On correlated reaction sets and coupled reaction sets in metabolic networks

Sayed-Amir Marashi*†‡ and Zhaleh Hosseini*

*Department of Biotechnology, College of Science
University of Tehran, Tehran, Iran

†Department of Bioinformatics, School of Computer Science
Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

‡marashi@ut.ac.ir

Received 13 August 2014
Revised 25 January 2015
Accepted 25 January 2015
Published 4 March 2015

Two reactions are in the same “correlated reaction set” (or “Co-Set”) if their fluxes are linearly correlated. On the other hand, two reactions are “coupled” if nonzero flux through one reaction implies nonzero flux through the other reaction. Flux correlation analysis has been previously used in the analysis of enzyme dysregulation and enzymopathy, while flux coupling analysis has been used to predict co-expression of genes and to model network evolution. The goal of this paper is to emphasize, through a few examples, that these two concepts are inherently different. In other words, except for the case of full coupling, which implies perfect correlation between two fluxes ($R^2 = 1$), there are no constraints on Pearson correlation coefficients (CC) in case of any other type of (un)coupling relations. In other words, Pearson CC can take any value between 0 and 1 in other cases. Furthermore, by analyzing genome-scale metabolic networks, we confirm that there are some examples in real networks of bacteria, yeast and human, which approve that flux coupling and flux correlation cannot be used interchangeably.

Keywords: Co-sets; flux coupling analysis (FCA); Monte Carlo sampling; uniform sampling; flux cone.

1. Introduction

There are two main strategies for modeling metabolic network fluxes. The first strategy is to use kinetic modeling. This strategy requires extensive knowledge on reaction kinetics and parameters. Unfortunately, it is not always possible to use this strategy for modeling metabolism, as complete knowledge about a metabolic system is rarely available. The second strategy is to use constraint-based modeling of metabolic fluxes. In the latter framework, the networks are supposed to function in steady state and the only constraints used are stoichiometric, thermodynamic

†Corresponding author.
(i.e. reversibility/irreversibility of reactions) and capacity constraints. In the present paper, we will discuss (and compare) the properties of two constraint-based concepts, namely flux coupling and flux correlation.

The term “coupled reaction set” was first used in 2004 to describe a set of metabolic reactions (fluxes) which are coupled to each other based on flux coupling analysis (FCA). If a nonzero flux through reaction \( i (v_i \neq 0) \) implies a nonzero flux through reaction \( j (v_j \neq 0) \), then \( i \) is said to be directionally coupled to \( j \) (or, \( i \rightarrow j \)). If \( i \rightarrow j \) and \( j \rightarrow i \), then reactions \( i \) and \( j \) are partially coupled (or, \( i \leftrightarrow j \)). Reactions \( i \) and \( j \) are called fully coupled (or, \( i \Leftarrow j \)) if they are partially coupled, and additionally, fluxes through reactions \( i \) and \( j \) are proportional \( (v_i/v_j = \text{constant}) \). Two unblocked fluxes are either (fully, partially or directionally) coupled, or uncoupled.

Coupling and uncoupling relationships have many biological implications. In case of full coupling, two reactions will always operate together with a fixed flux ratio. This may imply that their corresponding genes are located in the same operon and/or they are co-regulated/co-expressed. In case of directional coupling, the dependency between reactions is unidirectional. It suggests that the independent reaction (i.e. reaction \( j \) in an \( i \rightarrow j \) relation) should be more essential, expressed, regulated and conserved compared to the dependent reaction. Horizontal gene transfer has also been shown to be affected by constraints applied by directional coupling relations between transferred genes and genes already present in host genome. In addition, the set of equivalent knockouts for eliminating the activity of a particular reaction can be identified by finding the reactions to which the desired reaction is directionally coupled. This can be useful in designing genetic engineering experiments. On the other hand, uncoupled reactions have often compensatory roles in the absence of each other. In other words, the activity of only one is sufficient to have the function. Therefore, they can be regarded as alternative reactions, and from the essentiality point of view, they are synthetic lethal pairs. For instance, in some diseases like cancer, in which one of the two reactions is inactivated through mutation, the removal of the other reaction may result in cancerous cell death. In contrast, in normal cells (with no mutation), removal of the second reaction will not be harmful.

Shortly after the introduction of FCA, Price et al. used the term “correlated reaction set” (Co-Set), to describe a set of reactions with correlated fluxes. The Pearson correlation coefficient (CC) between every pair of reactions can be computed by Monte Carlo sampling of the flux space. For a pair of perfectly correlated fluxes \( i \) and \( j \) \( (|R_{ij}| = 1) \), the authors used the term “perfectly correlated” reactions.

Random sampling and flux correlation are used in various studies in order to explore different biological properties of metabolic networks. For example, it has been shown that uniformly sampled flux vectors can give insights into regulatory properties of different reactions in the network. Additionally, uniform random sampling can be used for comparing in silico flux distributions with experimental data, when additional constraints is introduced to model, based on experimental results. Calculation of pairwise CC between pairs of fluxes is a useful method for measuring the dependencies between pairs of fluxes under different conditions.
Such dependencies can be used to explore the effect of enzyme dysregulation and also enzymopathies.

Immediately after the introduction of the two concepts of “flux coupling” and “flux correlation”, Papin and his colleagues used the term Co-Sets to describe reactions with perfectly, partially and directionally related fluxes. Moreover, they emphasized that Co-Sets can be computed by pathways analysis, FCA or by analysis of correlated fluxes. After this paper, several other studies used similar terminology.

Here, we emphasize that using the term “Co-Set” can be confusing, as flux coupling and flux correlation are two different concepts. To demonstrate this, we show that uncoupled, directionally coupled and partially coupled fluxes can be perfectly correlated/completely uncorrelated.

To be precise, reactions $i$ and $j$ are defined to be in a perfect Co-Set if they have a fixed flux ratio. As a result, the definition of perfect Co-Sets is more restrictive than the definition of perfectly correlated reaction sets, since $|R_{ij}| = 1$ does not necessarily guarantee a fixed flux ratio, e.g. when $v_i = v_j = 0$ is not included in the feasible flux space.

2. Materials and Methods

2.1. Computing Pearson CC between pairs of fluxes

In the present study, for finding the Pearson CC, between a pair of fluxes in each example network, COBRA toolbox was used to sample 100,000 uniform random vectors over the flux space. COBRA toolbox uses parallelized artificial center hit and run algorithm (ACHR) to sample the flux space. Briefly, an initial point is selected in the feasible space of flux distributions. Then, the algorithm generates “warm-up” points. These points are stored in a matrix, and a centroid, $s$, is calculated for it. In the next step, a random vector, $x_n$, is chosen in the matrix and a new point is generated by moving toward the direction of $(s - x_n)$. This point is substituted randomly in the matrix. The procedure of determining the centroid and generating the new point is iterated until the desired number of points is achieved.

The CC is assumed to be equal to zero if $R^2 < 10^{-5}$. In the analysis of the small example networks (see Figs. 1–3) the lower bound (respectively upper bound) of fluxes is 0 (respectively 2), except those reactions with fixed fluxes (shown by asterisk). Note that the results do not depend on the upper bounds of fluxes, in the sense that changing the upper bounds will only change the flux ranges (on the axes of the plots) or the position of the lines in Figs. 1(d), 2(d) or 3(d) (see Sec. 3).

2.2. Computing maximal information coefficients between pairs of fluxes

In general, it is possible to have two variables which depend on each other, but their “linear” correlation is zero. One possibility to find the interdependence of such variables is to consider the mutual information of them. Mutual information between two variables $A$ and $B$, denoted by $I(A, B)$, determines the dependence between two
variables by measuring the information that they share. Mutual information is defined by the following equation:

\[
I(A, B) = \sum_{i=1}^{M_A} \sum_{j=1}^{M_B} p(a_i, b_j) \log \left( \frac{p(a_i, b_j)}{p(a_i)p(b_j)} \right),
\]

where \(p(a_i, b_j)\) is the joint probability that \(A\) is in the state of \(a_i\) and \(B\) is in the state of \(b_j\). Intuitively, \(I(A, B)\) measures the reduction in uncertainty about one variable after observing the other. For detailed description of mutual information please see Ref. 22.

In the present work, we used Minerva package\(^{23}\) in R environment\(^{24}\) to compute maximal information coefficients (MIC, a means to apply mutual information) to examine reaction pairs which are found to be linearly uncorrelated based on computing Pearson CC. MIC is a metric for measuring the association between two variables which is based on mutual information. It can capture a wide range of associations. Informally speaking, it is supposed that if two variables are related there is a grid that can partition the scatter plot of these two variables, such that the relationship can be captured. Therefore, all possible grids are examined until a maximum for grid resolution is achieved. The largest mutual information value which can be obtained based on these grids is calculated and normalized. This statistic which is referred as MIC is a value between zero and one.\(^{25}\)

### 2.3. Computing flux coupling relations

We calculated the flux coupling relations by means of CFCA tool.\(^{26}\) This tool is basically an extension of the previously reported tools for computing flux coupling relations.\(^{27,28}\) The main difference is to add the possibility to consider lower and upper bounds for each reaction in the network, while the previous tools assume no constraints on the lower and upper bounds. We used SoPlex solver in the OPTI Toolbox\(^{29}\) (as an “exact solver”) to solve the linear programs required in FCA.

### 2.4. Analysis of genome-scale metabolic networks

In the present study, to compare flux coupling relation and flux CC between real-world reactions, we studied seven different genome-scale metabolic networks of five biologically important microorganisms (see Table 1 for a summary of general biological properties of these networks).

*Pseudomonas aeruginosa* and *Mycobacterium tuberculosis* are major life-threatening pathogens, and therefore, their reconstructed metabolic networks can be used in identification of novel drug targets.\(^{30,31}\) *Escherichia coli* and *Pseudomonas putida* are two of the most important bacteria in industry. *In silico* modeling of the metabolism of these microorganisms can be employed in metabolic engineering applications for production of desired metabolites and in bioremediation processes.\(^{32,33}\) *Saccharomyces cerevisiae*, the baker’s yeast, is extensively used for production of
food, pharmaceuticals and biofuels. Different versions of yeast metabolic network have been used in metabolic engineering and in evolutionary studies.\textsuperscript{34}

For each metabolic network we considered five different conditions: (i) fixing the flux of glucose uptake; (ii) fixing the flux of oxygen uptake; (iii) fixing the fluxes of glucose and oxygen uptake; (iv) fixing the flux of fructose uptake; and finally (v) fixing the fluxes of fructose and oxygen uptake. It should be noted that the M. \textit{tuberculosis} model does not include fructose, and therefore, for this network cases (iv) and (v) are not applicable. In each case, we computed the number of instances where:

- Reactions are directionally coupled but perfectly correlated,
- Reactions are directionally coupled but completely uncorrelated,
- Reactions are partially coupled but perfectly correlated,
- Reactions are partially coupled but completely uncorrelated,
- Reactions are uncoupled but perfectly correlated.

3. Results and Discussion

3.1. Correlation between uncoupled fluxes

Figure 1(a) presents an example of uncoupled fluxes, where CC is zero (Fig. 1(b)). It might not be surprising to have uncorrelated uncoupled fluxes. Figure 1(c) is probably more interesting. In this example, flux through reaction 1 is always equal to 1. As a result, fluxes 2 and 3 are uncoupled, because $v_2 = 1$ (i.e. a nonzero flux through reaction 2) implies $v_3 = 0$, and additionally, $v_3 = 1$ implies $v_2 = 0$. However, these fluxes are perfectly correlated (Fig. 1(d)).

3.2. Correlation between directionally coupled fluxes

In Fig. 2(a), flux 3 is directionally coupled to flux 2. To the best of our knowledge, it was previously believed that flux coupling guarantees flux correlation ($R^2 \neq 0$). Figure 2(b) proves that this is not correct, where 2 and 3 are completely uncorrelated. Another example of directional coupling is shown in Fig. 2(c). Here, $v_3 = 1$, which results in a perfect correlation between fluxes 1 and 2 (Fig. 2(d)).

Note that the diamond-like pattern in Fig. 2(b) is obtained because reaction 3 cannot carry any flux independent of the flux of reaction 2. If $v_3$ is active in the forward direction, i.e. it converts B to C, then we have these two constraints:

$$v_2 + v_3 = v_4,$$
$$v_2 = v_3 + v_5.$$

Since $v_4 \leq 2$, from the first equation it is inferred that sum of $v_2$ and $v_3$ is equal or less than 2. Hence, the points in Fig. 2(b) should be under the line $v_2 + v_3 = 2$. From the second equation it is inferred that $v_2$ should always be more than $v_3$. Consequently, sampled points must be located under the line $v_2 = v_3$. Therefore, the points
are restricted in a triangle made by these three lines: $v_3 = 0, v_2 + v_3 = 2$ and $v_2 = v_3$. When $v_3$ is active in the reverse direction, the same reasoning can explain the existence of the lower part triangle in Fig. 2(b).

### 3.3. Correlation between partially coupled fluxes

We show that even partially coupled fluxes $i$ and $j$ (with $v_i \neq 0$ if and only if $v_j \neq 0$) do not necessarily belong to a Co-Set. It can be easily shown that fluxes 1 and 4 in Fig. 3(a) are partially coupled. However, the stoichiometric coefficients of reactions 2 and 3 are chosen such that the CC between fluxes 1 and 4 becomes zero (Fig. 3(b)).

Similar to the uncoupling and directional coupling cases, it is possible to find examples of partially coupled fluxes which are perfectly correlated. In Fig. 3(c),
fluxes 1 and 2 are partially coupled. If $v_3 = 1$, however, a perfect correlation is found between fluxes 1 and 2.

### 3.4. Flux coupling versus flux correlation in real biological networks

In the previous sections, we observed, through a few examples, that flux coupling and flux correlation are inherently different concepts. In other words, except for the case of full coupling, which implies perfect correlation between two fluxes ($R^2 = 1$), there are no constraints on Pearson CC in case of any other type of coupling/uncoupling relations. In other words, Pearson CC can take any value between 0 and 1 in other cases. Here, we show that such examples (which will be referred to as “exceptional cases”) can also be found in genome-scale metabolic networks.
We calculated the flux coupling relations by means of CFCA tool. For calculation of CC between pairs of reactions we sampled 100,000 points for each model. The results are shown in Fig. 4. For all of the metabolic networks, we found some exceptional cases at least under one of the growth conditions. Some of these cases are rather infrequent, e.g. directionally coupled reactions with perfect correlation, uncoupled reactions with perfect correlation, or partially coupled reactions with zero correlation. However, in all networks, we observed a large number of directionally coupled reaction pairs which are not correlated.

Figure 5(a) shows an example of directional coupling and perfect correlation in iND750. Here, d-sorbitol dehydrogenase reaction is directionally coupled to hexokinase reaction. As the flux through fructose exchange is fixed, the fluxes of the two mentioned reactions become perfectly correlated.
Fig. 4. Logarithm (base 10) of number of exceptional cases in seven different genome-scale metabolic networks (Table 1) under different growth conditions. For each metabolic network, we considered five different conditions: (i) Fixing the flux of glucose uptake, (ii) fixing the flux of oxygen uptake, (iii) fixing the fluxes of glucose and oxygen uptake, (iv) fixing the flux of fructose uptake, and finally (v) fixing the fluxes of fructose and oxygen uptake. The five exceptional cases in each horizontal axis are: (i) Directionally coupled/Perfectly correlated, (ii) Directionally coupled/Uncorrelated, (iii) Partially coupled/Perfectly correlated, (iv) Partially coupled/Uncorrelated, (v) Uncoupled/Perfectly correlated. It should be noted that the *M. tuberculosis* model does not include “fructose”, and therefore, for this network conditions (iv) and (v) are not applicable.
We should note that we also observed some cases in which none of the fluxes are fixed, but the possible values for one of the fluxes are too small compared to the other fluxes. For instance, in Fig. 5(b) a small subnetwork of iJN746 is shown. Here, due to the stoichiometric constraints imposed by the network, dihydrouracil dehydrogenase reaction can only carry fluxes in range of $0 - 10^{-4}$, while the range for other two

![Fig. 5. (a) Hexokinase and D-sorbitol dehydrogenase are directionally coupled and fully correlated, if the fructose exchange flux is fixed. (b) Pyrimidine-nucleoside phosphorylase and uracil phosphoribosyl transferase are directionally coupled and fully correlated, if the flux through dihydrouracil dehydrogenase reaction lies in a tight range ($0 - 10^{-4}$). (c) Glucose dehydrogenase and hexokinase are uncoupled but perfectly correlated, if glucose exchange reaction has a fixed flux. (d) Diphosphorylase and adenylyl transferase are directionally coupled but completely uncorrelated.](image)
reactions is 0–1000. In this case, CC between flux of pyrimidine-nucleoside phosphoribosyltransferase and flux of uracil phosphoribosyl transferase is 1.

For the case of uncoupled reactions with perfect correlation, we can mention a case from iJN746 (Fig. 5(c)). Here, flux through the glucose exchange reaction is fixed. In this case, glucose dehydrogenase and hexokinase reactions are uncoupled but perfectly correlated.

Fixing a flux to a particular value may dictate the existence of nonzero fluxes through other reactions. As an example, in iND750, when we fix the flux through oxygen, the glucose exchange flux becomes nonzero. These two reactions are partially coupled by definition (as both reactions always have a nonzero flux, while their flux ratio is not a constant value). However, the Pearson CC between these fluxes is zero. As a different case, we should mention ATP hydrolysis in iMO1086. When oxygen uptake flux is fixed to a nonzero value, H₂O exchange flux and ATP hydrolysis flux (ATP + H₂O → ADP + Pᵢ + H⁺) are always nonzero. However, the flux ratio for these two reactions is not fixed, and therefore, H₂O exchange and ATP hydrolysis are partially coupled. In this case, the computed Pearson CC between fluxes is almost zero (or more precisely, 2 × 10⁻⁶).

Finally, we observed some directionally coupled reaction pairs in which none of the fluxes are fixed but the CC is almost zero. An example of such cases is indicated in Fig. 5(d). In this figure, Amidase, diphosphorylase and pyrophosphorylase are directionally coupled to adenylyl transferase. Possible flux ranges for these four reactions are 0–23, 0–0.0025, 0–13.2 and 0–23.1, respectively. However, the Pearson CC between diphosphorylase and adenylyl transferase fluxes is 2.5 × 10⁻⁶.

3.5. **On the meaning of “(un)coupled” and “(un)correlated” reactions**

In the present paper, we show that CCs equal to zero/one may exist in case of (partial or directional) coupling and uncoupling. This means that only full coupling of fluxes implies perfect correlation, and none of the other coupling relations imply flux correlation, or vice versa. Therefore, we suggest that the terms “perfect”, “partial” and “directional” Co-Sets are misleading and should be avoided.

From the biological point of view, directional flux coupling means that one metabolic enzyme must be expressed if another metabolic enzyme is to be expressed in the cell. Partial and full flux coupling can be seen as the mutual dependence of such metabolic enzymes. On the other hand, flux correlation does not necessarily imply any biological constraint on the expression of the metabolic enzymes. From the biological point of view, flux correlation simply reflects the existence of some interrelationship between the enzyme activities. Such an interrelationship is presumably due to co-occurrence of these enzymes on some common biochemical pathway.

Here, we discuss an example which reflect the biological importance of the difference between flux coupling and flux correlation. In a previous study, sampling of the flux space was used to examine the properties of the human mitochondrial metabolic network in different conditions.¹³"
glucose diet in which the uptake rate of fatty acids was shown to be increased. It has been shown that the Pearson CC of many pairs of Co-Sets (sets of perfectly correlated fluxes) are highly different between this condition and normal physiological condition. Here, we calculated the flux coupling relationships between these reaction pairs in both conditions. Figure 6 shows one of these cases. Here, flux coupling relationships of the reaction pairs were identical between the two conditions. In this figure, Co-Set 1 consists of reactions for import of (R)-3-hydroxybutanoate to the mitochondria and its conversion to acetoacetate. Co-Set 2 consists of reactions for import of hexadecanoate (n-C16:0) to the cell, its conversion to palmitoyl carnitine and entrance of this metabolite to the mitochondria. Based on FCA, in both conditions these sets are found to be uncoupled. This means that expression of the enzymes of these sets are not dependent to each other, i.e. each of which can be expressed in the absence of the other. However, the Pearson CC of these sets are highly different from one condition to the other.\(^{13}\) As it is mentioned, the uptake rate of hexadecanoate and other fatty acids are increased in the high fat-low glucose diet condition. Additionally, the presence of NAD and NADH is essential for the reactions of both sets to carry a nonzero flux. The increased rate of fatty acids uptake in high fat-low glucose diet, results in differences in the available amount of these cofactors between the two sets in each condition. Consequently, since correlation of fluxes depends on the enzyme activities which in turn depends on the available metabolites, the flux correlation between these sets is different in the two studied conditions.

3.6. Mutual information versus Pearson CC

In Figs. 1(b), 2(b) and 3(b), \(R^2 = 0\) is based on the Pearson CC between fluxes. However, the (nonlinear) dependence between flux values can be easily observed in these examples. For instance, as it is explained in Sec. 3.2, although the reactions in
Fig. 2(b) are not linearly dependent, the diamond-like pattern shows that they are not completely independent. In fact, this pattern is generated because $v_2$ and $v_3$ are related by some linear constraints. While Pearson correlation is probably the most popular measure of dependence, our examples suggest that other measures of dependence, e.g. mutual information$^{22}$ are probably more relevant for studying the dependencies between fluxes. To show this, using Minerva package$^{23}$ in R, we calculated MIC$^{25}$ (a means to apply mutual information) for the reaction pairs in Figs. 1(b), 2(b) and 3(b). The MIC values are found to be 0.07, 0.13 and 0.24, respectively. Therefore, the observable nonlinear dependence of fluxes can be reflected if other measures of dependence, like MIC, are used in studying the flux correlations. We should note that mutual information has been previously used when Pearson CC could not reflect the dependence between variables (see Ref. 38 as an example).

### 3.7. Conclusions

In the present work, we showed that FCA and flux correlation analysis represent two inherently different concepts, since the existence of one does not dictate the existence of the other. Even in the simple case of full coupling and perfect correlation, we showed that these two concepts are not equivalent, i.e. perfect correlation does not imply full coupling. Additionally, we suggest avoiding the terms perfect, partial and directional “Co-Set”. This is because partial or directional coupling may not necessarily imply correlation, as we showed in our examples.

In the present paper, we presented a couple of hypothetical examples, each of which reflects one of the exceptional cases mentioned in the text. Furthermore, we explored real-world biological networks to find such examples, in order to stress on the importance of our analysis. We presented explicit examples in real-world genome-scale networks which approve that flux coupling and flux correlation cannot be used interchangeably.

One should note that some parts of our results depend on the fixing of flux values of some reactions. For instance, in case of Fig. 2(c), if $v_3$ is not fixed then the perfect correlation between $v_1$ and $v_2$ is not obtained ($v_1$ versus $v_2$ plot will be a triangle-like shape, instead of a straight line). However, in a real biological network, fluxes of many reactions may not be fixed. Therefore, in Sec. 3.4, where we analyzed real networks, we only fixed the fluxes of one or two exchange reactions in each simulation. We showed that even in these conditions we can find several cases which indicate the difference of flux coupling and flux correlation. In addition, we showed that even when none of the fluxes are fixed, interesting examples can be still found, e.g. when one of the fluxes shows a very tight range of variability (e.g. between 0 and 0.0001).

Finally, we showed that mutual information is a better measure of dependence to be applied in studying flux dependencies. Pearson CC, which has been widely used in the analysis of metabolic networks, is a proper measure of linear dependence.
However, for capturing the nonlinear dependence between fluxes, which is observed in many cases such as Figs. 2(b) and 3(b), other measures of dependence may be more relevant. In the present work, we showed that calculating mutual information using MIC can successfully detect the dependence between those fluxes with nonlinear dependence whose Pearson CC is zero.

**Acknowledgments**

We are thankful to L. David and A. Bockmayr (Freie Universität Berlin) for their fruitful discussions. The CFCA tool was kindly provided by L. David. We would like to acknowledge the financial support of University of Tehran for this research under grant number 28791/1/2.

**References**


**Sayed-Amir Marashi** is an Assistant Professor at Department of Biotechnology, College of Science, University of Tehran. He finished his doctoral studies at the Freie Universität Berlin in 2011. His research interests include protein-protein interaction network analysis and modeling metabolic networks and pathways.

**Zhaleh Hosseini** is currently an MSc student in direct PhD program of Biotechnology at University of Tehran, Iran. Her research interests include systems biology, and more specifically, modeling and analysis of metabolic networks.