

METHOD

Open Access

Path finding methods accounting for stoichiometry in metabolic networks

Jon Pey¹, Joaquín Prada¹, John E Beasley^{2*} and Francisco J Planes^{1*}

Abstract

Graph-based methods have been widely used for the analysis of biological networks. Their application to metabolic networks has been much discussed, in particular noting that an important weakness in such methods is that reaction stoichiometry is neglected. In this study, we show that reaction stoichiometry can be incorporated into path-finding approaches via mixed-integer linear programming. This major advance at the modeling level results in improved prediction of topological and functional properties in metabolic networks.

Background

The use of graph theory in the analysis of biological networks has been extensive in the past decade [1]. Particularly, in metabolic networks different relevant topics have been examined using the rich variety of graph-theoretic concepts, ranging from topological properties [2-5], evolutionary analysis [6-8], pathway analysis [9-13], transcriptional regulation [14-16], functional interpretation of 'omics' data [17-20] and prediction of novel drug targets [21-23].

Graph-based methods start by converting the metabolic network into an appropriate graph. Different representations are possible here: i) metabolite graphs, where nodes are metabolites and arcs represent reactions linking an input and output metabolite; ii) reaction graphs, in which nodes are reactions and arcs represent intermediate metabolites shared by reactions; iii) bipartite graphs, where nodes are reactions and metabolites, while arcs link metabolites to reactions (for substrates) and reactions to metabolites (for products). Note here that each type of graph can be either directed or undirected. A deeper introduction to such graphs can be found in Deville *et al.* [24].

Importantly, graph-based methods rely on the definition of connectivity based on paths, that is, two nodes in the graph are connected (or not) depending upon whether (or not) we have a path linking them. This

definition of connectivity is debatable, however, particularly when it is claimed that such a path is a competent metabolic pathway, as recently discussed [25-27]. In this context, the major criticism raised as to path-finding methods is that they neglect reaction stoichiometry and there is, therefore, no guarantee that any path found can operate in sustained steady-state.

The steady-state condition requires the definition of the boundary of the metabolic network under study. Metabolites inside the boundary of the network, typically called internal metabolites [28], must be in stoichiometric balance. Balancing does not apply to metabolites outside the boundaries of the system (external metabolites), which are typically input/output metabolites and (sometimes) cofactors. In other words, for internal metabolites, their production and consumption (if possible) must be captured with the reactions in the network under study.

The steady-state condition and its underlying boundary definition are critical for the performance of any method for analyzing a metabolic network and ignoring it may provide misleading insights. A nice illustration of this is the one presented in the work of de Figueiredo *et al.* [25], which (unsuccessfully) tested the ability of path-finding methods to answer the question as to whether (or not) fatty acids can be converted into sugars. Klamt *et al.* [29] also recently emphasized this issue for different biological networks.

Note here that elementary flux modes (and extreme pathways) represent a more general and elegant concept for metabolic pathways than paths [28,30]. Their computation is, however, much more expensive in large metabolic networks than paths and, though different efforts

* Correspondence: John.Beasley@brunel.ac.uk; fplanes@ceit.es

¹CEIT and TECNUN, University of Navarra, Manuel de Lardizabal 15, 20018 San Sebastian, Spain

²Mathematical Sciences Department, Brunel University, Kingston Lane, UB8 3PH, Uxbridge, UK

Full list of author information is available at the end of the article

have been made in this area [31-33], much research is still needed to make elementary flux modes a practical tool for the analysis of large metabolic networks.

Given the limitations discussed above, a novel theoretical concept termed flux paths is introduced here. A flux path is a simple path (in the graph-theoretical sense, so no nodes revisited) from a source metabolite to a target metabolite able to operate in sustained steady-state. In essence, flux paths incorporate reaction stoichiometry into traditional path-finding methods [4,7,34,35]. By means of this concept we show that the path structure of metabolic networks is substantially altered when stoichiometry is considered. In addition, we illustrate (with several examples) that flux paths offer new perspectives for the analysis of metabolic networks at the topological and functional levels. The determination of flux paths requires going beyond graph theory via mixed-integer linear programming. We present below details as to our mathematical optimization model for determining K-shortest flux paths between source and target metabolites.

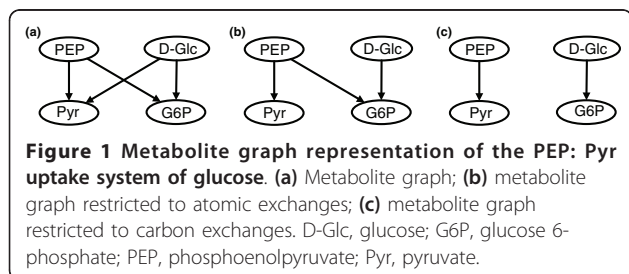
Results and discussion

Mathematical model

Assume we have a metabolic network that comprises R reactions and C metabolites. Note here that reversible reactions contribute two different reactions to the metabolic network. For this reason we can regard all fluxes as taking positive values. Let S_{cr} be the stoichiometric coefficient associated with metabolite c ($c = 1, \dots, C$) in reaction r ($r = 1, \dots, R$). As usual in the literature [28], input metabolites have a negative stoichiometric coefficient, whilst output metabolites have a positive stoichiometric coefficient.

We here used a metabolite (directed) graph representation of the network where nodes are metabolites and arcs link the input and output metabolites of each reaction. Figure 1a shows an example of the metabolite graph representation of the phosphoenolpyruvate (PEP): pyruvate (Pyr) phosphotransferase system for the uptake of glucose.

Suppose that we are concerned with finding a flux path from a source metabolite α to a target metabolite β . As mentioned above, a flux path is a simple path from the source metabolite α to the target metabolite β able to operate in steady-state. We present below our mathematical optimization model for flux paths.



Path finding constraints

We need to decide the arcs involved in the flux path from the source metabolite α to the target metabolite β . This fact is represented with a zero-one (binary) variable u_{ij} , where $u_{ij} = 1$ if the arc $i \rightarrow j$ linking metabolite i ($i = 1, \dots, C$) to metabolite j ($j = 1, \dots, C$) is active in the flux path, 0 otherwise.

Deletion of arcs from the metabolic graph is standard practice in path-finding methods [4,7,34,35]. We removed arcs not involving an effective carbon exchange. Carbon exchange is indeed essential for metabolic purposes. For this reason, we henceforth use the term carbon flux paths (CFPs).

Note here that a similar criterion has been used in [35]. In this work, however, input and output metabolites can have any type of atom or atom groups in common. This criterion is illustrated in Figure 1b, where PEP donates a phosphate group to glucose (D-Glc). The focus on carbon atoms makes our approach more restrictive, as observed in Figure 1c, which shows that there is only effective carbon exchange between D-Glc and glucose 6-phosphate (G6P), and PEP and Pyr.

Let d_{ijr} be a binary (0/1) coefficient establishing whether (or not) there exists an effective carbon exchange between input metabolite i ($S_{ir} < 0$) and output metabolite j ($S_{jr} > 0$) in reaction r . If $\sum_{r=1}^R d_{ijr} = 0$, so there is no reaction involving metabolites i to j in carbon exchange, then u_{ij} is also fixed to zero.

In the following lines we present constraints needed to obtain an appropriate directed path from the source metabolite (α) to the target metabolite (β). Equation 1 ensures that one arc leaves α and one arc enters β ; equation 2 that no arc enters α and no arc leaves β :

$$\sum_{j=1}^C u_{\alpha j} = 1 \quad \text{and} \quad \sum_{i=1}^C u_{i\beta} = 1 \quad (1)$$

$$\sum_{i=1}^C u_{i\alpha} = 0 \quad \text{and} \quad \sum_{j=1}^C u_{\beta j} = 0 \quad (2)$$

Equation 3 ensures that the number of arcs entering a metabolite k is equal to the number leaving; Equation 4 ensures that a metabolite cannot be revisited in the path:

$$\sum_{i=1}^C u_{ik} = \sum_{j=1}^C u_{kj}, \quad k = 1, \dots, C; \quad k \neq \alpha, \beta \quad (3)$$

$$\sum_{i=1}^C u_{ik} \leq 1, \quad k = 1, \dots, C \quad (4)$$

Stoichiometric constraints

Equations 1 to 4 define a simple path that preserves carbon exchange in each of its intermediate steps. We need to guarantee that this path can work in sustained steady-state. As will be shown below, to do this, it is required to find a steady-state flux distribution able to involve the path. We here introduce variables and constraints needed to define the steady-state flux space.

Any steady-state flux distribution satisfies Equation 5 for the set of internal metabolites (I). We denote v_r the non-negative (continuous) flux associated with each reaction, $r = 1, \dots, R$:

$$\sum_{r=1}^R S_{cr} v_r = 0, \quad \forall c \in I \quad (5)$$

External metabolites (E) are not subject to balancing constraints. If a specific growth medium (E_m) is introduced, however, metabolites not involved in such a medium can be produced, but cannot be consumed, as observed in Equation 6:

$$\sum_{r=1}^R S_{cr} v_r \geq 0, \quad \forall c \in E, c \notin E_m \quad (6)$$

For convenience, we introduced a zero-one (binary) variable z_r ($r = 1, \dots, R$), which defines the reactions involved in a steady-flux distribution, namely $z_r = 1$ if reaction r has a non-zero flux, 0 otherwise ($r = 1, \dots, R$). We need constraints relating the reaction variables z_r and the flux variables v_r . Equation 7 ensures that no flux traverses a reaction r if $z_r = 0$:

$$z_r \leq v_r, \quad r = 1, \dots, R \quad \text{and} \quad v_r \leq M z_r, \quad r = 1, \dots, R \quad (7)$$

In addition, it guarantees that v_r is non-zero if $z_r = 1$. Here we have scaled fluxes so that the maximum flux is M and the minimum (non-zero) flux is 1. This does not constitute an issue if we consider M sufficiently large.

As we split reversible reactions into two irreversible steps, we need to prevent a reaction and its reverse from appearing together in any steady-state flux distribution, as observed in Equation 8, where the set $B = \{(\lambda, \mu) \mid \text{reaction } \lambda \text{ and reaction } \mu \text{ are the reverse of each other}\}$:

$$z_\lambda + z_\mu \leq 1 \quad \forall (\lambda, \mu) \in B \quad (8)$$

Current path-finding approaches deal with this situation indirectly, namely by removing computed paths involving a reaction and its reverse.

Equations 5 to 8 define the steady-state flux space for a particular metabolic network.

Linking path finding and stoichiometric constraints

As noted above, it is required that the path defined by constraints 1 to 4 can operate in a steady-state flux distribution. For this purpose, we need to guarantee that if

we use an arc $i \rightarrow j$ in a path, then some reaction r with $d_{ijr} = 1$, that is, involving effective carbon exchange between i and j , is contained in the steady-state flux distribution. This is a critical point in our formulation, which makes it different from previous path-finding methods. With this condition we naturally link the topological and (steady-state) flux planes. This linking constraint is reflected in Equation 9:

$$\sum_{r=1, d_{ijr}=1}^R z_r \geq u_{ij} \quad i = 1, \dots, C; j = 1, \dots, C; i \neq j \quad (9)$$

Equation 9 ensures that if an arc $i \rightarrow j$ is active in the CFP (so $u_{ij} = 1$), then at least one reaction r containing this arc in carbon exchange (so $d_{ijr} = 1$) is forced to be active. By forcing z_r to be 1 there will be a non-zero flux associated with the reaction due to Equation 7. An important point to note from Equation 9 is that it allows reactions to be active even if they are not involved in the CFP. In other words reactions can be active with non-zero flux (to satisfy the requirements of steady-state, Equation 5) but without any of their input/output metabolites being involved in the CFP.

To illustrate constraint 9, consider the example metabolic network in Figure 2, which involves seven reactions and nine metabolites. The set of internal metabolites is $I = \{A, B, C, D, E, F\}$. Assume now that we are concerned with finding a CFP between metabolites A and F. We have only one possible path, namely $A \rightarrow B \rightarrow C \rightarrow E \rightarrow F$ ($u_{AB} = u_{BC} = u_{CE} = u_{EF} = 1$). Due to Equation 9, reactions 2, 3, 4 and 5 are active, that is, $z_2 = z_3 = z_4 = z_5 = 1$ and, therefore, via Equation 7, their flux will be non-zero. To balance such a path and satisfy the steady-state condition, Equation 5, we require three additional reactions off-path: reaction 1 for the production of A, reaction 7 to consume F and reaction 6 to produce D. If these off-path reactions are active, the

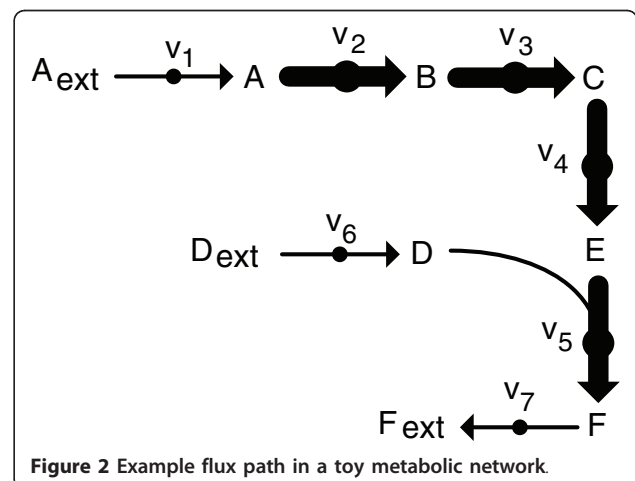


Figure 2 Example flux path in a toy metabolic network.

path from A to F is able to work in sustained steady-state and, therefore, it is a flux path, as denoted in the Background section. We are obviously considering that A_{ext} and D_{ext} are in the growth medium. If we remove one of these metabolites from the medium, though we still have a path from A to F at the graph-theoretical level, no flux path will exist, since the path cannot work in sustained steady-state.

Objective function

Equations 1 to 9 define the set of constraints for the determination of any CFP between metabolite α and metabolite β . However, our purpose here is to find the shortest CFP, as observed in Equation 10:

$$\text{Minimize } \sum_{i=1}^C \sum_{j=1, j \neq i}^C u_{ij} \quad (10)$$

Enumerating constraint

As in other path-finding approaches, we may be interested in computing not only the shortest CFP, but the k -shortest CFPs ($k = 1, \dots, K$). Since we have an objective relating to finding the shortest CFP, we need to add constraints eliminating previously found CFPs, as shown in Equation 11. In that constraint U_{ij}^k is the binary solution value for the u_{ij} variable in the k -shortest CFP:

$$\sum_{i=1}^C \sum_{j=1, j \neq i}^C U_{ij}^k u_{ij} \leq \sum_{i=1}^C \sum_{j=1, j \neq i}^C U_{ij}^{k-1} - 1 \quad k = 1, \dots, K - 1 \quad (11)$$

Validation

This section is organized as follows. By means of several well-documented examples, we first illustrate the biochemical relevance of particular constraints in our CFP approach. We then carry out a side-by-side comparison of our CFP approach with current path-finding approaches.

Path-finding comparison

As shown in the 'Mathematical model' section, the path-finding strategy used in our CFP approach is based on using arcs involving effective carbon exchange and imposing the reversibility constraint, Equation 8. In this sub-section we illustrate the importance of these factors and show that a path-finding approach incorporating them outperforms existing methods in the literature. For this analysis, the effect of stoichiometry is not considered, as is common in existing approaches. Its effect will be separately considered in detail in the next sub-section ('Effect of stoichiometry'). Therefore, for this analysis, Equations 5 and 6 were ignored.

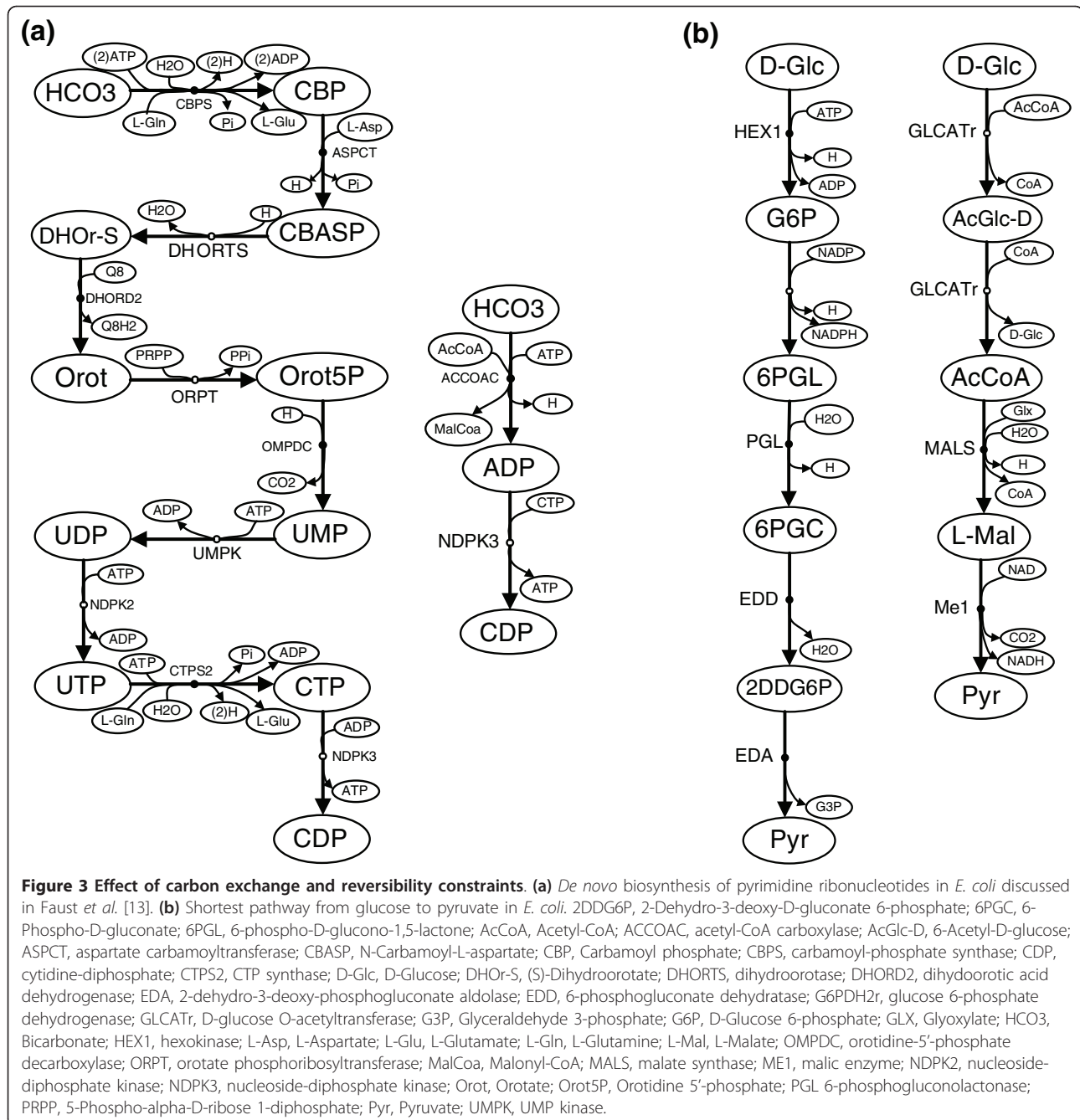
Effective carbon exchange Figure 3a shows two paths from bicarbonate (HCO_3) to cytidine-diphosphate (CDP) in *Escherichia coli*. The long path is a well-

known (canonical) metabolic pathway for *de novo* pyrimidine biosynthesis. The short path is a shortcut via ADP, which has no biological relevance. The removal of arcs not involving carbon exchange, as done in our CFP approach, considerably reduces the appearance of such non-meaningful paths. Indeed, when we applied our approach to find a CFP from HCO_3 to CDP to the genome-scale metabolic network of *E. coli* [36], the long pathway was directly recovered. Note here that we manually removed arcs not involving carbon exchange in the network of Feist *et al.* [36]. The resulting list of arcs can be found in Additional file 1. This same biochemical example was recently discussed in Faust *et al.* [13], under different strategies. In the best case scenario, they require additional information as to the intermediate metabolites to recover this pathway. The fact that our approach can recover the pathway without intermediate metabolite information shows how effective the carbon exchange constraint is.

Reversibility Path-finding methods typically split reversible reactions into two irreversible steps. In contrast to current approaches [13], in our CFP approach we prevent two such irreversible steps from being active in the same path, as observed in Equation 8. To illustrate the importance of this constraint, we analyzed the shortest path from D-Glc to Pyr in *E. coli*, which is the Entner-Doudoroff pathway, as shown in the left-hand side of Figure 3b. When we applied our CFP approach from D-Glc to Pyr without including Equation 8, we obtained the path in the right hand-side of Figure 3b (D-Glc \rightarrow AcGlc-D \rightarrow AcCoA \rightarrow L-Mal \rightarrow Pyr). This solution has no biochemical meaning, since the first and second step in that path is a cycle involving the forward and backward step of the reversible reaction catalyzed by D-glucose O-acetyltransferase (GLCATr: D-Glc + AcCoA \leftrightarrow AcGlc-D + CoA). By adding Equation 8 this path is removed from the solution space and our CFP approach directly obtains the Entner-Doudoroff pathway.

Side-by-side comparison In order to analyze the performance of any path-finding method, it is usual in the literature to evaluate its ability in recovering well-known metabolic pathways. For this purpose, we used a database of 40 reference *E. coli* (metabolic) pathways previously discussed in Planes and Beasley [37] (these 40 pathways are listed in Additional file 2).

The input metabolic graph was built from the genome-scale metabolic network of *E. coli* [36]. We computed the 100 shortest CFPs between the source and target metabolites of each of the 40 reference pathways. As mentioned above, stoichiometric constraints are not considered in this sub-section since the aim is to establish the effectiveness of carbon exchange when combined with reversibility in path finding. To compare the 100 shortest CFPs and the reference pathway, we used



the recovery rate. Recovery is a 0/1 parameter, being 1 if a CFP fully matches with the reference pathway, 0 otherwise.

A similar analysis was conducted for existing path-finding methods [4,7,34,35]. These methods make use of different strategies to provide biochemical meaning to the computed paths. For comparison, we classified these strategies into different groups: the first strategy (denoted 'topology') involves the use of an unadjusted metabolic graph; the second strategy (denoted 'hubs')

adjusts the metabolic graph by removing any arc involving a highly connected metabolite (hubs) [7,34] (we took the list of hubs from Planes and Beasley [37]); the third strategy (denoted 'connectivity') assigns weights to metabolites according to their connectivity in an unadjusted metabolic graph, where connectivity is defined to be the number of reactions involving a metabolite [9]. Finding K-shortest paths is substituted here by finding K- lightest paths, that is, the sum of weights of arcs involved in the path is minimized.

Note here that there are path-finding strategies that use structural atomic mapping information. These approaches can be classified into two different groups. In the first group atomic mapping is used to build the metabolic graph, that is, an input metabolite is linked to an output metabolite in a given reaction if they share an atom mapping. In other words, an arc between a given pair of input/output metabolites exists if they have atoms in common in at least one reaction. The work of Faust *et al.* [35], based on the RPAIR database [38], is a reference example for these approaches. The effective carbon exchange strategy used in our CFP approach also falls into this group. However, it is slightly more restrictive than the approach presented in Faust *et al.* [35], since we exclusively focus on carbon atoms, that is, an arc between a given pair of input/output metabolites exists if they have carbon atoms in common in at least one reaction.

In the second group atomic mapping is used to guarantee that the pathway target metabolite involves at least one atom from the source metabolite. This concept was first introduced by Arita *et al.* [39], and recently revisited in Blum and Kohlbacher [40], and Heath *et al.* [41]. We are aware that this type of approach is, in theory, more restrictive than the effective carbon exchange strategy used in our CFP approach, since we guarantee effective carbon exchange between intermediates in the path, but not between the source and target metabolites. Tracing an atom from source to target metabolite, however, requires detailed knowledge of carbon atom mappings for each reaction. Though active research is being undertaken into this topic, more effort is still needed to release a fully curated and complete database for atomic mappings in genome-scale metabolic networks, especially for those from the Biochemical Genetic and Genomic (BiGG) database [42], which we are using here. For completeness, we will include results for the most recent approach [41], denoted as atom mapping-based strategy. Results were extracted from the web service (named AtomMetaNetWeb) available from Kavraki's lab [43].

Figure 4 shows results obtained for each of the strategies discussed above. It can be observed that the hubs-based strategy increases the average recovery rate with respect to the unadjusted metabolic graph (topology) by around 20% on average. The atom mapping-based strategy is clearly less accurate than the hubs-based strategy, which reflects the point discussed above that current databases for atomic mappings require further development. In addition, the connectivity-based strategy substantially outperforms the hubs-based strategy - for example, for $k = 1$, 62.5% and 32.5% of reference pathways are recovered, respectively. Finally, our CFP approach outperforms the connectivity-based strategy. This analysis shows, therefore, that our CFP approach

(even without considering stoichiometry) outperforms existing path-finding methods.

Finally, note that other works [9,13] typically used the accuracy rate, instead of the recovery rate, for comparing the computed paths and reference pathways. We repeated the same analysis using this parameter. As observed in Additional file 2, a similar result to Figure 4 is obtained, which again shows that our CFP approach outperforms current methods.

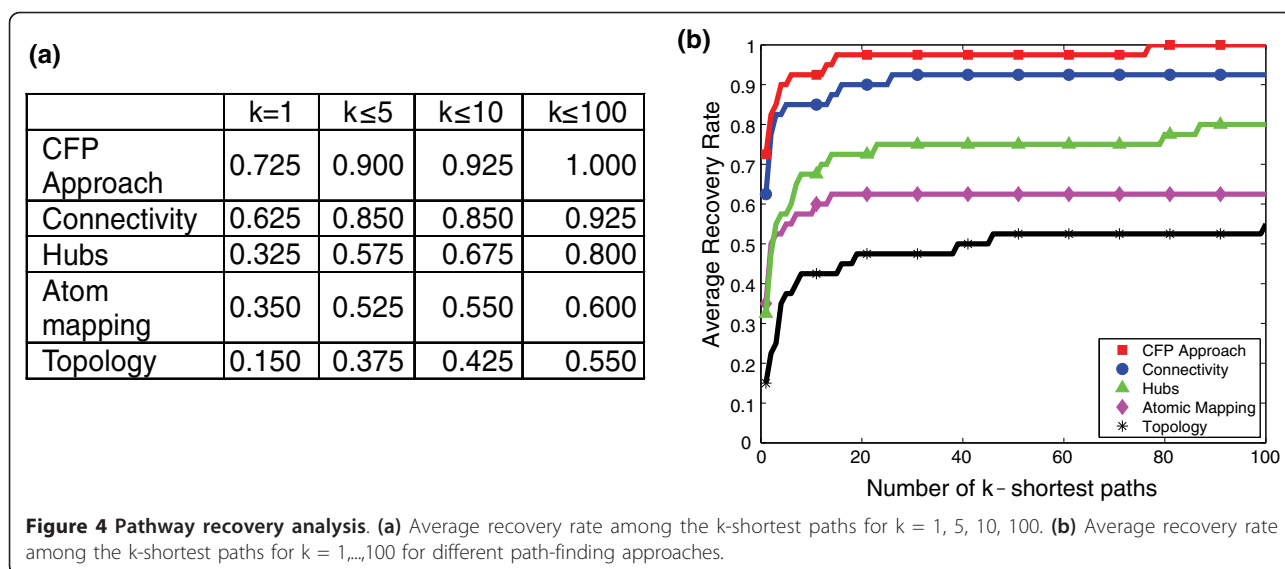
Effect of stoichiometry

To illustrate the effect of stoichiometry, we first analyze a previously considered example from the literature, which emphasizes the fact that some paths (at the graph-theoretical level) cannot perform in steady-state and therefore are not biologically meaningful. We then repeat the side-by-side comparison presented in Figure 4 when stoichiometry is considered. To emphasize its importance, we examine how the connectivity structure of several metabolites is altered when stoichiometry is considered.

Stoichiometry and infeasible paths Figure 5 shows a simplified network from that presented in de Figueiredo *et al.* [25], which considered the question as to whether (or not) fatty acids can be converted into sugars. This question is answered by finding pathways from acetyl-CoA (AcCoA) to G6P. In that work, two scenarios were analyzed, namely pathway structure from AcCoA to G6P in the presence and absence of the enzymes of the glyoxylate shunt (indicated by dashed lines in Figure 5). In the metabolic network in Figure 5, when the glyoxylate shunt is absent, no possible pathway can exist in a stoichiometric balance from AcCoA to G6P. As observed in de Figueiredo *et al.* [25], this fact is not properly captured by path-finding methods, since stoichiometry is not taken into account. In contrast, our CFP approach correctly answers this question, by finding no paths between AcCoA and G6P when the glyoxylate shunt is not active. This is due to the addition of constraint 9, which forces paths to be able to work in sustained steady-state.

Side-by-side comparison with stoichiometry We repeated the side-by-side comparison previously presented in Figure 4 for path-finding methods when stoichiometry is considered. Similarly, we used the 40 *E. coli* metabolic pathways discussed in Planes and Beasley [37], and the *E. coli* metabolic network in Feist *et al.* [36].

As we previously showed above (Figure 4) that our CFP approach (without considering stoichiometry, Equations 5 and 6) outperforms existing path-finding methods, we here compare the performance of our CFP approach with and without Equations 5 and 6 so as to evaluate the effect of stoichiometry. For this purpose, we analyzed our CFP approach in two different scenarios,



namely when we used a minimal medium based on glucose as a sole carbon source under oxic and anoxic conditions, respectively. See Additional file 3 for details.

It is important to note that the use of a specific minimal medium (as we do here) prevents some known metabolic pathways from functioning in *E. coli* due to stoichiometric constraints. For example, the tricarboxylic acid (TCA) cycle cannot work in anoxic conditions in *E. coli*. The ability to detect these false positives cannot be accomplished without the use of stoichiometry. In light of this, the definition of recovery (as used in Figure 4) is slightly modified here. Recovery rate is 1 if (under a given growth medium) the model recovers a feasible pathway or the model excludes from the solution space an infeasible pathway, 0 otherwise. For illustration, if our CFP approach (incorrectly) detects the TCA cycle in anoxic conditions, recovery would be zero. However, if our CFP approach correctly excludes the TCA cycle from the solution space, then recovery would be 1.

Figure 6a shows how recovery rate evolves over k-shortest CFPs ($k = 1, \dots, 100$) with/without stoichiometry in oxic conditions. We found that in these conditions, 6 out of 40 metabolic pathways cannot work in steady-state (Additional file 2). For example, the pathway for the degradation of 2,5-diketo-D-gluconate is not functionally feasible under these conditions since it cannot be synthesized from glucose in *E. coli* [44]. This logically cannot be captured without considering stoichiometry. This is reflected in Figure 6a, where average recovery rate among 100 shortest CFPs decreases to 0.85 without stoichiometry. The same analysis was repeated in anoxic conditions (Figure 6b), finding two additional pathways (TCA cycle and Allantoin degradation) not able to work in steady-state (given our growth medium). Figure 6c summarizes Figure 6a and Figure 6b for some particular

values ($k = 1, 5, 10$ and 100). See Additional file 2 for further details, including results when average accuracy rate was used instead of recovery rate. This analysis shows the importance of stoichiometry and its underlying boundary definition at the functional level.

Connectivity analysis and stoichiometry To emphasize the effect of stoichiometry, we examined the connectivity structure of oxaloacetate (OAA) in *E. coli*. OAA plays an important role in the regulation of carbon flux in most organisms. Again, for this study, we used the metabolic network presented in Feist *et al.* [36] and a minimal medium based on glucose as a sole carbon source and oxic conditions.

We determined CFPs from OAA to all reachable metabolites (obviously some metabolites may not be reachable via a CFP from OAA). In order to organize and compare the obtained results, we plotted a connectivity curve that shows the total number of connected metabolites when we move a specified number of reaction steps away from the source metabolite. To show the effect of stoichiometry, we plot the connectivity curves when stoichiometry is included (so including Equations 5 and 6) and when it is not included (so excluding Equations 5 and 6).

Figure 7a shows the connectivity curves for OAA. For example, in five reaction steps, OAA reaches 300 metabolites when stoichiometry is included and 400 metabolites otherwise. It can also be observed that, in any number of reaction steps, the number of metabolites reachable from OAA when stoichiometry is taken into account is 834, but 1,028 metabolites when it is not considered. These results clearly show the effect of considering stoichiometry. We repeated the same analysis in two structurally different metabolites, namely arginine (L-Arg), an amino acid, and phosphatidic acid (PA120),

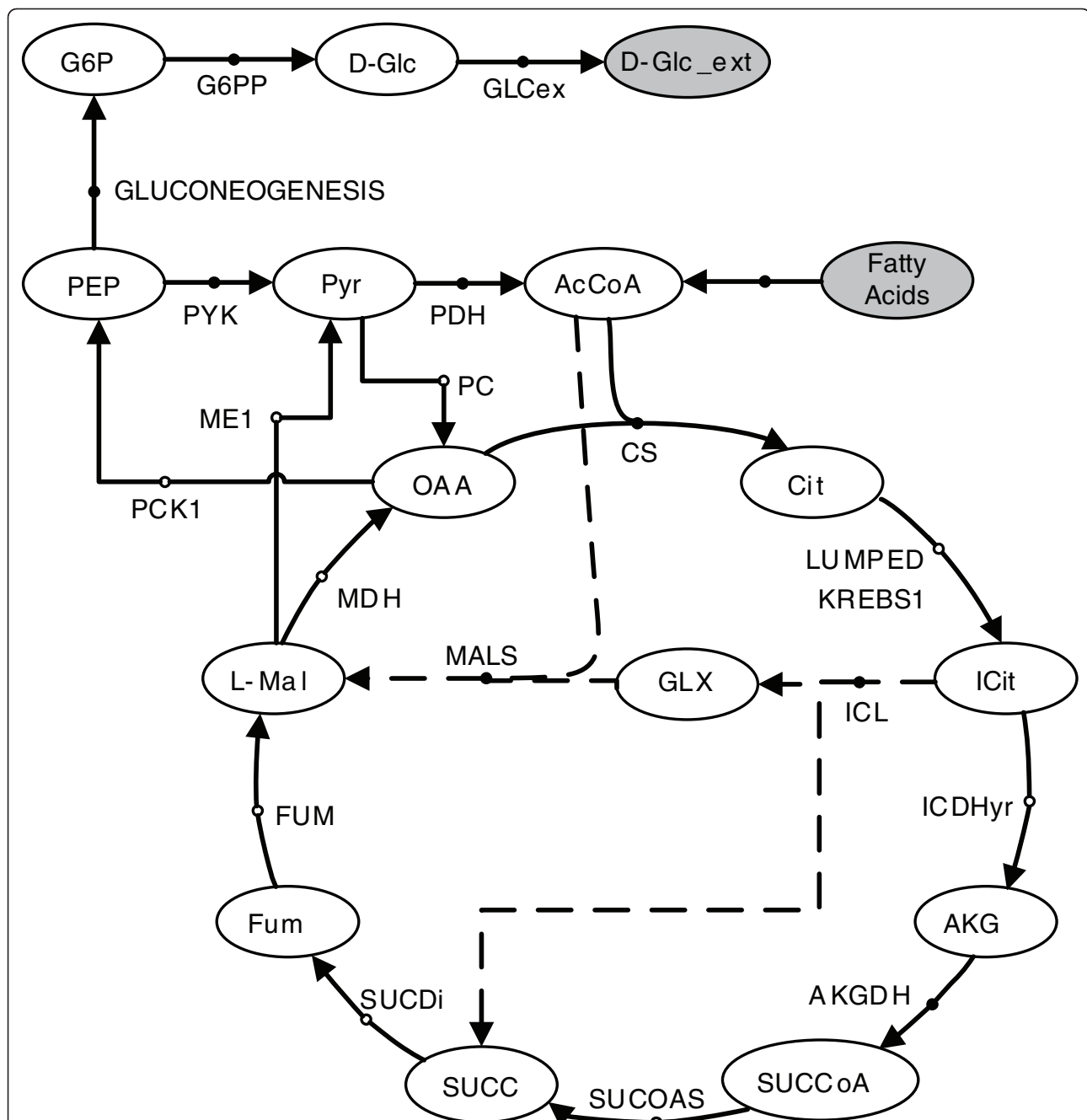
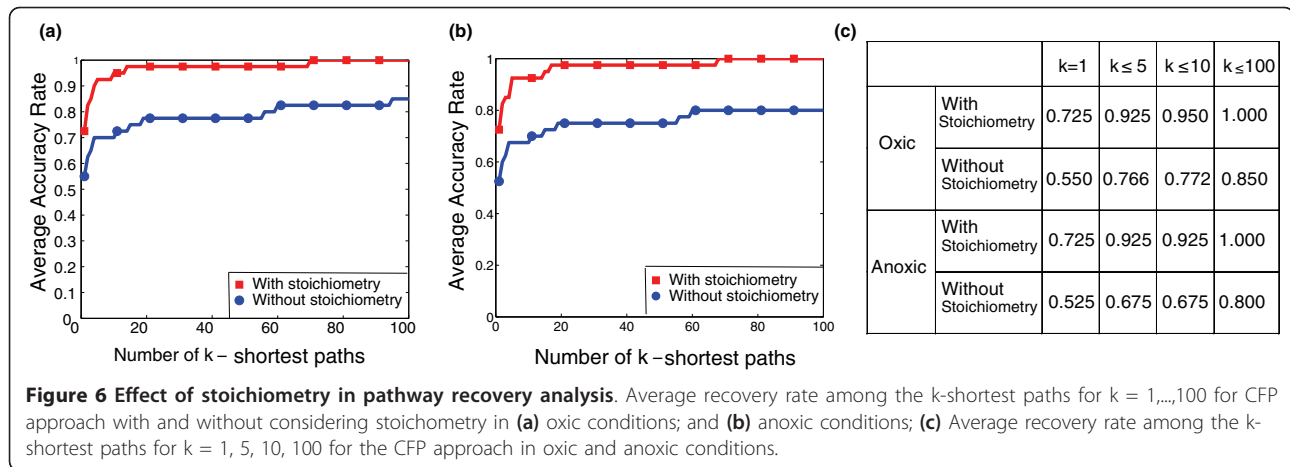


Figure 5 Simplified network from that presented in de Figueiredo *et al.* [25] considering the question of conversion of fatty acids to sugars. AcCoA, acetyl-CoA; AKG, 2-oxoglutarate; AKGDH, 2-oxoglutarate dehydrogenase; Cit, citrate; CS, citrate synthase; D-Glc, glucose; D-Glc-ext, external glucose; Fum, fumarate; FUM, fumarase; G6P, glucose 6-phosphate; G6PP, glucose-6-phosphate phosphatase; GLCex, Glucose Exchange; Glx, Glyoxylate; ICDHyr, isocitrate dehydrogenase; ICit, isocitrate; ICL, Isocitrate lyase; L-Mal, Malate; MALS, malate synthase; MDH, malate dehydrogenase; ME1, malic enzyme (NAD); OAA, oxaloacetate; PC, Pyruvate carboxylase; PCK1, phosphoenolpyruvate carboxykinase; PDH, pyruvate dehydrogenase; PEP, phosphoenolpyruvate; PYK, pyruvate kinase; Pyr, pyruvate; SUCC, succinyl; SUCCoA, succinyl-coenzyme A; SUCDi, succinate dehydrogenase; SUCOAS, succinyl-CoA synthetase.

an important lipid. We found a very similar behavior, as observed in Figure 7b,c. This analysis shows the importance of considering stoichiometry for the topological analysis of metabolic networks from a path-based perspective.

Application

It is usual to find K paths between a pair of key metabolites/reactions in, for example, the interpretation of 'omics' data [13,20]. Current path-finding methods do not take into account stoichiometric constraints for this



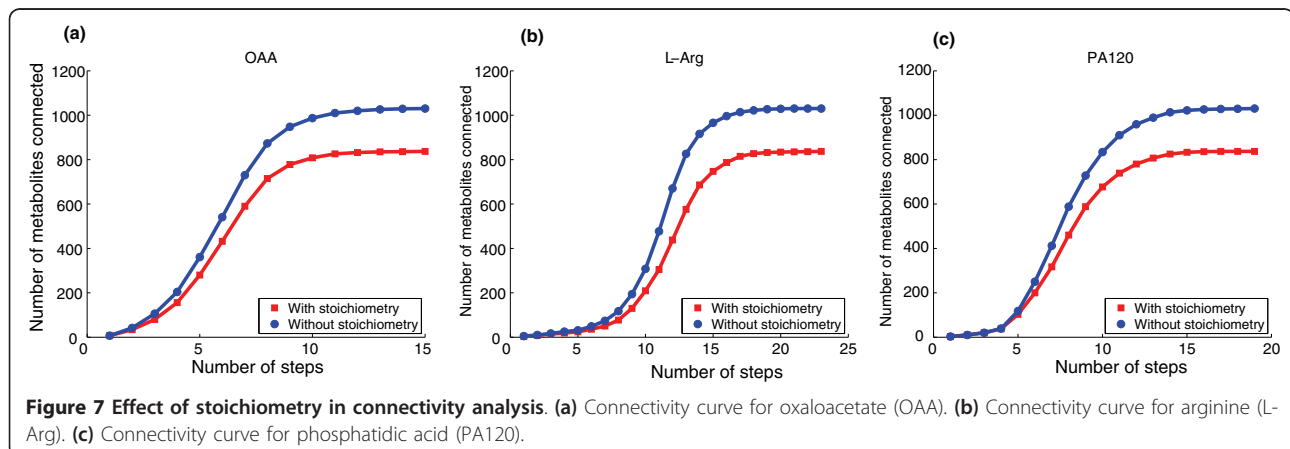
analysis. In the analysis presented below we show that the resulting K functional paths are strongly dependent on stoichiometric constraints. This fact is illustrated in this sub-section with the pathway analysis of Pyr-OAA metabolism.

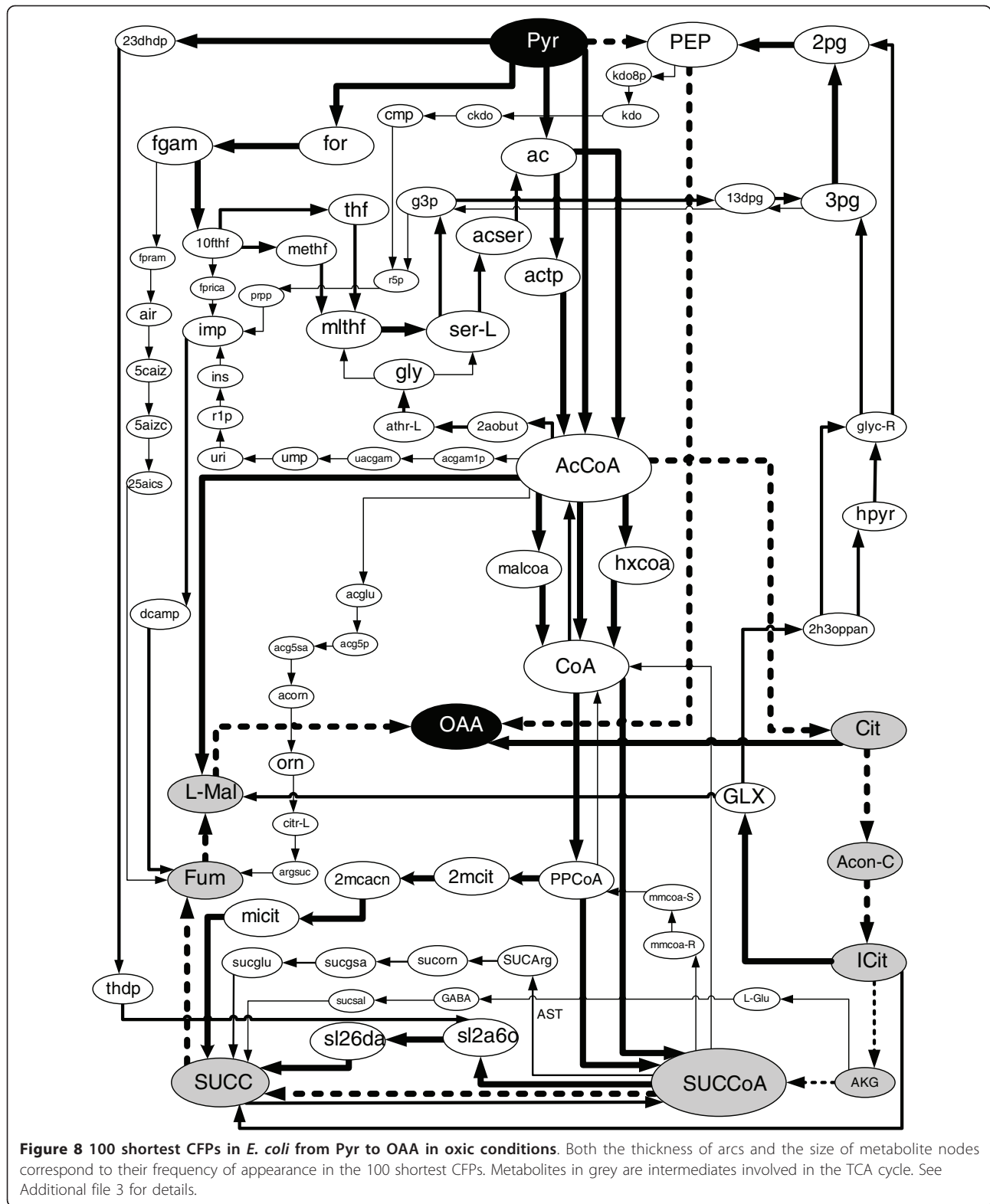
PEP, Pyr and OAA are important metabolites whose underlying inter-conversions control the carbon flux distribution in bacteria [45]. The performance of the PEP-Pyr-OAA node changes in different organisms and growth conditions. We focus here on the structure of CFPs from Pyr to OAA in *E. coli* in two different scenarios, namely in oxic and anoxic conditions. Pyr and OAA are linked by two fundamental metabolic processes. Firstly, Pyr (via PEP) can be carboxylated to OAA for the replenishment of TCA cycle intermediates or for anabolic purposes (for example, amino acid biosynthesis). This process is typically referred to as anaplerosis. In addition, Pyr and OAA are strongly related via the TCA cycle, which oxidizes carbon of Pyr to CO_2 and requires OAA to operate.

We calculated the 100 shortest CFPs in both scenarios using the metabolic network presented in Feist *et al.*

[36]. Again, we used the list of arcs presented in Additional file 1. In addition, we used a minimal medium based on glucose. See Additional file 3 for details as to the medium used.

Figure 8 shows the 100 shortest CFPs from Pyr to OAA in oxic conditions. Both fundamental metabolic processes described above between Pyr and OAA (anaplerotic route via PEP and the TCA cycle) are recovered (see dashed lines). In addition, different alternative routes to these processes are found. In particular, several bypasses to the TCA cycle can be observed in Figure 8. The glyoxylate (GLX) shunt was recovered, as well as the γ -aminobutyrate (GABA) shunt, whose role as an integral part of the TCA cycle was recently hypothesized [46]. We also determined a (theoretical, non-experimentally determined) bypass via propionyl-CoA (PPCoA), which was reported in a previous paper [31]. Interestingly, we also predicted a bypass to the TCA cycle via L-Arg catabolism. Though not shown in Figure 8, L-Arg is consumed in a reaction catalyzed by arginine succinyltransferase (AST; $\text{SUCCoA} + \text{L-Arg} \rightarrow \text{SUCArg}$). Several links to the TCA cycle with arginine-





L metabolism has been previously reported [47], although more research is needed to examine whether this detour is a functionally feasible alternative route to succinyl-CoA synthetase (SUCCoA) ($\text{ATP} + \text{CoA} + \text{SUCC} \leftrightarrow \text{ADP} + \text{P}_i + \text{SUCCoA}$).

Though the number of non-meaningful paths has been substantially reduced, it can be observed in Figure 8 that they still exist - for example, different routes via CoA. These false positives do not arise from the lack of stoichiometric balancing, but due to carbon exchange

constraints. Indeed, these routes exchange carbon atoms in each of their intermediate steps but do not exchange carbon atoms between Pyr and OAA. When the current limitations described above (in the discussion of atom mapping-based approaches) are addressed, such strategies may be an effective constraint to remove these false positives.

We repeated the same analysis in anoxic conditions (Figure 9). In this situation, the main variability in the 100 shortest CFPs is found in anaplerotic routes, since the TCA cycle is not active. This is due to the fact that the balancing of coenzyme Q (CoQ) and ubiquinol is not possible without oxygen and therefore enzyme succinate dehydrogenase ($\text{CoQ} + \text{SUCC} \rightarrow \text{Fum} + \text{CoQH}_2$) cannot work in sustained steady-state. This meant that several other reactions involved in the TCA cycle do not appear in the 100 shortest CFPs, namely isocitrate dehydrogenase ($\text{ICit} + \text{NADP} \leftrightarrow \text{AKG} + \text{CO}_2 + \text{NADPH}$), 2-oxoglutarate dehydrogenase ($\text{AKG} + \text{CoA} + \text{NAD} \rightarrow \text{CO}_2 + \text{NADH} + \text{SUCCoA}$) and SUCOAS ($\text{ATP} + \text{CoA} + \text{SUCC} \leftrightarrow \text{ADP} + \text{Pi} + \text{SUCCoA}$) are not in Figure 9. This is also the case for metabolite AKG, which is now not involved in the 100 shortest CFPs from Pyr to OAA, while in oxic conditions it appeared in five solutions. In addition, most of the bypasses previously mentioned in oxic conditions are not involved in Figure 9; indeed just the glyoxylate shunt is kept in the solution.

Finally, as observed in Figures 8 and 9, our CFP approach properly captures the metabolic changes induced when oxygen is removed from the medium. These changes cannot be captured if stoichiometric constraints are not considered, showing again the strength of our CFP approach.

Conclusions

Graph-based methods have been widely used for the analysis of metabolic networks, but suffer from the important weakness that reaction stoichiometry is neglected. In this paper we show that, using the novel concept of CFPs, reaction stoichiometry can be incorporated into path-finding approaches, which constitute a clear progress over the state of the art at the methodological level.

Our results show that, when stoichiometry is incorporated into path-finding methods, the resulting set of functional pathways is substantially altered, as observed in the analysis of the 40 reference pathways. This idea is also reflected in the analysis of aerobic and anaerobic Pyr-OAA metabolism, which emphasizes the importance of the steady-state condition and its underlying boundary definition for the analysis of metabolic networks. In addition, connectivity analysis revealed important differences when stoichiometry was considered, as we

illustrated with regard to a number of metabolites. In summary, CFPs open new avenues for analyzing metabolic networks at the topological and functional levels and constitute a major advance.

Though the incorporation of stoichiometry into a path-finding method is the main feature of our work, our CFP approach focuses on paths involving effective carbon exchange in each of their intermediate steps. The results we have presented confirm the relevance of this strategy when analyzing metabolic networks using a path-finding approach. Our public release of the manually curated *E. coli* database incorporating effective carbon exchange information (based on BiGG [42] and the work of Feist *et al.* [36]) represents a valuable dataset available for the scientific community, which can be used for further analysis.

It is important to mention that our CFP approach is formulated as a mixed-integer linear program, which cannot be solved using classical algorithms from graph theory and requires a branch and bound approach. Computational experience shows that the determination of CFPs is not expensive, namely in the order of milliseconds. This fact makes our approach an effective tool for addressing other relevant questions previously addressed by path-finding approaches.

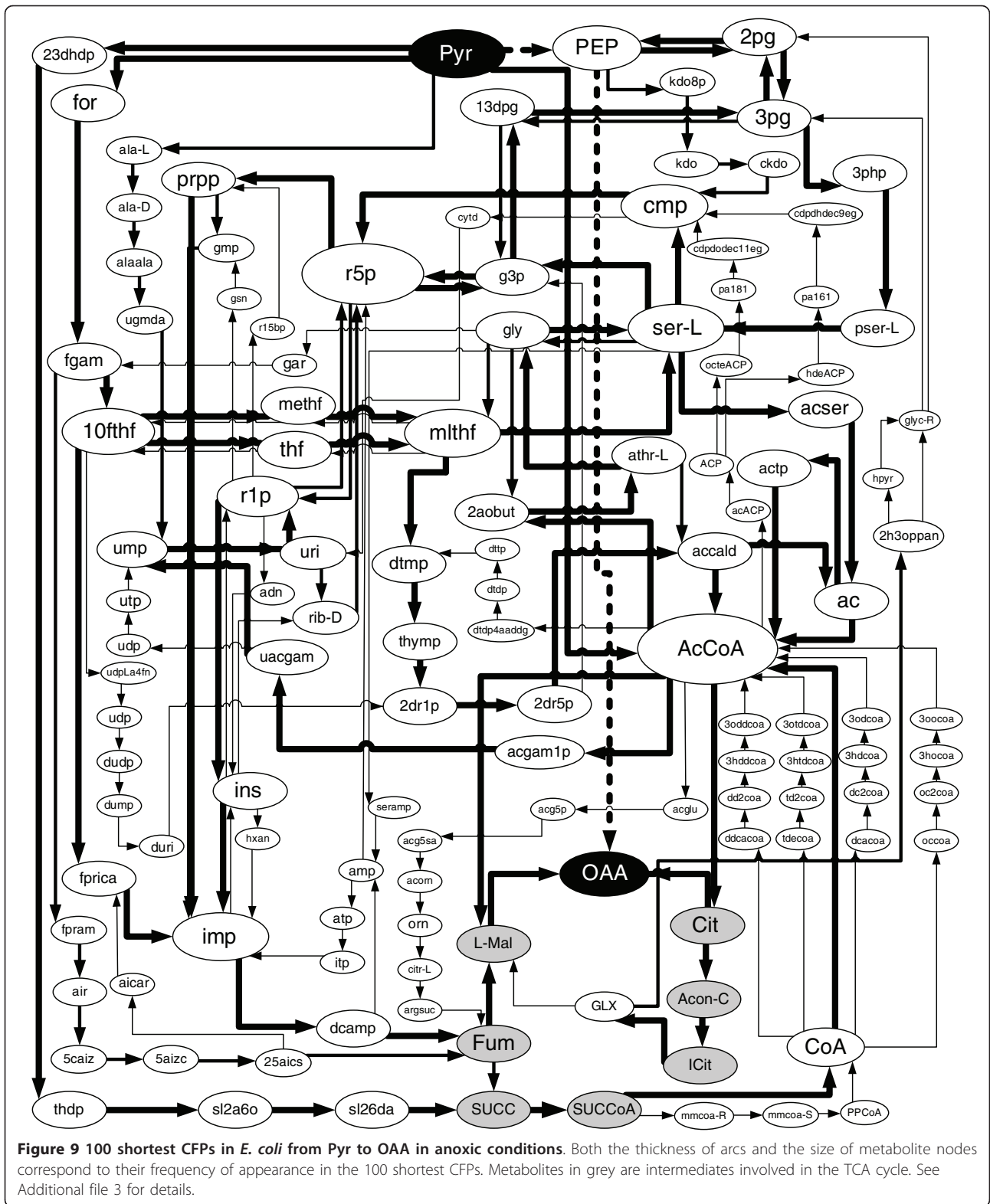
Our analysis of CFPs in aerobic Pyr-OAA metabolism allowed us to detect several bypasses to the TCA cycle. Some of these bypasses have been recently reported using a different pathway analysis technique, namely elementary flux patterns for the bypass via the GABA shunt [48] and generating flux modes for the bypass via PPCoA [31]. In addition, we found an alternative bypass to the TCA cycle via L-Arg. This novel pathway is currently theoretical (it should be treated with caution) and requires experimental validation; however, it shows the capability of our CFP approach to generate new hypothesis.

Finally, despite much debate in the field comparing the performance of path-finding methods and stoichiometric methods [25,27,49], this article shows that both approaches can work in a synergic fashion so as to explore the huge complexity in cellular metabolism.

Materials and methods

Equations 1 to 11 presented in the 'Mathematical model' sub-section define a mixed-integer linear problem and, algorithmically, such problems are solved by linear programming-based tree search. Modern software packages to perform this task, such as ILOG CPLEX, which we used, are well developed and highly sophisticated. ILOG CPLEX was run in a Matlab environment version 7.5 (R2007b).

The computation of the shortest CFP and the 100 shortest CFPs took us (on average) 300 ms and 2.5



minutes, respectively, on a 64-bit, 2.00 GHz PC with 12 Gb RAM. Analysis using regression indicated that, over the range of K values examined (up to $K = 250$), the total time for computing the K shortest CFPs was (approximately) proportional to $K^{1.4}$. This implies that the computation time of CFPs grows only as a low power of the number of paths (K) sought.

Additional material

Additional file 1: Database of carbon exchange arcs. PDF document containing a list of arcs involving effective carbon flux in the metabolic network of Feist et al. [36].

Additional file 2: Supporting data for side-by-side comparison. Word document containing a list of 40 reference pathways used in the side-by-side comparison, a side-by-side comparison using accuracy rate, and a discussion on infeasible pathways in Figure 6 [7,9,12,13,35-37,41,44,50-56].

Additional file 3: Supporting data for Figures 8 and 9. Details of the 100 shortest CFPs in oxic and anoxic conditions from Pyr to OAA.

Abbreviations

AcCoA: acetyl-CoA; AcGlc-D: 6-acetyl-D-glucose; AKG: 2-oxoglutarate; AST: Arginine succinyltransferase; BIGG: Biochemical Genetic and Genomic; CDP: cytidine-diphosphate; CFP: carbon flux path; CoA: coenzyme A; CoQ: coenzyme Q; D-Glc: glucose; Fum: fumarate; G6P: glucose 6-phosphate; GABA: gamma-aminobutyric acid; GLCATr: D-glucose O-acetyltransferase; GLX: Glyoxylate; PA120: phosphatidic acid; HCO₃: bicarbonate; ICit: isocitrate; L-Arg: arginine; L-Mal: acetyl-maltose; OAA: oxaloacetate; PEP: phosphoenolpyruvate; PPCoA: propionyl-CoA; Pyr: pyruvate; SUCArg: Succinyl-L-arginine; SUCC: Succinate; SUCCoA: succinyl-coenzyme A; SUCOAS: succinyl-CoA synthetase; TCA: tricarboxylic acid.

Acknowledgements

The work of JPe was supported by the Basque Government. The authors would like to thank two anonymous referees for their valuable comments improving the manuscript.

Author details

¹CEIT and TECNUN, University of Navarra, Manuel de Lardizabal 15, 20018 San Sebastian, Spain. ²Mathematical Sciences Department, Brunel University, Kingston Lane, UB8 3PH, Uxbridge, UK.

Authors' contributions

JPe developed and implemented the method, wrote the manuscript and performed analyses; JPr developed the method and carbon exchange database; JEB developed the method and wrote the manuscript; FJP conceived the study, developed the method and wrote the manuscript. All authors discussed the results, and read, commented and approved the final manuscript.

Received: 1 March 2011 Revised: 14 May 2011 Accepted: 27 May 2011
Published: 27 May 2011

References

- Aittokallio T, Schwikowski B: Graph-based methods for analysing networks in cell biology. *Brief Bioinform* 2006, **7**:243-255.
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL: The large-scale organization of metabolic networks. *Nature* 2000, **407**:651-654.
- Ma HW, Zeng AP: The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 2003, **19**:1423-1430.
- Arita M: The metabolic world of *Escherichia coli* is not small. *Proc Natl Acad Sci USA* 2004, **101**:1543-1547.
- Montanez R, Medina MA, Sole RV, Rodriguez-Caso C: When metabolism meets topology: Reconciling metabolite and reaction networks. *Bioessays* 2010, **32**:246-256.
- Fell DA, Wagner A: The small world of metabolism. *Nat Biotechnol* 2000, **18**:1121-1122.
- Wagner A, Fell DA: The small world inside large metabolic networks. *Proc Biol Sci* 2001, **268**:1803-1810.
- Yamada T, Bork P: Evolution of biomolecular networks: lessons from metabolic and protein interactions. *Nat Rev Mol Cell Biol* 2009, **10**:791-803.
- Croes D, Couche F, Wodak SJ, van Helden J: Inferring meaningful pathways in weighted metabolic networks. *J Mol Biol* 2006, **356**:222-236.
- Arita M: Metabolic reconstruction using shortest paths. *Simulation Practice Theory* 2000, **8**:109-125.
- Rahman SA, Advani P, Schunk R, Schrader R, Schomburg D: Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). *Bioinformatics* 2005, **21**:1189-1193.
- Planes FJ, Beasley JE: Path finding approaches and metabolic pathways. *Discrete Appl Mathematics* 2009, **157**:2244-2256.
- Faust K, Dupont P, Callut J, van Helden J: Pathway discovery in metabolic networks by subgraph extraction. *Bioinformatics* 2010, **26**:1211-1218.
- Patil KR, Nielsen J: Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci USA* 2005, **102**:2685-2689.
- Zelezniak A, Pers TH, Soares S, Patti ME, Patil KR: Metabolic network topology reveals transcriptional regulatory signatures of type 2 diabetes. *PLoS Comput Biol* 2010, **6**:e1000729.
- Kharchenko P, Church GM, Vitkup D: Expression dynamics of a cellular metabolic network. *Mol Syst Biol* 2005, **1**:2005.0016.
- Antonov AV, Dietmann S, Wong P, Mewes HW: TICL - a web tool for network-based interpretation of compound lists inferred by high-throughput metabolomics. *FEBS J* 2009, **276**:2084-2094.
- Antonov AV, Dietmann S, Mewes HW: KEGG spider: interpretation of genomics data in the context of the global gene metabolic network. *Genome Biol* 2008, **9**:R179.
- Jourdan F, Breiting R, Barrett MP, Gilbert D: MetaNetter: inference and visualization of high-resolution metabolomic networks. *Bioinformatics* 2008, **24**:143-145.
- Cottret L, Wildridge D, Vinson F, Barrett MP, Charles H, Sagot MF, Jourdan F: MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res* 2010, **38**:W132-137.
- Rahman SA, Schomburg D: Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks. *Bioinformatics* 2006, **22**:1767-1774.
- Fatumo S, Plaimas K, Mallm JP, Schramm G, Adebisi E, Oswald M, Eils R, König R: Estimating novel potential drug targets of *Plasmodium falciparum* by analysing the metabolic network of knock-out strains in silico. *Infect Genet Evol* 2009, **9**:351-358.
- Guimera R, Sales-Pardo M, Amaral LA: A network-based method for target selection in metabolic networks. *Bioinformatics* 2007, **23**:1616-1622.
- Deville Y, Gilbert D, van Helden J, Wodak SJ: An overview of data models for the analysis of biochemical pathways. *Brief Bioinform* 2003, **4**:246-259.
- de Figueiredo LF, Schuster S, Kaleta C, Fell DA: Can sugars be produced from fatty acids? A test case for pathway analysis tools. *Bioinformatics* 2008, **24**:2615-2621.
- de Figueiredo LF, Schuster S, Kaleta C, Fell DA: Response to comment on "Can sugars be produced from fatty acids? A test case for pathway analysis tools". *Bioinformatics* 2009, **25**:3330-3331.
- Faust K, Croes D, van Helden J: In response to "Can sugars be produced from fatty acids? A test case for pathway analysis tools". *Bioinformatics* 2009, **25**:3202-3205.
- Schuster S, Fell DA, Dandekar T: A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol* 2000, **18**:326-332.
- Klamt S, Haus UU, Theis F: Hypergraphs and cellular networks. *PLoS Comput Biol* 2009, **5**:e1000385.
- Schilling CH, Letscher D, Palsson BO: Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol* 2000, **203**:229-248.
- Rezola A, de Figueiredo LF, Brock M, Pey J, Podhorski A, Wittmann C, Schuster S, Bockmayr A, Planes FJ: Exploring metabolic pathways in

- genome-scale networks via generating flux modes. *Bioinformatics* 2011, **27**:534-540.
32. Terzer M, Stelling J: Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics* 2008, **24**:2229-2235.
 33. de Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S, Planes FJ: Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics* 2009, **25**:3158-3165.
 34. van Helden J, Wernisch L, Gilbert D, Wodak SJ: Graph-based analysis of metabolic networks. *Bioinformatics Genome Analysis* 2002, **38**:245-274.
 35. Faust K, Croes D, van Helden J: Metabolic pathfinding using RPAIR annotation. *J Mol Biol* 2009, **388**:390-414.
 36. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO: A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007, **3**:121.
 37. Planes FJ, Beasley JE: An optimization model for metabolic pathways. *Bioinformatics* 2009, **25**:2723-2729.
 38. Kotera M, Hattori M, Oh M, Yamamoto M, Komeno T, Yabuzaki Y, Tonomura K, Goto S, Kanehisa M: RPAIR: a reactant-pair database representing chemical changes in enzymatic reactions. *Genome Informatics* 2004, **15**:P062.
 39. Arita M: *In silico* atomic tracing by substrate-product relationships in *Escherichia coli* intermediary metabolism. *Genome Res* 2003, **13**:2455-2466.
 40. Blum T, Kohlbacher O: Using atom mapping rules for an improved detection of relevant routes in weighted metabolic networks. *J Comput Biol* 2008, **15**:565-576.
 41. Heath AP, Bennett GN, Kavradi LE: Finding metabolic pathways using atom tracking. *Bioinformatics* 2010, **26**:1548-1555.
 42. Schellenberger J, Park JO, Conrad TM, Palsson BO: BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* 2010, **11**:213.
 43. AtomMetaNetWeb.. [<http://www.kavrakilab.org/atommetanetweb/#home>].
 44. Eschenfeldt WH, Stols L, Rosenbaum H, Khambatta ZS, Quate-Randall E, Wu S, Kilgore DC, Trent JD, Donnelly MI: DNA from uncultured organisms as a source of 2,5-diketo-D-gluconic acid reductases. *Appl Environ Microbiol* 2001, **67**:4206-4214.
 45. Sauer U, Eikmanns BJ: The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol Rev* 2005, **29**:765-794.
 46. Fait A, Fromm H, Walter D, Galili G, Fernie AR: Highway or byway: the metabolic role of the GABA shunt in plants. *Trends Plant Sci* 2008, **13**:14-19.
 47. Lu CD: Pathways and regulation of bacterial arginine metabolism and perspectives for obtaining arginine overproducing strains. *Appl Microbiol Biotechnol* 2006, **70**:261-272.
 48. Kaleta C, de Figueiredo LF, Schuster S: Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. *Genome Res* 2009, **19**:1872-1883.
 49. Planes FJ, Beasley JE: A critical examination of stoichiometric and path-finding approaches to metabolic pathways. *Brief Bioinform* 2008, **9**:422-436.
 50. Baldoma L, Aguilar J: Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: aerobic-anaerobic regulation of L-lactaldehyde dissimilation. *J Bacteriol* 1988, **170**:416-421.
 51. Becker DJ, Lowe JB: Fucose: biosynthesis and biological function in mammals. *Glycobiology* 2003, **13**:41R-53R.
 52. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 2006, **34**:D354-357.
 53. Keseler IM, Bonavides-Martinez C, Collado-Vides J, Gama-Castro S, Gunsalus RP, Johnson DA, Krummenacker M, Nolan LM, Paley S, Paulsen IT, Peralta-Gil M, Santos-Zavaleta A, Shearer AG, Karp PD: EcoCyc: a comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res* 2009, **37**:D464-470.
 54. Kornberg H: Krebs and his trinity of cycles. *Nat Rev Mol Cell Biol* 2000, **1**:225-228.
 55. O'Donovan GA, Neuhard J: Pyrimidine metabolism in microorganisms. *Bacteriol Rev* 1970, **34**:278-343.
 56. Xi H, Schneider BL, Reitzer L: Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. *J Bacteriol* 2000, **182**:5332-5341.

doi:10.1186/gb-2011-12-5-r49

Cite this article as: Pey et al.: Path finding methods accounting for stoichiometry in metabolic networks. *Genome Biology* 2011 **12**:R49.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

