In silico prediction of elementary mode fluxes

Jean-Marc Schwartz[1], Peter Neal Taylor[2]

[1] Faculty of Life Sciences, University of Manchester, Manchester, M13 9PT, UK
[2] School of Computing Science, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

Abstract. How perturbations such as enzyme deletions or changes in nutrient composition affect the distribution of fluxes in a metabolic system remains poorly understood. Elementary modes are a mathematical concept enabling the decomposition of complex metabolic networks into minimal chains of reactions that operate at steady state. They were proven to be a useful tool for analyzing the range of possible functional modes of metabolic systems. We here report an investigation into how the usage of elementary modes changes with substrate input and gene mutation. The key method of this study involved the computation of an elementary mode flux value for each elementary mode corresponding to different flux distributions in the central metabolism of *E. coli*. Our results indicate that elementary modes which have no external outputs (excluding co-factors) tend to carry higher flux. Using an entirely *in silico* approach, we also propose a novel method for estimating elementary mode flux values based on an uptake score and show that its predictions are in better agreement with experimental findings than previous estimations based on biomass production.

1 Introduction

The use of stoichiometry based approaches in systems biology has the advantage that kinetic parameters, which are often difficult to obtain experimentally, are not required. Elementary modes analysis is one form of stoichiometric analysis technique for metabolic networks. An elementary flux mode (EM) represents a steady-state metabolic flux distribution, with the proportions of fluxes in different reactions adequately fixed to satisfy stoichiometric consistency and mass conservation, and which is non-decomposable [1]. This means that the set of biochemical reactions composing an EM cannot be split whilst still maintaining the same rate of turnover of molecules through the system. It is therefore often described as a "minimal set" of reactions.

At steady state each elementary mode can be assigned an elementary mode flux value. Poolman *et al.* [2] introduced the term "elementary mode flux" (EMF) to refer to a numerical value representative of the turnover rate of molecules through a particular elementary mode, rather than a particular reaction. Calculation of EMF values requires the use of a mathematical function which gives the set of EMFs (**w**) given the matrix of EMs (**E**) and the observed set of fluxes through each reaction (**v**). Formally, $\mathbf{v} = \mathbf{E} \cdot \mathbf{w}$; however, this relation can generally not be inverted to give **w** as a function of **v** as the number of EMs (rows in **E**) is usually far larger than the number of reac-

tions, making **E** non-invertible. It is therefore necessary to define a function (*f*) to calculate EMF values as

$$\mathbf{w} = f(\mathbf{E}, \mathbf{v}) \quad (1)$$

Several functions have been proposed and implemented to solve this problem. Poolman *et al.* [2] used the concept of the Moore-Penrose generalized inverse, Schwartz & Kanehisa [3] used a quadratic programming approach, and Chan & Ji proposed a mixed integer linear approach [4]. Further to this, Wiback *et al.* [5] showed that, under the analysis of extreme pathways (a particular subset of elementary modes), an α-spectrum can be defined as the allowable range of extreme pathway contributions to a given flux distribution.

A vector of flux values for each reaction (**v**) is required for the calculation of EMFs. Obtaining such values experimentally can be expensive and time consuming. On the other hand, modeling using simulated fluxes is not always possible due to a lack of kinetic data. Stelling *et al.* [6] defined a formal method to estimate the efficiency of EMs and applied this principle to relate network structure to function. They introduced the concept of control effective flux (CEF) values, which measure how important reactions are to the operation of the network as a whole. Such techniques require only the network topology in the form of a stoichiometric matrix and can therefore be implemented entirely *in silico*. The availability of a full data set of flux values in both wild-type and a pyruvate kinase (pykF) knockout mutant of *E. coli* by Kurata *et al.* [7] and Emmerling *et al.* [8] provides an opportunity for the use of such data to help understanding the effect of such perturbation on metabolic functions.

The primary goal of this work was to determine EMF values for the central metabolic network of *E. coli* for different substrate inputs in both wild type and pykF knockout mutant and investigate the principles directing the redistribution of fluxes. Further to this, we adapted the control effective flux method proposed by Stelling *et al.* [6] to EMF analysis in order to predict, without experimental data, which Ems would have significant activity. We additionally introduce a new method for the prediction of EMFs based on the calculation of an uptake score and we compare the performance of this approach with the concept of efficiency as defined by Stelling *et al.*

2 Methods

The JaPathways package was used for the computation of elementary modes and elementary mode fluxes [9,10]. For the decomposition into EMFs to be successful it is essential that fluxes are strictly conserved in the original metabolic network. Since experimentally measured fluxes have limited precision there are cases where the original flux distribution is not entirely consistent. For example, one discrepancy noted was that of the flux of the AcCoA to Acetate reaction in the distribution provided by Kurata *et al.* [7]. Flux conservation implied that the flux should be 20 whereas the actual value provided was 19. Similar cases were encountered in the Emmerling data [8]. These values were adjusted before performing the EMF calculation to allow for a consistent decomposition.

To obtain EMF values for *E. coli*, JaPathways was executed using the experimental reaction flux data for wild type and for pykF knockout mutant provided by Kurata *et al.* [7]. The method was repeated for the network and corresponding flux data described by Emmerling [8]. To verify the EMs produced by JaPathways the same stoichiometric data was used with CellNetAnalyzer [11]. Compression techniques were applied to reduce the size of the system of equations as described by Gagneur & Klamt [12]. EMs and EMFs were compared using the compressed and the uncompressed version of the stoichiometric data file and the output checked to be identical. The compressed network is represented along with the original uncompressed network in Figure 1.

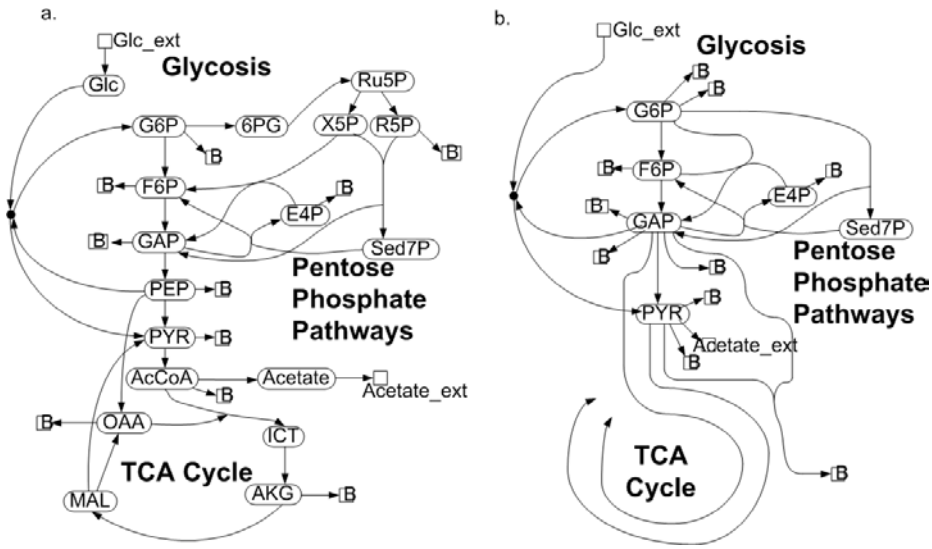


Fig. 1. The *E. coli* metabolic network (a) before and (b) after compression. External metabolites (represented by squares, biomass represented by 'B') are not included in the stoichiometric matrix but are shown here for better clarity. Filled circles indicate reactions with multiple products and substrates.

Three different methods were used to estimate EMFs using an entirely *in silico* approach. First, control effective flux values as defined by Stelling [6] were used as input to JaPathways, along with the stoichiometric matrix. These values were then compared to the EMF values obtained using the experimental data to test this *in silico* approach. Second, efficiency values (e_j) for each individual EM (j) were calculated using the method described by Stelling [6]. The absolute flux of each EM (i.e. the sum of the stoichiometric coefficients in the mode) was divided by the absolute stoichiometric flux of extracellular glucose (Glc_ext). This gave a value (a_j) representative of the absolute flux of each EM, normalized by the input. The absolute amount of normalized biomass (b_j) produced by the EM was then divided by a_j. This value is being taken as representative of efficiency. Essentially this method optimizes for a small absolute flux (few reactions with little turnover) and high biomass production.

$$e_j = \frac{b_j}{a_j} \tag{2}$$

A third method, which we introduce here, calculates what we call an 'uptake' score (u_j). We begin by using the same method as above to calculate a_j. However, we differ from Stelling's approach by then dividing these summed values by the amount of Glc_ext required (g_j) which is the stoichiometric uptake flux in the corresponding EM. In essence one can infer that the more Glc_ext is required the more costly the EM is and therefore the denominator becomes larger ensuring a smaller estimated value. Whereas Stelling's method optimizes for biomass production, our method uses a similar technique to optimize for input consumption.

$$u_j = \frac{a_j}{g_j} \tag{3}$$

3 Results

3.1 Elementary mode fluxes

The EM with the highest flux value for wild type is shown in Figure 2 (EM23). It has an EMF value of 7.78, which is 26% greater than the EM with the second highest EMF value of 5.73. The EM with the highest EMF value in the pykF knockout mutant is EM57 with a value of 3.29, which is 23% greater than the second highest EMF which has a value of 2.54 (Figure 2).

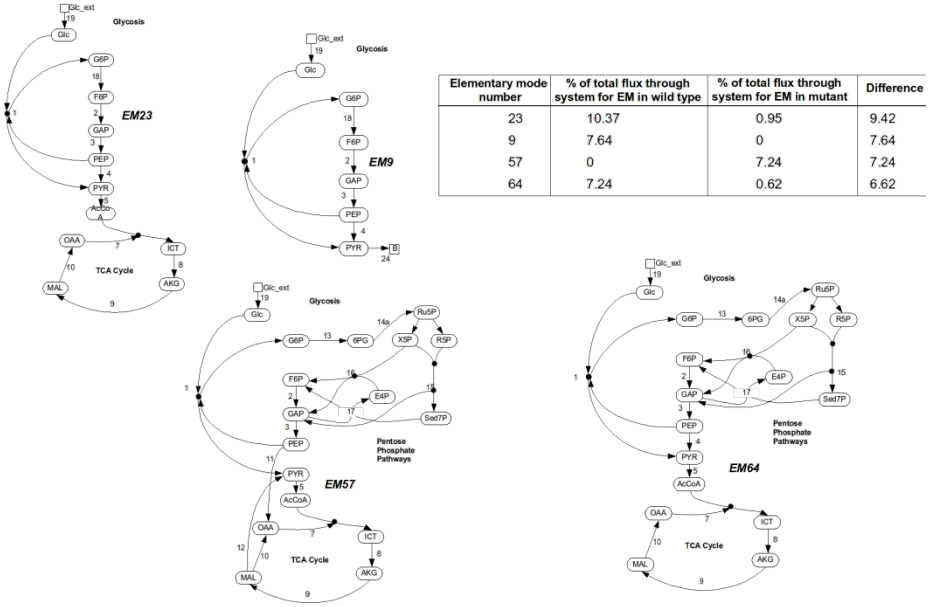
Overall, 30 of the 73 EMs found had an EMF value greater than zero in the wild type and 46 had an EMF value greater than zero in the pykF knockout mutant. Figure 3 shows the distribution of EMF values amongst the EMs for both wild type and mutant. It shows that the mutant has a broader distribution of EMF values; the mutant tends to use more EMs but with smaller flux through each one.

3.2 Throughput of the TCA cycle and pentose phosphate pathway

Of the 73 EMs found, 24 include the tricarboxylic acid (TCA) cycle and 44 include the pentose phosphate (PP) pathway; 14 of these include both the TCA cycle and pentose phosphate pathway. The remaining 19 EMs contain neither of them.

In wild type the sum of the EMF values for EMs involving the pentose phosphate pathways represents 20% of the total produced flux through the whole network. EMs using the TCA cycle account for 51% of the total flux through the network.

In the pykF mutant the sum of the EMF values for EMs including the pentose phosphate pathway represents 63% of the total produced flux through the whole network. This is an increase of 43% over wild type. The sum of the EMF values for EMs involving the TCA cycle in the pykF knockout mutant accounts for 68% of the total flux through the system, an increase of 17% over wild type.



Elementary mode number	% of total flux through system for EM in wild type	% of total flux through system for EM in mutant	Difference
23	10.37	0.95	9.42
9	7.64	0	7.64
57	0	7.24	7.24
64	7.24	0.62	6.62

Fig. 2. The four elementary modes which have the biggest change in elementary mode flux value between wild type and mutant based on the Kurata *et al.* data. EM57 has an increase in flux in the mutant whereas EM23, EM9 and EM64 have a decreased elementary mode flux value. Numbers next to reaction arrows correspond to reaction ID.

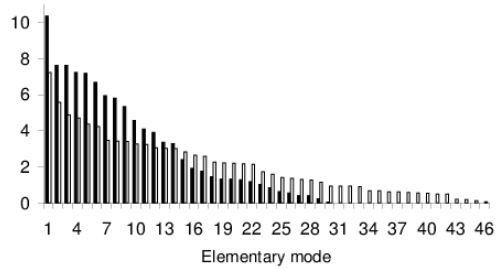


Fig. 3. Distribution of elementary mode flux values for wild type (black bars) and pykF mutant (white bars) *E. coli* as a percentage of the sum of all elementary mode flux values.

3.3 Efficiency and control effective flux

The availability of EMF values calculated from experimental data provided an opportunity to test and compare several techniques for *in silico* prediction of such values. Firstly efficiency values calculated as proposed by Stelling *et al.* [6] are shown in comparison to EMF values derived from the experimental data in Figure 4a. Many of the EMs are not efficient according to Stelling's definition of efficiency because they do not produce biomass and therefore have values of zero; however our results show that many "inefficient" Ems do indeed carry fluxes greater than zero.

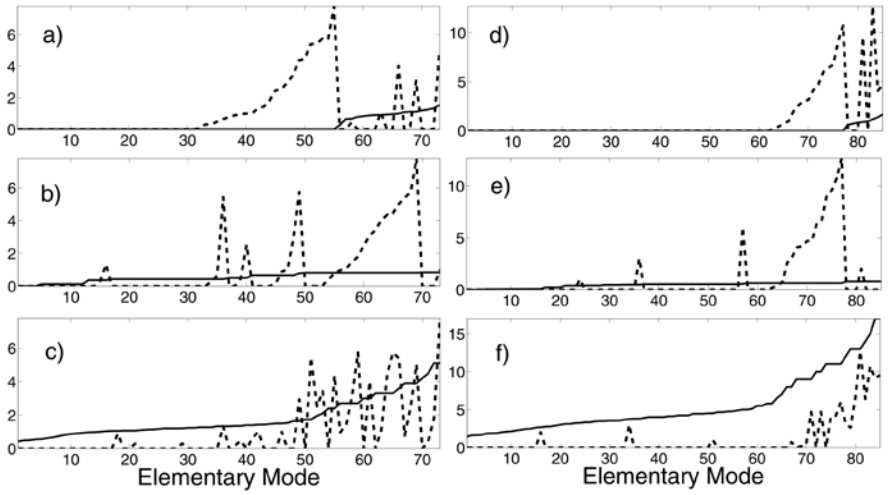


Fig. 4. Comparison between elementary mode fluxes (dashed lines) and different *in silico* scores (solid lines) for the set of 73 elementary modes from the network of Kurata *et al.* (a-c) and for the set of 83 elementary modes calculated using the nitrogen limited conditions described by Emmerling *et al.* (d-f). (a,d) Solid lines represent efficiency values calculated using Stelling's definition of efficiency. (b,e) Solid lines represent control effective flux (CEF) values as defined by Stelling and decomposed by JaPathways. (c,f) Solid lines represent uptake scores (equation 3). Elementary modes are sorted in each graph according to increasing scores. Values for solid lines in (a,d) and (c,f) have been multiplied by 10 and 0.3 respectively for scale.

Secondly, control effective fluxes (CEF), generally described as measuring the importance of a reaction, were used as input to the EMF computation algorithm in the same way as the experimental flux values were. A comparison between the EMF values for these two situations is shown in Figure 4b. Qualitatively speaking the approximation is better with many of the EMs with positive fluxes having higher values.

Finally, our approach using a modified version of efficiency (which we term the uptake score) to optimize for incoming Glc_ext rather than biomass production, is shown in Figure 4c. This new method appears most accurate and correctly predicts the EM with the highest value. The method also provides a better estimate for many other EMs; all EMs carrying high flux are predicted to have a high uptake score. There appears to be a stronger qualitative correlation between the uptake score and the EMF value.

Since the flux distribution strongly depends on environmental conditions we repeated this analysis using the alternative set of experimental conditions described by Emmerling *et al.* [8]. They experimentally measured reaction fluxes in the central metabolic network of *E. coli* in glucose limited and nitrogen limited conditions. In glucose limited conditions the reaction fluxes (and consequently the elementary mode fluxes) are highly similar to those described by Kurata *et al.* (data not shown). In nitrogen limited conditions though, some of the reaction fluxes are significantly different. The direction of the flux is inverted in the transketolase II reaction bringing more flux into the pentose phosphate pathway. Because EMs in the Kurata network are

irreversible, additional EMs are required to account for this opposite directionality of the flux. This results in a new set of 85 elementary modes. There is again a stronger correlation between the uptake score and the EMF values (Figure 4f) compared to previous methods (Figure 4d and e).

4 Discussion

We showed in this study that the quantification of flux distributions by elementary mode fluxes can cast new light on the principles directing the redistribution of fluxes in an enzymatic knockout. These principles may be used to investigate in a more systematic manner, without experimental data, the effect of gene deletions on metabolic activity and growth. Of the complete set of 73 EMs in the central metabolic network of *E. coli*, there are 24 which involve the TCA cycle. This represents a minority when compared to the number of EMs involving the PP pathway (44). Without the consideration of flux information, these numbers could suggest that the PP pathway should carry a greater proportion of the total flux through the system, but there is no clear correlation between the number of EMs and the actual flux carried by the reactions they contain. To consider only the topology of EMs is thus unlikely to be sufficient to reach biological conclusions about the organization of fluxes. Similarly, to consider only the flux values of individual reactions can also lead to incorrect conclusions. For example, the flux data from Kurata *et al.* [7] shows reaction 9 (α-ketoglutarate to malate) to have a flux of 73 in the pykF mutant, whereas reaction 13 (glucose-6-phosphate to 6-phosphogluconate) has a greater flux of 79 in the mutant. This flux data could give the incorrect impression that the PP pathway carries a greater proportion of the total flux through the system than the TCA cycle.

The integration of flux data with EMs via the computation of EMFs enables a higher level of analysis. Considering EMF values implies that, instead of considering fluxes in terms of individual reactions, flux levels are considered in terms of entire processes or paths. EMFs indicate that the EMs which contain the TCA cycle carry a greater proportion of the total system flux than those containing the PP pathway in both wild type and mutant, utilizing 51% and 68% of the total throughput of the system respectively. The TCA cycle is crucial for life and converts carbohydrates, fats and proteins into water and carbon dioxide to generate energy for the organism. As the TCA cycle is critical, one could expect that in wild type the simplest and most optimal path through the system which incorporates the TCA cycle would also carry the highest flux. This is indeed the case with EM23, as shown in Figure 2 when optimality is considered to be adapted for uptake glucose (Glc_ext), rather than for biomass production. This EM has no external outputs other than cofactors and is therefore extremely optimized in substrate usage. Furthermore, our estimation based on uptake score shows a better agreement with a range of EM flux values than previous estimations based on biomass production. We demonstrated the repeatability of this result in different environmental conditions, showing that the uptake score has better predictive power in carbon limited and nitrogen limited experiments. These results are also consistent with other observations made in yeast [13] which showed that low

yield pathways synthesizing important products can be favored over pathways maximizing molar yield. The maximization of molar yield thus does not seem to be a universal principle. These observations show that EMFs are able to quantify results that could be qualitatively deducted by biological intuition, adding justification to the use of a quadratic decomposition approach to quantify elementary mode fluxes.

A possible explanation for the large increase in the percentage of total throughput of the system for EMs including the PP pathway in the pykF mutant (from 20% up to 63%, Table 1) may be that many of them serve to supply pyruvate through alternative pathways. Many of the alternative paths that can provide pyruvate to the TCA cycle use the PP pathway. The pykF gene codes for an enzyme that catalyzes an essential glycolytic reaction leading to the production of pyruvate, which therefore must be supplied by alternative routes in the mutant. One may expect the large change from 20% to 63% to have a catastrophic effect on the organism, especially since the PP pathway accounts for 35-40% of the total NADPH generated [14]; however, the deletion of the glucose-6-phosphate 1-dehydrogenase (zwf) gene which is crucial to the PP pathway has been shown to have negligible effects on the specific growth rate and biomass yield [15]. This means that the organism can cope with large variations in the throughput of the PP pathway.

There exists a hereditary disorder known as pyruvate kinase deficiency which can be caused by a mutation to the pykF gene. Sufferers of the disorder have symptoms including lethargy and fatigue. The extra usage of other EMs in *E. coli* shows that, in order to get enough throughput through the TCA cycle, much energy is used by including reactions from less efficient modes through the PP pathway. It may be worth investigating the possibility of increasing the throughput of EMs that include the TCA cycle and not the PP pathway, since the regulation of these enzymes can be very expensive in terms of energy.

Acknowledgement

PNT was supported by the Biotechnology and Biological Sciences Research Council (BBSRC).

References

1. Schuster, S., Hilgetag, C., Woods, J.H., Fell, D.A.: Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. J. Math. Biol. 45, 153–181 (2002).
2. Poolman, M.G., Venkatesh, K.V., Pidcock, M.K., Fell, D.A.: A method for the determination of flux in elementary modes, and its application to *Lactobacillus rhamnosus*. Biotechnol. Bioeng. 88, 601–612 (2004).
3. Schwartz, J.M., Kanehisa, M.: A quadratic programming approach for decomposing steady-state metabolic flux distributions onto elementary modes. Bioinformatics 21 Suppl 2, ii204–ii205 (2005).
4. Chan, S.H.J., Ji, P.: Decomposing flux distributions into elementary flux modes in genome-scale metabolic networks. Bioinformatics 27, 2256–2262 (2011).

5. Wiback, S.J., Mahadevan, R., Palsson, B.Ø.: Reconstructing metabolic flux vectors from extreme pathways: defining the alpha-spectrum. J. Theor. Biol. 224, 313–324 (2003).
6. Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E.D.: Metabolic network structure determines key aspects of functionality and regulation. Nature 420, 190–193 (2002).
7. Kurata, H., Zhao, Q., Okuda, R., Shimizu, K.: Integration of enzyme activities into metabolic flux distributions by elementary mode analysis. BMC Syst. Biol. 1, 31 (2007).
8. Emmerling, M., Dauner, M., Ponti, A., Fiaux, J., Hochuli, M., Szyperski, T., Wüthrich, K., Bailey, J.E., Sauer, U.: Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. J. Bacteriol. 184, 152–164 (2002).
9. Schwartz, J.M., Kanehisa, M.: A quadratic programming approach for decomposing steady-state metabolic flux distributions onto elementary modes. Bioinformatics 21(Suppl. 2), ii204–ii205 (2005).
10. Geenen, S., Taylor, P.N., Snoep, J.L., Wilson, I.D., Kenna, J.G., Westerhoff, H.V.: Systems biology tools for toxicology. Arch. Toxicol. 86, 1251–1271 (2012).
11. Klamt, S., Saez-Rodriguez, J., Gilles, E.D.: Structural and functional analysis of cellular - networks with CellNetAnalyzer. BMC Syst. Biol. 1, 2 (2007).
12. Gagneur, J., Klamt, S.: Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinformatics 5, 175 (2004).
13. Schuster, S., Pfeiffer, T., Fell, D.A.: Is maximization of molar yield in metabolic networks favoured by evolution? J. Theor. Biol. 252, 497–504 (2008).
14. Sauer, U., Canonaco, F., Heri, S., Perrenoud, A., Fischer, E.: The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. J. Biol. Chem. 279, 6613–6619 (2004).
15. Trinh, C.T., Carlson, R., Wlaschin, A., Srienc, F.: Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. Metab. Eng. 8, 628–638 (2006).