

Systems biology

tEFMA: computing thermodynamically feasible elementary flux modes in metabolic networks

Matthias P. Gerstl^{1,2}, Christian Jungreuthmayer^{1,2,*} and Jürgen Zanghellini^{1,2,*}

¹Austrian Centre of Industrial Biotechnology, Vienna, Austria and ²Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria

*To whom correspondence should be addressed.

Associate Editor: Igor Jurisica

Received on October 17, 2014; revised on February 2, 2015; accepted on February 15, 2015

Abstract

Summary: Elementary flux modes (EFMs) are important structural tools for the analysis of metabolic networks. It is known that many topologically feasible EFMs are biologically irrelevant. Therefore, tools are needed to find the relevant ones. We present thermodynamic tEFM analysis (tEFMA) which uses the cellular metabolome to avoid the enumeration of thermodynamically infeasible EFMs. Specifically, given a metabolic network and a not necessarily complete metabolome, tEFMA efficiently returns the full set of thermodynamically feasible EFMs consistent with the metabolome. Compared with standard approaches, tEFMA strongly reduces the memory consumption and the overall runtime. Thus tEFMA provides a new way to analyze unbiasedly hitherto inaccessible large-scale metabolic networks.

Availability and implementation: <https://github.com/mpgerstl/tEFMA>

Contact: christian.jungreuthmayer@boku.ac.at or juergen.zanghellini@boku.ac.at

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

An *elementary flux mode* (EFM) is a steady-state pathway that consists of an indivisible set of reactions (Schuster *et al.*, 1999). EFMs permit a system wide analysis of metabolism and can be used to identify engineering targets (Zanghellini *et al.*, 2013). Different approaches were published to calculate EFMs in large metabolic models. For instance, Hunt *et al.* (2014) split a metabolic network into subnetworks and calculated the EFMs for each subnetwork in parallel. Others only enumerated the set of the shortest EFMs (De Figueiredo *et al.*, 2009) or randomly sampled the full set of EFMs (Machado *et al.*, 2012). A complete EFM analysis for large genome-scale models is still out of reach as the number of EFMs increases dramatically with the size of the metabolic network (Klamt and Stelling, 2002) requiring inhibitingly large computing resources.

To tackle this problem the biological relevance of EFMs needs to be taken into account as only a small set of EFMs are biologically relevant (Schwartz and Kanehisa, 2006). Therefore, tools are necessary to find those subsets of (biologically) relevant EFMs.

Jungreuthmayer *et al.* (2013) enumerated only those EFMs that were consistent with a given transcriptional regulatory network. Here, we use a network-embedded thermodynamic (NET) analysis (Kümmel *et al.*, 2006) to detect and remove thermodynamically infeasible EFMs already during their enumeration.

Thermodynamic EFM analysis (tEFMA) applies the NET principle and uses metabolite concentrations and the compounds' Gibbs energy of formation to determine the energy surface of an EFM. At a constant pressure a biochemical reaction only proceeds spontaneously if its Gibbs free energy of reaction is negative (Alberty, 2003). If the thermodynamic information is available for a reaction, it can be tested for its thermodynamic feasibility. tEFMA does not only analyze the feasibility of an isolated reaction, but considers the Gibbs free energy of reaction for all reactions contributing to an EFM. tEFMA is computationally much more efficient than a sequential approach, where an ordinary EFM analysis is followed by a NET analysis (Jol *et al.*, 2012). The reason for this is that tEFMA checks the feasibility of (intermediate) EFMs continuously and

eliminates infeasible ones during the enumeration process. This reduces the computational costs significantly.

2 Implementation

We extended the open source program 'efmtool' (Terzer and Stelling, 2008) and implemented the usage of metabolomics data. efmtool uses the double description method (Fukuda and Prodon, 1996) to enumerate EFMs. This method computes EFMs iteratively. At each iteration, a new reaction is processed and EFMs are updated accordingly. The process stops when all reactions of the network have been processed. We integrated our extension at the beginning of each iteration step. Before EFMs are updated and new ones are created, the program tests each intermediate EFM for its thermodynamical feasibility using a linear program:

$$\min 0 \quad (1a)$$

$$\text{s.t. } \Delta_r G_i \leq 0, \quad \forall i \in \text{supp}(e_k) \quad (1b)$$

$$\Delta_r G_i = \sum_{j=1}^m S_{ji} \Delta_f G_j' \quad (1c)$$

$$\Delta_f G_j' = \Delta_f G_j'^0 + RT \ln(c_j/c_0), \quad c_0 = 1 \text{ M} \quad (1d)$$

$$c_j^{\min} \leq c_j \leq c_j^{\max}. \quad (1e)$$

According to the second law of thermodynamics every active reaction i has a negative Gibbs energy $\Delta_r G_i$. This applies not only for a single reaction but also in context of a whole pathway, i.e. an (intermediate) EFM, e_k (Eq. 1b). $\Delta_r G$ is calculated using the transformed Gibbs energies, $\Delta_f G_j'$, of the participating metabolites in respect to their stoichiometric coefficient, S_{ji} (Eq. 1c). $\Delta_f G'$ values dependent on the standard transformed Gibbs energy of formation, $\Delta_f G_j'^0$, the molar gas constant, R , the temperature, T , and the concentration of the metabolite, c_j , (Eq. 1d), which are allowed to vary within a given concentration range (Eq. 1e). As any solution showing the thermodynamic feasibility of an EFM is fine, we do not need to search for an optimum (Eq. 1a). As soon as an (intermediate) infeasible EFM is found it can be removed from the analysis without impacting feasible EFMs.

The new features are accessible via additional command line arguments: `-conc` takes the concentration file. Each line in this file contains the speciesID of the network, the name of the species in the thermodynamic property file, and the minimum and maximum concentration in the following format: `speciesID;thermodynamic name;minimum;maximum`; The thermodynamic property file is taken by the argument `-thermo`. Each line contains the $\Delta_f G^0$ value, the charge and the number of H atoms for each charge state of the metabolite in the following format: `thermodynamic name-($\Delta_f G_0^0, z_0, H_0$), ($\Delta_f G_1^0, z_1, H_1$), ...`. Optional parameters are `-cmin` and `-cmax` for default minimum and maximum concentration values. `-ph`, `-temperature` and `-ionstrength` for defining pH, temperature and ionic strength of the environment, respectively. With the parameter `-thermothreads` the number of threads for the linear program can be chosen. Infeasible patterns, the linear problem and its variables are written to the files given by `-pattern`, `-lpfile` and `-lpvar`.

3 Applications

We tested our tool on an *Escherichia coli* core model (Orth et al., 2010), consisting of 53 intracellular metabolites and 155 irreversible

Table 1. Comparison between tEFMA and an ordinary EFM analysis (EFMA)

| Method | Number of modes | | Runtime (h) | | RAM (GB) | |
|--------|-----------------|--------------------|-------------|-----------|----------|--|
| tEFMA | glucose | 94 911 298 (35%) | 7.1 (25%) | 21 (23%) | | |
| | glycerol | 131 112 724 (48%) | 14.0 (49%) | 43 (48%) | | |
| | acetate | 147 201 012 (54%) | 13.7 (48%) | 48 (53%) | | |
| EFMA | | 271 494 722 (100%) | 28.8 (100%) | 90 (100%) | | |

Numbers in brackets list the percentage compared to EFMA.

reactions. We used published metabolite concentrations for *E. coli* grown on glucose, glycerol and acetate (Bennett et al., 2009). $\Delta_f G^0$ values were retrieved from eQuilibrator (Flamholz et al., 2012). Measured concentration ranges were available for 28 metabolites. For 23 metabolites no measured values were available and default ranges (10^{-7} –1 M) were applied. For two metabolites no $\Delta_r G$ were available. Reactions to which these two metabolites contributed were not checked for thermodynamic consistency to avoid false positives. Thus, uncharacterized reactions were assumed to be thermodynamically consistent with all other reactions. As a reference, we used the complete set of EFMs without considering any thermodynamic data. We note that any extension of the core model used here will only add more EFMs to the system but will not change the (in)feasibility of already existing EFMs. All runs were performed on a computer with 2 Intel Xeon CPUs (12 cores each) and a total of 378 GB of RAM using 10 threads for EFM enumeration and 18 threads for thermodynamic checks.

Table 1 compares the performance of tEFMA with an ordinary EFM analysis. The number of feasible EFMs decreased by up to 65%, the runtime decreased by up to 75% and the RAM usage decreased by up to 77%. We verified that a sequential approach, i.e. an ordinary EFM analysis followed by a NET analysis (Jol et al., 2012), identified an identical set of infeasible EFMs, yet without harvesting any computational savings. More importantly, we were able to interpret all infeasible EFMs predicted by tEFMA in terms of known infeasible pathways (see Supplementary Table 1). tEFMA did not falsely identify any feasible pathway to be infeasible (Gerstl et al., 2015).

4 Conclusion

By integrating the metabolite concentrations into the EFM enumeration we killed two birds with one stone. First, we strongly reduced the memory usage and the runtime of the EFM enumeration. Second, we got a step further to only calculate the set of biological relevant EFMs. Both points are essential to reach the final goal of calculating EFMs in large, genome-scale models. We stress that tEFMA greatly reduces the computational costs. Thus an EFM analysis of medium-scale models become possible already on current high-end personal computers and no longer requires a dedicated high performance computing environment.

Acknowledgements

This work has been supported by the Austrian BMVFW, BMVIT, SFG, Standortagentur Tirol and ZIT through the Austrian FFG-COMET-Funding Program.

Conflict of Interest: none declared.

References

Alberty, R.A. (2003) *Thermodynamics of Biochemical Reactions*. Wiley, New Jersey.

- Bennett, B.D. *et al.* (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.*, **5**, 593–599.
- De Figueiredo, L.F. *et al.* (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics*, **25**, 3158–3165.
- Flamholz, A. *et al.* (2012) eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Res.*, **40**, D770–D775.
- Fukuda, K. and Prodon, A. (1996) Double description method revisited. In: Deza, M. *et al.* (eds.) *Combinatorics and Computer Science, volume 1120 of Lecture Notes in Computer Science*. Springer, Berlin, pp. 91–111.
- Gerstl, M.P. *et al.* (2015) Metabolomics integrated elementary flux mode analysis in large metabolic networks. *Sci. Rep.* **5**, 893; DOI: 10.1038/srep08930.
- Hunt, K.A. *et al.* (2014) Complete enumeration of elementary flux modes through scalable demand-based subnetwork definition. *Bioinformatics*, **30**, 1569–1578.
- Jol, S.J. *et al.* (2012) System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes. *PLoS Comput. Biol.*, **8**, e1002415.
- Jungreuthmayer, C. *et al.* (2013) regEfmtool: speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic. *Biosystems*, **113**, 37–39.
- Klamt, S. and Stelling, J. (2002) Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.*, **29**, 233–236.
- Kümmel, A. *et al.* (2006) Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol. Syst. Biol.*, **2**, 2006.0034.
- Machado, D. *et al.* (2012) Random sampling of elementary flux modes in large-scale metabolic networks. *Bioinformatics*, **28**, i515–i521.
- Orth, J.D. *et al.* (2010) Reconstruction and Use of Microbial Metabolic Networks: the Core *Escherichia coli* Metabolic Model as an Educational Guide, EcoSal Plus; doi:10.1128/ecosalplus.10.2.1.
- 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.
- Schwartz, J.-M. and Kanehisa, M. (2006) Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. *BMC Bioinf.*, **7**, 1–20.
- Terzer, M. and Stelling, J. (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, **24**, 2229–2235.
- Zanghellini, J. *et al.* (2013) Elementary flux modes in a nutshell: properties, calculation and applications. *Biotechnol. J.*, **8**, 1009–1016.