Organising metabolic networks: cycles in flux distributions

Maurício Vieira Kritz[1], Marcelo Trindade dos Santos[1], Sebastián Urrita[2], Jean-Marc Schwartz[3]

[1] LNCC/MCT, Av. Getúlio Vargas, 333, 25651-075, Petrópolis, RJ, Brazil

[2] Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Prédio do ICEx - Pampulha, 31270-010, Belo Horizonte, MG, Brazil

[3] Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester, M13 9PT, UK

Abstract

Metabolic networks are among the most widely studied biological systems. The topology and interconnections of metabolic reactions have been well described for many species, but are not sufficient to understand how their activity is regulated in living organisms. The principles directing the dynamic organisation of reaction fluxes remain poorly understood. Cyclic structures are thought to play a central role in the homeostasis of biological systems and in their resilience to a changing environment. In this work, we investigate the role of fluxes of matter cycling in metabolic networks. First, we introduce a methodology for the computation of

cyclic and acyclic fluxes in metabolic networks, adapted from an algorithm initially developed to study cyclic fluxes in trophic networks. Subsequently, we apply this methodology to the analysis of three metabolic systems, including the central metabolism of wild type and a deletion mutant of *Escherichia coli*, erythrocyte metabolism and the central metabolism of the bacterium *Methylobacterium extorquens*. The role of cycles in driving and maintaining the performance of metabolic functions upon perturbations is unveiled through these examples. This methodology may be used to further investigate the role of cycles in living organisms, their pro-activity and organisational invariance, leading to a better understanding of biological entailment and information processing.

Keywords: systems biology; organisation; flux; cycle.

1. Introduction

Biological systems are highly complex and dynamic by nature. From the scale of molecules to that of ecosystems, numerous components and processes interact, and these interactions create the biological functions that allow entities to live, reproduce and grow. The challenge of making sense of this complex organisation is not new, but it is becoming all the more crucial in the post-genome era. With the development of omics technologies and systems biology, large amounts of biological data are produced each day, using various experimental techniques. However the integration and interpretation of these data is proving to be very challenging and a large effort is needed in developing new methods for analysing and interpreting such complex data.

Metabolic networks are among the best characterised and most widely studied cellular interaction networks. The present availability of extensive data is allowing the construction of genome-scale metabolic networks for an increasing number of species, generally through a careful human-driven curation process (Feist et al., 2007; Heinemann et al., 2005; Herrgård et al., 2008; Ma et al., 2007). The topological properties of metabolic networks have been investigated in great details, revealing scale-free, modular and hierarchical properties (Jeong et al., 2000; Ravasz et al., 2002; Sales-Pardo et al., 2007).

These networks, however, primarily reflect our knowledge about the possible biochemical reactions in a given organism. The reactions and substrates that

compose them are not active all the time or present everywhere in the cell. Despite the rich knowledge already gained about the topology and connectivity of metabolic reactions, the principles regulating the dynamic activity of metabolic networks remain poorly understood. It is now widely accepted that the regulation of metabolic networks is distributed, and it is becoming ever clearer that reactions occur at different localisations and rates in a cell at any given time (Binder et al., 2008; Bluthgen & Platt, 2008; Fell & Poolman, 2008). The distribution of fluxes in a metabolic network cannot be understood by studying the properties of individual enzymes or rate-limiting steps, but it arises from the set of complex interactions between interconnected reactions, regulated at the transcriptional, translational, signalling and metabolic levels (Heinrich & Rapoport, 1974; Kacser & Burns, 1995; Rossell et al., 2005). So far, many efforts to understand the behaviour of large metabolic systems have taken a 'linear' view, essentially considering stoichiometrically consistent sets of reactions that link one or several source compounds to one or several products. Examples of such approaches include analyses by elementary modes, extreme pathways (Gagneur & Klamt, 2004; Papin et al., 2003; Schwartz & Kanehisa, 2006; Teixeira et al., 2007), as well as expansions of sets of source compounds and their metabolic scopes (Handorf et al., 2005; Raymond & Segrè, 2006).

Thus, the topology of metabolic networks is not sufficient. To improve our knowledge about the localisation of reactions and the distribution of substrate concentrations in cells, it is necessary to enhance our understanding about their

dynamic activity and their characteristics as living entities. However, the presently available methods still impose severe constrains on observing chemical activity distributed in space and time. One possibility for advancing our knowledge with respect to cell dynamics, then, is to investigate the distribution of flows that overlays the possible chemical interactions reflected by metabolic networks; that is, to search for knowledge about how much of a substrate present in a cell may be distributed among the reactions in its scope. What is the capacity of a metabolic network to retain and distribute substrate concentrations? How do fluxes split among the many pathways of a network and supply the substrates and energy needed by the cell at any given time? One manner of retaining substrates and making fluxes available is to keep them cycling.

Notwithstanding, cyclic structures have been often neglected in metabolic network studies. For a long time, metabolic cycles were characterised as 'futile', as it was thought that they could only result in unnecessary energy dissipation and should have been repressed by evolution (Rohwer & Botha, 2001; Schilling et al., 2000; Schuster et al., 2000). However, it is known that cyclic structures play a central role in the homeostasis of biological systems at several scales, as well as in their resilience and apt responses to environmental stimuli (Gleiss et al., 2001; Kun et al., 2008; Ma'ayan et al., 2008). This aspect has been investigated both in macroscopic and microscopic biological systems, but is far from being extensively addressed.

One feature distinguishing biological systems from physico-chemical systems is the nature of entailment. For a biochemical system the cause does not necessarily precede the effect in time (Wolkenhauer, 2001). Also, living entities embed all information required for their own functional activity, which is a necessary but not sufficient requirement for their *organisational invariance* (Cornish-Bowden & Cárdenas, 2007; Letelier, 2006). Cycles have been shown to play a major role in both embedding information and organisational invariance, since they disrupt the arrow of time. Thus, we ought to develop methods for analysing biological data from several perspectives in order to get a better understanding of living phenomena.

The concept of cyclic decomposition in networks was described in the context of trophic networks by Ulanowicz (1983). Metabolic networks, however, distinguish themselves from trophic networks in several manners. Aside the computational complexity of enumerating cycles in graph structures, there is the problem of interpreting and manipulating them properly in the context of metabolism. Our purpose here is to present a cyclic decomposition methodology for metabolic networks based on that of Ulanowicz, and to illustrate its relevance by applying it to the analysis of three examples of interest. This approach is expected to enhance our knowledge of cellular dynamics by decomposing a metabolic network, with a given flux distribution, into flux cycles and a residual acyclic flow graph.

We are working under the following premises, supported by non-quantitative observations, which may not be directly seen in the arguments but are subjacent to the whole approach. First, we are assuming that the available metabolic networks represent possible reactions and their interconnections, which may or not take place at a given steady-state. Second, reactions connected in the network may not be functionally related if the occur at different localisations. Third, the available data about metabolic fluxes reflect mean values over populations of cells that may be in different steady-states. Although they are not usually made explicit, these assumptions underlie the majority of current studies of metabolic networks.

The approach presented here allows for investigations about the organisation of metabolic networks based on the decomposition of a flux distribution into cyclic and acyclic fluxes. Each example reveals different properties of the decomposition and different manners of thinking the organisation of the cell. The decomposition algorithm and methodology are described in the next section. Examples and results obtained are presented in the third section. In the fourth section, we discuss this approach and some of its implications.

2. Methods and algorithms

The cycle decomposition algorithm consists of two phases. The first phase finds all existing cycles of a network; this is an NP-complete problem whose results do not depend, however, on any flux values. The second phase uses fluxes or other values associated to arcs to gradually extract the identified cycles from the graph, leaving

a residual acyclic graph in the case of open networks. A first distinction about metabolic and trophic networks is that the former are indeed hypergraphs while the second are graphs. This is circumvented here by considering the representation of hypergraphs as bipartite graphs and is discussed in the first subsection. The second subsection presents the details of our decomposition method and the last section discusses characteristics and other possibilities for inspecting the cycle and flux structure of a metabolic network.

a) Representation of metabolic networks

Strictly speaking, metabolic networks are hypergraphs, since reactions are in general associated with several substrates and products. They may be represented in at least three interchangeable forms. In the first form, metabolites are represented as nodes and the reactions as edges or arcs (which are directed edges) if reactions have a preferred direction. In the second form, reactions are depicted as nodes while metabolites are depicted as edges, which is the dual form of the first in terms of hypergraphs. In the third form, both metabolites and reactions are represented as two different types of nodes, and arcs connect them in accordance with biochemistry laws. The latter is essentially the representation of hypergraphs as bipartite graphs. The most general representation is the latest, the other two may be obtained from it (Figure 1). Moreover, there is a one to one association between cycles in each of these representations.

In the sequel, the directed bipartite graph representation will be used for metabolic networks. An arc from a metabolite into a reaction means that the metabolite is a substrate for the reaction, and an arc from a reaction into a metabolite means that the latter is a product of the reaction. If a reaction is reversible, arcs in both directions may be used. Arcs and nodes may be labelled with indicative values. Usually, metabolic networks have fluxes attributed to reactions and concentrations to metabolites. While employing the bipartite representation, we have migrated this information to the bipartite arcs by means of the stoichiometry of each reaction, in order to apply the decomposition method.

b) Fluxes and mass conservation

Since we are working in steady-state conditions, it is important that flux values and the decomposition algorithm conform to mass conservation laws. Mass particles flow from one reaction to another or are exchanged with the environment. Therefore, to apply the cycle decomposition methodology to metabolic networks, the values associated to arcs of the hypergraph should reflect conserved quantities.

To accomplish this we convert the molar flux $v(R)$ of each reaction R into mass fluxes associated to each arc, either incoming or outgoing, incident to R. An arc $'a'$ (or an edge $'e'$) and a node $'n'$ are said to be incident if $'n'$ is a node belonging to $'a'$. The conversion is done proportionally to the molar masses and

stoichiometric coefficients of each metabolite associated to the reaction, in the following manner.

Let $A_i, 1 \le i \le m,$ denote the substrates of reaction R and $B_j, 1 \le j \le p,$ denote the products of this reaction. Then, the mass flux $f(A_i)$ associated to substrate arc (A_i, R) is:

$$f(A_i) = a_i \times M(A_i) \times v(R), 1 \le i \le m,$$

where a_i is the stoichiometric coefficient of A_i in R, $M(A_i)$ is the molar mass of A_i, and $v(R)$ the molar reaction flux. Likewise, the mass flux of the product arc (B_i, R) of R is given by:

$$f(B_j) = b_j \times M(B_j) \times v(R), 1 \le j \le p,$$

where b_j is the stoichiometric coefficient of B_j in R, $M(B_j)$ is the molar mass of B_j, and $v(R)$ the molar reaction flux.

In a given metabolic model, cofactors do not necessarily need to be represented explicitly. In this case, fluxes through some reactions may be apparently unbalanced, because a part of the mass flux has been exported to or imported from the environment through cofactors. To cope with this apparent unbalance of mass flux we associate to a reaction node R a *gateway* (an arc and a node), that represents mass exchange with the environment, whenever required. Moreover, sequences of reactions may be represented as a single reaction R_s. In this case, all

co-factors exchanged in the sequence and not explicitly represented are summed up into a single gateway.

c) Computing cycles

We use Tarjan's algorithm (Tarjan, 1973) to solve the cycle enumeration problem for the direct bipartite graph representation of metabolic networks. Tarjan's algorithm requires as input a directed graph $\mathbf{G} = \{\mathsf{N}, \mathsf{A}\}$ with nodes enumerated from 1 to *n*, the number of elements in N, and an adjacency list $Adj(n)$ for each $n \in \mathsf{N}$ .The *adjacency list* $Adj(n)$ is a list containing all nodes n' for which $(n, n') \in \mathsf{A}$. A *path* P is defined as a sequence of arcs $(n_1, n_2), (n_2, n_3), ..., (n_{i-1}, n_i) \in \mathsf{N}$, such that the terminal node of an arc is the initial node of the next one. Paths will be represented, without loss of generality, by their set of nodes $p_j = (n_{j_1}, n_{j_2}, ..., n_{j_k})$. A path P is called *elementary* if all its nodes occur only once in P . An *elementary cycle* c_j is defined as an *elementary path* p_j in which the first node n_{j_1} and last node n_{j_k} coincide. The following description of a generic cycle finding algorithm justifies our choice of Tarjan's algorithm, that is fully described in Appendix A.

General searches for cycles in a graph can be performed by an unconstrained backtracking algorithm; this means exploring all possible elementary paths on the graph and verifying which paths are elementary cycles. Given $\mathbf{G} = \{\mathsf{N}, \mathsf{A}\}$ with its

nodes enumerated from 1 to n and its adjacency list $Adj(n)$, an unconstrained algorithm proceeds as follows:

Start from any given node n_i, chose an arc $a \in Adj(n_i)$ traversing from node n_i to node $n_h, i < h$. Continue traversing to another node $n_k, h < k$, via the adjacency list of n_h.

Whenever n_k is adjacent to n_i an elementary cycle $p_j = \left(n_{j_1}, n_{j_2}, ..., n_{j_k}\right)$ has been found and is enumerated.

Continue until there are no more subsequent nodes. Then return one node back, choosing another arc to traverse.

Stop when all elementary paths $p_j = \left(n_{j_1}, n_{j_2}, ..., n_{j_k}\right)$, such that $n_{j_{i-1}} < n_{j_i}$ for all $2 \le i \le k$ have being examined.

This basic procedure explores many more paths than necessary and has exponential computational complexity. For an efficient cycle enumeration there must be a pruning method to avoid futile searches. Tarjan's algorithm provides such an efficient pruning method (see a pseudo-code of the algorithm in Appendix A), theoretically requiring $O\left((N + A)(C + 1)\right)$ run time steps, where N, A and C are the total number of nodes, arcs and cycles, respectively. It is thus bilinear in

these preceding quantities. In the name of simplicity, the algorithm does not take into account graphs with self-loops or multiple arcs, conditions that are naturally satisfied by the bipartite representation of hypergraphs that reflect metabolic networks.

d) Network decomposition and residual acyclic graphs

The second phase of the method is the decomposition of the network by subtracting cycles based on the mass flux values up to a point where there are no more cycles to be subtracted. The algorithm proceeds as follows (Figure 2).

Let $C = \{c_0, c_1, c_2, ..., c_q\}$ be the set of elementary cycles resulting from phase 1, where $c_i = [a_{i0}, a_{i1}, ..., a_{ik_i}]$ for $0 \le i \le q$, and $a_{ij}, 0 \le j \le k_i$, are the arcs composing each cycle c_i. Then, the procedure is as follows:

Step 1. Find the *critical arc* ($\boldsymbol{ca}$) of C, which is defined as the arc with the minimum flux value $f(\boldsymbol{ca})$ among the arcs of all cycles in C. That is,

$$f(\boldsymbol{ca}) = \min_{0 \le i \le q} \min_{0 \le j \le k_i} f(a_{ij})$$

Step 2. Find the set $\mathbf{N}(\boldsymbol{ca})$ of elementary cycles in C that contain this critical arc $\boldsymbol{ca}$. The set $\mathbf{N}(\boldsymbol{ca})$ is called the *nexus* of $\boldsymbol{ca}$ and is a subset of C.

Step 3. Assign probabilities to each cycle in $\mathbf{N}(\boldsymbol{ca})$ as follows (Figure 3):

1. Let $a_{ij} = (n_{in}, n_{out})_{ij}$ be any arc of a cycle c_i in $\mathbf{N}(\boldsymbol{ca})$.

2. Define $P(a_{ij}) = f(a_{ij}) \div f_{in}(a_{ij})$, where $f(a_{ij})$ is the flux through arc a_{ij} and $f_{in}(a_{ij})$ is the total flux at its first node n_{in}. The ratio $P(a_{ij}) < 1$ designates the portion of flux entering the first arc node n_{in} and remaining in arc a_{ij}.

3. Assign to all cycles c_i in $\mathbf{N}(\boldsymbol{ca})$ the probability $P(c_i) = \prod_{0 \le j \le k_i} P(a_{ij})$.

The value $P(c_i)$ can be interpreted as the probability that a given mass amount *m* in cycle c_i flows through all arcs of this cycle, returning to the initial node; that is, the probability that *m* remains in the cycle. This sub-procedure distributes the flux of the critical arc $\boldsymbol{ca}$ among the cycles of nexus $\mathbf{N}(\boldsymbol{ca})$ according to the cycle probabilities $P(c_i)$.

Step 4. Each cycle in nexus $\mathbf{N}(\boldsymbol{ca})$ now has a flux value $f(c_i) = \mu \times P(c_i) \times f(\boldsymbol{ca})$, where $\mu = \left(\sum_i P(c_i)\right)^{-1}$ is a normalisation factor. The flux amount $f(c_i)$ of each cycle is then subtracted from the flux at all arcs a_{ij} in cycle c_i, for all cycles c_i in nexus $\mathbf{N}(\boldsymbol{ca})$; that is $f(a_{ij}) \leftarrow f(a_{ij}) - f(c_i)$ for all $0 \le j \le k_i$ and all c_i in $\mathbf{N}(\boldsymbol{ca})$.

After this subtraction, the flux of the critical arc ca in $\mathbf{N}(ca)$, $f(ca)$, becomes zero. The arc ca is then removed from the network and all cycles in the nexus $\mathbf{N}(ca)$ become open paths.

Step 5. If C is empty, **STOP**. Otherwise, restart from **Step 1**, with another critical arc ca and its nexus $\mathbf{N}(ca)$.

e) Key characteristics of the decomposition

This decomposition has the following characteristics:

- The enumeration of cycles of a network (graph) is unique and does not depend on flux values. Cycles are enumerated only once.
- The decomposition result, however, particularly the final acyclic graph, does depend on the values of fluxes.
- The heuristics that distributes the flux through the critical arc according to the probability of a given mass to remain on a cycle is meaningful in the case of metabolic networks, as much as for ecological networks.
- The heuristics employed reflects our current knowledge of metabolism. The final result, though, may depend on the choice of the heuristics (Ulanowicz, 1983).
- The sub-algorithm that associates probabilities to each cycle in a nexus depends on a choice of probability distribution that also reflects current knowledge; namely, that there is very little information about the distribution of substrate masses in a cell.

The choice of a heuristics essentially defines one algorithm. Other heuristics are possible but, given the presently available knowledge, the above solution is the most natural one. Therefore, the foregoing method is in fact a class of algorithms.

3. Results

We applied this cycle decomposition algorithm to three different examples of metabolic networks of growing complexity.

a) Central metabolism of *E. coli*

The first case under study is a model of the central metabolism of the bacterium *Escherichia coli* published by Kurata et al. (2007). The authors constructed a model that combines glycolysis, the pentose phosphate pathway and the tricarboxylic acid (TCA) cycle, and measured the metabolic steady-state fluxes in these pathways in both wild-type and pyruvate kinase knockout (*pykF*) mutant cells. In the latter, the pyruvate kinase reaction that links phosphoenolpyruvate (PEP) and pyruvate (PYR) is deleted. The decomposition in cycles of the network is shown for both wild-type (Figure 4) and *pykF* knockout mutant (Figure 5). All reactions in these figures are colour coded to indicate the intensity of flux carried by reactions.

As expected, the cycle enumeration algorithm identified 16 cycles in both cases. A comparison of fluxes of individual reactions clearly shows that the flux in the

pyruvate kinase reaction (R4) is depleted in the mutant, but it is difficult to assess the effect of the deletion on the global organisation of fluxes by considering only individual fluxes. The cycle decomposition however reveals several additional properties. First, the structure of the acyclic graph is unaffected by the deletion; the cell maintains its global growth regime, continuing to process glucose into biomass compounds and energy. Second, the intensity of fluxes changes in parts of the acyclic graph, because the deletion of pyruvate kinase results in a reduction of acyclic flux in the entire branch from glucose-6-phosphate (Glc6P) to pyruvate (PYR). Third, the inspection of the set of cycles reveals that most of them maintain the same flux level in the wild-type and mutant. A notable exception is the cycle running through glucose-6-phosphate (Glc6P), fructose-6-phosphate (Fru6P), glyceraldehyde-phosphate (GAP) and phosphoenolpyruvate (PEP) (Figure 5b). This cycle does not contain the mutated reaction and yet, interestingly, its activity has decreased by a factor of 12 as a result of the pyruvate kinase mutation. The quantification of cyclic mass fluxes thus reveals a more fundamental disturbance in the cell's functional organisation than simply a decrease of flux in an individual branch. The recycling of matter from phosphoenolpyruvate to glucose-6-phosphate is the fundamental engine driving glycolysis and allowing it to produce energy with a limited input of additional glucose. When this recycling process is hampered, the efficiency of the cell's metabolism is fundamentally altered, since larger amounts of new glucose have to be imported to maintain the same metabolic activity. This example illustrates how the analysis of cyclic mass fluxes is able to cast new light on the organisation of cellular processes.

b) Erythrocyte metabolism

We applied the same algorithm to a model of central erythrocyte metabolism built by Holzhütter (2004), which contains glycolysis and the pentose phosphate pathway (Figure 6a). In contrast to the previous example, all cofactors were explicitly represented in this example. There were 848 cycles identified by the enumeration algorithm. The decomposition reveals that the cycles carrying the highest flux values are indeed those involving cofactors: in this case the NAD/NADH cycle and the ATP/ADP cycle. Almost all cycles carrying significant fluxes contain at least one of these four cofactors. The only exception is the erythrose-4-phosphate/glyceraldehyde-phosphate cycle. The acyclic graph shows one dominant route carrying a large amount of flux, which runs from glucose to lactose.

These observations raise some important points about the role of cofactors in metabolic networks. It is well known that cofactors are essential energy providers to metabolic reactions (Morowitz & Smith, 2007). These molecules are usually heavier than small metabolites; it is thus not surprising that they carry the highest flux of matter. As already shown by the example of the pyruvate kinase deletion mutant, this observation reinforces the fact that recycling of matter is an efficient way to drive cellular processes at minimal expenses, since it reduces the amount of new compounds needed to be input into the system to keep cellular metabolism running. At the same time, this result raises the question of whether mass is the best indicator in terms of biomass output and energy production of a metabolic

network. While larger molecules in principle have a higher potential to provide energy and elementary molecules for cellular anabolism, there is no absolute dependency between the two. Intense cofactor cycles may obscure other cyclic processes present in cellular activity. Depending on the cellular process under investigation, it may be instructive to distinguish between different levels of cyclic activity and to represent this by means of a proper model of organisation.

c) Central metabolism of *Methylobacterium extorquens*

Our third example is a model of the central metabolism of *Methylobacterium extorquens* AM1 presented by Holzhütter (2004). The model covers the pathways of formaldehyde metabolism, glycolysis and gluconeogenesis, tricarboxylic acid (TCA) cycle, pentose phosphate shunt, serine cycle, poly b-hydroxy butyrate synthesis, respiration and oxidative phosphorylation of the bacterium (Figure 7a). The distribution of fluxes was calculated by Holzhütter (2004) relying upon the principle of flux minimisation and subsequently validated by 13C label tracing and mass spectroscopy measurements. Cofactors were not explicitly represented in this example. In this case, 16 cycles were enumerated by the algorithm. This model is significantly larger than the previous two examples (78 fluxes and 77 metabolites), yet the computation of cycles could still be carried out in a few seconds on a common desktop computer. If cofactors were to be included however, the number of cycles would rise over two million and the enumeration algorithm would need several hours to complete.

The two cycles carrying the largest of flux values are the tetrahydromethanopterin (H4MPT) cycle and the tetrahydrofolate (H4F) cycle. They correspond to two pools of folate that drive the metabolism of the bacterium (this metabolism processes formaldehyde produced out of methanol). Interestingly, the acyclic graph also shows an intense flux carried from acetoacetyl-CoA to succinate-CoA, entering and exiting the system via cofactors; the cofactor entering via R46 is acetyl-CoA, the cofactor exiting via R27 is CoA. This branch constitutes in fact the main part of a cycle, which could be closed by the pyruvate dehydrogenase reaction transforming pyruvate and CoA into acetyl-CoA. However, this reaction carries no flux in the observed distribution, effectively breaking the cycle that would recycle CoA into Acetyl-CoA. The bacterium is thus apparently consuming acetyl-CoA without replacing it from internal carbon sources, heavily relying on external sources of Acetyl-CoA. This observation casts doubts onto whether the flux distribution under consideration is biologically viable.

4. Discussion

As the reductionist approach that has dominated biology until now is progressively being complemented by a more integrated understanding of biological systems, cyclic structures are thought to play a more fundamental role in the organisation and origin of life than previously thought. Cycles of chemical reactions are thought to be one of the determining characteristics of living systems (Cornish-Bowden & Cárdenas, 2008). Ordered cycles are also believed to contribute to dynamic stability (Ma'ayan et al., 2008). Cycles help keeping the organisational

characteristics of a system invariant. It is important to note that the cycles considered in this study are not stoichiometrically closed. Stoichiometric cycles, which have been described in other works (Schilling et al., 2000; Wright & Wagner, 2008), represent closed sets of chemical reactions that do not exchange matter or energy with their environment. Such cycles are believed to be thermodynamically unfeasible. The cycles considered here on the contrary represent cyclic flows of mass transferred between different molecules. Even though the flow of mass is conserved within each cycle, several cycles may overlap, exchanging mass with each other. They are driven by external sources of mass and energy, which may enter a cycle in the form of a certain molecular species and leave it under a different form. A classical example of mass cycle in ecology is the carbon cycle, which provides a representation of carbon exchanges between the biomass, the ocean and the atmosphere; carbon atoms are embedded into different molecular forms in each part of the cycle. Similarly, mass cycles in metabolism represent flows of matter that are reorganised by living organisms into different chemical forms, while participating in different metabolic processes and being exchanged between different molecules.

The inclusion of cofactors drastically influences the number of cycles in a network and the applicability of Tarjan's algorithm and this decomposition method. The enumeration of cycles is theoretically of order $O((N+A)(C+1))$ in time, where N, A and C are the total number of nodes, arcs and cycles of a graph $\mathbf{G}$, respectively. Because of their ubiquity as metabolites in biochemical reactions, a

single pair of cofactors like ATP/ADP may be attached to many functionally unrelated reactions and add thousands of arcs to a metabolic network. This leads to a considerable increase in the number of network cycles, that do not necessarily correspond to occurring cycles of biochemical reactions. If cofactors are filtered from the complete network, our method may also be applied to genome-scale models; otherwise, it would require large scale computing resources or additional refinements, e.g. a parallelisation procedure. We however believe that a more fruitful way to extend this methodology to complete models at the genome-scale would be to find biologically grounded methods to gradually and selectively include cofactors and repeat the decomposition in an iterative manner. A related approach to tackle genome-scale models may consist in a hierarchisation of the network representation and decomposition. Biologically related subparts of the network may be condensed into reaction-like nodes at a higher level of representation, enabling cycles to be determined at different levels of this hierarchy. However the question of ubiquitous metabolites that may interact at different levels remains to be solved.

The consideration of spacio-temporal information offers a perspective for solving such problems. As already noted in the introduction, the localisation of reactions is also of great importance to the comprehension of cellular organisation and biochemical flows. Till now it has been challenging to both obtain and embed this information into models. Nevertheless, there are indications that reactions associated in a metabolic network may occur in different places inside a cell

(Binder et al., 2008). Therefore, substrates attached to each reaction in a metabolic network may occupy different cellular compartments or even specific regions of space within a single compartment. Systems of equations associated to metabolic reactions describe the overall dynamical behaviour of many instances of reactions of the same type and represent universal conservation laws. To render their localisation explicit would require information about space-time distributions and fluctuations, for which data are largely unavailable. Such information may nevertheless lead to important progress in our understanding of cellular organisation in the future.

5. Conclusion

Systems are precise, formal whenever possible, descriptions of an object of study. A system is not a model but a step towards it. In physics and chemistry, a system is primarily attached to the choice of a region in space-time and parameter space where the phenomenon of interest occurs. System biology focuses on the description of the elements intervening in the phenomenon and their interactions. In many senses it is an outcome (Kitano, 2000) or revival (Wolkenhauer, 2001) of General Systems Theory, which is also associated with circuits, signals, networks, observability and control. There are thus two conceptions of a system: that associated to space and time and that associated to elements and their interactions.

These two concepts are facets of the same thing. Components of a general system need to be close together to interact, while chemical and biological components only interact when they are of the appropriate type, even when occupying a sufficiently small neighbourhood in space or colliding. Concepts inherited from both approaches must be taken into account when interpreting biological results. Reaction networks typically reflect connections between reacting substrates. They contain intensive information about possible interaction among the many substrates. They conceal extensive information about where these substrates react within the cell and what percentage of the total volume of each is performing a given reaction. Numbers associated to network arcs or reaction nodes only reflect a mean, instantaneous state, usually related to steady-state regimes.

In this work we presented a methodology for studying the role of cycles in the organisation of mass fluxes in metabolic networks. Once a network is properly represented, the algorithm unveils cyclic and acyclic flows of matter through the network, leading towards a joint treatment of both system perspectives. This methodology was applied to three metabolic network models, showing that it unveils how disturbances in flux distributions due to perturbations, like mutations and environmental changes, affect the biochemical behaviour of the cell. These effects could not be identified only by inspecting the original graph and flux distribution. This methodology can be used to further investigate the importance of cycles in living organisms, their pro-activity and organisational invariance,

leading to a better understanding of biological entailment and information processing.

6. Acknowledgements

We would like to gratefully thank the PCI/LNCC/MCT program, under contracts number 170089/2008-8 and 170114/2008-2, for financial support. MVK, MST and JMS conceived and performed the research; SU implemented Tarjan's algorithm in C++. All authors read and approved the final manuscript.

7. References

- Binder, B., Goede, A., Holzhütter, H.G., 2008. De novo formation of organelles in time and space. ECMTB 08 – European Conference on Mathematical and Theoretical Biology (Edinburgh, 29 June - 4 July 2008).
- Bluthgen, N., Platt, R., 2008. What makes a good oscillator? ECMTB 08 – European Conference on Mathematical and Theoretical Biology (Edinburgh, 29 June - 4 July 2008).
- Cornish-Bowden, A., Cárdenas, M.L., 2007. Organizational invariance in (M,R)-systems. Chem. Biodivers. 4, 2396-2406.
- Cornish-Bowden, A., Cárdenas, M.L., 2008. Self-organization at the origin of life. J. Theor. Biol. 252, 411-418, doi:10.1016/j.jtbi.2007.07.035.
- Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for

1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3, 121, doi:10.1038/msb4100155.

- Fell, D.A., Poolman, M.G., 2008. Modelling the photosynthetic Calvin cycle. ECMTB 08 – European Conference on Mathematical and Theoretical Biology (Edinburgh, 29 June - 4 July 2008).
- Gagneur, J., Klamt, S., 2004. Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinformatics 5, 175, doi:10.1186/1471-2105-5-175.
- Gleiss, P.M., Stadler, P.F., Wagner, A., Fell, D.A., 2001. Relevant cycles in chemical reaction networks. Adv. Complex Syst. 4, 207-226.
- Handorf, T., Ebenhöh, O., Heinrich, R., 2005. Expanding metabolic networks: scopes of compounds, robustness, and evolution. J. Mol. Evol. 61, 498-512, doi:10.1007/s00239-005-0027-1.
- Heinemann, M., Kümmel, A., Ruinatscha, R., Panke, S., 2005. In silico genome-scale reconstruction and validation of the *Staphylococcus aureus* metabolic network. Biotechnol. Bioeng. 92, 850-864.
- Heinrich, R., Rapoport, T.A., 1974. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. Eur. J. Biochem. 42, 89-95.
- Herrgård, M.J., Swainston, N., Dobson, P., Dunn, W.B., Arga, K.Y., et al., 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat. Biotechnol. 26, 1155-1160.

- Holzhütter, H.G., 2004. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. Eur. J. Biochem. 271, 2905-2922, doi:10.1111/j.1432-1033.2004.04213.x.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.L., 2000. The large-scale organization of metabolic networks. Nature 407, 651-654.
- Kacser, H., Burns, J.A., 1995. The control of flux. Biochem. Soc. Trans. 23, 341-366.
- Kitano, H., 2000. Perspectives on systems biology. New Generation Computing 18, 199-216.
- Kun, Á., Papp, B., Szathmáry, E., 2008. Computational identification of obligatorily autocatalytic replicators embedded in metabolic networks. Genome Biol. 9, R51, doi:10.1186/gb-2008-9-3-r51.
- Kurata, H., Zhao, Q., Okuda, R., Shimizu, K., 2007. Integration of enzyme activities into metabolic flux distributions by elementary mode analysis. BMC Syst. Biol. 1, 31, doi:10.1186/1752-0509-1-31.
- Letelier, J.C., Soto-Andrade, J., Guíñez Abarzúa, F., Cornish-Bowden, A., Cárdenas, M.L., 2006. Organizational invariance and metabolic closure: analysis in terms of (M,R)-systems. J. Theor. Biol. 238, 949–961, doi:10.1016/j.jtbi.2005.07.007.
- Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O., Goryanin, I., 2007. The Edinburgh human metabolic network reconstruction and its functional analysis. Mol Syst Biol 3, 135.

- Ma'ayan, A., Cecchi, G.A., Wagner, J., Ravi Rao, A., Iyengar, R., Stolovitzky, G., 2008. Ordered cyclic motifs contribute to dynamic stability in biological and engineered networks. Proc. Natl. Acad. Sci. U.S.A. 105, 19235-19240.
- Morowitz, H., Smith, E., 2007. Energy flow and the organization of life. Complexity 13, 51-59.
- Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A., Palsson, B.Ø., 2003. Metabolic pathways in the post-genome era. Trends Biochem. Sci. 28, 250-258.
- Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.L., 2002. Hierarchical organization of modularity in metabolic networks. Science 297, 1551-1555.
- Raymond, J., Segrè, D., 2006. The effect of oxygen on biochemical networks and the evolution of complex life. Science 311, 1764-1767.
- Rohwer, J.M., Botha, F.C., 2006. Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data. Biochem. J. 358, 437-445.
- Rossell, S., van der Weijden, C.C., Lindenbergh, A., van Tuijl, A., Francke, C., Bakker, B.M., Westerhoff, H.V., 2006. Unraveling the complexity of flux regulation: a new method demonstrated for nutrient starvation in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U.S.A. 103, 2166–2171.
- Sales-Pardo, M., Guimera, R., Moreira, A.A., Amaral, L.A., 2007. Extracting the hierarchical organization of complex systems. Proc. Natl. Acad. Sci. U.S.A. 104, 15224-15229.
- Schilling, C.H., Letscher, D., Palsson, B.Ø., 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic

function from a pathway-oriented perspective. J. Theor. Biol. 203, 229-248, doi:10.1006/jtbi.2000.1073.

- Schuster, S., Fell, D.A., Dandekar, T., 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat. Biotechnol. 18, 326-332.
- Schwartz, J.M., Kanehisa, M., 2006. Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Bioinformatics 7, 186, doi:10.1186/1471-2105-7-186.
- Tarjan, R.E., 1973. Enumeration of the elementary circuits of a directed graph. SIAM J. Comput. 2, 211-216.
- Teixeira, A.P., Alves, C., Alves, P.M., Carrondo, M.J.T., Oliveira, R., 2007. Hybrid elementary flux analysis/nonparametric modeling: application for bioprocess control. BMC Bioinformatics 8, 30.
- Ulanowicz, R.E., 1983. Identifying the structure of cycling in ecosystems. Math. Biosci. 65, 219-237.
- Wolkenhauer, O., 2001. Systems biology: the reincarnation of systems theory applied to biology? Brief. Bioinformatics 2, 258–270.
- Wright, J., Wagner, A., 2008. Exhaustive identification of steady state cycles in large stoichiometric networks. BMC Syst. Biol. 2, 61, doi:10.1186/1752-0509-2-61.

Appendix A

We here present a pseudo-code describing Tarjan's algorithm (Tarjan, 1973). Given a graph **G** with nodes $\mathbf{n_i}$, where $\mathbf{1 \leq i \leq N}$, and the adjacency lists **A(i)** for each node, the algorithm searches the paths in **G** for cycles starting from any node **s**. The path **p** currently being considered in the search is stored on a **path_stack** that has **s** as its bottom element. Any other node **j** of **G** entering the path **p** satisfies **s<j**. Another stack, named **marked_stack,** stores a flag. A vertex **I** at the top of **path_stack** is "**marked**" if (1) it belongs to the elementary path **p** (see subsection **2.c**) or (2) if every other possible elementary path connecting **i** to **s** intersects **p** at a node different from **s**.

Input:

A graph **G** of size **n**, given by an array **A** of adjacency lists.

Restriction 1:

For each node index **s**, the algorithm generates elementary paths starting at **s** containing no nodes with an index smaller than **s** (**s<i**).

Restriction 2:

Once a node **i** has been used in a path **p** it can only be used in another path if

1. it has been removed from stack **path_stack** and
2. it has been removed from stack **marked_stack**.

A node **i** becomes unmarked when a path from **i** to **s** is found, such that it does not intersect **p** in any node other than **s**. This restriction drastically reduces the search space.

Output:

If the top node index **i** of the stack is adjacent to its bottom node with index **s,** **path** is returned, containing an enumerated cycle.


```
Procedure CYCLE_ENUMERATION (integer n, array of lists A(1:n)) {

    Procedure BACKTRACK (integer n, boolean f) {
        boolean g;
        f := false;
        # place n on path_stack
        path_stack(n) := true;
        # place n on marked_stack
        marked_stack(n) := true;
        foreach w in A(n) {
            if w < s {
                delete w from A(n);
            }
            else if w=s {
                f := true;
```

```
                return path_stack with an enumerated cycle
            }
            else if not marked_stack(w) {
                BACKTRACK (w, g);
                f := f || g;
            }
        }
        If f=true {
            pop marked_stack until top of marked_stack = n;
        }
        delete n from marked_stack
        marked_stack(n) := false;
        # end of BACKTRACK
}

# start the enumeration of cycles
for (i:=1 until n) {
        marked_stack(i) := false;
}
for (s:=1 until n) {
        BACKTRACK(s, flag);
        delete all nodes from marked_stack;
}
```

```
}
```

Figure legends

Figure 1: Bipartite representation of metabolic networks. The figure represents the network given by (i) R1: A+B->C; (ii) R2: B+C->D; (iii) R3: D->F.

Figure 2: Decomposition algorithm. See detailed explanations in the Methods section.

Figure 3: Probability assignment to arcs and cycles. As an illustration, considering the nexus $N = \{C_1, C_2, C_3\}$ the probability for arc a_{11} is calculated as follows: $P(a_{11}) = f(a_{11}) / (f(a_{11}) + f(a_{21}) + f(a_{31}) + f(a_j))$. Thus, $P(C_1) = P(a_{10})*P(a_{11})*P(a_{12})*P(a_{13})$. $P(C_2)$ and $P(C_3)$ are calculated in the same way. As a result, the proportions of the critical arc flux $f(a_{10})$ to be subtracted from each cycle in the nexus N are determined.

Figure 4: Decomposition in cycles of a model of the central metabolism of *Escherichia coli* (wild-type). Cofactors are not explicitly represented in this model and are indicated by yellow triangles. The colour of each reaction indicates the mass flux it carries. The full set of cycles is represented on the right-hand side, where the colour indicates the flux value carried by each cycle.

Figure 5: Decomposition in cycles of a model of central metabolism of *Escherichia coli* (*pykF* knockout mutant). Cofactors are not explicitly represented in this model

and are indicated by yellow triangles. The colour of each reaction indicates the mass flux it carries. The full set of cycles is represented on the right-hand side, where the colour indicates the flux value carried by each cycle.

Figure 6: Decomposition in cycles of a model of erythrocyte metabolism. All cofactors are explicitly represented in this model. The colour of each reaction indicates the mass flux it carries. Only cycles carrying the highest flux are represented on the right-hand side, where the colour indicates the flux value carried by each cycle.

Figure 7: Decomposition in cycles of a metabolic model of *Methylobacterium extorquens*. Cofactors are not explicitly described in this model and are indicated by yellow triangles. The colour of each reaction indicates the mass flux it carries. Only cycles carrying the highest flux are represented on the right-hand side, where the colour indicates the flux value carried by each cycle.

Figure 1

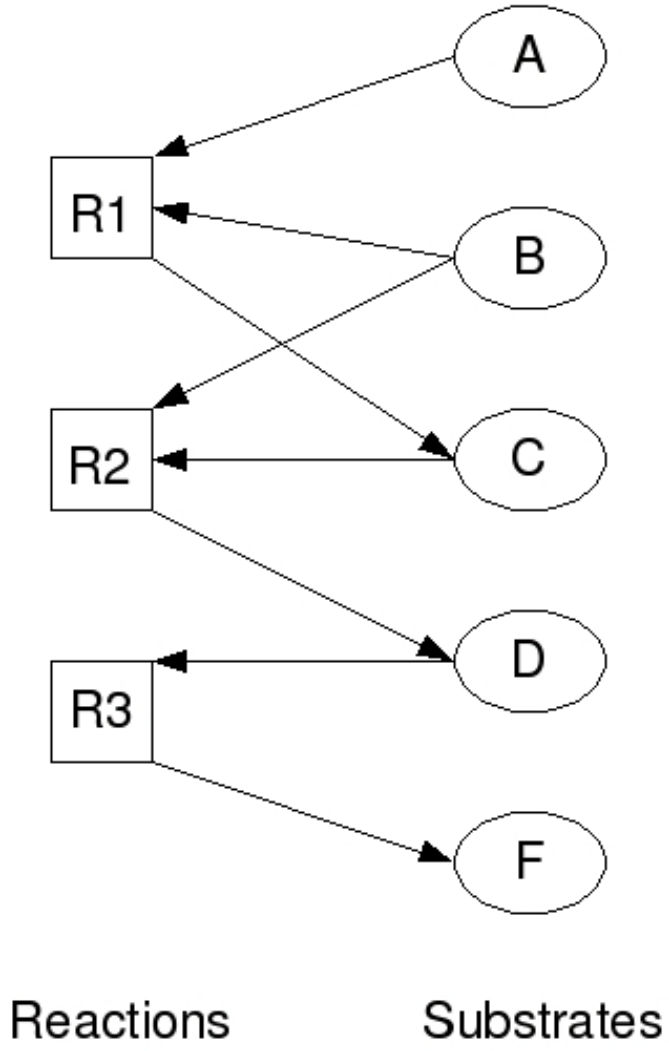

Figure 2

Decomposition Algorithm

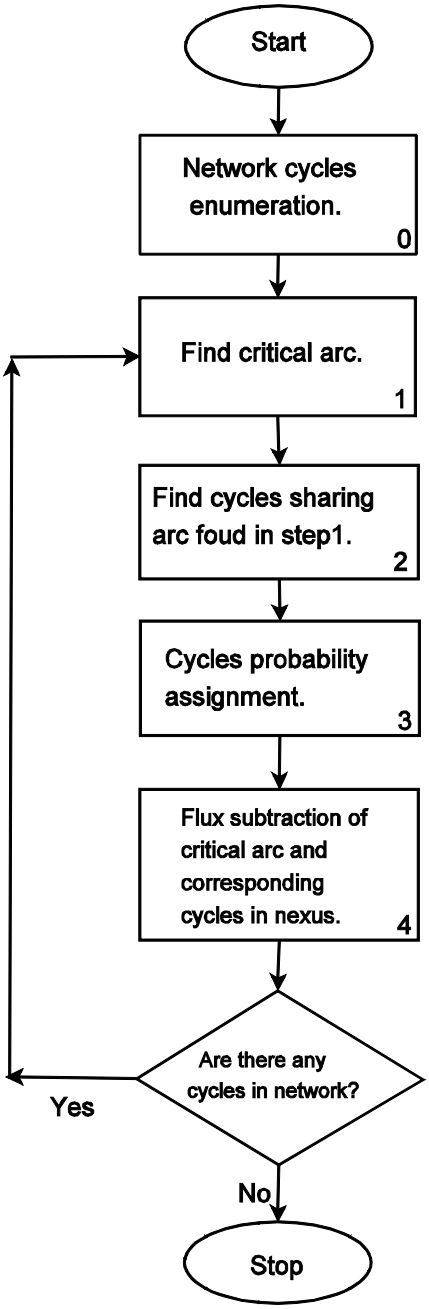

Figure 3

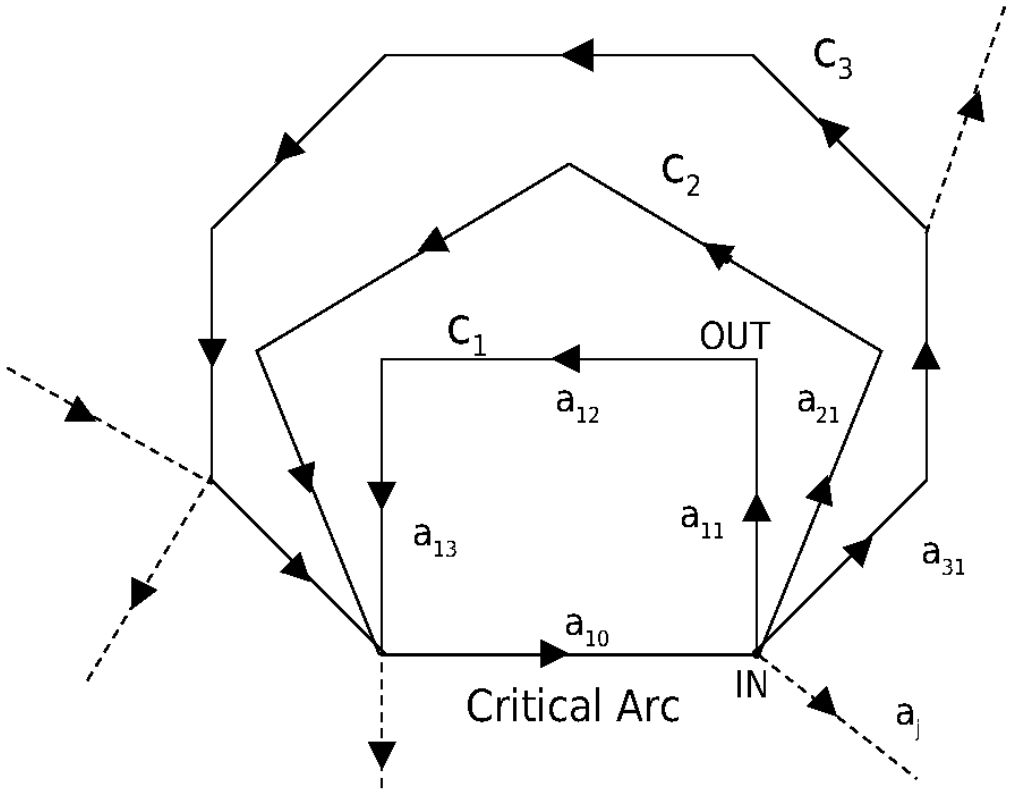

Figure 4

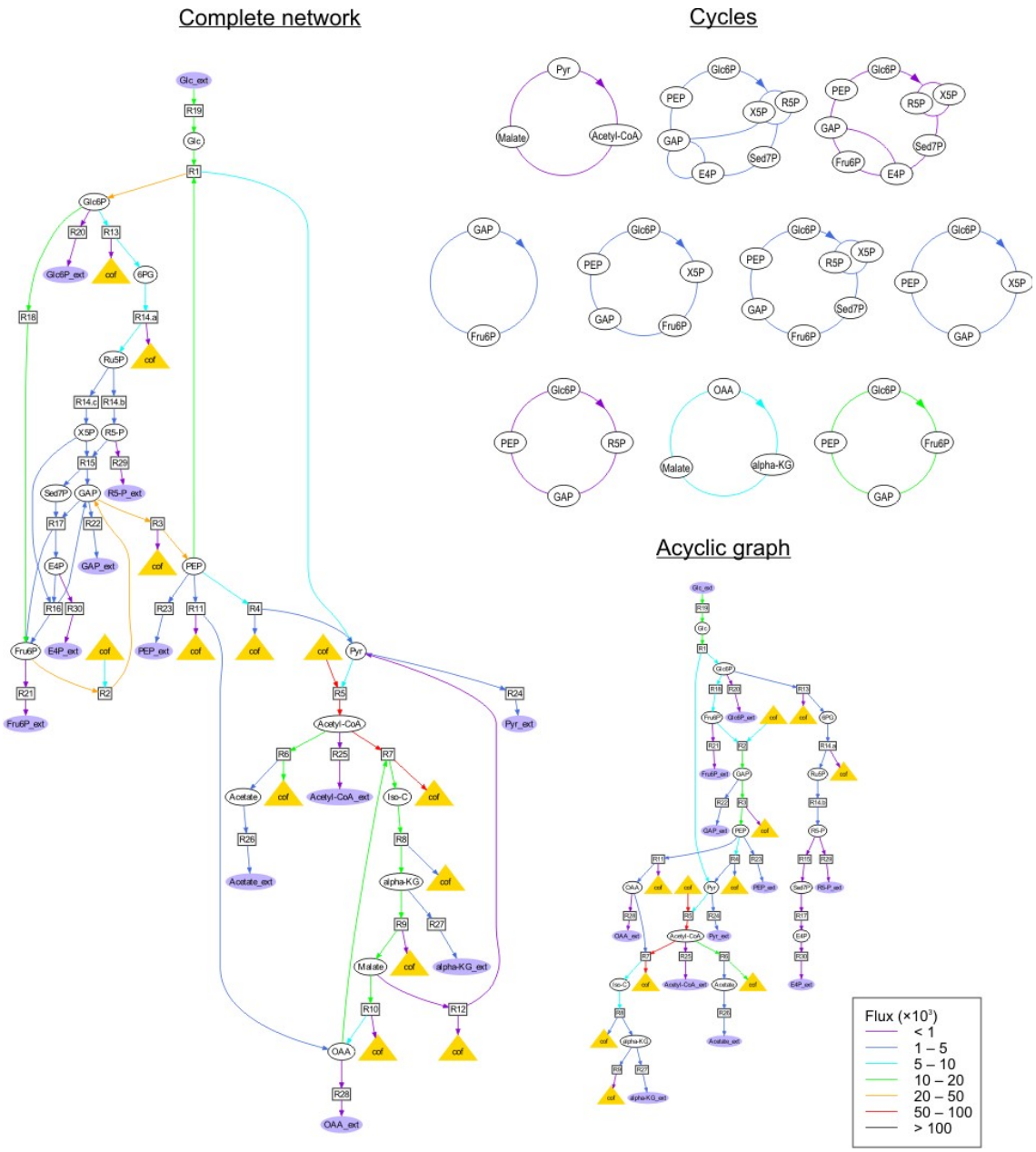

Figure 5

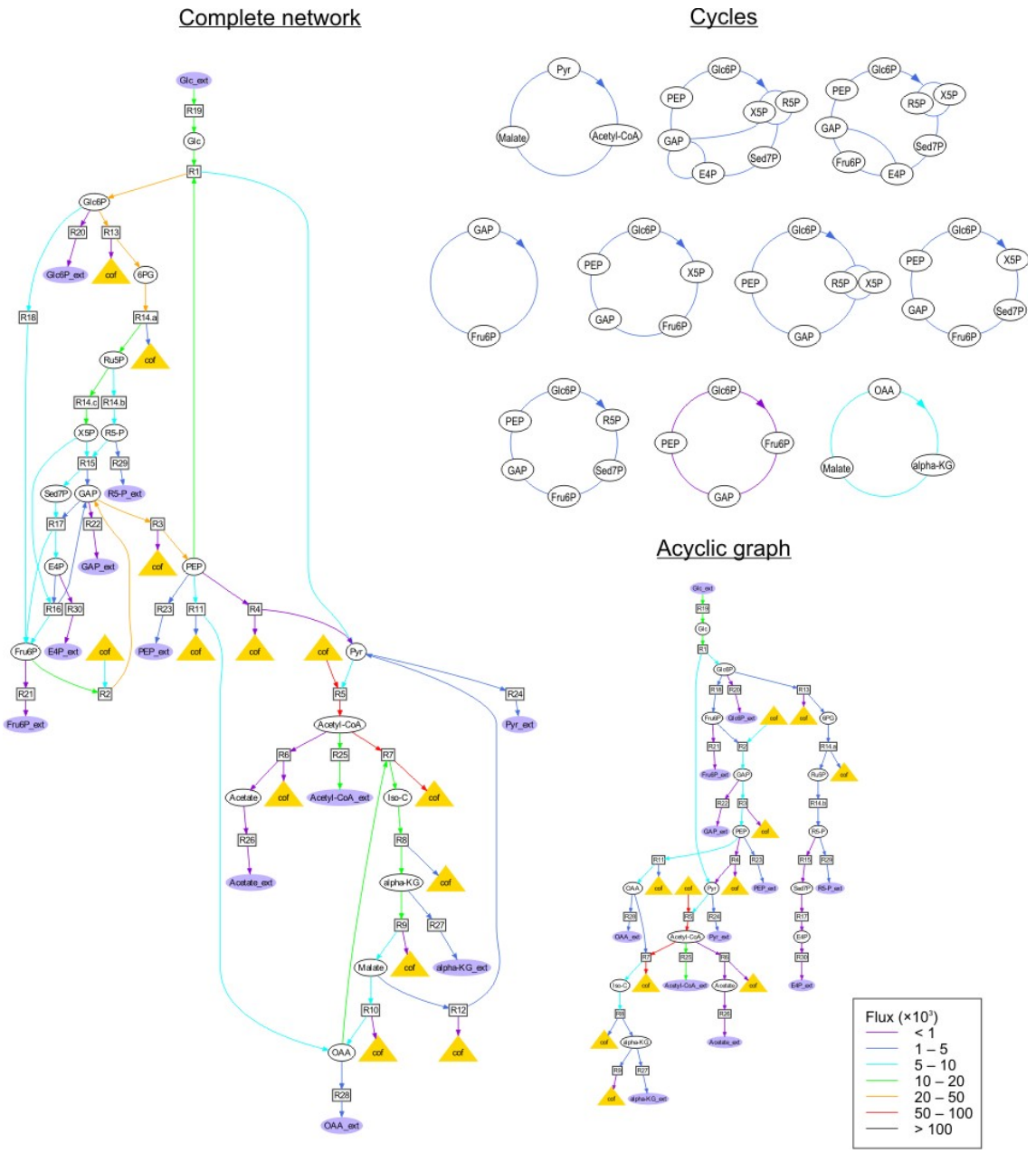

Figure 6

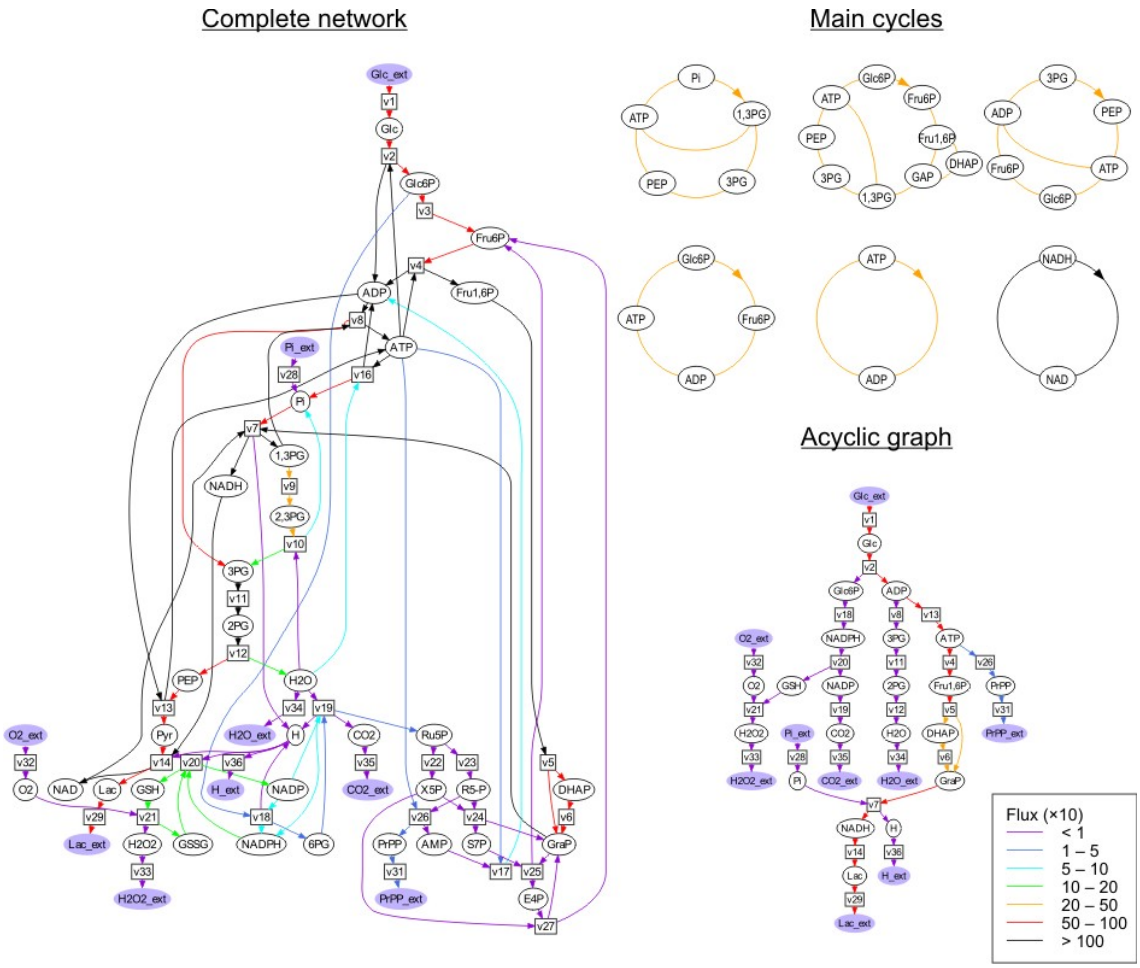

Figure 7