Enhancing Elementary Flux Modes Analysis Using Filtering Techniques in an Integrated Environment

Paulo Maia, Marcellinus Pont, Jean-François Tomb, Isabel Rocha, and Miguel Rocha

Abstract. Elementary Flux Modes (EFMs) have been claimed as one of the most promising approaches for pathway analysis. These are a set of vectors that emerge from the stoichiometric matrix of a biochemical network through the use of convex analysis. The computation of all EFMs of a given network is an NP-hard problem and existing algorithms do not scale well. Moreover, the analysis of results is difficult given the thousands or millions of possible modes generated. In this work, we propose a new plugin, running on top of the OptFlux Metabolic Engineering workbench, whose aims are to ease the analysis of these results and explore synergies among EFM analysis, phenotype simulation and strain optimization.

1 Introduction

Over the last few years, a growing number of genome-scale metabolic models for different organisms have been reconstructed, based on the information contained in their annotated sequenced genomes, on the application of Bioinformatics tools and on physiological data. Given the major difficulties in obtaining the kinetic parameters for the whole set of reactions and also in reaching reliable regulatory information, most methods rely on the analysis of the network stoichiometry, using a constraint-based approach to reach some conclusions and to achieve the phenotypic simulation of the system.

Paulo Maia · Isabel Rocha
Dep. Biological Engineering / IBB-CEB, University of Minho, Braga, Portugal
e-mail: {paulo.maia,irocha}@deb.uminho.pt

Marcellinus Pont · Jean-François Tomb
E.I. DuPont De Nemours & Co., Inc, Dellaware, USA
e-mail: {marcellinus.pont,jean-francois.tomb}@usa.dupont.com

Miguel Rocha
Dep. Informatics/ CCTC, University of Minho, Braga, Portugal
e-mail: mrocha@di.uminho.pt

Metabolic models and stoichiometric network analysis have been successfully used to address tasks such as assessing the network's capabilities (e.g. determining maximal product yields), network design (e.g. studying the effects of adding/removing reactions), analysis of functional pathways, analysis of network consistency, flexibility or robustness, among others [13]. One of the major application fields of stoichiometric network analysis has been Metabolic Engineering (ME), a field that deals with designing organisms with enhanced capabilities regarding the productivities of desired compounds [12]. Recently, we have developed *OptFlux* [6], a new computational platform for ME, based on stoichiometric network analysis and constraint-based approaches. The platform implements the major methods related to phenotypical simulation using steady state approaches, such as Flux Balance Analysis (FBA) [3], where a flux distribution that obeys the constraints and maximizes a pre-defined biomass flux is obtained. Also, the platform provides algorithms for strain optimization (e.g. identifying the set of reactions to delete from a model in order to maximize a given objective function).

In contrast to this approach, and though also based on stoichiometric network analysis, the field of Pathway Analysis (PA) characterizes the complete space of admissible flux distributions. PA allows the analysis of meaningful routes involved in metabolic networks. In the recent past, two closely related approaches were developed; namely, Elementary Flux Modes (EFMs) [8] and Extreme Pathways (EPs) [7]. Both approaches aim to dissect metabolic networks into basic functional units and provide means to understand their behavior.

Given the importance of PA within the stoichiometric analysis framework, this paper reports on the development of a software tool that addresses the major tasks within PA. This tool is developed within the *OptFlux* project as a new plug-in. The main aims of this work are (i) to enhance the ME capabilities of the OptFlux platform using EFM analysis and (ii) to provide a graphical user interface to one of the most effective libraries in the computation of EFMs.

2 Elementary Modes

In this paper, we will use a purely structural analysis of biochemical networks, where the network is represented by: q internal compounds (metabolites), m reactions and a stoichiometric matrix N with dimensions $q \times m$. The rows of N correspond to the compounds, while the columns represent reactions; the elements of the matrix in row i and column j represent the stoichiometric coefficient of compound i in reaction j. The framework also allows the definition of external metabolites, thought to be sinks or sources, which lie outside the system. Also, reactions are defined to be reversible or irreversible. Figure 1 provides a simple example extracted from [13] with 6 internal metabolites and 10 reactions; also, this network has 4 external metabolites.

In a steady-state situation, the mass-balance in the network can be represented by the equation:

$$N\mathbf{v} = 0 \tag{1}$$
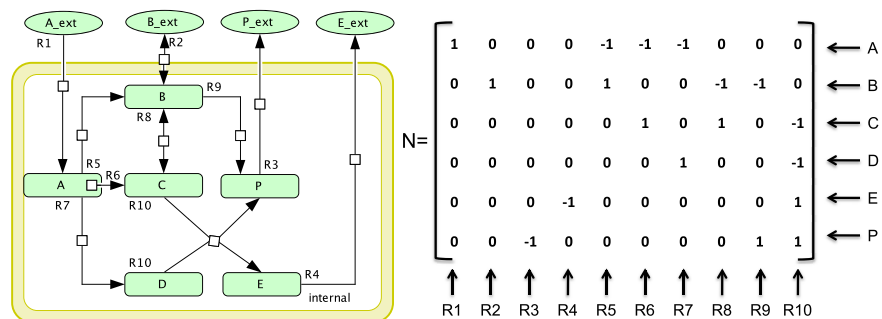
Fig. 1 The simple network used in this case study. $\mathbf{N} : q \times m$ is the stoichiometric matrix where each row corresponds to one species and each column to one of the reactions. The reversibility information is given by $\mathbf{rev} = \{R2, R8\}$ and $\mathbf{irrev} = \{R1, R3, R4, R5, R6, R7, R9, R10\}$, where $\mathbf{rev}$ is the set of reversible reactions and $\mathbf{irrev}$ is the set of irreversible ones. Also notice that $\mathbf{rev} \cap \mathbf{irrev} = \emptyset$

where $\mathbf{v}$ is a vector with the fluxes through the set of reactions. Irreversible reactions are represented by including constraints in the form:

$$v_i \geq 0 \tag{2}$$

The set of all flux vectors that obey the constraints imposed by equations 1 and 2 represent a convex polyhedral cone or a flux cone in stoichiometric studies.

In this context, Elementary Flux Modes (EFMs) are flux vectors composed of q elements $(e_1, e_2, \ldots, e_q)$, that fulfill the following conditions [8]:

1. **Pseudo steady state:** $Ne = 0$, i.e. all EFMs obey equation 1;
2. **Feasibility:** all irreversible reactions proceed in the forward direction, i.e. $e_i \geq 0$ if reaction i is irreversible (equation 2);
3. **Non-decomposability:** EFMs represent the minimal functional units in the network, therefore no reaction can be deleted from an EFM while maintaing a valid flux distribution (that satisfies equations 1 and 2)

This definition imposes the following important properties: (i) there is a unique set of EFMs for a given network; (ii) all feasible steady state flux distributions (that satisfy equations 1 and 2) are a nonnegative superposition of the set of EFMs in the network; (iii) when removing a reaction from the network, the set of EFMs for the new network is equal to the one for the original network but removing all EFMs that include the reaction that was deleted.

Extreme Pathways (EPs) are closely related to EFMs. They are a subset of the EFMs calculated over a reformulated network. Given the similarity of the two concepts, EPs will not be further addressed in this paper nor in the current version of the plug-in.

EFMs are a very useful tool for the analysis of a metabolic network, since they provide the portfolio of all elementary functional units. Thus, EFM analysis can be

used, among others, in the following tasks: (i) identify all routes (pathways) that convert a substrate into a given product and also identify the ones with maximal yield; (ii) identify the importance of reactions in a given context and the correlated reactions; (iii) predict the effect of reaction deletions and viability of resultant mutants.

Since the number of EFMs grows exponentially with the network size, the computation of the whole set of EFMs for a metabolic network is a very hard problem. Several algorithms have been proposed to calculate EFMs (or EPs). They are mostly based on approaches that solve the equivalent problem of extreme ray enumeration from computational geometry. The developed algorithms that solve this problem are typically based on the double description method and can be grouped into two major approaches: the canonical basis approach [8] and the nullspace approach [16], which introduced simplifications in the algorithm and lead to an improved performance.

To further boost the performance, binary vectors were proposed [1] to store the processed reactions in each EFM, reducing memory demands and facilitating set operations. Also, rank computations were introduced to test elementarity in a divide and conquer strategy. Finally, Terzer and Stelling [14] introduced bit pattern trees to index subsets during elementarity testing and also the concept of candidate narrowing using a recursive enumeration approach.

When discussing software tools for EFM calculation, the METATOOL application is a reference [15]. METATOOL was developed in MatLab and has been quite popular since its first version in 1998. It is also included in other more comprehensive packages such as YANASquare [9] or CellNetAnalyzer [5]. More recently, the EFMTool was developed [14] using the Java language (but including a wrapper allowing it to run over MatLab). This is a library that implements a very efficient algorithm for EFM enumeration. This most effective approach is currently available through a command line interface. A major objective of our work is to provide a graphical user interface to enable easy access to the EFMTool.

3 Efm4Optflux – An EMA Plugin for the OptFlux Workbench

The need for a rational approach to analyze results generated by the EFM calculation algorithms was the driver behind the creation of a plug-in within the OptFlux ME platform. OptFlux is a freely available open-source software. It was developed in a modular fashion to facilitate the addition of new features (plug-in based architecture), it is compatible with the Systems Biology Markup Language (SBML) [2] and the layout information of Cell Designer [4]).

In the current version (2.0), the software accommodates several tools that have been developed for the analysis of metabolic models. It incorporates methods for phenotype simulation, such as Flux Balance Analysis, Minimization of Metabolic Adjustment [10] and Regulatory on/off minimization of metabolic flux changes [11], as well as strain optimization algorithms, such as Evolutionary Algorithms and Simulated Annealing. It also packages a suitable model visualization tool.

In this work, we present an extension to OptFlux, a plug-in that provides graphical access to the EFMTool calculation and analysis capabilities. The plugin delivers several interesting characteristics such as a simple graphical user interface (GUI) and state-of-the-art EFM calculation. It also allows filtering of the results based on an intuitive form, allowing the definition of patterns with presence/absence of external metabolites. The organization of the filtered results is done by grouping the EFMs with unique net conversions.

After the calculation of the EFMs, each EFM vector is multiplied by the submatrix component of the full stoichiometric matrix containing only the rows relative to the external metabolites, thus returning the vectors of net conversions. These are scanned and only unique conversions are maintained. All the results are kept in disk in order to ease the memory burden of some EFM computations. The EFMs are kept in one file, the net conversions in another and a lookup table is generated, in a third file, to correlate between them.

Furthermore, for each net conversion, the greatest common divisor is calculated to improve the reading of the conversion equation. To do so, all the coefficients have to be integers and therefore the EFM calculation is limited to using big integer arithmetics. In the filtering step, the software filters EFMs based on the presence/absence of external metabolites in the net conversions. Moreover, it can also order by yield, providing an input and output metabolite is provided for the computation. When the filters have been defined, the software scans the definition files for compatible conversions and their related EFMs.

The user can browse through the filtered conversions in an intuitive table that presents the conversion equation and yields and provides access to the related EFMs. The visualization of these EFMs is presented in a column-wise table, where each column corresponds to an EFM and each line to a reaction of the model.

Moreover, the user can export the EFM values to CellDesigner, if a valid CellDesigner SBML file had been previously loaded. The values for each reaction in the mode vector are represented by the thickness of the lines in the CellDesigner layout. This thickness varies between a user-defined range and it is relative to the value.

4 Case Study

We present the plugin capabilities, following a simple example from [13]. This network is presented in Figure 1. This example is provided in the project homepage (www.optflux.org) as a CellDesigner SBML file.

In order to compute the EFMs for this network, the following steps were required:

1. Load the file *Stelling_toy.xml* using the **New Project Wizard**. Select SBML, CellDesigner SBML and default options until the end of the wizard;
2. To execute the EFMs computation, access **Plugins → Elementary Modes → Compute Elementary Modes** and use all the default options. This computation is instantaneous in any regular PC;

3. Clicking the new datatype placed on the clipboard will launch a generic viewer. Press **Proceed to filtering** and a filtering dialog will appear. In this dialog we selected every filter as *Don't Care* and the yield as $A\_ext \rightarrow P\_ext$.
4. A *FilteredEFMResults* instance will be placed on the clipboard. Clicking it will launch the viewer (Figure 2).

| Total of 5 conversions matching your criteria					
Conversion... ▲	Conversion Equation	BCY	SPY	SBY	EMs
0	A_ext ==> P_ext	0.0	1.00000	0.0	ⓘ ...
1	A_ext ==> B_ext	0.0	0.00000	0.0	ⓘ ...
2	B_ext + A_ext ==> P_ext + E_ext	0.0	1.00000	0.0	ⓘ ...
3	2 A_ext ==> P_ext + E_ext	0.0	0.50000	0.0	ⓘ ...
4	B_ext ==> P_ext	0.0	0.00000	0.0	ⓘ ...

Fig. 2 Screenshot of the Filtered Results Viewer. A total of 5 unique conversions were discovered by the algorithm. These are listed together with the conversion equation and respective yields. BCY = Biomass/product Coupled Yield; SPY = Substrate to Product Yield; SBY = Substrate to Biomass Yield

The computation resulted in a total of 8 EFMs and 5 unique net conversions. The EFMs were exported to CellDesigner and the results were compared and validated with the ones obtained in [13]. A graphical representation of the layout in CellDesigner is presented in Figure 3.

5 Conclusion

Despite several incursions regarding the computation of EFMs, no real efforts have been made to ease analysis of the results. The proposed plug-in provides a seamless integration with the OptFlux ME workbench, thus providing a rational interface and smart filters to analyze the large number of EFMs that usually generated in such analysis approaches. There is, however, still much space for improvement, namely regarding the arithmetics allowed by the platform. The limitation to integers, though consistent with the display mode, is nevertheless, a limitation. Other ways of filtering and sorting the calculated EFMs are also currently under development.
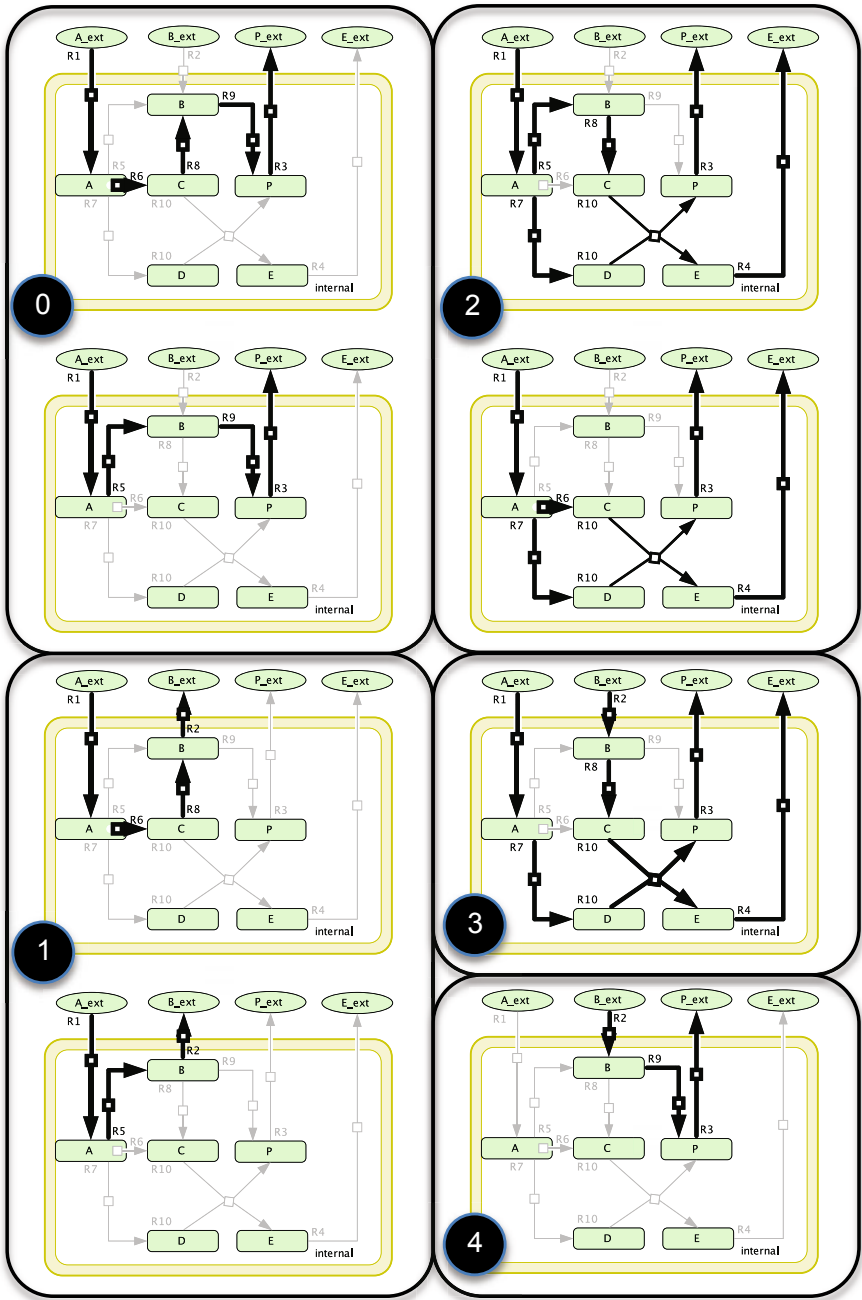
Fig. 3 Graphical representation of the EMs associated with each unique net conversion. The numbers are the Conversion IDs depicted in Figure 2. The conversions are the following: **0** : $A\_ext \rightarrow P\_ext$, **1** : $A\_ext \rightarrow B\_ext$, **2** : $2A\_ext \rightarrow P\_ext + E\_ext$, **3** : $B\_ext + A\_ext \rightarrow P\_ext + E\_ext$, **4** : $B\_ext \rightarrow P\_ext$

References

1. Gagneur, J., Klamt, S.: Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinformatics 5(175) (2004)
2. Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., et al.: The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19(4), 524 (2003)
3. Kauffman, K.J., Prakash, P., Edwards, J.S.: Advances in flux balance analysis. Curr. Opin. Biotechnol. 14, 491–496 (2003)
4. Kitano, H., Funahashi, A., Matsuoka, Y., Oda, K.: Using process diagrams for the graphical representation of biological networks. Nature biotechnology 23(8), 961–966 (2005)
5. Klamt, S., Saez-Rodriguez, J., Gilles, E.: Structural and functional analysis of cellular networks with cellnetanalyzer. BMC Systems Biology 1(2) (2007)
6. Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J.P., Nielsen, J., Patil, K.R., Ferreira, E.C., Rocha, M.: Optflux: an open-source software platform for in silico metabolic engineering. BMC Systems Biology (in Press, 2010)
7. Schilling, C.H., Letscher, D., Palsson, B.Ø.: Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. Journal of Theoretical Biology 203(3), 229–248 (2000)
8. Schuster, S., Hilgetag, C.: On elementary flux modes in biochemical reaction systems at steady state. J. Biol. Syst. 2(2), 165–182 (1994)
9. Schwarz, R., Liang, C., Kaleta, C., Kuhnel, M., Hoffmann, E., Kuznetsov, S., Hecker, M., Griffiths, G., Schuster, S., Dandekar, T.: Integrated network reconstruction, visualization and analysis using yanasquare. BMC Bioinformatics 8(313) (2007)
10. Segre, D., Vitkup, D., Church, G.M.: Analysis of optimality in natural and perturbed metabolic networks. Proceedings of the National Academy of Sciences 99(23), 15112 (2002)
11. Shlomi, T., Berkman, O., Ruppin, E.: Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proceedings of the National Academy of Sciences of the United States of America 102(21), 7695 (2005)
12. Stephanopoulos, G., Aristidou, A.A., Nielsen, J.: Metabolic engineering principles and methodologies. Academic Press, San Diego (1998)
13. Szallasi, Z., Stelling, J., Periwal, V.: System modeling in cell biology: from concepts to nuts and bolts. MIT Press, Cambridge (2006)
14. Terzer, M., Stelling, J.: Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24(19), 2229–2235 (2008)
15. von Kamp, A., Schuster, S.: Metatool 5.0: fast and flexible elementary modes analysis. Bioinformatics 22(15), 1930–1931 (2006)
16. Wagner, C.: Nullspace approach to determine the elementary modes of chemical reaction systems. J. Phys. Chem. B 108, 2425–2431 (2004)