Supplemental Material for

Control of Fluxes in Metabolic Networks

Georg Basler[1,2,8], Zoran Nikoloski[3,8], Abdelhalim Larhlimi[4], Albert-László Barabási[5−7,9], Yang-Yu Liu[6−7,9]

[1] Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720, USA
[2] Department of Environmental Protection, Estación Experimental del Zaidín CSIC, Granada, 18008 Spain
[3] Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, Potsdam, 14476, Germany
[4] Laboratoire d'Informatique de Nantes Atlantique, Universit de Nantes, Nantes, 44322 France
[5] Center for Complex Network Research and Departments of Physics, Computer Science and Biology, Northeastern University, Boston, MA 02115, USA
[6] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[7] Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA
[8] These authors contributed equally to this work.
[9] Corresponding authors; email: alb@neu.edu; yyl@channing.harvard.edu

Contents

Flux coupling analysis

Let $C = \{v \in \mathbb{R}^n | Sv = 0, v_{irrev} \geq 0\}$ be the steady-state flux cone of a given stoichiometric matrix S with n reactions, where $irrev$ are the irreversible reactions with flux v_{irrev}. Here, we restrict our analysis to a subspace $F \subset C$ by bounding the fluxes and by the additional requirement of a (non-zero) exchange of matter with the environment: $F = \{v \in \mathbb{R}^n | Sv = 0, v_{irrev} \geq 0, lb \leq v \leq ub, \exists i \in E : v_i \neq 0\}$, where lb, ub are the lower and upper flux bounds, respectively, and E is the set of exchange reactions. We refer to $v \in F$ as *feasible flux distributions*.

For a feasible flux distribution $v \in F$, we denote v_i, v_j the fluxes of reactions i, j. We denote $i^* = \max_{v \in F}\{v_i\}$ the maximum possible flux of reaction i (Mahadevan and Schilling, 2003). Let i and j be *unblocked* reactions, i.e., there exist v, $w \in F$ with $v_i \neq 0$ and $w_j \neq 0$. Further, let i and j be *non-essential*, i.e., there exist v, $w \in F$ with $v_i = 0$, and $w_j = 0$ (cf. *Blocked reactions* and *Essential reactions*). A pair of reactions (i, j) is called *coupled*, if any of the following holds for all $v \in F$:

1. Full coupling: $v_i = \lambda v_j, \lambda \neq 0$.
2. Partial coupling: $v_i = 0 \Leftrightarrow v_j = 0$.
3. Directional coupling: $v_i \neq 0 \Rightarrow v_j \neq 0$.
4. Anti-coupling: $v_i = 0 \Rightarrow v_j \neq 0$.
5. Inhibitive coupling: $v_i = i^* \Rightarrow v_j = 0$.

If none of the above holds, the reaction pair (i, j) is called *uncoupled*. Note that full, partial, and directional couplings have been studied in literature (Burgard et al., 2004), while anti- and inhibitive couplings are novel concepts introduced here.

We refer to a reaction i as *active* in v, denoted by $\sigma_i^v = 1$, if $v_i \neq 0$, and *inactive* in v, denoted by $\sigma_i^v = 0$, otherwise. We use the coupling relations to control the status of reactions: if i and j are fully coupled, controlling the value for v_i can be used to specify the value for v_j, and vice versa. Similarly, if i and j are either fully or partially coupled, activation of i leads to activation of j, while deactivation of i leads to deactivation of j, and vice versa. If i is directionally coupled to j, then activation of i leads to activation of j, and deactivation of j leads to deactivation of i, but not vice versa. If i and j are anti-coupled, then deactivation of i leads to activation of j, and vice versa. Finally, if i is inhibitively coupled to j, then maximizing v_i leads to deactivation of j.

The above definitions include four possibilities of qualitative coupling between pairs of active and inactive reactions:

1. An active reaction implies that all reactions to which it is fully, partially, or directionally coupled are active.
2. An inactive reaction implies that all reactions to which it is fully, partially, or reversed directionally coupled are inactive.
3. An inactive reaction implies that all reactions to which it is anti-coupled are active.
4. A reaction with maximal flux implies that all reactions to which it is inhibitively coupled are inactive.

Full coupling is a special case of partial coupling, and partial coupling is a special case of directional coupling. Inhibitive coupling may hold for a pair of reactions concomitantly with other couplings. Moreover, full and partial coupling are equivalence relations. Directional coupling is transitive, and anti-coupling is symmetric. The transitivity relations and their implications on the computational performance for the calculation of driver reactions are detailed in *Efficient calculation of driver reactions*.

We further note that the definition of anti-coupling requires that $\vec{0} \notin F$ (otherwise, $v_i = 0 \Rightarrow v_j \neq 0, \forall v \in F$, would not hold for any i, j). Similarly, if flux distributions consisting only of futile cycles (i.e., flux distributions in which only internal reactions have a non-zero flux) are allowed, then $v_i = 0, v_j = 0$ holds for

any two reactions i and j not part of the futile cycle, which would render anti-coupling impractical. Therefore, we restrict the flux space to flux distributions which have a non-zero exchange of matter with the environment using the constraint $\exists i \in E : v_i \neq 0$, where E is the set of exchange reactions. If i and j are anti-coupled, then, for each $v \in F$, either $v_i \neq 0$ or $v_j \neq 0$. Thus, anti-coupling is similar to the concept of synthetic lethality, by which only the simultaneous knockout of two genes is lethal. Here, the simultaneous deactivation of two anti-coupled reactions i, j would violate the constraints imposed on F, and therefore at least one of i or j must be active at steady state. We observe that anti-coupling is rare in large metabolic networks (cf. Supplemental Table 1), which is a natural consequence of their redundancy: if there are at least three flux distributions with disjoint active reactions, $v_1, v_2, v_3 \in F$ with $\{i : \sigma_i^{v_1} = 1\} \cap \{j : \sigma_j^{v_2} = 1\} \cap \{k : \sigma_k^{v_3} = 1\} = \emptyset$, then no anti-coupled pair of reactions exists, since there is a feasible flux distribution $v \in F$ with $v_l = v_m = 0$ for each pair of reactions l, m.

Inhibitive coupling represents the trade-off between the fluxes of reactions (or, in general, pathways) which share common substrates or have common products. For example, two reactions which share a substrate compete for its consumption, as the increase of the flux of one reaction leads to a decrease of substrate concentrations, and, thus, decrease of the other reaction's flux. Although we focus on qualitative coupling relations, we point out that a similar definition may also be used to represent quantitative relations, since the gradual increase of a reaction flux may lead to a parallel decrease of the flux of a reaction to which it is inhibitively coupled. Note that the intuitive definition based on active and inactive reactions, i.e., $v_i \neq 0 \Rightarrow v_j = 0$, does not hold for any two unblocked reactions i and j, because $v, w \in F$ with $v_i > 0$ and $w_j > 0$ implies that $x = v + w \in F$ with $x_i > 0$ and $x_j > 0$ (and, similarly, for negative fluxes of reversible reactions). Since inhibitive coupling relies on the maximum possible fluxes of reactions, it is necessary to specify finite upper bounds for each (irreversible) flux to ensure that the problem is bounded. Nevertheless, we point out that finite upper bounds of fluxes are a realistic assumption due to the saturation of enzyme activities, and demonstrate that the proposed definition of inhibitive coupling does not depend on the choice of uniform upper bounds (see *Inhibitive coupling does not depend on the choice of uniform upper bounds*).

Linear programs

To perform the flux coupling analysis with the five coupling relations, we first determine the blocked reactions, and fully, partially, and directionally coupled reaction pairs using the F2C2 tool (Larhlimi et al., 2012). Then, we determine the essential reactions, anti- and inhibitively coupled reactions pairs via linear programs. For completeness, we list all the linear programs used in the flux coupling analysis as follows. All calculations were performed using MATLAB™ and Tomlab™.

Blocked reactions

A reaction i is called blocked, if it has a zero flux in each feasible flux distribution, i.e., $v_i = 0, \forall v \in C$. Blocked reactions can be calculated by subsequently minimizing and maximizing the following linear program (Burgard et al., 2004):

$$\begin{aligned} \min \,|\, \max \quad & v_i \\ \text{s.t.} \quad & Sv = 0 \\ & lb \leq v \leq ub \end{aligned}$$

where $S \in \mathbb{R}^{m \times n}$ is the stoichiometric matrix representing m metabolites and n reactions, $lb, ub \in \mathbb{R}^n$ are the lower and upper flux bounds, and $v \in \mathbb{R}^n$ is a decision variable representing a flux distribution. If both minimum and maximum solutions are zero, then i is blocked, and discarded from the subsequent analyses.

Full, partial, and directional couplings

To determine full, partial, and directional couplings efficiently, reactions are first classified as *fully reversible* or *pseudo-irreversible* (Larhlimi et al., 2012): a reversible reaction i is called *fully reversible*, if there is a flux distribution $v \in C$ with $v_i \neq 0$ and $\forall j \in irrev : v_j = 0$. A reversible reaction i is called *pseudo-irreversible*, if it is not fully reversible.

In the following, full, partial, and directional couplings are determined for the three possible cases:

1. The reactions i and j are both fully reversible or both pseudo-irreversible. In this case, if i is directionally coupled to j, then i and j are also partially and fully coupled (Larhlimi et al., 2012). The following linear program determines whether i, j are fully coupled:

$$\begin{aligned} \max \;\; & v_i \\ \text{s.t.} \;\; & v_j = 0 \\ & 0 \leq v_i \leq 1 \\ & Sv = 0 \end{aligned}$$

   i and j are fully coupled if and only if the optimal solution is $v_i = 0$.

2. The reaction i is pseudo-irreversible, and the reaction j is neither fully reversible nor pseudo-irreversible. In this case, i can only be directionally coupled to j, but they cannot be fully or partially coupled. To determine whether i is directionally coupled to j, the following linear program is subsequently minimized and maximized:

$$\begin{aligned} \min \,|\, \max \;\; & v_i \\ \text{s.t.} \;\; & v_j = 0 \\ & -1 \leq v_i \leq 1 \\ & Sv = 0 \end{aligned}$$

   i is directionally coupled to j if and only if both minimum and maximum solutions are zero.

3. Neither i nor j is fully reversible or pseudo-irreversible. Then the following linear program is minimized separately for v_i and v_j:

$$\begin{aligned} \tau_{i|j} = \min \;\; & v_{i|j} \\ \text{s.t.} \;\; & v_{j|i} = 1 \\ & Sv = 0 \end{aligned}$$

   i is directionally coupled to j if and only if $\tau_j \neq 0$; j is directionally coupled to i, if and only if $\tau_i \neq 0$. Finally, i and j are fully coupled, if and only if $\tau_i \neq 0$, $\tau_j \neq 0$, and $\tau_i = 1/\tau_j$.

Essential reactions

To facilitate the definition of new coupling types, we restrict the flux space by the additional requirement of a (non-zero) exchange of matter with the environment, i.e., $\exists i \in E : v_i \neq 0$, where E is the set of exchange reactions. This constraint also facilitates the definition of essential reactions, i.e., reactions, which are active in all feasible flux distributions $v \in F$. To simplify the problem formulation, in the following we represent a reversible reaction i by two irreversible reactions, denoted i^+ and i^-, with non-negative fluxes $v_{i^+}, v_{i^-} \geq 0$.

Thereby, the new stoichiometric matrix S' has an additional column, S'_{*,i^-}, for each reversible reaction i, with $S'_{*,i^+} = -S'_{*,i^-} = S_{*,i}$.

To test whether a reaction i is essential, we determine the feasibility of $v \in F$ with $v_i = 0$. The mixed-integer linear program for determining the essentiality of a reaction i is:

$$\begin{aligned} v_i &= 0 \\ S'v &= 0 \\ lb \leq v &\leq ub \\ x_k \cdot \alpha \leq &v_k, \text{ for all } k \in E \\ \sum_{k \in E} x_k &\geq 1 \end{aligned}$$

where $S' \in \mathbb{R}^{m \times n}$, $lb, ub \in \mathbb{R}^n_{\geq 0}$, α, and the set E of exchange reactions are given, and $v \in \mathbb{R}^n$, $x \in \{0,1\}^n$ are decision variables. The parameter α specifies the minimum required flux of at least one exchange reaction. Here, we use $lb = (0)^n$, $ub = (1,000)^n$ and $\alpha = 1$. These values have been used previously for similar constraints (Schellenberger et al., 2011). If the program has no solution, then no $v \in F$ with $v_i = 0$ exists, and thus i is essential. As all reactions are both directionally and anti-coupled to an essential reaction by definition (i.e., if i is essential, then $v_j \neq 0 \Rightarrow v_i \neq 0$ and $v_j = 0 \Rightarrow v_i \neq 0$ for all reactions j), we neglect essential reactions from the subsequent calculation of coupling relations.

Anti-coupling

To determine whether a pair of reactions i, j is anti-coupled, we test the feasibility of $v \in F$ with $v_i = 0$ and $v_j = 0$ using the following mixed integer linear program:

$$\begin{aligned} v_i &= 0 \\ v_j &= 0 \\ S'v &= 0 \\ lb \leq v &\leq ub \\ x_k \cdot \alpha \leq &v_k, \text{ for all } k \in E \\ \sum_{k \in E} x_k &\geq 1 \end{aligned}$$

where S', v, lb, ub, x, α, and E are defined as before. If the program has no solution, then i and j are anti-coupled. Note that, due to the symmetry of anti-coupling, this program needs to be solved only once for each unordered reaction pair $\{i, j\}$ with $i \neq j$.

Inhibitive coupling

To determine inhibitively coupled reaction pairs, we first calculate the maximum fluxes i^* of all reactions i by maximizing v_i over all $v \in F$ (Mahadevan and Schilling, 2003). Then, we determine the feasibility of $v_i = i^*$ and $v_j \neq 0$ using the following linear program:

$$v_i = i^* \tag{1}$$
$$v_j \geq \epsilon_1 \tag{2}$$
$$S'v = 0 \tag{3}$$
$$lb \leq v \leq ub \tag{4}$$

where S', v, lb, ub are defined as before, i^* is given, and ϵ_1 is a positive number close to zero, which allows the problem to be formulated as linear program. The implementation requires additional threshold values to

allow for flexibility of the numerical solutions. Thus, we implement constraint (1) as $v_i \geq i^* \cdot (1 - \epsilon_2)$ with $\epsilon_2 = 10^{-3}$, and use $\epsilon_1 = j^* \cdot 10^{-2}$, where j^* is the maximum flux of j. Note that changing these parameters by several orders of magnitude does not notably affect the results.

If the program has no solution, then i and j are inhibitively coupled. The program must be solved for each ordered pair of reactions, and for each direction of a reversible reaction separately, making this procedure the most time consuming step. Note that the constraint $\exists i \in E : v_i \neq 0$ is not required here, because it does not affect the maximum fluxes of reactions, rendering the program faster than a corresponding mixed integer program. Further, as a reaction in S' may represent either direction of a reversible reaction in S, we consider two reactions i and j inhibitively coupled in S, if one of the following holds:

1. If i is reversible and j is irreversible, then i is inhibitively coupled to j in S, if either i^+ or i^- is inhibitively coupled to j in S'.
2. If i is irreversible and j is reversible, then i is inhibitively coupled to j in S, if i is inhibitively coupled to both j^+ and j^- in S'.
3. If both i and j are reversible, then i is inhibitively coupled to j in S, if either i^+ or i^- is inhibitively coupled to both j^+ and j^- in S'.

Thus, reversible reactions i and j are inhibitively coupled, if maximizing the flux of i in either direction leads to inhibition of both directions of j.

Analogy between flux coupling and mass balance

We point out that the flux coupling relations can be derived from the mass balance equations of a metabolic network. As an example, consider the mass balance equations of the metabolic network from Figure 1A:

$$v_{R_1} = v_{E_1} - v_{R_4} \tag{1}$$
$$v_{R_2} = v_{R_1} + v_{E_2} \tag{2}$$
$$v_{R_3} = v_{E_3} = v_{R_1} + v_{R_2} \tag{3}$$
$$v_{R_4} = v_{R_5} = v_{E_4} \tag{4}$$

where $v_{R_1}, \ldots, v_{R_5}, v_{E_1}, \ldots, v_{E_4} \geq 0$. Now, we observe that

1. R_3 and E_3 are fully coupled, since $v_{R_3}/v_{E_3} = 1$ (Eq. 3).
2. R_4, R_5 and E_4 are fully coupled, since $v_{R_4}/v_{R_5} = v_{R_4}/v_{E_4} = v_{R_5}/v_{E_4} = 1$ (Eq. 4).
3. R_1 is directionally coupled to R_2 and R_3.
   Proof: Since $v_{R_2} = v_{R_1} + v_{E_2}$ (Eq. 2) and $v_{R_3} = v_{R_1} + v_{R_2}$ (Eq. 3), we have $v_{R_3} \geq v_{R_2} \geq v_{R_1}$. Then, $v_{R_1} > 0$ implies $v_{R_3} \geq v_{R_2} > 0$.
4. R_2 and R_3 are partially coupled.
   Proof: From $v_{R_2} = 0$ follows $v_{R_1} = 0$ (since R_1 is directionally coupled to R_2). Then, $v_{R_3} = v_{R_1} + v_{R_2} = 0$. If $v_{R_2} > 0$, then from $v_{R_3} = v_{R_1} + v_{R_2} \geq v_{R_2}$ follows $v_{R_3} > 0$. Hence, $v_{R_2} = 0 \Leftrightarrow v_{R_3} = 0$.
5. R_2 and R_4 are anti-coupled.
   Proof: If v_{R_2}=0, then from Eq. (2) follows v_{R_1}=v_{E_2} = 0. Then, from Eq. (3) follows v_{R_3}=v_{E_3}=0. Now, if we assume v_{R_4}=0, then v_{R_4}=v_{R_5}=v_{E_4}=0, then $v_{E_1}, v_{E_2}, v_{E_3}, v_{E_4} = 0$, which violates $\exists i \in E : v_i \neq 0$, and thus, $v \notin F$. Hence, $v_{R_4} > 0$.
6. R_3 and R_4 are anti-coupled.
   Proof: If $v_{R_3} = v_{R_1} + v_{R_2} = 0$, then, since R_1 is directionally coupled to R_3, v_{R_2}=0. Since R_2 and R_4 are anti-coupled, it follows that $v_{R_4} > 0$.

7. R_4 is inhibitively coupled to R_1.
   From flux variability analysis, we obtain $R_4^* = E_1^* = 1,000$. Now, let $v_{R_4} = R_4^*$. Then, from Eq. (1), it follows that $v_{E_1} = 1,000$ and $v_{R_1} = 0$.

Similarly, we may derive the remaining reaction couplings of the metabolic network (Supplemental Figure 1).

Efficient calculation of driver reactions

The problem of calculating a minimum out-dominating set is NP-hard (Garey and Johnson, 1979). Nevertheless, we find that its calculation is computationally inexpensive for large networks, which is a consequence of the transitivity of coupling relations. In the following, we first demonstrate that the minimum out-dominating sets in a *transitive digraph*, i.e., a directed graph $G = (V, E)$, for which $(i, j), (j, k) \in E$ implies $(i, k) \in E$, can be determined in polynomial time. Next, we show that this leads to a substantial reduction of the computational effort for calculating driver reactions in a control graph.

Let G be a transitive digraph. The vertices of G with an in-degree of zero, called *source vertices*, are included in any out-dominating set of G. We denote the source vertices of G by V^{source}. Furthermore, we refer to the union of V^{source} with its out-neighborhood, $V^{\text{source}} \cup N^+(V^{\text{source}})$, as the vertex set *dominated by* V^{source} and denote this set by $D(V^{\text{source}})$.

Let $H = G[V \setminus D(V^{\text{source}})]$ be the subgraph of G induced by $V \setminus D(V^{\text{source}})$. From the transitivity of G follows that the induced subgraph H is also transitive. Furthermore, $(i, j) \in E$ with $i \in D(V^{\text{source}})$ implies $j \in D(V^{\text{source}})$ due to transitivity. Since all source vertices of G are contained in $D(V^{\text{source}})$, each vertex of H must have an in-degree of at least one.

Theorem 1. *Each weakly connected component*[1] *of H is dominated by one vertex.*
Proof: Let C be the set of vertices of a weakly connected component of H. Since all vertices $v \in C$ have a non-zero in-degree, C does not contain any isolated vertices and there are at least two vertices in C with a non-zero out-degree. Let V^{sink} denote the vertices of G with an out-degree of zero, called *sink vertices*. Then, there is a closed walk $v_1, \ldots, v_k, v_1$ through all non-sink vertices $v_1, \ldots, v_k \in C \setminus V^{\text{sink}}$. Due to the transitivity of H, it follows that $(i, j), (j, i) \in E(H)$ for all non-sink vertices $i \neq j, i, j = 1, \ldots k$. Hence, any vertex $l \in C$ with non-zero out-degree dominates $C \setminus V^{\text{sink}}$. Moreover, due to the transitivity of H, it also follows that $(l, i) \in E(H)$ for all vertices $i \in C \cap V^{\text{sink}}$. □

Let $\mathcal{C}^H = \{C_1, C_2, \ldots\}$ denote the vertex sets of the weakly connected components of H.

Corollary. *A minimum out-dominating set $\mathcal{D}^{out}$ of the transitive digraph $G = (V, E)$ is given by the source vertices V^{source} of G and one vertex with non-zero out-degree from each weakly connected component of $G[V \setminus D(V^{\text{source}}]$:*

$$\mathcal{D}^{out}(G) = V^{\text{source}} \cup Z,$$

where

$$Z \in \prod_{C_i \in \mathcal{C}^H} C_i \setminus V^{\text{sink}}.$$

Proof: As shown before, the source vertices V^{source} of G belong to any out-dominating set of G, and they dominate $D(V^{\text{source}}) = V^{\text{source}} \cup N^+(V^{\text{source}})$. The remaining vertex set, $V \setminus D(V^{\text{source}})$, is dominated by one vertex with non-zero out-degree from each weakly connected component of its induced subgraph (Theorem 1). From the transitivity of G, it follows that there is no edge $(i, j) \in E$ with $i \in D(V^{\text{source}})$ and $j \notin D(V^{\text{source}})$. If there is an edge $(i, j) \in E$ with $i \notin D(V^{\text{source}})$ and $j \in D(V^{\text{source}})$, then it does not affect the minimum out-dominating sets, since j is already dominated by a vertex in V^{source}. Hence, a minimum out-dominating set of G is given by the union of V^{source} and the minimum out-dominating sets of each weakly connected component of the subgraph of G induced by $V \setminus D(V^{\text{source}})$. □

[1] A weakly connected component is a maximal subgraph, for which the underlying undirected graph is connected.

Since the set of vertices with zero in-degree, an induced subgraph, and the weakly connected components can be determined in polynomial time, a minimum out-dominating set of G can also be determined in polynomial time.

We now demonstrate that the transitivity of coupling relations may substantially improve the computational performance of calculating a minimum out-dominating set in a control graph. Let $G = (V, E)$ be a CG obtained by integrating a FCG with a reaction activity pattern σ. We denote an edge $(i, j) \in E$ by $i \overset{\sigma_i\sigma_j}{\rightarrow} j$, resulting in four types of edges: $\overset{00}{\rightarrow}$, $\overset{01}{\rightarrow}$, $\overset{10}{\rightarrow}$, and $\overset{11}{\rightarrow}$, corresponding to different types of coupling (cf. Methods).

Theorem 2. *For a control graph $G = (V, E)$ with $i, j, k \in V$ and $i \overset{\sigma_i\sigma_j}{\rightarrow} j$, $j \overset{\sigma_j\sigma_k}{\rightarrow} k$, the following transitivity relations hold:*

1. *If $\sigma_i = \sigma_j = \sigma_k$, then $i \overset{\sigma_i\sigma_k}{\rightarrow} k$.*
2. *If $\sigma_i = 0$ and $\sigma_k = 1$, then $i \overset{01}{\rightarrow} k$.*
3. *If $\sigma_i = 1$ and $\sigma_j = \sigma_k = 0$, then $i \overset{10}{\rightarrow} k$.*

Proof:

1. From $i \overset{\sigma_i\sigma_j}{\rightarrow} j$ and $j \overset{\sigma_j\sigma_k}{\rightarrow} k$ with $\sigma_i = \sigma_j = \sigma_k$, it follows that i, j and j, k are fully, partially, or directionally coupled. Directional coupling may hold from i to j and from j to k (if $\sigma_i = \sigma_j = \sigma_k = 1$), or from k to j and from j to i (if $\sigma_i = \sigma_j = \sigma_k = 0$). Then, from the transitivity of full, partial, and directional coupling, it follows that i and k are fully, partially, or directionally coupled and $i \overset{\sigma_i\sigma_k}{\rightarrow} k$.
2. From $i \overset{\sigma_i\sigma_j}{\rightarrow} j$ with $\sigma_i = \sigma_j = 0$, it follows that i and j are fully or partially coupled, or j is directionally coupled to i. Moreover, from $j \overset{\sigma_j\sigma_k}{\rightarrow} k$ with $\sigma_k = 1$, it follows that j and k are anti-coupled. Then, i and k are anti-coupled, and $i \overset{\sigma_i\sigma_k}{\rightarrow} k$. If $\sigma_i = 0$ and $\sigma_j = \sigma_k = 1$, then i and j are anti-coupled, and j, k are fully or partially coupled, or j is directionally coupled to k. Therefore, i and k are anti-coupled and $i \overset{\sigma_i\sigma_k}{\rightarrow} k$.
3. From $i \overset{\sigma_i\sigma_j}{\rightarrow} j$ and $j \overset{\sigma_j\sigma_k}{\rightarrow} k$ with $\sigma_i = 1$ and $\sigma_j = 0$, it follows that i is inhibitively coupled to j. Moreover, from $\sigma_j = \sigma_k = 0$, it follows that j and k are fully or partially coupled, or k is directionally coupled to j. Consequently, i is inhibitively coupled to k and $i \overset{\sigma_i\sigma_k}{\rightarrow} k$. □

The cases 1 and 2 from Theorem 2 cover all possibilities of full, partial, directional, and anti-coupling between three reactions i, j, and k. Therefore, a control graph obtained from a metabolic network which only has couplings of these types is a transitive digraph. Consequently, its driver reactions can be determined in polynomial time (see Corollary above). However, since inhibitive coupling is not transitive, this claim does not hold for a metabolic network which has inhibitive couplings. Nevertheless, if all vertices $j \in V$ with $j \overset{10}{\rightarrow} k$ have an in-degree of zero, then the control graph is transitive, which follows from case 3 of Theorem 2. Thus, for a general control graph G, the problem of determining the driver reactions is intractable only for the subgraph of G induced by the vertices $\{i \in V : deg^{in}(i) > 0, i \overset{01}{\rightarrow} j, j \in V\}$. Indeed, we observe that calculation of driver reactions for the largest of the analyzed metabolic networks (glandular breast tissue), which has 98,781 inhibitive couplings, takes less than 1 second per reaction activity pattern on a standard desktop PC, rendering the approach applicable to large metabolic networks.

Sensitivity to missing reactions

Since flux coupling analysis was found to be sensitive to missing reactions (Marashi and Bockmayr, 2011), we assess to what extent this affects the calculated driver reactions. To this end, we generate networks by removing 5, 15, or 25% of the unblocked reactions from the genome-scale (Orth et al., 2011) and central metabolic (Orth et al., 2010) networks of *E. coli* at random (100 samples per network and percentage). Since the removal of a reaction may lead to the blocking of other reactions, and thus their effective removal, we approximate the

percentage of removed reactions by removing one reaction at a time, and determining the resulting number of blocked reactions. Hence, we obtain networks, where at least 5, 15, or 25% of the reactions have been removed.

We determine the flux coupling, reaction activity patterns, and driver reactions for the generated networks. Next, we compare the driver reactions in the original genome-scale and central metabolic networks of *E. coli* to those in the corresponding networks obtained by removing reactions. To this end, for each reaction we determine the fraction of activity patterns in which it is a driver, and compare the fractions of the original network to those obtained by removing reactions using Euclidean distances and Pearson correlation coefficients.

When removing 5% of the reactions in the central metabolic network of *E. coli*, we obtain a mean Euclidean distance of 2.5 and mean Pearson correlation coefficient of 0.74 (p values $\leq 1.0 \cdot 10^{-6}$). When removing 15% of the reactions, we obtain a mean Euclidean distance of 3.4 and mean Pearson correlation coefficient of 0.55 (p values $\leq 9.6 \cdot 10^{-3}$). For networks with 25% of reactions removed, the mean Euclidean distance is 3.5, and the mean Pearson correlation coefficient is 0.49 (p values ≤ 0.055).

For the genome-scale metabolic network of *E. coli*, we obtain a mean Euclidean distance of 7.7 and mean Pearson correlation coefficient of 0.89 (p values $\leq 5.4 \cdot 10^{-289}$) when removing 15% of its reactions. When removing 25% of the reactions, we obtain a mean Euclidean distance of 8.7 and mean Pearson correlation coefficient of 0.85 (p values $\leq 3.5 \cdot 10^{-217}$). These values demonstrate that the determined driver reactions are not significantly affected when removing up to 15% from the central metabolic network, and up to 25% of the reactions from the genome-scale metabolic network of *E. coli*.

Inhibitive coupling does not depend on the choice of uniform upper bounds

In the following, we show that inhibitive coupling does not depend on the choice of the upper flux bounds $ub \in \mathbb{R}^n$, as long as $lb = 0$ and ub is uniform, i.e., the upper bound is the same for all reaction fluxes. We employ the duality principles of linear programming to show that, given two reactions i and j, if i is inhibitively coupled to j for any upper bound $x > 0$, then this holds for all upper bounds.

We denote by e^i the n-dimensional unit vector where $e^i_i = 1$, and $e^i_j = 0$ for all $j \neq i$. Let $\mathbf{1}$ be the all-ones vector, i.e., $\mathbf{1} = \Sigma_{i=1}^{n} e^i$. For a positive scalar $a > 0$, let $C(a) = \{v \in \mathbb{R}^n | Sv = 0, 0 \leq v \leq a \cdot \mathbf{1}\}$ be the flux space where all reaction fluxes are limited by the same upper bound a. In the following, let $x > 0$ and $y > 0$ be two positive scalars and let i and j be two reactions.

Let's assume that i is inhibitively coupled to j in $C(x)$, i.e.,

$$\forall v \in C(x) : v_i = \max_{w \in C(x)} \{w_i\} \;\Rightarrow\; v_j = 0.$$

We show that i is also inhibitively coupled to j in $C(y)$.

Let $v' \in C(y)$ such that $v'_i = \max_{w \in C(y)} \{w_i\}$. We need to show that $v'_j = 0$. To this end, we consider the following primal linear program:

$$\begin{aligned} c_1 = \max\ & {e^i}^T w \\ \text{s.t. } & Sw = 0, \\ & 0 \leq w \leq y \cdot \mathbf{1}. \end{aligned}$$

The corresponding dual problem is then:

$$\begin{aligned} c_2 = \min\ & y \cdot \mathbf{1}^T u \\ \text{s.t. } & u + S^T t \geq e^i, \\ & u \geq 0. \end{aligned}$$

The *strong duality* property states that $c_1 = c_2$. Accordingly, there exist two vectors $u \in \mathbb{R}^n$ and $t \in \mathbb{R}^m$ such that $v'_i = y \cdot \mathbf{1}^T u$, $u + S^T t \geq e^i$ and $u \geq 0$.

By choosing $v = v' \cdot x/y$, we have $Sv = 0$ and $0 \leq v \leq x \cdot \mathbf{1}$, and consequently $v \in C(x)$. We also have $v_i = v'_i \cdot x/y = x \cdot \mathbf{1}^T u$. Using again the *strong duality* property, we have $v_i = \max_{w \in C(x)}\{w_i\}$, and since i is inhibitively coupled to j in $C(x)$, $v_j = 0$. Finally, since $v = v' \cdot x/y$ with $x, y > 0$ and $v_j = 0$, we have $v'_j = 0$, and so the claim follows. $\square$

References

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

Garey MR and Johnson DS. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA.

Larhlimi A, David L, Selbig J, and Bockmayr A. 2012. F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. *BMC Bioinformatics* **13**: 57.

Mahadevan R and Schilling CH. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* **5**: 264–276.

Marashi SA and Bockmayr A. 2011. Flux coupling analysis of metabolic networks is sensitive to missing reactions. *Biosystems* **103**: 57–66.

Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, and Palsson BØ. 2011. A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. *Mol Syst Biol* **7**: 535.

Orth JD, Fleming RMT, and Palsson BØ. 2010. Reconstruction and use of microbial metabolic networks: the core Escherichia coli metabolic model as an educational guide. *EcoSal Plus* **28**: 245–248.

Schellenberger J, Lewis NE, and Palsson BØ. 2011. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys J* **100**: 544–553.