I. Troch, F. Breitenecker, eds. ISBN 978-3-901608-35-3

DETECTING METABOLIC CONVERSIONS IN GENOME-SCALE METABOLIC NETWORKS ON THE BASIS OF ELEMENTARY FLUX PATTERNS IN SUBNETWORKS

Christoph Kaleta, Friedrich-Schiller-Universität Jena, Germany
Luís Filipe de Figueiredo, Friedrich-Schiller-Universität Jena, Germany
Stefan Schuster, Friedrich-Schiller-Universität Jena, Germany

Corresponding author: Christoph Kaleta
Department of Bioinformatics
Friedrich-Schiller-Universität Jena – D-07743 Jena, Ernst-Abbe-Platz 2 – Germany
Email: [redacted]

Abstract. Elementary modes (EMs) represent a powerful means in the analysis of metabolic networks and their characteristic properties. However, EM analysis cannot be applied to genome-scale systems since their number is growing exponentially with network size. Thus, EMs can only be computed in networks usually not corresponding to the entire known system and hence they might not reflect the true metabolic capabilities of the entire system. Here we present a novel concept, elementary flux patterns, which allows to circumvent these problems. Within a large or genome-scale metabolic network elementary flux patterns are defined as sets of reactions that represent the basic routes of any steady-state flux of the genome scale network through a particular subsystem. Equipped with this method we analyze the EMs obtained for two networks within the central metabolism of *Escherichia coli* that have been studied previously. We integrate these networks as subnetworks into a publicly available genome scale metabolic model of *E. coli*. Thus, we find that 6 of the 16 elementary modes of the first system cannot be present in any steady-state flux of the genome-scale system. For the second system, all EMs are part of global steady-state fluxes. Furthermore we analyze the elementary flux patterns of the subnetworks and find several alternative routes on which intermediates of the subnetworks can be produced from species in the growth media. We conclude that the concept of elementary flux patterns offers two major advantages. First, elementary flux patterns more faithfully reflect the possible steady-state fluxes through a subsystem of a large metabolic network by taking into account the entire system. Second, they enable the application of tools from EM analysis to genome-scale metabolic networks.

1 Introduction

Elementary modes (EM) have seen a wide array of applications in Systems biology and biotechnical engineering. They have been used to assess network robustness [15, 1], to find pathways with optimal yields for certain metabolic species [8], to study enzyme deficiencies [11], to find possible targets for the engineering of metabolic networks [6] and to analyze the effect of such an engineering [13]. Due to the growing number of publicly available genome-scale metabolic networks [2, 3, 4] it becomes desirable to apply EM analysis to such systems.

The computation of EMs in larger metabolic networks meets with the principal problem that their number grows exponentially with network size. Thus, owing to constraints in memory and computation time, they become difficult to analyze [7] or even impossible to enumerate. In consequence, EM analysis has to be applied to smaller networks containing reactions of interest and not the entire known system [9, 10, 12, 11]. The integration of the model into the remaining system is achieved by the use of exchange fluxes, corresponding to the production or consumption of a species by a large set of reactions of the remaining model, as well as external species. These species are considered to be buffered by reactions of the complete system and hence they can be excluded from the steady-state condition. Examples for such species are energy or redox currency species like ATP, $FADH_2$ or NADH.

The exclusion of these currency species from the steady-state condition is easily justified by the large number of reactions in which they participate. However, due to the constraints in EM computation usually only a small part of the known network is considerd and hence most of its species are implicitly defined as external. In consequence, the subnetwork is considered almost independent from the remaining model. However, the topology of a network can give rise to dependencies between the production and consumption of different external species. In such a case, some EMs of the subnetwork need to be combined with other EMs to allow a global steady-state flux or might be even not part of any steady-state flux.

Elementary flux patterns, a concept to be introduced in this work, circumvents this problem by explicitly considering the possible fluxes of the entire network, when analyzing the steady-state fluxes of a subnetwork. Elementary flux patterns correspond to the basic routes of each steady state using reactions from a specific subnetwork. Owing to their definition, they allow to apply concepts building on EMs to genome-scale metabolic networks without the drawbacks encountered in EM analysis. Since, elementary flux patterns can be mapped to EMs on the genome scale, it is even possible to analyze “genome-scale” EMs.

This work is sectioned into three main parts. First we introduce the concepts central to this work in Section 2. These concepts are applied to a genome-scale metabolic network of *E. coli* in Section 3. Finally, we conclude in Section 4.

2 Central concepts

For a metabolic network consisting of n reactions among m metabolites or species the $m \times n$ stoichiometric matrix $\mathbf{M}$ indicates the consumption and production of each species in each reaction. An entry M_{ij} is negative if species i is consumed by reaction j and positive if it is produced by j.

2.1 Elementary modes

Elementary modes are defined as minimal reaction sets that can operate at steady state with all reactions obeying the irreversibility constraint [12]. The term "minimal" refers to that there is no subset of reactions that could also operate at steady state.

An EM is a flux vector $\mathbf{v}$ of length n which is a solution to the equation system

$$\mathbf{M} \cdot \mathbf{v} = 0 \tag{1}$$

Furthermore, $\mathbf{v}$ has to obey thermodynamic constraints. That is, for a reaction considered to be irreversible, the corresponding entry in $\mathbf{v}$ has to be non-negative. For simplicity we split reversible reactions into forward and back directions and hence the irreversibility constraint becomes

$$\mathbf{v} \geq 0 \tag{2}$$

Formally, the solution space of equations 1 and 2 can be visualized as a convex polyhedral cone in the n-dimensional flux space. The spanning vectors or extreme rays of this cone correspond to the EMs, and hence every steady-state flux can be written as a non-negative linear combination of EMs.

As mentioned above the analysis with EMs usually involves the introduction of external species. External species are considered to be buffered by reactions outside of the model. Thus, the steady-state condition can be relaxed by removing the corresponding rows from $\mathbf{M}$. However, doing this, all information on their production and consumption by reactions outside of the subsystem under consideration is lost.

2.2 Elementary flux patterns

Flux patterns are defined as sets of reaction indices in a subnetwork of a large metabolic model. For simplicity we assume that the k first reactions of the system belong to the subnetwork. These reactions correspond to the k first columns of the stoichiometric matrix $\mathbf{M}$. Now, a flux pattern s is a set of indices i with $1 \leq i \leq k$ fulfilling the following condition

$$\exists v \in \mathbb{R}^n : \mathbf{v} \geq 0 \text{ and } \mathbf{M} \cdot \mathbf{v} = 0 \text{ and } \forall i \in s : v_i > 0 \text{ and } \forall j \in \{1..k\} \setminus s \ : \ v_j = 0 \tag{3}$$

Thus, a flux pattern can be understood as a set of reactions, or more precisely a set of reaction indices, of the subnetwork that is part of a steady-state flux $\mathbf{v}$ of the entire network. We require that only the indices s of the subsystem have a non-zero flux in $\mathbf{v}$, while the remaining fluxes are constrained to zero.

Now elementary flux patterns can be defined as the building blocks of the set of flux patterns S of a subnetwork. A flux pattern s is called elementary if

$$\nexists s'_1, \ldots, s'_l \subseteq S \setminus s : \bigcup_{1 \leq i \leq l} s'_i = s \tag{4}$$

Hence, s is called elementary if it cannot be written as set union of other flux patterns of a subnetwork. As outlined in Section 2.3, this definition is less restrictive than this in the concept of EMs since one elementary flux pattern can be subset of another.

It is important to mention that elementary flux patterns are applicable even if the subnetwork is not connected. The subnetwork can consist of independent units of reactions, which do not interface to each other through common substrates or products.

The computation of elementary flux patterns proceeds by iteratively solving a mixed-integer linear program (MILP). This MILP returns in each iteration a flux pattern that is not a combination of previously found elementary flux patterns and has the least number of reactions. Thus, it is guaranteed that only elementary flux patterns are found. Due to the constraints of the MILP, it becomes infeasible if no further elementary flux pattern exists.

Elementary Modes		Corresponding Elementary flux patterns	
Number	Reactions	Numbers	Reactions
1	$\rightarrow S_1 \rightarrow S_2 \rightarrow S_5 \rightarrow S_8 \rightarrow S_{10} \rightarrow$	1	$\rightarrow S_1,\ S_{10} \rightarrow$
2	$\rightarrow S_1 \rightarrow S_2 \rightarrow S_6 \rightarrow S_8 \rightarrow S_{10} \rightarrow$	1	$\rightarrow S_1,\ S_{10} \rightarrow$
3	$\rightarrow S_1 \rightarrow S_2 \rightarrow S_6 \rightarrow S_9 \rightarrow S_{10} \rightarrow$	2	$\rightarrow S_1,\ S_9 \rightarrow S_{10} \rightarrow$
4	$\rightarrow S_1 \rightarrow S_3 \rightarrow S_5 \rightarrow S_8 \rightarrow S_{10} \rightarrow$	1	$\rightarrow S_1,\ S_{10} \rightarrow$
5	$\rightarrow S_1 \rightarrow S_3 \rightarrow S_6 \rightarrow S_8 \rightarrow S_{10} \rightarrow$	1	$\rightarrow S_1,\ S_{10} \rightarrow$
6	$\rightarrow S_1 \rightarrow S_3 \rightarrow S_6 \rightarrow S_9 \rightarrow S_{10} \rightarrow$	2	$\rightarrow S_1,\ S_9 \rightarrow S_{10} \rightarrow$
7	$\rightarrow S_1 \rightarrow S_4 \rightarrow S_6 \rightarrow S_8 \rightarrow S_{10} \rightarrow$	3	$\rightarrow S_1 \rightarrow S_4 \rightarrow S_6,\ S_{10} \rightarrow$
8	$\rightarrow S_1 \rightarrow S_4 \rightarrow S_6 \rightarrow S_9 \rightarrow S_{10} \rightarrow$	2+3	$\rightarrow S_1 \rightarrow S_4 \rightarrow S_6,\ S_9 \rightarrow S_{10} \rightarrow$
9	$\rightarrow S_1 \rightarrow S_4 \rightarrow S_7 \rightarrow S_9 \rightarrow S_{10} \rightarrow$	4	$\rightarrow S_1 \rightarrow S_4 \rightarrow S_7 \rightarrow S_9 \rightarrow S_{10} \rightarrow$

Table 1: List of EMs and elementary flux patterns in the example network. The third column gives the indices of the elementary flux patterns that represent the flux of the EM in the subnetwork. The forth column gives the reactions of the elementary flux pattern.

2.3 Comparison of EMs and elementary flux patterns

The most obvious difference between EMs and elementary flux patterns is that the latter are defined as set of indices rather than vectors of fluxes. An EM represents a particular flux distribution in a network, in which flux proportions are considered. In contrast, a flux pattern can correspond to several flux proportions within the genome-scale system. However, most of the applications of EMs only require the set of non-zero indices of the EMs [5].

From the definition of elementary flux patterns in equation 3 and 4 important conclusions about a flux vector $\mathbf{v} \in \mathbb{R}^n$ fulfilling the flux pattern condition for a certain elementary flux pattern s_v can be drawn. $\mathbf{v}$ fulfills the steady-state condition and hence it can be written as a positive linear combination of a set of h elementary modes e_1,...,e_h. We can map each of these elementary modes to a corresponding flux pattern s_{e_1},..., s_{e_h} in the subsystem by identifying indices with non-zero fluxes in the elementary modes. It might appear that some elementary mode does not use any reaction of the subsystem. Thus, the corresponding flux pattern corresponds to the empty set. Since $s_{e_1} \cup \cdots \cup s_{e_h} = s_v$ we can conclude that at least one $s_{e^*} \in \{s_{e_1} \cdots s_{e_h}\}$ is equal to s_v, else s_v would not be elementary but could be written as a combination of other flux patterns derived from some of the flux patterns s_{e_1},..., s_{e_h}. Hence, for each elementary flux pattern s_v there exists at least one flux $\mathbf{v}$ that is an elementary mode and fulfills the flux pattern condition for s_v. In consequence, elementary flux patterns can be used to study EMs that go through the entire system and use reactions of the subsystem.

2.4 Example

Next, we will outline the above presented concepts in more detail by way of a small example network comprising 10 species and 17 reactions. 7 of these reactions are considered to belong to the subsystem.

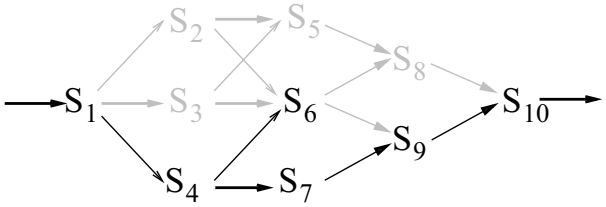

Figure 1: Simple example network. Black arrows indicate the reactions in the subnetwork, species drawn in gray are assumed external in Section 2.5.

The entire network contains 9 EMs and 4 elementary flux patterns in the subnetwork (Table 1). The lower number of EMs in contrast to elementary flux patterns is due to the alternative reactions that lead from S_1 to S_8. Four of the possible paths do not belong to the subsystem. This is exemplified by elementary flux pattern 1 which only contains the production of S_1 and the consumption of S_{10}. On the scale of the entire system, this elementary flux pattern represents the fluxes of the EMs 1, 2, 4 and 5. This example demonstrates how elementary flux patterns can ease the analysis by "abstracting" fluxes through parts of the system that are of no interest. In a natural system this may be relevant, for example, if we want to know which reactions of the tricarboxylic acid cycle are operative in EMs of the entire network without analyzing that network completely. Another important point arises from EM 3. In this case, the corresponding flux pattern is the union of the elementary flux patterns 2 and 3.

2.5 Example with external species

In a next step, we analyzed the results we obtained when defining the species not belonging to the subsystem as external. The results can be found in Table 2.

Here the potential problems arising if species are considered external become obvious. Thus, only EMs 9 and 10

Elementary Modes	
Number	**Reactions**
1	$\rightarrow S_1 \rightarrow [S_2]$
2	$\rightarrow S_1 \rightarrow [S_3]$
3	$[S_8] \rightarrow S_{10} \rightarrow$
4	$[S_2] \rightarrow S_6 \rightarrow [S_8]$
5	$[S_3] \rightarrow S_6 \rightarrow [S_8]$
6	$[S_2] \rightarrow S_6 \rightarrow S_9 \rightarrow S_{10} \rightarrow$
7	$[S_3] \rightarrow S_6 \rightarrow S_9 \rightarrow S_{10} \rightarrow$
8	$\rightarrow S_1 \rightarrow S_4 \rightarrow S_6 \rightarrow [S_8]$
9	$\rightarrow S_1 \rightarrow S_4 \rightarrow S_6 \rightarrow S_9 \rightarrow S_{10} \rightarrow$
10	$\rightarrow S_1 \rightarrow S_4 \rightarrow S_7 \rightarrow S_9 \rightarrow S_{10} \rightarrow$

Table 2: List of EMs when setting the species S_2, S_3, S_5 and S_8 of the example to external. EMs only representing reactions among external species have been omitted for clarity. External species are denoted with square brackets.

represent a true flux through the reaction network. The reason for this discrepancy arises from the fact that only two of the EMs contain the influx of S_1 and the outflux of S_{10} which is necessary for any steady-state flux in this network. This is exemplified by EM 1. In this EM the definition of S_2 to be external leads to the assumption that there is a pathway that uses up S_2 independently from the reactions in the subnetwork. However, this assumption is false since each flux in the system can only be in steady state if the outflux of S_{10} is used.

In the above example, only monomolecular reactions are included. In the more general situation where reactions of higher molecularity occur, another interesting case can arise. Consider, for example, the following system

$$\rightarrow A + 2B$$
$$2A + B \rightarrow$$

with species A defined as external. In this system we find an elementary mode using an equal flux of reaction 1 and 2, while there is no elementary mode if no species is set to external. In the previous examples, we could still obtain a steady-state flux by combining infeasible elementary modes with others. In this system in contrast, we find an elementary mode even though there is no steady-state flux at all. In consequence, we also do not find any elementary flux pattern for whatever subsystem chosen.

2.6 Checking the feasibility of elementary modes

Next we will outline how it can be checked, whether an elementary mode $\mathbf{e} \in \mathbb{R}^k$ is part of a steady-state flux $\mathbf{v} \in \mathbb{R}^n$ through the entire network. This check can be formulated in the terms of the following linear program

(1) $\mathbf{M} \cdot \mathbf{v} = 0$
(2) $\forall i \in \{1,...,k\} : \; e_i = v_i \cdot c$
(3) $\mathbf{v} \geq 0$
(4) $c \geq 1$

with the variables $\mathbf{v}$ and c. Since we only want to test the existence of $\mathbf{v}$ it is sufficient to check the feasibility of the linear program. Thus, we determine whether there is a steady state flux $\mathbf{v}$ in the complete system that uses a multiple of the fluxes of the elementary mode in the subsystem. Likewise, the flux through reactions having a zero entry in $\mathbf{e}$ is constrained to zero. We can multiply $\mathbf{e}$ with a non-zero number c, since an elementary mode only gives flux ratios and thus can be scaled. Furthermore, we would normally only require that $c > 0$, but since there is no upper bound for $\mathbf{v}$ we can translate this constraint into $c \geq 1$.

3 Application

In this Section we will outline the concept of elementary flux patterns by way of two example subsystems from the metabolism of *Escherichia coli* that have been studied previously using EMs. The first system is a model of the tricarboxylic acid cycle (TCA cycle), glyoxylate shunt and associated reactions and has been analyzed in [9]. The second system is an integrated model of glycolysis and pentose phosphate pathway presented by [10]. We will study both systems in context of a genome-scale model of *E. coli* that has been presented by [3]. The genome-scale model was modified by adding an outflow for each external species, which was subsequently set to internal. Inflows for glucose and other basic compounds were added to allow for the production of biomass (see Section 2.3for a complete list). The final network contains 1972 species and 3559 reactions.

We mapped the systems that have been used for EM analysis to the corresponding subsystems in the genome-scale model by identifying those reactions in the genome-scale model that belong to the particular system under study.

#	Reaction	Functional Assignment
1	$Arg + SuccCoA \rightarrow CoA + Sucarg$	Arginine degradation
2	$Hom + SuccCoA \rightarrow CoA + Suchms$	Methionine synthesis
3	$SuccCoA \rightarrow MmCoA$	Propionate utilization
4	$H2O + SuccCoA + Thdp \rightarrow CoA + Sl2a6o$	Lysine synthesis

Table 3: List of reactions consuming succinyl-CoA in the network of [3]. A list of abbreviations can be found in Appendices B and C.

We analyzed these systems under two aspects. First, we checked whether the EMs that have been obtained in the original publications indeed belonged to a steady-state flux in the genome-scale model. Thus, we checked for each EM if there exists a steady-state flux through the entire system that uses exactly the flux rations of the EM in the subsystem. This can be done using a linear program outlined in Section 2.6. Second, we analyzed the elementary flux patterns the subsystems give rise to.

3.1 Case study I: TCA cycle

Next, we will analyze a model of the tricarboxylic acid cycle (TCA cycle), glyoxylate shunt and associated reactions in *E. coli* that has been previously analyzed using EMs [9].

We did not incorporate the output fluxes used in the EM analysis in the subsystem. These reactions can be interpreted as abstractions of other reactions in a larger network that are not modelled in the network analyzed with EMs. As outlined in Section 2.5, using such fluxes underlies the assumption that the species consumed by them can be used up by reactions outside of the subsystem without requiring an additional flux through it. Since elementary flux patterns explicitly take into account the complete network, we do not need to add such abstract reactions.

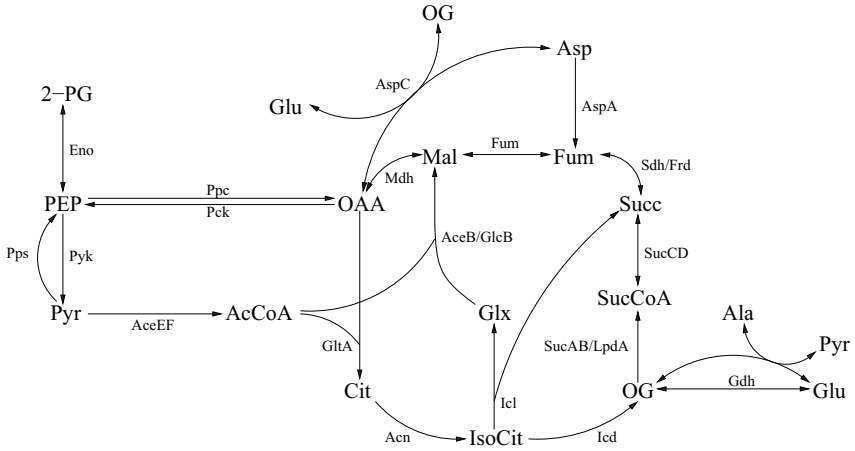

Figure 2: Scheme of part of the central metabolism of *E. coli*, which is studied as a subnetwork of a genome-scale network in case study I. Irreversible reactions are indicated by unidirectional arrows. A list of abbreviations can be found in Appendices A and C.

Analysis of EMs In a first step we analyzed the 16 EMs that have been identified in [9]. We found that only 10 out of the 16 EMs are indeed part of a steady-state flux within the entire system. An analysis of the remaining EMs reveals that they all produce succinyl-CoA. In the genome-scale system we found 4 reactions consuming this species (Figure 3). However, 3 of these reactions require an additional species that can only be produced from intermediates of the subsystem to be part of a steady-state flux.

An example is entry 4 in Table 3 containing the consumption of succinyl-CoA during lysine synthesis. Later in that pathway, succinate is released. However, this reaction also consumes Thdp which can be produced only from oxaloacetate. Oxaloacetate in turn can only be produced by reactions within the subsystem. A similar line of arguments can be applied to entry number 3 in Table 3. Even though this reaction does not require an additional educt, the further steps in propionate utilization lead to a conversion of oxaloacetate to pyruvate. Again, this reaction can only be in steady state if additional reactions of the subsystem are used to replenish oxaloacetate.

Analysis of elementary flux patterns In a next step we analyzed the elementary flux patterns of the subsystem. This system gives rise to 75 elementary flux patterns. In contrast, [9] found only 16 EMs. The reason for this difference is the larger number of possible inputs and outputs of the subsystem to the entire system. On the other hand, this reflects the central role of the TCA cycle in metabolism.

In some cases, elementary flux patterns are cryptic when only considering the reactions of the subsystem they

contain. Therefore, the analysis of the EMs of the entire system containing them yields further insight. One example is the reaction of malate to oxaloacetate catalyzed by malate dehydrogenase (Mdh) and malate:quinone oxidoreductase (Mqo). We took into account the network analyzed in [9] and hence we added only the reaction catalyzed by Mdh to the subsystem. In consequence, an elementary flux pattern producing malate and consuming oxaloacetate not necessarily incorporates the reaction catalyzed by Mdh to close the gap between both species. The EM associated to such a flux pattern can alternatively use Mqo not present in the subsystem. As a result we find many elementary flux patterns not containing the reaction of Mdh. In such a case, the associated EM uses Mqo instead.

Entry points into the subsystem As a starting point for our analysis we considered glycolysis as the principal pathway for the production of TCA cycle intermediates. Nevertheless, many pairs of elementary flux patterns only differ in a few steps of the routes by which these intermediates are produced. In each such pair, one elementary flux pattern uses phosphoglycerate as entry point, while the other uses pyruvate. In the entire network there are many possible routes on which pyruvate can be produced from glucose without the use of reactions of glycolysis that are part of the subsystem (e.g., the Entner-Doudoroff pathway, Figure 4 B).

One of the principal routes on which intermediates of glycolysis can enter the TCA cycle proceeds via the pyruvate dehydrogenase producing acetyl-CoA from pyruvate. However, the consumption of acetyl-CoA does not allow a positive production rate of any of the species of the TCA cycle [14]. If TCA cycle intermediates are to be consumed they need to be replenished using either PEP carboxykinase to produce oxaloacetate from phosphoenolpyruvare or the glyoxylate bypass to produce malate and succinate from citrate and acetyl-CoA. In this context, we found an elementary flux pattern that was capable of producing phosphoenolpyruvate from acetyl-CoA via the glyoxylate bypass. This indicates the third entry-point into the subsystem via the intermediate of acetyl-CoA. However, acetyl-CoA is not produced from reactions of the subsystem but by another pathway. An example is the above mentioned Entner-Doudoroff pathway producing pyruvate and a subsequent conversion of pyruvate into formate and acetyl-CoA using the pyruvate formate lyase (not present in the subsystem).

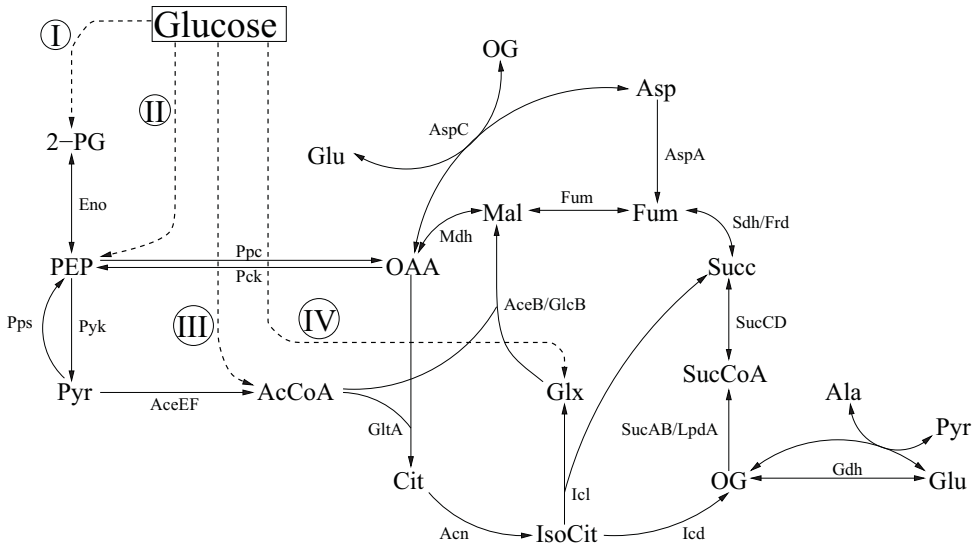

Figure 3: Entry points into the TCA cycle. **(I)** Glycolysis. **(II)** Entner-Doudoroff-pathway. **(III)** Entner-Doudoroff-pathway and production of acetyl-CoA from pyuvate, e.g., using pyruvate formate lyase. **(IV)** Production of glyoxylate using an alternative pathway through nucleotide synthesis (Figure 4 A). Dashed arrows correspond to schematic reactions. A list of abbreviations can be found in Appendices B and C.

Isocitrate lyase is an essential enzyme in the glyoxylate bypass. Nevertheless, we found several elementary flux patterns that used up glyoxylate but did not incorporate isocitrate lyase. Most interestingly whenever we encountered such a case, aspartate aminotransferase was operative. Aspartate aminotransferase produces aspartate and oxoglutarate from glutamate and oxaloacetate. The consumption of glutamate can be counter-balanced by the amination of oxoglutarate. Aspartate in turn is essential for a reaction in the synthesis of the purine base inosine. Using several reactions this compound is subsequently converted into glyoxylate via the intermediate of inosine-monophosphate and urate (Figure 4 A). This pathway corresponds to the fourth entry point into the subsystem.

3.2 Case study II: Glycolysis and pentose phosphate pathway

Next, we will analyze a model of glycolysis and pentose phosphate pathway that has been studied in [10]. The corresponding subsystem is depicted in Figure 5.

Analysis of EMs In a first step we analyzed the 7 EMs that have been found by [10]. In contrast to the previous example we found that all EMs are part of steady-state fluxes within the genome-scale system and thus were indeed

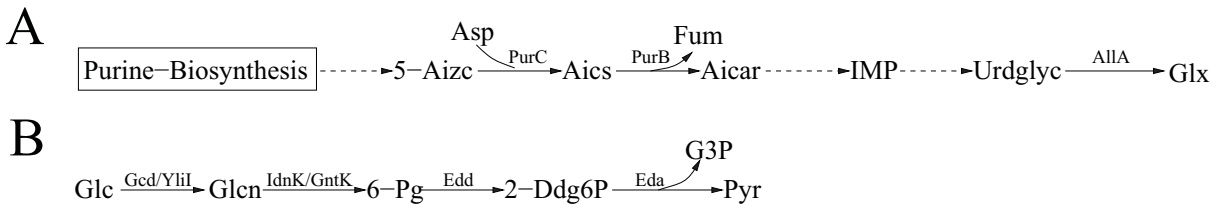

Figure 4: Alternative pathways in the central metabolism of *E. coli*. Dashed arrows represent condensed reactions. **A** Alternative glyoxylate producing pathway. **B** Entner-Doudoroff pathway. A list of abbreviations can be found in Appendices B and C.

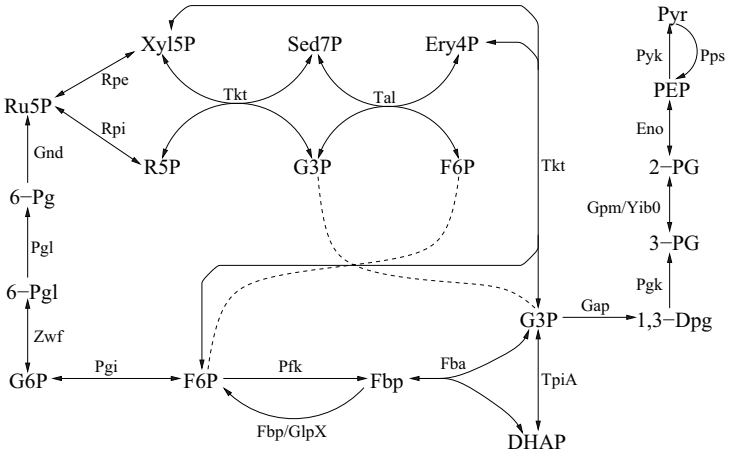

Figure 5: Scheme of glycolysis and pentose phosphate pathway in *E. coli* which is studied as the second subnetwork. Species connected by dashed lines are identical. A list of abbreviations can be found in Appendices B and C.

feasible.

Analysis of elementary flux patterns This subsystem gives rise to 87 elementary flux patterns. Thus, the increase from the 7 EMs analyzed in [10] is even greater than in case study I. This large increase is due to the smaller number of exchange reactions that have been used in [10]. Only an outflow of pyruvate and ribose 5-phosphate had been added to model the production of TCA cycle intermediates, respectively, nucleotides. However, there is a drain from several additional species of this subsystem to other compounds that are necessary for *E. coli* to survive, for instance phosphoenolpyruvate, utilized in the production of aromatic amino acids, or fructose 6-phosphate, for the production of cell-membrane constituents.

Entry points into the subsystem Similar to the first case-study we examined on which pathways glucose can be converted into intermediates of the subsystem. Thus, we examined on which pathways the EMs associated to the elementary flux patterns used up glucose before they used reactions from the subsystem.

Even though glycolysis is the principal pathway of glucose metabolism there is a large number of alternative pathways that can yield intermediates of glycolysis and pentose phosphate pathway from glucose (Figure 6). In the usual pathway, glucose is phosphorylated either by a phosphotransferase system (PTS) or by glucokinase. The alternative pathways use a route over gluconate or fructose. In the former case gluconate is converted into the pentose phosphate pathway intermediate 6-phospho-D-gluconate. Subsequently, it can either enter glycolysis through the pentose phosphate pathway or use the Entner-Doudoroff pathway yielding glyceraldehyde 3-phosphate and pyruvate. Using the route via fructose we find two further pathways. Fructose is converted into fructose 6-phosphate which is an intermediate of glycolysis. Subsequently, an alternative route to the reaction of fructose 6-phosphate to glyceraldehyde 3-phosphate and dihydroxyacetone catalyzed by phosphofructokinase and fructose-bisphosphate aldolase can be used. This alternative pathway uses fructose 6-phosphate aldolase to produce dihydroxyacetone and glyceraldehyde 3-phosphate. Dihydroxyacetone is subsequently phosphorylated to dihydroxyacetone phosphate.

4 Conclusion

In this work we introduced the concept of elementary flux patterns which allows a much more reliable analysis of fluxes in genome-scale metabolic networks. It enables a consistent application of many of the concepts from EM analysis to such networks. These applications include the determination of minimal feasible growth media and the analysis of the outcome of knockout experiments. Even though the analysis concentrates on fluxes within a subnetwork, elementary flux patterns are tightly coupled to EMs on the genome scale. Hence, in contrast to EM analysis,

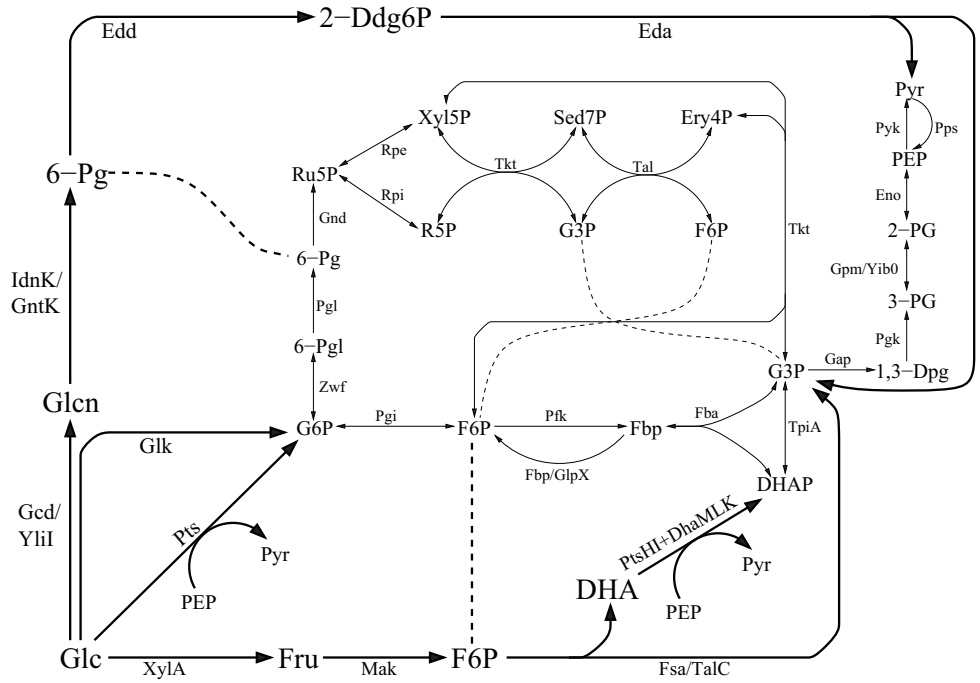

Figure 6: Entry points into the subsystem. Pathways for the production of glycolysis and pentose phosphate pathway intermediates from glucose in the growth media (bold lines). Where possible, identical species have been connected by dashed lines. The reversibility of reactions leading into the subsystem has been omitted for clarity. A list of abbreviations can be found in Appendices B and C.

the information about the complete network is not ignored. In consequence, problems due to the introduction of external species and exchange fluxes are overcome. By analyzing the EMs obtained from a previously published work, we demonstrated that an improper definition of external species can lead to wrong results. Thus, EMs can be incomplete, i.e., additional reactions are necessary to guarantee a steady state, or might not even be part of any steady-state flux in the entire system.

Using elementary flux patterns several important issues in the analysis of metabolic networks are addressed. First, even though a flux pattern only represents a set of reactions rather than specific fluxes, most of the applications of EMs only require these sets [5]. Thus, many of the applications of EMs can be easily used in combination with elementary flux patterns instead.

Second, by a non-constrained topology of the reactions of the subsystem it is, for example, possible to analyze certain parts of a pathway in detail, while others can be left out. In consequence, the emphasis of the analysis can be scaled as desired. Thus, also the combinatorial explosion seen in the analysis using EMs can be avoided.

Third, by considering the entire network, the information on all possible pathways connecting two species is maintained. In Section 3.1 we started with 2-phosphoglycerate as principal source for intermediates of the subsystem. Nevertheless, we found 4 possible routes that were able to provide other intermediates including a not yet known pathway for the production of glyoxylate. In the context of a subsystem comprising glycolysis and pentose phosphate pathway we found 6 different pathways on which intermediates of the subsystem can be produced from glucose. The incorporation of all the knowledge contained within a genome-scale network into flux analysis is of importance, since a comprehensive knowledge about the (to-date known) 3359 reactions and 1972 species in the metabolism of *E. coli* is difficult to maintain. In consequnce, since elementary flux patterns take into account all this information, they more faithfully reflect the metabolic capabilities of an organism.

5 Funding

Financial support from the German Ministry for Research and Education (BMBF) to C.K. within the framework of the Forsys Partner initiative and from the Fundação Calouste Gulbenkian, FCT and Siemens SA Portugal (PhD grant SFRH/BD/32961/2006) to L.F.F. is gratefully acknowledged.

6 References

[1] J. Behre, T. Wilhelm, A. von Kamp, E. Ruppin, and S. Schuster. Structural robustness of metabolic networks with respect to multiple knockouts. *J Theor Biol*, 252(3):433–441, Jun 2008.

[2] I. Borodina and J. Nielsen. From genomes to in silico cells via metabolic networks. *Curr Opin Biotechnol*, 16(3):350–355, Jun 2005.

[3] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. Ø. Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 orfs and thermodynamic information. *Mol Syst Biol*, 3:121, 2007.

[4] A. M. Feist and B. Ø. Palsson. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotechnol*, 26(6):659–667, Jun 2008.

[5] J. Gagneur and S. Klamt. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, 5:175, 2004.

[6] S. Klamt and E. D. Gilles. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234, Jan 2004.

[7] S. Klamt and J. Stelling. Combinatorial complexity of pathway analysis in metabolic networks. *Mol Biol Rep*, 29(1-2):233–236, 2002.

[8] J. O. Krömer, C. Wittmann, H. Schröder, and E. Heinzle. Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. *Metab Eng*, 8(4):353–369, Jul 2006.

[9] S. Schuster, T. Dandekar, and D. A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol*, 17(2):53–60, Feb 1999.

[10] S. Schuster, D. A. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol*, 18(3):326–332, Mar 2000.

[11] S. Schuster and D. Kenanov. Adenine and adenosine salvage pathways in erythrocytes and the role of S-adenosylhomocysteine hydrolase. A theoretical study using elementary flux modes. *FEBS J*, 272(20):5278–5290, Oct 2005.

[12] S. Schuster, S. Klamt, W. Weckwerth, and T. Pfeiffer. Use of network analysis of metabolic systems in bioengineering. *Bioproc. Biosyst. Eng.*, 24:363–372, 2002.

[13] J. Schwender, F. Goffman, J. B. Ohlrogge, and Y. Shachar-Hill. Rubisco without the calvin cycle improves the carbon efficiency of developing green seeds. *Nature*, 432(7018):779–782, Dec 2004.

[14] E. O. Weinman, E. H. Srisower, and I. L. Chaikoff. Conversion of fatty acids to carbohydrate; application of isotopes to this problem and role of the Krebs cycle as a synthetic pathway. *Physiol Rev*, 37(2):252–272, Apr 1957.

[15] T. Wilhelm, J. Behre, and S. Schuster. Analysis of structural robustness of metabolic networks. *Syst Biol (Stevenage)*, 1(1):114–120, Jun 2004.

Appendices

A Input-Species

The following species are provided as input to the model. They are necessary to produce all biomass metabolites

D-glucose, ammonium, nitrate, sulfate, Fe^{2+}, Fe^{3+}, CO_2, H^+, potassium, calcium, cobalt, molybdate, sodium, phosphate, oxygen, water, chloride, Cu^{2+}, Mg^{2+}, Mn^{2+} and Zn^{2+}

B Abbreviations

Table 4: List of abbreviated species names.

Abbreviation	Species
1,3-Dpg	3-phospho-D-glyceroyl phosphate
2-Ddg6P	2-dehydro 3-deoxy-D-gluconate 6-phosphate
2-PG	D-glycerate 2-phosphate
3-PG	D-glycerate 3-phosphate
5-Dglcn	5-dehydro-D-gluconate
5-Aizc	5-amino 1,5-phospho-D-ribosyl-imidazole 4-carboxylate
6-Pg	6-phospho-D-gluconate
AcCoA	acetyl-CoA
Aicar	5-Amino 1,5-Phospho-D-ribosyl-imidazole 4-carboxamide
Aics	S-2,5-amino 1,5-phospho-D-ribosyl-imidazole 4-carboxamidosuccinate
Ala	L-alanine
Asp	L-aspartate
Cit	citrate
CoA	coenzyme A
DHA	dihydroxyacetone
DHAP	dihydroxyacetone phosphate
Ery4P	D-erythrose 4-phosphate
F6P	D-fructose 6-phosphate
Fbp	D-fructose 1,6-bisphosphate
Fru	D-fructose
Fum	fumarate
G3P	glyceraldehyde 3-phosphate
G6P	D-glucose 6-phosphate
Glc	D-glucose
Glcn	Gluconate
Glu	L-glutamate
Glx	glyoxylate
H2O	water
IMP	inosine monophosphate
IsoCit	isocitrate
Mal	L-malate
MmCoA	R-methylmalonyl-CoA
OAA	oxaloacetate
OG	2-oxoglutarate
PEP	phosphoenolpyruvate
Pyr	pyruvate

Abbreviation	Species
Q8	Ubiquinone-8
Q8H2	Ubiquinol-8
R5P	α-D-Ribose 5-phosphate
Ru5P	D-ribulose 5-phosphate
Sed7P	sedoheptulose 7-phosphate
Sl2a6o	N-succinyl 2-L-amino 6-oxoheptanedioate
SucArg	N2-succinyl-L-arginine
Succ	succinate
SucCoa	succinyl-CoA
Suchms	O-succinyl-L-homoserine
Thdp	2,3,4,5-Tetrahydrodipicolinate
Urdglyc	ureidoglycolate
Xyl5P	D-xylulose 5-phosphate

C Enzyme names

Table 5: List of abbreviated enzyme names.

Abbreviation	Species
AceB/GlcB	malate synthase
AceEF	pyruvate dehdrogenase
Acn	aconitase
AllA	ureidoglycolate hydrolase
AspC	aspartate aminotransferase
Eda	2-keto-3-deoxy-6-phosphogluconate aldolase
Edd	6-phosphogluconate dehydratase
Eno	enolase
Fba	fructose-bisphosphate aldolase
Fbp/GlpX	fructose-1,6-bisphophatase
Fsa/TalC	Fructose 6-phosphate aldolase
Fum	fumarase
Gap	glyceraldehyde-3-phosphate dehydrogenase
Gcd/YliI	glucose dehydrogenase / aldose sugar dehydrogenase
Gdh	glutamate dehydrogenase
GltA	citrate synthase
Gnd	6-phosphogluconate dehydrogenase
Gnt/IdnT	gluconate transporter / GntP family L-idonate transporter
Gpm	phosphoglycerate mutase
Icd	isocitrate dehydrogenase
Icl	isocitrate lyase
IdnK/GntK	D-gluconate kinase
IdnO	5-keto-D-gluconate reductase
IdnT	GntP family L-idonate transporter
Mak	manno(fructo)kinase
Mdh	malate dehydrogenase
Mqo	malate dehydrogenase
Pck	phosphoenolpyruvate carboxykinase
Pfk	phosphofructokinase

Abbreviation	Species
Pgi	glucose-6-phosphate isomerase
Pgk	phosphoglycerate kinase
Pgl	6-phosphogluconolactonase
Ppc	phosphoenolpyruvate carboxylase
Pps	phosphoenolpyruvate synthase
Pts	phosphotransferase system
PtsHI+DhaMLK	Dihydroxyacetone phosphotransferase
PurB	adenylosuccinate lyase
PurC	phosphoribosyaminoimidazole-succinocarboxamide synthetase
Pyk	pyruvate kinase
Rpe	D-ribulose-5-phosphate 3-epimerase
Rpi	ribose-5-phosphate isomerase
Sdh/Frd	succinate dehydrogenase / fumarate reductase
SucAB/LpdA	a-ketoglutarate dehydrogenase / dihydrolipoamide dehydrogenase
SucCD	succinyl-CoA synthetase
Tal	transaldolase
Tkt	transketolase
Tpi	triosephosphate isomerase
XylA	D-xylose isomerase
YbiV	HAD phosphatase
Zwf	glucose-6-phosphate dehydrogenase