(\phi^{-1}(v)) \Rightarrow \tilde{R}_j \in \text{supp}(\phi^{-1}(v)) \Rightarrow r(\tilde{R}_j) \subseteq \text{supp}(v) \Rightarrow R_2 \in \text{supp}(v) \Rightarrow v_2 \neq 0.$$

Since $v \in \mathcal{C}$ was arbitrary, this proves one direction of the theorem.

To show the other direction of the theorem, assume that $\tilde{v}_i \neq 0$ for some arbitrary $\tilde{v} \in \tilde{\mathcal{C}}$. Additionally, assume $R_1 \in r(\tilde{R}_i)$ and $R_2 \in r(\tilde{R}_j) \setminus \bigcup_{k \neq j} r(\tilde{R}_k)$ are such that $R_1 \longrightarrow R_2$. Then,

$$\tilde{v}_i \neq 0 \Rightarrow \tilde{R}_i \in \text{supp}(\tilde{v}) \Rightarrow R_1 \in \text{supp}(\phi(\tilde{v})) \Rightarrow R_2 \in \text{supp}(\phi(\tilde{v})) \Rightarrow \tilde{R}_j \in \text{supp}(\tilde{v}) \Rightarrow \tilde{v}_j \neq 0.$$

Therefore, $\tilde{R}_i \longrightarrow \tilde{R}_j$ as was desired. □

As an introductory example, suppose that in $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathcal{I})$, R_i is blocked. Let S^i and $S^{(i)}$ denote the ith column of S and the resulting submatrix after removing it from S, respectively. Trivially, $\tilde{\mathcal{N}} = (\mathcal{M}, \mathcal{R} \setminus \{R_i\}, S^{(i)}, \mathcal{I} \setminus \{R_i\})$ is a reduction of $\mathcal{N}$ by letting

$$\phi(\tilde{v}) = (\tilde{v}_1, \ldots, \tilde{v}_{i-1}, 0, \tilde{v}_i, \ldots, \tilde{v}_{n-1})$$

and $r(R_d) = \{R_d\}$ for all $d \neq i$. In a similar manner, if $R_i \notin \mathcal{I}$ and it is only blocked in the reverse direction, then $\tilde{\mathcal{N}} = (\mathcal{M}, \mathcal{R}, S, \mathcal{I} \cup \{R_i\})$ is a reduction of $\mathcal{N}$ by letting $\phi(\tilde{v}) = \tilde{v}$ and $r(R_d) = \{R_d\}$ for all d. On the other hand, if $R_i \notin \mathcal{I}$ and it is only blocked in the forward direction, then $\tilde{\mathcal{N}} = (\mathcal{M}, \mathcal{R}, \tilde{S}, \mathcal{I} \cup \{R_i\})$ where

$$\tilde{S}^d = \begin{cases} -S^i & \text{if } d = i \\ S^d & \text{otherwise} \end{cases}$$

is a reduction of $\mathcal{N}$ by letting

$$\phi(\tilde{v}) = (\tilde{v}_1, \ldots, \tilde{v}_{i-1}, -\tilde{v}_i, \tilde{v}_{i+1}, \ldots, \tilde{v}_n)$$

and $r(R_d) = \{R_d\}$ for all d. By applying the reduction theorem, $\mathcal{N}$ and $\tilde{\mathcal{N}}$ have exactly the same set of flux coupling relations in all these three cases.

For another example, suppose that in $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathcal{I})$, distinct R_i and R_j are fully coupled with (1) as their corresponding FCE. If we define $\phi(\tilde{v}) = (\tilde{v}_1, \ldots, \tilde{v}_{i-1}, v_i, \tilde{v}_i, \ldots, \tilde{v}_{n-1})$ where

$$v_i = \begin{cases} c\tilde{v}_j & \text{if } i > j \\ c\tilde{v}_{j-1} & \text{otherwise} \end{cases}$$

for the surjection, and

$$r(R_d) = \begin{cases} \{R_i, R_j\} & \text{if } d = j \\ \{R_d\} & \text{otherwise} \end{cases}$$

for the reduction map, then the metabolic network specified by $(\mathcal{M}, \mathcal{R} \setminus \{R_i\}, \hat{S}^{(i)}, \mathcal{I} \setminus \{R_i\})$, where

$$\hat{S}^d = \begin{cases} S^j + cS^i & \text{if } d = j \\ S^d & \text{otherwise} \end{cases}$$

is a reduction of $\mathcal{N}$.

By applying the reduction theorem, all of the flux coupling relations of this reduced network are valid in the original network as well. Furthermore, if in the reduced network R_k is directionally, partially, or fully coupled to R_j or vice versa, then in the original network R_k is directionally, partially, or fully coupled to R_i or vice versa, and these are all the additional flux coupling relations of $\mathcal{N}$.

Arguably, fully coupled reactions are the most comprehensible ones as FCE gives an entirely straightforward relation between their activity rates. QFCA provides DCE and EDCE relating the rates of directionally coupled reactions, which naturally generalize FCE given by the full coupling definition. Later on, we intend to generalize the latter example from FCE to DCE. The point that makes DCE especially suitable for this purpose is that, just like FCE, it both maintains the steady-state conditions and respects the irreversibility constraints because of the positivity of the coefficients of (2).

Suppose that in a metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathcal{I})$, all the blocked reactions have already been removed and all the fully coupled reactions have already been merged so that there exists no more previously investigated reductions applicable to further reduce $\mathcal{N}$. For any $R_j \in \mathcal{R}$ such that $\mathcal{D}_j \neq \emptyset$ and (2) holds, we refer to $\tilde{\mathcal{N}} = (\mathcal{M}, \mathcal{R} \setminus \{R_j\}, \hat{S}^{(j)}, \mathcal{I} \setminus \{R_j\})$ where

$$\hat{S}^d = \begin{cases} S^d + c_d S^j & \text{if } R_d \in \mathcal{D}_j \\ S^d & \text{otherwise} \end{cases}$$

as a *DCE-induced reduction* of $\mathcal{N}$ if $\phi(\tilde{v}) = (\tilde{v}_1, \ldots, \tilde{v}_{j-1}, v_j, \tilde{v}_j, \ldots, \tilde{v}_{n-1})$, where

$$v_j = \sum_{d<j : R_d \in \mathcal{D}_j} c_d \tilde{v}_d + \sum_{d>j : R_d \in \mathcal{D}_j} c_d \tilde{v}_{d-1},$$

and

$$r(R_d) = \begin{cases} \{R_d, R_j\} & \text{if } R_d \in \mathcal{D}_j \\ \{R_d\} & \text{otherwise} \end{cases}$$

is its corresponding reduction map.

One can easily check to see that this is actually a legitimate metabolic network reduction. Therefore by applying the reduction theorem, we conclude that all the flux coupling relations of the reduced network are valid in the original network as well. Moreover, we already know which reactions are directionally coupled to R_j. But then again, how can we tell whether $R_j \longrightarrow R_i$ or not for some $R_i \in \mathcal{R} \setminus \{R_j\}$, without undoing the changes which were applied to $\mathcal{N}$? The following Lemma answers this question.

Lemma 5.2. *Let $r^{-1}(R_i)$ denote the set of reactions $\tilde{R}_d \in \tilde{\mathcal{R}}$ such that $R_i \in r(\tilde{R}_d)$. For an arbitrary pair $R_i, R_j \in \mathcal{R}$, if each of the reactions in $r^{-1}(R_i)$ is directionally coupled to some reaction in $r^{-1}(R_j)$, then $R_i \longrightarrow R_j$.*

Remark. As an immediate corollary for any arbitrary metabolic network reduction, $R_i \longrightarrow R_j$ if $r^{-1}(R_i) \subseteq r^{-1}(R_j)$.

Proof. If $v_i \neq 0$ for some fixed $v \in \mathcal{C}$, then we have

$$v_i \neq 0 \Rightarrow R_i \in \operatorname{supp}(v) \Rightarrow r^{-1}(R_i) \cap \operatorname{supp}(\phi^{-1}(v)) \neq \emptyset.$$

Without loss of generality, assume that $\tilde{R}_p \in r^{-1}(R_i) \cap \operatorname{supp}(\phi^{-1}(v))$. Since $\tilde{R}_p \in r^{-1}(R_i)$, there exists $\tilde{R}_q \in r^{-1}(R_j)$ such that $\tilde{R}_p \longrightarrow \tilde{R}_q$. Therefore,

$$\tilde{R}_q \in \operatorname{supp}(\phi^{-1}(v)) \Rightarrow r(\tilde{R}_q) \subseteq \operatorname{supp}(v) \Rightarrow R_j \in \operatorname{supp}(v) \Rightarrow v_j \neq 0.$$

Since $v \in \mathcal{C}$ was arbitrary, the proof of the lemma is complete. □

By applying the above Lemma to the DCE-induced reduction corresponding to R_j, if all of the reactions in $r^{-1}(R_j)$ are directionally coupled to R_i, then $R_j \longrightarrow R_i$. Yet the converse is also true, because directional coupling is transitive, thus $\mathcal{D}_j \longrightarrow R_j \longrightarrow R_i$. Hence in order to determine if $R_j \longrightarrow R_i$ holds in the original metabolic network, it is enough to check if $r^{-1}(R_j) \longrightarrow R_i$ holds in the DCE-reduced metabolic network. This derives all the extra flux coupling relations of $\mathcal{N}$ as was promised before.