

## Can sugars be produced from fatty acids? A test case for pathway analysis tools

Luis F. de Figueiredo<sup>1,\*</sup>, Stefan Schuster<sup>1,\*</sup>, Christoph Kaleta<sup>2</sup> and David A. Fell<sup>3</sup>

<sup>1</sup>Department of Bioinformatics, <sup>2</sup>Bio Systems Analysis Group, Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany and <sup>3</sup>School of Life Sciences, Oxford Brookes University, Headington, Oxford, OX3 0BP, UK

Received on July 24, 2008; revised and accepted on September 18, 2008

Advance Access publication September 19, 2008

Associate Editor: Alfonso Valencia

### ABSTRACT

**Motivation:** In recent years, several methods have been proposed for determining metabolic pathways in an automated way based on network topology. The aim of this work is to analyse these methods by tackling a concrete example relevant in biochemistry. It concerns the question whether even-chain fatty acids, being the most important constituents of lipids, can be converted into sugars at steady state. It was proved five decades ago that this conversion using the Krebs cycle is impossible unless the enzymes of the glyoxylate shunt (or alternative bypasses) are present in the system. Using this example, we can compare the various methods in pathway analysis.

**Results:** Elementary modes analysis (EMA) of a set of enzymes corresponding to the Krebs cycle, glycolysis and gluconeogenesis supports the scientific evidence showing that there is no pathway capable of converting acetyl-CoA to glucose at steady state. This conversion is possible after the addition of isocitrate lyase and malate synthase (forming the glyoxylate shunt) to the system. Dealing with the same example, we compare EMA with two tools based on graph theory available online, PathFinding and Pathway Hunter Tool. These automated network generating tools do not succeed in predicting the conversions known from experiment. They sometimes generate unbalanced paths and reveal problems identifying side metabolites that are not responsible for the carbon net flux. This shows that, for metabolic pathway analysis, it is important to consider the topology (including bimolecular reactions) and stoichiometry of metabolic systems, as is done in EMA.

**Contact:** ldpf@minet.uni-jena.de; schuster@minet.uni-jena.de

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

### 1 INTRODUCTION

While the conversion of carbohydrates into fatty acids is experimentally well established, the existence of the converse transformation has long been discussed in biochemistry. This question was posed around the turn of the 19th century by Chaveau. While Pflüger stated that fat was the main source of sugar in diabetes, Lusk wrote that this was a figment of imagination (cf. Weinman

*et al.*, 1957). This controversy intensified in 1922 with the discovery of insulin and the extended work on diabetes.

In the 1950s, experiments using a new method involving isotopically labelled compounds started to reveal the mechanism by which carbons of fatty acids are incorporated in carbohydrates. Experiments showed that labelled carbons arrived at glucose when the system was supplied with <sup>14</sup>C-labelled fatty acids. The Krebs cycle (tricarboxylic acid cycle) seemed to play a key role in this process (Weinman *et al.*, 1957). Nevertheless, these experiments were not conclusive because the Krebs cycle, as other metabolic pathways, does not operate alone and the net synthesis was yet to be proved.

In 1957, the question around the net synthesis of carbohydrates from fatty acids being the most important constituents of lipids started being answered by Weinman *et al.*, who formulated an algebraical treatment of the problem and proved that fatty acids cannot give rise to a net gain of carbohydrate running along the Krebs cycle. The main conclusions from their work was that fatty acids can enter in the metabolite pool of the Krebs cycle but the net synthesis of glucose is due to an influx of other intermediates in the Krebs cycle, such as amino acids or lactic acid (Weinman *et al.*, 1957).

Since fatty acids in living organisms usually contain an even number of carbon atoms with the most common numbers being 16 and 18 (cf. Stryer, 1995), Weinman only analysed that case. Here, we will do the same, by considering acetyl-CoA (AcCoA) as the initial substrate. AcCoA results from the degradation of even-chain fatty acids and ketogenic amino acids (cf. Stryer, 1995). In the case where odd-chain fatty acids occur, such as in some plants and marine organisms (cf. Voet and Voet, 2004), a minor fraction of the products of  $\beta$ -oxidation of these acids is propionyl-CoA, which can, via succinyl-CoA, be converted to pyruvate and, thus, to glucose. Moreover, for both chain lengths, glucose can be produced from glycerol, which is part of phospholipids and triglycerides.

Also in 1957, Kornberg and Madsen (1957) published a paper describing the discovery of the ‘glyoxylate bypass’, an alternative route from isocitrate to malate. The key enzymes in this pathway are isocitrate lyase (formerly called isocitritase), which cleaves isocitrate, and malate synthase (formerly called malate synthetase), which catalyzes the condensation of AcCoA and glyoxylate to malate. This new route enables the conversion of acetate—and

\*To whom correspondence should be addressed.

therefore fatty acids—to carbohydrates with a stoichiometry of 1 mol of oxaloacetate (OAA) per 2 mol of AcCoA.

This discovery reopened the question about the possibility of transforming fatty acids into sugars, though in another perspective. It was connected to the new question of whether the glyoxylate cycle is present in humans. The first experiments showed the presence of the glyoxylate cycle in microbes and plants (Kornberg and Beevers, 1957; Kornberg and Madsen, 1957). Madsen was the first to report that the glyoxylate cycle is not present in animal tissue even under conditions in which one might expect it to occur like hibernating mammals and chick embryos because these must use their fat reservoirs (Madsen, 1958). The only clade of animals where the glyoxylate shunt was detected is that of the nematodes, where a bifunctional malate synthase/isocitrate lyase enzyme occurs (Liu *et al.*, 1995). The question around the presence of the glyoxylate cycle in animal tissues remains open since some authors claim the presence of isocitrate lyase and malate synthase (Davis and Goodman, 1992; Ganguli and Chakraverty, 1961; Goodman *et al.*, 1980; Jones, 1980; Morgunov *et al.*, 2005; Popov *et al.*, 2005) although the coding sequence of these enzymes in humans remains unknown and there is no homology with known sequences. Kondrashov *et al.* (2006) found the sequence of malate synthase, but not isocitrate lyase, in some animals besides nematodes.

Today, there is increased knowledge of biochemical networks, and genome-scale metabolic models have been established. But are we able to really handle such networks? Researchers pay special attention to topological properties of the metabolic model in order to redefine what metabolic pathways are. Recently, several methods have been proposed for determining metabolic pathways in an automated way based on network topology (Beasley and Planes, 2007; Croes *et al.*, 2005, 2006; Rahman *et al.*, 2005; Schuster *et al.*, 1999, 2000). It is of interest to see whether these methods can help answering the question posed in the title of this article and, in particular for didactic purposes in biochemistry, to revisit the study by Weinman *et al.* (1957).

The term ‘elementary flux mode’ refers to a minimal group of enzymes that can operate at steady state with all the irreversible reactions used in the right direction (Schuster *et al.*, 1999, 2000). If only the enzymes belonging to one elementary mode (EM) are operative and, thereafter, one of the enzymes is inhibited, then the remaining enzymes can no longer be operational because the system cannot any longer maintain a steady state. Several software tools were established for computing EMs, for example, METATOOL 5.0 (von Kamp and Schuster, 2006). Elementary modes analysis (EMA) has been applied to various systems (Cakir *et al.*, 2004; Carlson and Srien, 2004; Poolman *et al.*, 2003; Schwartz *et al.*, 2007; Stelling *et al.*, 2002; Wilhelm *et al.*, 2004). Also the Krebs cycle, glyoxylate shunt and adjacent reactions have been analysed by that method earlier, though not with the objective of the present article (Schuster *et al.*, 1999). A concept related to that of EMs is that of extreme pathways (Schilling *et al.*, 2000). A comparison of the two concepts was made by Klamt and Stelling (2003).

Any stationary flux distribution in the living cell is a linear combination of EMs (Schuster *et al.*, 1999). Therefore, if there is no EM consuming a given substrate or synthesizing a desired product, then we can conclude that there is no stationary flux distribution that would be able to consume that substrate or leading to that product.

Graph theory is another approach to studying metabolic networks based on the concept that these networks can be described as a simple graph (where nodes and edges represent metabolites and reactions, respectively) or as a bipartite graph (where two or more nodes, metabolites, connect to a common node of a second type, representing a reaction/enzyme). While EMA and graph-theoretical analyses of metabolic networks use the same input information, they usually produce different, complementary outputs (which should be consistent, though). A general comparison between the two methods can be found in Planes and Beasley (2008).

Based on graph-theoretical approaches, several computer programs have been presented. Pathway Hunter Tool (PHT; Rahman *et al.*, 2005) and PathFinding (Croes *et al.*, 2005, 2006) are freely available web tools, which can be used to reconstruct and analyse the shortest path connecting two metabolites. PHT uses a fingerprint algorithm to calculate the similarity between two molecules and in this way automatically assigns side metabolites (like ATP, ADP, water). Then a breadth-first-search algorithm calculates the shortest path between the seed and sink metabolites.

The approach underlying PathFinding (Croes *et al.*, 2005, 2006) is based on the connectivity of metabolites which is used to calculate the weight of paths between two metabolites or two reactions. Metabolites with a high connectivity will reduce the score of the path. The reason is that cofactors such as ATP are usually highly connected and should not be considered as intermediates on metabolic paths.

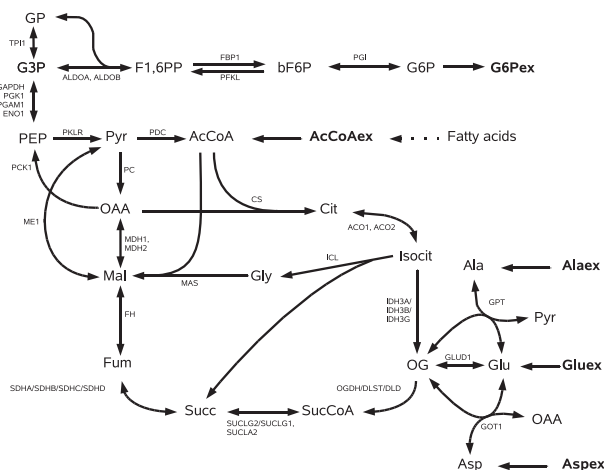
Here, we compare different tools for pathway finding, Metatool, PHT and PathFinding, by applying them to carbon metabolism in view of the question whether sugars can be produced from even-chain fatty acids. In Section 2, the reaction scheme to be analysed will be outlined. In Section 3, the results of the various tools will be presented and compared. The EMA of the lipid-sugar system considerably extends a preliminary analysis presented recently (Schuster and Fell, 2007). A final conclusion will be given in Section 4.

## 2 METHODS

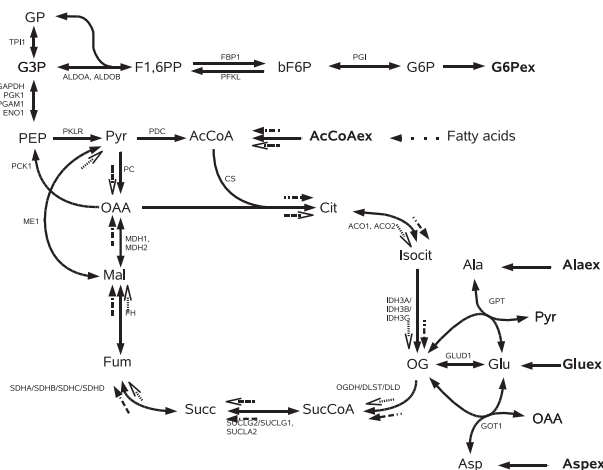
The system under study is composed of reactions present in the Krebs cycle, which is the pathway for the oxidation of AcCoA and, thus, even-chain fatty acids, and the reactions in glycolysis and gluconeogenesis, responsible for the catabolism and anabolism, respectively, of glucose. The initial model draft was reconstructed on the basis of the human model present in the KEGG database (Aoki-Kinoshita, 2006). The model was refined and completed with some anaerobic reactions using biochemistry textbooks (Michal, 1999; Nelson and Cox, 2000; Voet and Voet, 2004). The hypothesis of a carbon net flux using amino acids was also tested and external reactions were added to the first model enabling the influx of glutamate, aspartate and alanine. Using the first model as a template, a second model was generated by adding reactions catalyzed by isocitrate lyase and malate synthase to test the hypothesis that the glyoxylate cycle enables a net flux of carbons from AcCoA to  $\alpha$ -D-glucose-6-phosphate (G6P), see Figure 1.

For the methods of EMs, the reader is referred to Schuster *et al.* (1999, 2000) and Gagneur and Klamt (2004). For computing EMs, we used the program METATOOL 5.0 (von Kamp and Schuster, 2006), which implements an algorithm proposed by Urbanczik and Wagner (2005). The reaction list of the complete model containing the glyoxylate cycle is represented in the Supplementary Material.

In order to study automated pathway generating tools, we queried PathFinding (Croes *et al.*, 2005, 2006) and PHT (Rahman *et al.*, 2005)



**Fig. 1.** Complete model of human glycolysis/gluconeogenesis and Krebs cycle, containing reactions of the glyoxylate cycle. For abbreviations of metabolites, see Supplementary Material. Enzyme abbreviations are according to the database ExPASy. When separated by commas or slashes, they correspond to isoenzymes or multi-enzyme complexes, respectively. The external metabolites, that is, the source and sink metabolites are represented in bold-face or are omitted (e.g. the cofactors such as ATP, ADP and NAD). Reactions between PEP and G3P are lumped into one reaction.



**Fig. 2.** Two EMs of the model without glyoxylate cycle and no entry of external amino acids, where AcCoA is consumed (empty and full dashed arrows). The mode shown in full dashed arrows is the usual Krebs cycle.

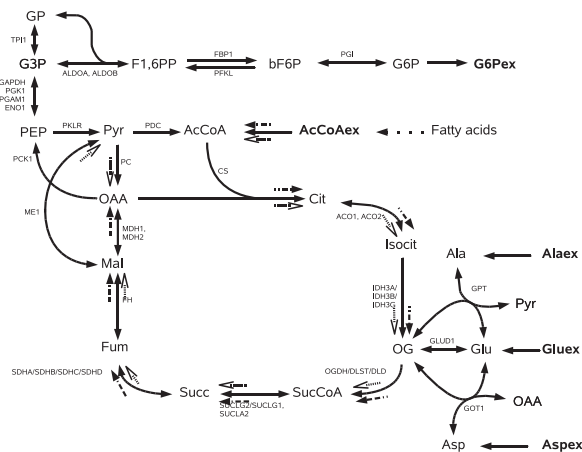
for a possible connection between AcCoA (KEGG entry C00024) and G6P (KEGG entry C00668).

### 3 RESULTS AND DISCUSSION

#### 3.1 EMs analysis

The first model containing no glyoxylate cycle, and with no influx of amino acids, resulted in six EMs. None of these produces G6P. Two of these consume AcCoA, go along the Krebs cycle, produce GTP, NADH and CO<sub>2</sub> (Fig. 2). The absence of EMs producing G6P and, thus, of an enzyme set able to synthesize G6P from AcCoA at steady state supports the hypothesis that it is impossible to synthesize glucose from fatty acids using the Krebs cycle and the gluconeogenic reactions only. This can be understood by inspecting Figure 2. To consume 1 mol of AcCoA, 1 mol of OAA is needed. Going around the Krebs cycle, this produces 1 mol of OAA. To produce G6P via PEP, one more mole of OAA would be needed. This cannot be formed at steady state, though. Another explanation is that two carbons enter the Krebs cycle by AcCoA and two leave it in the form of CO<sub>2</sub> (not shown in the Figures). Therefore, no carbon net flux can go to glucose. Nevertheless, if AcCoA is radioactively labeled, some of the labeled carbons flow to G6P because there is a connected route linking AcCoA with G6P and because some carbon atoms are actually transferred along the entire route. For example, if carbon 1 in acetate is labelled, then tracer is detected at carbons 3 and 4 in glucose (Weinman *et al.*, 1957).

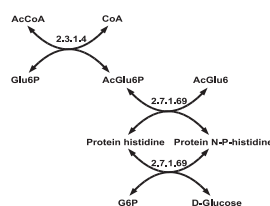
Then, we allowed for a carbon influx into the Krebs cycle from an additional source, for example, amino acids because this had also been analysed in Weinman *et al.* (1957). To simulate this, we added external reactions that enable the influx of the glucogenic amino acids glutamate, aspartate or alanine into the system (extended first model). This increased the number of EMs to 18. Among these,



**Fig. 3.** Two EMs of the model with glyoxylate cycle and no external reaction of amino acids, where AcCoA is consumed (empty and full dashed arrows) and G6P produced. The mode shown in full dashed arrows is the usual glyoxylate cycle.

five modes connect one of the amino acids each to G6P using at least OAA or 2-oxoglutarate as intermediaries (Supplementary Material). Thus, glucogenic amino acids can really generate a carbon flux towards G6P synthesis. The number of modes using an influx of AcCoA remained the same and none of those modes could synthesize G6P.

The second model contains the glyoxylate cycle, yet no influx from amino acids. This model gives rise to 11 EMs, two of which convert AcCoA to G6P, using isocitrate lyase and malate synthase in the glyoxylate shunt. Moreover, these two modes use part of the Krebs cycle (Fig. 3). The two modes differ in the use of the malic enzyme (ME1) and pyruvate carboxylase (PC) versus malate dehydrogenase (MDH). These results reinforce the hypothesis that the synthesis of glucose from fatty acids through the Krebs cycle is possible in the presence of enzymes from the glyoxylate cycle.



**Fig. 4.** Best scored path obtained from PathFinding with query 1. It involves part of the phosphotransferase system (PTS) present in bacteria only and is, thus, not relevant for humans. Moreover, no atom is transferred from AcCoA to G6P on that route. Weight = 178.0.

### 3.2 Graph-theoretical analysis

PathFinding at first glance has one major disadvantage when compared with PHT because it does not have an option to choose between different organisms. Therefore, we filtered the results by choosing, from the output, those enzymes that are present in humans. That information can easily be obtained from KEGG.

We queried PathFinding (April, 2008) to indicate 50 paths leading from AcCoA to G6P, and PathFinding is indeed able to detect that many. Figure 4 shows the path with the best score. However, from the molecular point of view, this path is not valid because it consumes D-glucose to produce G6P and, in the second and third reactions, only orthophosphate is transferred. In fact, not a single atom from AcCoA is transferred to G6P on that route. All of the paths generated for the first query are not present in humans, as results from a check with KEGG data.

Now we tried to find paths present in humans by splitting the path into two, choosing an intermediary metabolite that would connect both paths. The first metabolite chosen was (s)-malate (KEGG entry C00149) which takes part in the Krebs cycle and in the malate–aspartate shuttle as a precursor of OAA. Other metabolites chosen were phosphoenolpyruvate (PEP, KEGG entry C00074) and pyruvate (KEGG entry C00022) which are central metabolites in glycolysis and gluconeogenesis (Table 1). The only query that did not output any result was query 6 (data not shown). The number of the paths (within the output list) present in human is represented in Table 1. The only paths connecting AcCoA to G6P were obtained combining query 4 with query 5, using PEP as intermediary.

Regarding the weight range of the paths, the lower the weight is, the more significant should be the path. For the paths shown in Table 1, the weight range seems to be in an acceptable range because the weights of the two paths resulting from query 7, which correspond to gluconeogenesis, are 210 and 211.

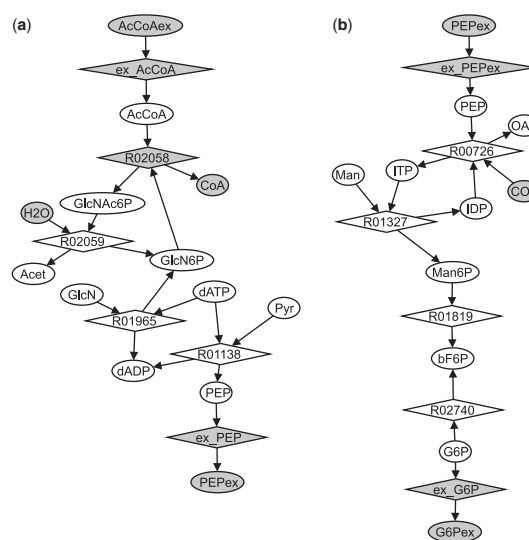
In the results of PathFinding, the connection between different reactions is established by cofactors, such as ITP, IDP, dATP and dADP. However, these compounds are not responsible for the carbon net flux. Figure 5 represents one of the possible connections between AcCoA to G6P when PEP is predefined as an obligatory intermediate. All the other possible paths are combinations between paths of queries 4 and 5. Additionally, it can be noted in Figure 5 that neither of the depicted paths is balanced at steady state.

PHT is easier to handle due to the organism selection option which enables one to choose only paths present in humans. Two other features of this algorithm are ‘Atom Mapper’ (molecular local similarity) and ‘Atom Tracer’ (molecular global similarity), which can be used to improve the results quality though they

**Table 1.** Queries of PathFinding and retrieved paths present in humans

Query	Start	Stop	Paths present in humans	
			Number	Weight Range
1	AcCoA	G6P	—	(178–203)
2	AcCoA	Mal	—	(85–204)
3	Mal	G6P	43	101
4	AcCoA	PEP	2; 16; 17	199–206
5	PEP	G6P	6; 16; 17; 35	78–87
6	AcCoA	Pyr	—	—
7	Pyr	G6P	44; 47	210–211

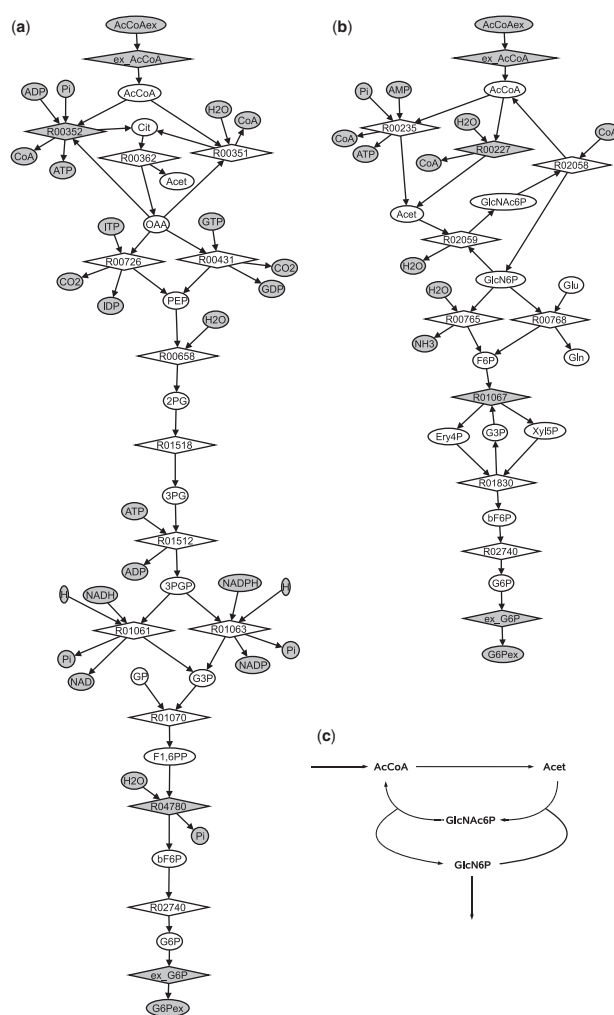
Weight ranges of paths not present in humans are given in parentheses. Tool options: Maximum weight = 2500; Maximum metabolic steps = 50; Mode = Weighted; Number of pathways = 50.



**Fig. 5.** (a) Path 2 from the results of PathFinding query 4; (b) Path 16 from the results of PathFinding query 5. Grey ellipses, external metabolites; white ellipses, internal metabolites; grey diamonds, irreversible reactions; white diamonds, reversible reactions. External reactions *ex\_AcCoA*, *ex\_PEP*, *ex\_G6P* added *a posteriori* for EMA.

may not work properly when metabolites do not have a defined structure, like macromolecules. In our analysis, activating both features simultaneously did not produce any paths. For this reason, we tested different combinations of these molecular similarity options (Supplementary Material). In Figure 6, the results obtained with PHT (April, 2008) by switching the ‘Atom Mapper’ *on* and leaving ‘Atom Tracer’ switched *off* and by switching both options *off* are shown.

The results from PHT are better regarding side metabolites because the chemical structure information is used to identify them (Fig. 6). The results of this algorithm were analysed by EMs. However, no such mode could be found, that is, there is no enzyme set capable of converting AcCoA to G6P. The path in Figure 6a resembles gluconeogenesis but is not balanced at steady state. This can clearly be seen in the figure because glycero phosphate (GP)



**Fig. 6.** (a) First result obtained by PHT for the conversion of AcCoA to G6P, with the features ‘Atom Mapper’ and ‘Atom Tracer’ switched *off*. (b) Result for the same query with the features ‘Atom Mapper’ and ‘Atom Tracer’ switched *on* and *off*, respectively. Grey ellipses, external metabolites; white ellipses, internal metabolites; grey diamonds, irreversible reactions, white diamonds, reversible reactions. External reactions *ex\_AcCoA* and *ex\_G6P* added *a posteriori* for EMA. (c) Scheme of part of the path shown in (b). There is no steady-state conversion of AcCoA to GlcN6P. For simplicity, the external metabolites were removed and reactions R00227 and R00235 were combined into one.

would be consumed in that path but not replenished. This imbalance can be resolved by including triose-phosphate isomerase, which interconverts G3P and GP. However, EMA shows that even in that case, transforming AcCoA to G6P is impossible at steady state because OAA is not balanced.

The path in Figure 6b can be shortened if we take into account the different levels of specificity at which substances are indicated in the KEGG database. In reaction R01067, generic D-fructose 6-phosphate (F6P) is indicated, while in reactions R01830 and R02740, β-D-fructose 6-phosphate (βF6P) is given. Even if reaction R01067 uses both the α and β forms of F6P, the detour via reactions R01067 and R01830 (both of which refer to transketolase,

EC 2.2.1.1) is unnecessary, since βF6P spontaneously anomerises to a mixture of α and β F6P (cf. Stryer, 1995). That means, F6P can be converted directly to G6P by phosphoglucose-isomerase (R02740). Moreover, from the structure of the path in Figure 6b, it is possible to identify the cycle schematically represented in Figure 6c. Equal amounts of D-glucosamine 6-phosphate (GlcN6P) are produced and consumed in the cycle, so that no drain to synthesize F6P is possible. Therefore, it cannot function as a pathway at steady state because GlcN6P cannot be balanced.

Looking carefully at the metabolite chemical structure in the cycle shown in Figures 6b and c, it can be seen that the atoms from the acetyl group transferred from AcCoA are not present in GlcN6P, which is connected to the rest of the path linking to G6P.

To demonstrate the generality of our results, we have checked another example, which concerns the question whether a pathway connecting G6P with pyruvate in bacteria lacking phosphofructokinase and G6P dehydrogenase exists. Pollack *et al.* (1997) proposed that such a pathway would exist in *Mycoplasma hominis*. Since *M.hominis* is not completely sequenced, its metabolism is not available from KEGG or similar databases. However, the completely sequenced *Bordetella pertussis* is comparable because in its genome, genes for phosphofructokinase and for the enzymes of the oxidative pentose pathway were not found (Armstrong and Gross, 2007). For (whatever) bacteria lacking the above-mentioned enzymes, an EMA had been performed in Schuster *et al.* (1999). It shows that G6P cannot then be converted to G3P at steady state by the glycolysis/pentose phosphate pathway and, thus, neither to pyruvate, although there is a connected route between them via the non-oxidative pentose phosphate pathway. Interestingly, both PHT and PathFinding output such a route (results given in the Supplementary Material).

## 4 CONCLUSIONS

It has long been considered that given an input of AcCoA from the breakdown of fatty acids or ketogenic amino acids, it is impossible for animals (except nematodes) to achieve net synthesis of glucose from this precursor by the Krebs cycle and gluconeogenesis. Although <sup>14</sup>C-labelled isotopes can pass along this apparent pathway, animals cannot make glucose from two-carbon precursors in substantial amounts at a sustained steady state.

By applying the method of EMs, we have here substantiated this fact and that, when the set of enzymes involved in the glyoxylate shunt are added, the system can synthesize glucose out of AcCoA. Green plants, many bacteria (cf. Stryer, 1995) and fungi (cf. Deacon, 2006) harbour that shunt and are indeed capable of converting AcCoA into glucose at steady state.

We have elaborated on an earlier sketch of a pathway analysis of the lipid-to-sugar transformation (Schuster and Fell, 2007). Among other extensions, we have here studied the possibility of amino acid consumption, have compared several path finding methods and have given a historical review of the subject. It should be noted that we have restricted our analysis to the Krebs cycle (optionally allowing the influx of amino acids), glyoxylate shunt and gluconeogenesis. It cannot be excluded that a conversion of fatty acids into sugars is found when larger (perhaps genome-scale) metabolic networks in animals are studied. Indeed, already Weinman *et al.* (1957)



mentioned the possibility of a conversion via acetone or acetoacetyl-CoA (see below), and this has been supported by subsequent studies (Hetenyi and Ferrarotto, 1985; Reichard *et al.*, 1979).

Moreover, there appear to be various pathways alternative to the glyoxylate shunt or even the Krebs cycle in some bacteria (cf. Ensign, 2006). For example, in *Rhodobacter sphaeroides*, 2 mol of AcCoA can be condensed to acetoacetyl-CoA and converted further to malate and succinate in a series of condensation, rearrangement and carboxylation reactions (cf. Ensign, 2006). In some Archaeans, such as *Ignicoccus hospitalis*, AcCoA can be carboxylated by pyruvate synthase to give pyruvate (Jahn *et al.*, 2007).

The stoichiometry plays an important role in the question under study putting in evidence a molecular constraint. The acetyl group, which is a two-carbon group, enters the Krebs cycle as AcCoA and in two successive reactions, catalyzed by isocitrate dehydrogenase and  $\alpha$ -ketoglutarate dehydrogenase, two carbons are converted into carbon dioxide and leave the cycle, although these are not the same atoms (Weinman *et al.*, 1957). Thus, the net carbon balance of an entire turn of the Krebs cycle is zero and the only way to synthesize glucose is to circumvent these decarboxylations or add a carbon source other than AcCoA. Another explanation of the role of the glyoxylate shunt is that it balances synthesis and use of OAA (see Section 3.1). In the absence of the glyoxylate shunt, a net flux of carbons from other carbon sources, like glucogenic amino acids, to G6P via the Krebs cycle is possible, in agreement with the work by Weinman *et al.* (1957).

The results presented above also indicate that automated pathway analysis is difficult. This is due to errors in metabolic databases (Poolman *et al.*, 2006), to ontological problems such as pointed out in Section 3.2 for the  $\alpha$  and  $\beta$  forms of F6P, and to combinatorial explosion in large networks (Klamt and Stelling, 2002). Therefore, we advocate that, at the present stage, metabolic networks constructed by extraction from databases should be checked carefully.

The information about network properties obtained by EMA is complementary to that derived from graph theory-based methods because of the high frequency of reactions with more than one substrate or product (e.g. bimolecular reactions) in metabolic networks. Due to the presence of such reactions, connectedness of a network does not necessarily imply a steady-state flow. Metabolic networks are more complicated than graphs in the sense of graph theory. Mathematically, they are hypergraphs.

Several authors have used graph-theoretical concepts to define metabolic pathways (Croes *et al.*, 2005, 2006; Jeong *et al.*, 2000; Ma and Zeng, 2003; Ma *et al.*, 2004; Rahman *et al.*, 2005; Seo *et al.*, 2001). In large-scale networks, these methods are indeed easier to apply than stoichiometric methods. However, paths traced on graphs may not be competent metabolic pathways. This is illustrated by the example of conversion of fatty acids into sugars. To make a distinction between (a) connected routes in the sense of graph theory and (b) pathways that are able to carry a net flux at steady state, a distinction in terminology appears to be necessary and helpful. The terms path and pathway could be used for (a) and (b), respectively (cf. Beasley and Planes, 2007; Planes and Beasley, 2008). Routes detected by graph theory are of interest, for example, for the flow of radioactive tracer.

We here critically examined two tools for finding paths, PathFinding and PHT. They did succeed in finding paths connecting AcCoA to glucose. However, none of them is a biochemically

relevant pathway. Though the paths generated by these algorithms are connected they cannot, at steady state, synthesize G6P out of AcCoA and some do not even realize an overall transfer of carbon atoms. This example illustrates that if only the connectedness of the graph is considered and the stoichiometric constraints are neglected, then it is likely that non-functional pathways will be postulated. Another example is monosaccharide metabolism in *M. hominis* and *B. pertussis*, for which graph-theoretical methods again predict invalid pathways from G6P to pyruvate.

Another drawback of the graph-theoretical approaches mentioned above (methods using bipartite graphs excepted) is that cycles cannot be easily obtained because they search for linear paths that connect metabolite A to metabolite B not taking into account metabolites that are not synthesized by the path. Nevertheless, as the paths found in the results of PathFinding and PHT (see Section 3.2) show, certain types of cycles can be obtained. One type can occur where there is more than one reaction synthesizing the same product using the same substrate (like the reaction converting 3PGP into G3P in Fig. 6a). Another type can be obtained when one of the substrates in a path (such as GlcNAc6P in Fig. 6b) occurs as a product of a reaction further down in the path. As observed in the above results of the programs PHT and PathFinding, it is not possible to obtain a non-trivial cycle or cyclic pathways like the Krebs cycle using these algorithms, probably also due to the fact that they search for the shortest pathway or the pathway with the lowest weight. It is a well-known biochemical fact that complex metabolisms involve cyclic pathways, such as the Krebs cycle or the urea cycle. Therefore, algorithms for detecting them are useful.

One option for using graph-theoretical methods also for detecting pathways is to use the theory of Petri nets, which are bipartite graphs. Metabolites and reactions are then represented by two different types of nodes (cf. Koch *et al.*, 2005; Zevedei-Oancea and Schuster, 2003). Another option (used here) is to choose an algebraic treatment such as in EMA, which properly takes into account stoichiometry. The problem of combinatorial explosion in large networks could be solved by using linear programming approaches, by which only specific pathways are computed (Beasley and Planes, 2007; Feist and Palsson, 2008; Fell and Small, 1986). However, a fully automated solution cannot easily be achieved by such approaches either because the proper definition of side metabolites is context dependent.

## ACKNOWLEDGEMENTS

The authors would like to thank three anonymous referees for very helpful comments. The groups in which L.F.F., S.S. and C.K. work belong to the Jena Centre for Bioinformatics (JCB). The scientific atmosphere in the JCB has significantly stimulated this work.

*Funding:* Portuguese entities: Fundação Calouste Gulbenkian, FCT and Siemens SA Portugal (PhD grant SFRH/BD/32961/2006 to L.F.F.). Biotechnology and Biosciences Research Council (BB/E00203X/1 to D.A.F.).

*Conflict of Interest:* none declared.

## REFERENCES

- Aoki-Kinoshita, K.F. (2006) Overview of KEGG applications to omics-related research. *J. Pesticide Sci.*, **31**, 296–299.

- Armstrong,S.K. and Gross,R. (2007) Primary metabolism and physiology of *Bordetella* species. In Loch,C. (ed.) *Bordetella: Molecular Microbiology*, ch. 8. 1 edn. Horizon Bioscience, Norwich, pp. 165–190.
- Beasley,J.E. and Planes,F.J. (2007) Recovering metabolic pathways via optimization. *Bioinformatics*, **23**, 92–98.
- Cakir,T. et al. (2004) Metabolic pathway analysis of enzyme-deficient human red blood cells. *BioSystems*, **78**, 49–67.
- Carlson,R. and Srien,F. (2004) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: identification of reactions. *Biotechnol. Bioeng.*, **85**, 1–19.
- Croes,D. et al. (2005) Metabolic PathFinding: inferring relevant pathways in biochemical networks. *Nucleic Acids Res.*, **33**(Web Server issue), W326–W330.
- Croes,D. et al. (2006) Inferring meaningful pathways in weighted metabolic networks. *J. Mol. Biol.*, **356**, 222–236.
- Davis,W.L. and Goodman,D.B. (1992) Evidence for the glyoxylate cycle in human liver. *Anat. Rec.*, **234**, 461–468.
- Deacon,J. (ed.) (2006) *Fungal Biology*. 4th edn. Blackwell Publishing, Oxford.
- Ensign,S.A. (2006) Revisiting the glyoxylate cycle: alternate pathways for microbial acetate assimilation. *Mol. Microbiol.*, **61**, 274–276.
- Feist,A.M. and Palsson,B.O. (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.*, **26**, 659–667.
- Fell,D.A. and Small,J.R. (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem. J.*, **238**, 781–786.
- Gagneur,J. and Klamt,S. (2004) Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, **5**, 175.
- Ganguli,N. and Chakraverty,K. (1961) Evidence for malic synthetase in animal tissues. *J. Am. Chem. Soc.*, **83**, 2581–2583.
- Goodman,D.B. et al. (1980) Glyoxylate cycle in toad urinary bladder: possible stimulation by aldosterone. *Proc. Natl Acad. Sci. USA*, **77**, 1521–1525.
- Hetenyi,G. and Ferrarotto,C. (1985) Gluconeogenesis from acetone in starved rats. *Biochem. J.*, **231**, 151–155.
- Jahn,U. et al. (2007) Insights into the autotrophic CO<sub>2</sub> fixation pathway of the archaeon *Ignicoccus hospitalis*: comprehensive analysis of the central carbon metabolism. *J. Bacteriol.*, **189**, 4108–4119.
- Jeong,H. et al. (2000) The large-scale organization of metabolic networks. *Nature*, **407**, 651–654.
- Jones,C.T. (1980) Is there a glyoxylate cycle in the liver of the fetal guinea pig? *Biochem. Biophys. Res. Commun.*, **95**, 849–856.
- Klamt,S. and Stelling,J. (2002) Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.*, **29**, 233–236.
- Klamt,S. and Stelling,J. (2003) Two approaches for metabolic pathway analysis? *Trends Biotechnol.*, **21**, 64–69.
- Koch,I. et al. (2005) Application of Ketri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. *Bioinformatics*, **21**, 1219–1226.
- Kondrashov,F.A. et al. (2006) Evolution of glyoxylate cycle enzymes in metazoa: evidence of multiple horizontal transfer events and pseudogene formation. *Biol. Direct*, **1**, 31.
- Kornberg,H.L. and Bevers,H. (1957) The glyoxylate cycle as a stage in the conversion of fat to carbohydrate in castor beans. *Biochim. Biophys. Acta*, **26**, 531–537.
- Kornberg,H.L. and Madsen,N.B. (1957) Synthesis of C<sub>4</sub>-dicarboxylic acids from acetate by a glyoxylate bypass of the tricarboxylic acid cycle. *Biochim. Biophys. Acta*, **24**, 651–653.
- Liu,F. et al. (1995) Bifunctional glyoxylate cycle protein of *Caenorhabditis elegans*: a developmentally regulated protein of intestine and muscle. *Dev. Biol.*, **169**, 399–414.
- Ma,H. and Zeng,A.-P. (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, **19**, 270–277.
- Ma,H.-W. et al. (2004) Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. *Bioinformatics*, **20**, 1870–1876.
- Madsen,N.B. (1958) Test for isocitritase and malate synthetase in animal tissues. *Biochim. Biophys. Acta*, **27**, 199–201.
- Michal,G. (1999) *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*. 1st edn. Spektrum Akademischer Verlag, Heidelberg, Berlin.
- Morgunov,I.G. et al. (2005) Evidence of the glyoxylate cycle in the liver of newborn rats. *Med. Sci. Monit.*, **11**, BR57–BR60.
- Nelson,D.L. and Cox,M.M. (2000) *Lehninger Principles of Biochemistry*. 3rd edn. W. H. Freeman, New York.
- Planes,F.J. and Beasley,J.E. (2008) A critical examination of stoichiometric and path-finding approaches to metabolic pathways. *Brief. Bioinformatics*, **9**, 422–436.
- Pollack,J.D. et al. (1997) The comparative metabolism of the mollicutes (*Mycoplasmata*): the utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cells. *Crit. Rev. Microbiol.*, **23**, 269–354.
- Poolman,M.G. et al. (2003) Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. *Eur. J. Biochem.*, **270**, 430–439.
- Poolman,M.G. et al. (2006) Challenges to be faced in the reconstruction of metabolic networks from public databases. *IEE Proc. Syst. Biol.*, **153**, 379–384.
- Popov,V.N. et al. (2005) Comparative analysis of the glyoxylate cycle clue enzyme isocitrate lyases from organisms of different systemic groups. *J. Evol. Biochem. Physiol.*, **41**, 507–513.
- Rahman,S.A. et al. (2005) Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). *Bioinformatics*, **21**, 1189–1193.
- Reichard,G.A. et al. (1979) Plasma acetone metabolism in the fasting human. *J. Clin. Invest.*, **63**, 619–626.
- Schilling,C.H. et al. (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.
- Schuster,S. and Fell,D. (2007) Modelling and simulating metabolic networks. In Lengauer,T. (ed.) *Bioinformatics: From Genomes to Therapies*. Vol. 2. Wiley-VCH, Weinheim, pp. 755–806.
- Schuster,S. et al. (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.*, **17**, 53–60.
- Schuster,S. et al. (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. et al. (2007) Observing metabolic functions at the genome scale. *Genome Biol.*, **8**, R123.
- Seo,H. et al. (2001) Graph-theoretical identification of pathways for biochemical reactions. *Biotechnol. Lett.*, **23**, 1551–1557.
- Stelling,J. et al. (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature*, **420**, 190–193.
- Stryer,L. (1995) *Biochemistry*. 4th edn. W.H. Freeman and Company, New York.
- Urbanczik,R. and Wagner,C. (2005) An improved algorithm for stoichiometric network analysis: theory and applications. *Bioinformatics*, **21**, 1203–1210.
- Voet,D. and Voet,J. (2004) *Biochemistry*. 3rd edn. Wiley, USA.
- von Kamp,A. and Schuster,S. (2006) Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, **22**, 1930–1931.
- Weinman,E.O. et al. (1957) Conversion of fatty acids to carbohydrate: application of isotopes to this problem and role of the Krebs cycle as a synthetic pathway. *Physiol. Rev.*, **37**, 252–272.
- Wilhelm,T. et al. (2004) Analysis of structural robustness of metabolic networks. *Syst. Biol. (Stevenage)*, **1**, 114–120.
- Zevedei-Oancea,I. and Schuster,S. (2003) Topological analysis of metabolic networks based on Petri Net theory. *In Silico Biol.*, **3**, 323–345.