Can the Whole Be Less Than the Sum of its Parts? Pathway Analysis in Genome-Scale Metabolic Networks using Elementary Flux Patterns

- Supplementary Material -

Christoph Kaleta[1,2], Luís Filipe de Figueiredo[1], and Stefan Schuster[1]

[1] Department of Bioinformatics, Friedrich Schiller University Jena, Ernst-Abbe-Platz 2, D-07743 Jena, Germany
[2] Bio Systems Analysis Group, Jena Centre for Bioinformatics (JCB) and School of Mathematics and Computer Science, Friedrich Schiller University Jena, D-07743 Jena, Germany

Corresponding Author:
Christoph Kaleta
Department of Bioinformatics, Friedrich Schiller University Jena
Ernst-Abbe-Platz 2, D-07743 Jena, Germany
Phone: +49-3641-949583
Fax: +49-3641-946452
E-Mail: Christoph.Kaleta@uni-jena.de

S1 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+}

S2 Abbreviations

Table 1: List of abbreviated species names.

Abbreviation	**Species**
2-H3OP	2-hydroxy 3-oxopropanoate
2-PG	D-glycerate 2-phosphate
2,3-G6P	2-dehydro 3-deoxy-D-gluconate 6-phosphate
3PG	3-phospho-D-glycerate
5-Aizc	5-amino 1,5-phospho-D-ribosyl-imidazole 4-carboxylate
6-Pgc	6-phospho-D-gluconate
Ac	acetate
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
Arg	L-arginine
Asp	L-aspartate
Cit	citrate
CoA	coenzyme A
Cys	cysteine
DHAP	dihydroxyacetone phosphate
FDP	D-fructose 1,6-bisphosphate
Fum	fumarate
G3P	glyceraldehyde 3-phosphate
G6P	D-glucose 6-phosphate
GABA	4-Aminobutanoate
Gl	glycerol
Glc	D-glucose
Glcn	Gluconate
Glu	L-glutamate
Glx	glyoxylate

Abbreviation	**Species**
Glyc	R-Glycerate
H20	water
His	Histidine
Hpyr	Hydroxypyruvate
IMP	inosine monophosphate
ICit	isocitrate
Mal	L-malate
MmCoA	R-methylmalonyl-CoA
NH4	ammonium
OAA	oxaloacetate
OG	2-oxoglutarate
PEP	phosphoenolpyruvate
Prpp	5-phospho-alpha-D-ribose 1-diphosphate
Pyr	pyruvate
R5P	α-D-ribose 5-phosphate
Ru5P	D-ribulose 5-phosphate
Sl2a6o	N-succinyl 2-L-amino 6-oxoheptanedioate
SO4	sulfate
SucArg	N2-succinyl-L-arginine
Succ	succinate
SucCoa	succinyl-CoA
Suchms	O-succinyl-L-homoserine
Sucsal	succinic-semialdehyde
Thdp	2,3,4,5-tetrahydrodipicolinate
Urdglyc	ureidoglycolate
X5P	D-xylulose 5-phosphate

S3 Enzyme names

Table 2: List of abbreviated enzyme names.

Abbreviation	**Species**
AceB/GlcB	malate synthase
AceEF	pyruvate dehdrogenase
Acn	aconitase
AllA	ureidoglycolate hydrolase
AspA	aspartate ammonia-lyase

Abbreviation	**Species**
AspC	aspartate aminotransferase
Eda	2-keto-3-deoxy-6-phosphogluconate aldolase
Edd	6-phosphogluconate dehydratase
Eno	enolase
Fum	fumarase
GabD	succinate-semialdehyde dehydrogenase
GarK	glycerate kinase I
GarR/GlxR	tartronate semialdehyde reductase
Gcd/YliI	glucose dehydrogenase / aldose sugar dehydrogenase
Gcl	glyoxylate carboligase
Gdh	glutamate dehydrogenase
GltA	citrate synthase
GlxK	glycerate kinase II
Hyi	hydroxypyruvate isomerase
Icd	isocitrate dehydrogenase
Icl	isocitrate lyase
IdnK/GntK	D-gluconate kinase
Mdh	malate dehydrogenase
Mqo	malate dehydrogenase
Pck	phosphoenolpyruvate carboxykinase
Ppc	phosphoenolpyruvate carboxylase
Pps	phosphoenolpyruvate synthase
PurB	adenylosuccinate lyase
PurC	phosphoribosyaminoimidazole-succinocarboxamide synthetase
PuuE/GabT	4-aminobutyrate aminotransferase
Pyk	pyruvate kinase
Sdh/Frd	succinate dehydrogenase / fumarate reductase
SucAB/LpdA	a-ketoglutarate dehydrogenase / dihydrolipoamide dehydrogenase
SucCD	succinyl-CoA synthetase

S4 Runtime complexity of the computation of elementary flux patterns

For the analysis of the runtime complexity of the algorithm for the computation of elementary flux patterns it is necessary to take a closer look on the mixed-integer linear program (MILP) used for the computation of elementary flux patterns. The MILP contains twice the number of reactions as integer variables. It has been shown by Lenstra (1983) that the runtime of solving a MILP is exponential in the number of integer variables and polynomial in the number of real variables. We have to solve the MILP as many times as there are elementary flux patterns. A trivial upper boundary of the number of flux patterns is 2^k which is also an upper boundary for the number of elementary flux patterns. Multiplying two exponential functions yields another exponential function. Hence, the upper boundary for the running time of the presented algorithm is exponential in the number of reactions in the subsystem and polynomial in the number of reactions of the entire system.

S5 Relationship between elementary modes and elementary flux patterns

Here, we want to examine the properties of a flux vector $\mathbf{v} \in \mathrm{R}^n$ fulfilling the flux pattern condition for a certain elementary flux pattern s_v in more detail. Thus, we first examine the special case in which a subsystem encompasses the entire system. Second, we show that each elementary flux pattern is part of at least one elementary mode in the entire system. Third, we demonstrate how such an elementary mode can be found. Forth, we outline a linear program that allows to determine whether an elementary mode of the subsystem is part of a steady-state flux of the entire system. We will use the same notation as in the main document. Thus, for a metabolic network with n reactions among m metabolites or species $\mathbf{M}$ defines the $m \times n$ stoichiometric matrix of which the first k columns, *i.e.,* reactions, are assumed to belong to the subsystem.

S5.1 A subsystem encompassing the entire system

If the subnetwork encompasses the entire network ($k = n$), each elementary flux pattern corresponds to an elementary mode of the network. This can be shown through the definition of a flux pattern. Given a flux $\mathbf{v}$ in such a system that obeys the condition $\mathbf{Mv} \geq 0$ and $\mathbf{v} \geq 0$ (*i.e.*, it is a steady state flux), we identify r as the set of reactions having a non-zero flux in $\mathbf{v}$. As it can be easily seen, r fulfills the flux pattern condition (conditions 3-6 of the main document) since the subsystem encompasses the whole system. Hence, we can find a flux pattern to each steady state flux of the network. Each steady state can be written as a positive linear combination of elementary modes and each flux pattern as a set-union of elementary flux patterns. Thus, each elementary flux pattern s uniquely corresponds to the non-zero indices of an elementary mode $\mathbf{e}$ and $\mathbf{e}$ is a flux-vector fulfilling the flux pattern condition for s.

S5.2 Elementary modes associated to elementary flux patterns

Since $\mathbf{v}$ fulfills the steady-state condition, it can be written as a positive linear combination of a set of h elementary modes e_1,...,e_h. Each of these elementary modes can be assigned to one corresponding flux pattern in the subsystem s_{e_1},..., s_{e_h} by identifying those reactions of the elementary mode having a non-zero flux in the subsystem. If some elementary mode does not use any reaction of the subsystem, the corresponding flux pattern might be equal to the empty set. Since $s_{e_1} \cup \cdots \cup s_{e_h} = s_v$ and s_v fulfills the elementarity condition, at least one $s_{e^*} \in \{s_{e_1} \cdots s_{e_h}\}$ needs to be equal to s_v. Consequently, for each elementary flux pattern there is at least one elementary mode e^* in the complete system, being a solution vector of the flux pattern condition. Thus, each elementary flux pattern is part of at least one global flux corresponding to an elementary mode of the complete system. This elementary mode can be computed using an approach outlined in Section S5.3. Therefore, elementary flux patterns can also be used to study elementary modes in genome-scale metabolic networks using routes through the predefined subsystem.

S5.3 Finding elementary modes containing an elementary flux pattern

Next, we will outline how an elementary mode fulfilling the flux pattern condition for an elementary flux pattern s can be found.

As outlined in Section 2.3 we can start with a flux vector $\mathbf{v}$ that fulfills the flux pattern condition for s, *i.e.*, a vector $\mathbf{v}$ that is a feasible solution of the linear program outlined in the Appendix. It is necessary to decompose $\mathbf{v}$ into elementary modes in order to obtain an elementary mode $\mathbf{e}*$ containing exactly the reactions of s in the subsystem. Thus, we want $\mathbf{v}$ already to use as few reactions as possible and, hence, to be the linear combination of only few elementary modes. This can be achieved by adding an objective function to the linear program used to check whether a reaction set is a flux pattern (see the Appendix for more details).
We obtain an initial $\mathbf{v}$ by solving the linear program

$$min \sum_{i=1}^{n} v_i$$

subject to

(1) $\mathbf{M} \cdot \mathbf{v} = 0$

(2) $\mathbf{v} \geq 0$

(3) $\forall i \in s : v_i \geq 1$

(4) $\forall j \in \{1, .., k\} \setminus s : \ v_j = 0$

Hence, we find a $\mathbf{v}$ that fulfills the flux pattern condition for s and has the least overall flux, if a minimal flux of 1 for the reactions of the flux pattern is required. This approach is similar to flux-minimization as described in Holzhütter (2004). Practical experience shows that the solution of this linear program often returns an elementary mode immediately. However, this is not necessarily the case. Thus, we need to decompose $\mathbf{v}$ into the set of elementary modes of which it is a positive linear combination. This can be done by computing the elementary modes of the system just made up by the reactions having a non-zero flux in $\mathbf{v}$. Note that this approach might also find additional elementary modes that are not necessary to decompose $\mathbf{v}$. In this case we just need to identify an elementary mode $\mathbf{e}^*$ corresponding to a flux vector fulfilling the flux pattern condition for s. This flux pattern is necessarily contained within the set of elementary modes we obtain (*cf.* Section 2.3 of the main document). Doing this we can even analyze elementary modes of the complete system, *i.e.*, “genome-scale” elementary modes using reactions from the subsystem.

S5.4 Checking the feasibility of elementary modes

In order to check whether an elementary mode $\mathbf{e}$ of the subsystem is a part of a steady-state flux in the entire system we need to check if there exists a steady-state flux $\mathbf{v}$ in the complete system such that the flux ratios of $\mathbf{v}$ in the subsystem are a multiple of $\mathbf{e}$.

First, we have to derive $\mathbf{e}'$ from $\mathbf{e}$ by removing indices corresponding to exchange reactions that have been added to the subsystem for elementary mode analysis. Next, we check whether the following condition is fulfilled

$$\exists \mathbf{v} \in \mathbb{R}^n, \ c > 0 : \ \mathbf{v} \geq 0 \text{ and } \mathbf{M} \cdot \mathbf{v} = 0 \text{ and } \forall i \in \{1..k\} : \ v_i = c \cdot e'_i$$

This can be easily transformed into a linear program almost identical to the one used to determine whether a set of reactions fulfills the flux pattern condition (see the Appendix for more details).

S6 Complete MILP and proof of elementarity of the computed flux pattern

Given the stoichiometric matrix $\mathbf{M}$ of which the first k columns correspond to the subsystem and the set of previously found elementary flux patterns S, the MILP reads

$$min \sum_{i=1}^{k} b_i$$

subject to

$$
\begin{array}{ll}
\text{(LP 1)} & \mathbf{M} \cdot \mathbf{v} = 0 \\
\text{(MILP 1)} & \forall i \in \{1,..,k\}: \ b_i \leq v_i \leq c \cdot b_i \\
\text{(MILP 2)} & \forall i \in \{1,..,k\}: \ b_i - h_i \geq 0 \\
\text{(MILP 3)} & \forall s' \in S: \ \sum_{i \in s'} b_i + h_i \leq |s'| \\
\text{(MILP 4)} & \sum_{i=1}^{k} h_i \geq 1 \\
\text{(LP 2)} & \mathbf{v} \geq 0 \\
\text{(MILP 5)} & \mathbf{b}, \mathbf{h} \in \{0,1\}^k
\end{array}
$$

with the real variables $\mathbf{v}$ and the binary variables $\mathbf{b}$ and $\mathbf{h}$.

By minimizing the number of elements of $\Theta(\mathbf{b})$ in the objective function, we achieve that we find an elementary flux pattern. If $\Theta(\mathbf{b})$ is not elementary it can be written as a union of a set of elementary flux patterns S' ($\Theta(\mathbf{b}) \notin S'$) that contain at least one element $\Theta(\mathbf{t})$ that has not been found by the previous iterations. Else $\mathbf{b}$ could not be part of a feasible solution of the MILP and constraints (2) to (4) would be violated. Furthermore, $\Theta(\mathbf{t})$ is, by definition, a proper subset of $\Theta(\mathbf{b})$ and hence $\sum_{i=1}^{k} t_i < \sum_{i=1}^{k} b_i$. Thus, the objective function is not minimal. Hence, $\Theta(\mathbf{b})$ has to be an elementary flux pattern.

S7 Relationship between flux coupling analysis and elementary flux pattern analysis

Next, we want to outline how the coupling of reactions as determined in flux coupling analysis (Burgard et al. 2004) can be derived from calculating elementary flux patterns. Given two reactions i and j flux coupling analysis consists of determining the relationships between the fluxes through both reaction in a steady-state flux. There are three types of coupling (Burgard et al. 2004). First, i is called directionally coupled to j if a non-zero flux in i implies a non-zero flux in j, but not necessarily the reverse. That is, reaction i requires a non-zero flux in j to appear in a steady-state flux. Second, i and j are called partially coupled if a non-zero flux in i implies a non-zero flux in j and vice versa. In such a case, i is directionally coupled to j and j is directionally coupled to i. Third, i and j are called fully coupled if a non-zero flux in i implies a specific flux in j. That is, $\frac{v_i}{v_j} = const.$ for every steady-state flux $\mathbf{v} \in \mathbb{R}^n$.

In this context we can compute the coupling between two reactions by defining the reactions i and j as the subsystem. Assuming that i and j can appear in a steady-state flux, we obtain one of the following sets of flux patterns S

1. $S = \{\{i\}, \{j\}\}$: in this case both reactions can have a non-zero flux independent from each other. Hence, they are not coupled.

2. $S = \{\{i\}, \{i, j\}\}$: in this case reaction j can only have a non-zero flux if reaction i has a non-zero flux. Thus, j is directionally coupled to i.
3. $S = \{\{i, j\}, \{j\}\}$: similar to 2., Here, i is directionally coupled to j.
4. $S = \{\{i, j\}\}$: in this case, j can only have a non-zero flux if reaction i has a non-zero flux and vice versa. Thus, both reactions are partially coupled.

References

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

Holzhütter HG 2004. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur J Biochem* **271**(14): 2905–2922.

Lenstra H 1983. Integer programming with a fixed number of variables. *Math Oper Res* **8**: 538–548.