



HAL
open science

Computing Thermodynamically Consistent Elementary Flux Modes with Answer Set Programming

Emma Crisci, Maxime Mahout, Sabine Peres

► **To cite this version:**

Emma Crisci, Maxime Mahout, Sabine Peres. Computing Thermodynamically Consistent Elementary Flux Modes with Answer Set Programming. CMSB 2024 - 22nd International Conference Computational Methods in Systems Biology, Sep 2024, Pise, Italy. pp.80-88, 10.1007/978-3-031-71671-3_7. hal-04707880

HAL Id: hal-04707880

<https://hal.science/hal-04707880v1>

Submitted on 5 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Computing thermodynamically consistent Elementary Flux Modes with Answer Set Programming

Emma Crisci^{1,2}, Maxime Mahout³, and Sabine Peres^{1,2}

¹ UMR CNRS 5558, Laboratoire de Biométrie et de Biologie Évolutive, Université Claude Bernard Lyon 1, 69100 Villeurbanne, France.

² ERABLE, INRIA Lyon Centre, 69100 Villeurbanne, France.

³ LIFEWARE, INRIA Saclay, Palaiseau, France

Abstract. Elementary Flux Modes (EFM) allow the description of the minimal sets of reactions in a metabolic network under steady-state conditions, representing unique and feasible pathways. They fully characterize the solution space but a combinatorial explosion prevents their calculation when the network is large. Furthermore, it is not necessary to calculate all of them, as many of them are not biologically relevant. Therefore, the software *aspefm*, which combines the use of Answer Set Programming and Linear Programming, proposes to integrate different types of constraints in the EFM computation such as equilibrium constants, Boolean regulatory rules, growth yields and growth medium. The addition of constraints makes it possible to cut off research pathways that lead to non-relevant EFMs. The computation of the EFMs of interest significantly reduces the computational time and saves space. In this article, we have added thermodynamic constraints in terms of the Gibbs energy of reactions, which constrain metabolite concentrations within a chosen interval. This constraint is added as a theory propagator and it reduces the enumeration during the computation. We applied our tool to the central carbon metabolism of *E. coli* and showed that the Gibbs energy constraints suppress a large number of non-relevant EFMs.

Keywords: Metabolic Networks · Elementary flux modes · Logic programming · Thermodynamics.

1 Introduction

A metabolic network is a set of successive chemical reactions, catalysed by enzymes, that result in the production of metabolites required by the cell. The concept of elementary flux mode (EFM) plays a crucial role in their analysis from a pathway-focused point of view. An EFM is defined as a smallest sub-network (with respect to reactions set inclusion) that enables the metabolic system to operate at steady state with all irreversible reactions proceeding in the appropriate direction [16]. A major challenge in the computation of EFMs is the large

number of EFMs associated with large biochemical networks. One approach to calculating EFMs is to add biological knowledge to restrict the search to biologically relevant EFMs. The idea is not to enumerate the whole set and then select the EFMs of interest, but to restrict the search to biologically feasible solutions. In the double description method [12] which is used by *efmtool*, only the support of monotone constraints such as negative transcriptomic Boolean regulations [9], thermodynamic equilibrium constants [15, 14] and the Gibbs energy [8, 7] could be integrated into the algorithm. Such constraints significantly reduce the complexity of enumerating EFMs, but may not be sufficient to aid computation in a genome-scale metabolic model. In previous works, we proposed constraint logic programming methods based on SAT Boolean satisfiability to enumerate subsets of EFM of biological interest [13, 11], but their efficiency suffered from the complexity of minimizing the size of the solutions searched. Recently, we developed *aspefm* [10], a highly efficient software based on Answer Set Programming (ASP), to search for relevant EFMs respecting different kind of biological constraints, such as transcriptomic regulations, equilibrium constants, growth yields and growth medium.

In this article, we develop the implementation of a new *aspefm* extension, called *DeltaGChecker*. *DeltaGChecker* specifically checks whether the Gibbs free energies, or thermodynamics, of reactions are possible. With this new constraint type, we aim to reduce the total number of EFMs and the complexity of computations. We test our *DeltaGChecker* on one small-scale metabolic networks based on *Escherichia coli*. We observe that the number of EFMs retrieved is particularly dependent on the minimum and maximum bounds chosen for the metabolite concentration variables and that the extension does not lose significant amounts of computational time.

2 Computation of EFMs compatible with thermodynamic constraints

A metabolic network is represented by a stoichiometric matrix $S \in \mathbb{R}^{|\mathcal{M}| \times |\mathcal{R}|}$, where $|\mathcal{M}|$ and $|\mathcal{R}|$ are the numbers of metabolites and reactions, respectively. Each element $S_{m,r}$ denotes the stoichiometric coefficient of metabolite m in reaction r . The steady-state condition is expressed as $S \cdot v = 0$, where $v \in \mathbb{R}^{|\mathcal{R}|}$ is a flux vector. A flux mode v satisfies two conditions $S \cdot v = 0$ and $v_i \geq 0$ for all $i \in irrev$, where *irrev* is the set of indices corresponding to irreversible reactions. An Elementary Flux Mode (EFM) e is a flux mode with minimal support, where the support is defined as $supp(e) = \{r \in \mathbb{N} : e_r \neq 0\}$. Depending on the size of the metabolic network, there may be a combinatorial explosion in the search for EFMs. Even if it could be possible to enumerate billions of EFMs, for some networks, analysing in post-processing all the EFMs would be time-consuming. *aspefm* allow us to directly find the EFMs of interest, by exploring only solutions that satisfy the constraints. Several types of constraints can be added, including the so-called linear constraints, referring to the field of linear programming.

By analogy, our thermodynamics extension *DeltaGChecker* checks the Gibbs free energies of reactions by using linear programming. Linear programming allows the algorithm to check whether a given support of an EFM in construction (the set of active reactions in the EFM) satisfies these thermodynamic constraints. To determine if an EFM is thermodynamically feasible, the extension finds a set of metabolite concentrations that satisfies the thermodynamic constraints.

2.1 Thermodynamics of chemical reactions and Formulation in LP

Thermodynamics is the study of energy transformations that occur within mechanical systems. In thermodynamics, chemical reactions are analysed from an energetic perspective, in order to predict whether a given reaction is favorable or unfavorable.

In this article, we will focus on the second law of thermodynamics, expressed through the Gibbs free energy equation, which provides a criterion for determining the spontaneity and direction of chemical reactions [1].

The Gibbs free energy of formation ($\Delta_F G'$) of metabolites is calculated using the Gibbs free energies of formation in their standard states ($\Delta_F G'^{\circ}$), the molar gas constant, R and the temperature, T . The standard state is thus at the metabolite level. The free energy of formation denoted $\Delta_F G'$ is the Gibbs energy associated with the formation of one mole of a substance from its constituents in their standard states at a reference pressure and temperature. In other words, it is the energy required to form one mole of a substance from its most stable components. The sign of $\Delta_R G_r$ gives the direction of the reaction r . If the calculated $\Delta_R G_r$ is strictly negative, the reaction is thermodynamically favorable under current conditions. Conversely, if $\Delta_R G_r$ is positive, the reaction is unfavorable or requires additional energy to make it possible. Thus, no EFM can therefore have reactions in that direction. Finally, if the energy $\Delta_R G_r$ is null, this means that the reactants and products are in thermodynamic equilibrium.

The thermodynamic constraints can be expressed as a linear program (1). The principle behind this linear program is to find a set of metabolite concentrations such that thermodynamics is respected throughout the metabolic network. The objective function will thus be expressed as the minimisation of 0, *i.e.* no attempt is made to minimise or maximise a function. Constraints on chemical reactions are tested for a set of active reactions in the metabolic network, support of a potential EFM e .

$$\left\{ \begin{array}{l} \min \quad 0 \\ \text{s.t} \quad \Delta_R G_r \leq 0, \quad \forall r \in \text{supp}(e) \\ \Delta_R G_r = \sum_{m=1}^M S_{m,r} \Delta_F G'_m \\ \Delta_F G'_m = \Delta_F G'_m{}^{\circ} + RT \ln(C_m/C_0), \quad C_0 = 1M \\ \ln(C_m^{\min}/C_0) \leq \ln(C_m/C_0) \leq \ln(C_m^{\max}/C_0) \end{array} \right. \quad (1)$$

Finally, the linear program verifies that all reactions contributing to an EFM (present in the support), are simultaneously feasible and consistent with a metabolome with given error bounds C_m^{min} and C_m^{max} .

The integration of Gibbs free energy into EFM calculations has already been implemented as an extension to *efmtool* [17] in the thermodynamics EFM analysis (tEFMA) [7, 8]. *tEFMA* proposes to add the same thermodynamic constraints in the same LP form. *tEFMA* uses the double description algorithm [12], which is an incremental method and does not enumerate solutions one-by-one. Moreover, it can integrate only support monotone constraints during the computation. Filtering EFMs with multiple constraints during enumeration is promising for finding EFMs of interest in large networks. In this article, we integrate this linear program into *aspefm*, which allows additional constraints to be combined simultaneously and can enumerate solutions one-by-one. The *aspefm* tool defines a logic program in ASP to be solved using a modified version of the *clingo*[LP] solver, making use of high-performance state-of-the-art SAT-solving techniques, as well as expressive IBM CPLEX linear constraints [10].

2.2 *DeltaGChecker* extension to *aspefm*

Thermodynamic constraints require additional data, including the standard Gibbs free energies of formation $\Delta_F G_m^{\circ}$ and the allowable concentration ranges for each metabolite, and can be found in the database as Equilibrator [5]. The minimum and maximum concentrations for each metabolite are user-defined. Users can base their choices on their own experimental data or from relevant databases. To retrieve the thermodynamic data, we developed a script using the well-documented Equilibrator API. In particular, we used the following standard conditions recommended by Equilibrator: a pH of 7.4, an ionic strength of 0.25M and a temperature of 298.15K [2]. More details on input data requirements, API scripts, file transformation utilities, and code for *aspefm* and the new *DeltaGChecker* extension are available in our GitLab repository: <https://gitlab.inria.fr/erable/aspefm>.

The *DeltaGChecker* extension is directly integrated into the EFMs calculation of *aspefm*. In the SAT resolution algorithm of *aspefm*, EFMs are constructed reaction by reaction, as partial supports. A partial support is defined here as an unfinished sketch of an EFM support in the constraint propagation. By evaluating the thermodynamic constraints on partial supports, the *DeltaGChecker* extension ensures that any added reactions in the EFM retain their its feasibility. The linear program solver CPLEX (1) is called for each partial support constructed and the feasibility of the $\Delta_R G_r$ energies for the calculated metabolite concentration variables constrained by the min/max concentration bounds is checked [4]. If the linear program has no solutions, then the thermodynamics for that partial set of active reactions is considered not respected. The invalidation of thermodynamics leads to the addition of 'nogoods' in the SAT resolution, which cut off the search for certain EFM search branches. In the case of *aspefm* and *DeltaGChecker*, nogoods are added for sets of reactions that are incompat-

ible with each other. For rest of the resolution, the branches explored will not consider the forbidden sets of reactions together.

3 Application to central carbon metabolism of *E.coli*

The aim, with a simple model as *E.coli* from the published model [3], is to show that the extension encoding thermodynamic constraints can restrict us to only relevant EFMs by eliminating those for which it is impossible to find a metabolite concentration solution. When validating the thermodynamic extension, the idea is to specify for each metabolite an interval around the reference value associated with the metabolite (to be found in the literature) and to compare the number of associated EFMs. In other words, if the percentage is x , then each metabolite will be assumed to have a concentration within a range of values between :

$$\begin{cases} 0 \leq x \leq 100, & \text{for all } x \text{ in percentage} \\ C_{\text{ref}} \times (1 - \frac{x}{100}) \leq C_{\text{ref}} \leq C_{\text{ref}} \times (1 + \frac{x}{100}) \end{cases} \quad (2)$$

Thus, the higher the percentage, the less the metabolites are constrained in their concentration as they have a wider range of values, *i.e.* the solver has more freedom in choosing the optimal concentration set to find. Increasing the percentage range directly reduces the constraints on thermodynamics. In the rest of the study, we will compare the EFMs found in *aspefm* with and without the use of thermodynamics, to show how the new extension reduces the number of relevant EFMs. The metabolic network of *E.coli* consists of 55 reactions, 46 metabolites and 1658 EFMs without any constraint.

3.1 Selection of EFMs of thermodynamic interest in *E.coli*

In this *E. coli* model, some metabolite concentrations were unknown. We first set their reference values to 10^{-5} M. Relevant EFMs (satisfying thermodynamic constraints) were found only from 50% onwards, increasing to 99% where we reached 1658 EFMs, matching the number without thermodynamic constraints (Fig. 1).

Instead of setting a reference concentration for the metabolite that does not have one, we then set the predefined interval from 10^{-7} to 1 M as in *tEFMA*. It can be seen that with the interval set for metabolite concentrations, more EFMs are obtained than with the use of a reference concentration, *e.g.* for 75 percent, 100 times more EFMs are obtained with the imposed interval. This is because the fixed interval is large, so less thermodynamic stress is applied.

Even though about 15 percent of the metabolite concentrations were unknown, we were successfully able to limit the computation to thermodynamically consistent EFMs by setting default reference concentrations or default bounds. Therefore, we highlight that knowing the concentrations for all the metabolites in the metabolic model might not always be necessary. In summary, the *DeltaGChecker* extension efficiently reduces the amount to only relevant thermodynamics EFMs by incorporating thermodynamic constraints. However, for

these constraints to be effective, the concentration intervals provided must be sufficiently stringent.

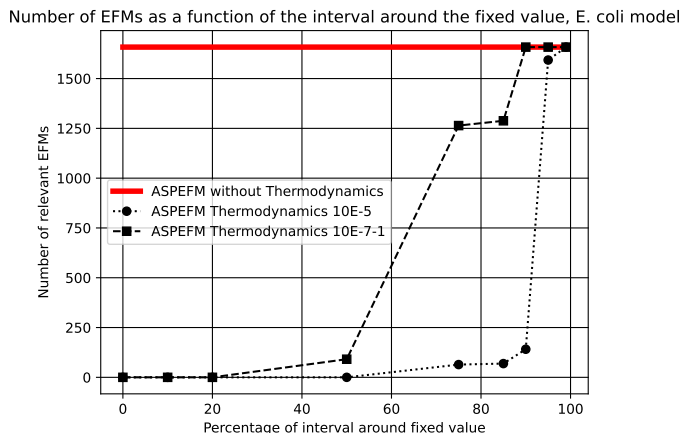


Fig. 1. Evolution of the number of EFMs selected as a function of metabolite concentration constraints, in *E.coli*. Abscissa are percentages of interval around the reference concentration values.

4 Conclusion

The *DeltaGChecker* extension gives *aspefm* a strong advantage in finding relevant EFMs. The addition of thermodynamic constraints significantly reduces the number of EFMs in most cases, provided the system has enough constraints. Our *DeltaGChecker* extension determines, for a potential EFM support, whether there exists a set of metabolite concentrations such that the system satisfies the thermodynamic constraints. The extension has been tested on the *E. coli* model, and we have shown that it reduces the number of EFMs, by an order of magnitude, allowing faster post-processing analysis.

It should be noted that at the scale of the applications studied, the aim is mainly to reduce the number of EFMs, rather than to reduce the computation time at all costs. For larger networks, where the combinatorial explosion is strong, it will be important to know whether the extension actually multiplies computation time or reduces it. For this reason, we have performed additional analysis on the computational time (Appendix 5.1). We have concluded that the extension has a certain cost in computation time, but this time could probably be worthwhile on larger networks. In the case of large networks, where in some cases the aim is simply to enumerate EFMs of interest, the idea of using thermodynamic constraints to avoid traversing certain paths, and thus reduce computation, seems promising. Further extensions will be developed in the future, in particular to incorporate kinetic and enzyme concentration constraints.

5 Appendix

5.1 Computation times on *E.coli*

The thermodynamic extension makes it possible to reduce the number of EFMs of interest, and therefore the post-processing search, but it's interesting to look at and compare what the extension can bring to processing. The extension reduces the number of paths used to search for EFMs, and therefore saves time on paths that we know to be thermodynamically infeasible. On the other hand, the extension will waste time on path traversal, because for each EFM support tested, it will be necessary to check whether the support has a concentration set solution that validates the thermodynamics, which implies running the CPLEX solver. So, for the extension to be cost-effective in terms of computation time, the absence of some non-thermodynamically feasible EFM paths would have to be beneficial in terms of the computation time added to test whether each partial support has a thermodynamically feasible solution.

This figure shows the same percentages used in the previous study (Fig. 2). It was seen that the lower the percentage, the more the extension restricted the number of EFMs of interest, so that the extension must have had an impact on the paths taken by the EFMs. As the percentage increased, the number of EFMs of interest increased, approaching the number found by *aspefm* without the addition of the extension, due to the fact that the thermodynamic constraints were not strong enough. We note that in the case of the *E.coli* model, computation times of *aspefm* oscillates between 25.6 and 28.4 seconds while not increasing with the percentage as would be expected; time within and outside the extension remains approximately constant. Surprisingly, filtering out more solutions had no effect on the computation time. This result might be explained by the fact that the network is not large enough, i.e. the time saved by not traversing certain paths is not visible at this scale. Overall, when removing the times taken by the extension, we find that the differences in time are almost all attributable to the extension, with a time difference of around ~ 2 seconds (on average 17.51 against 15.56).

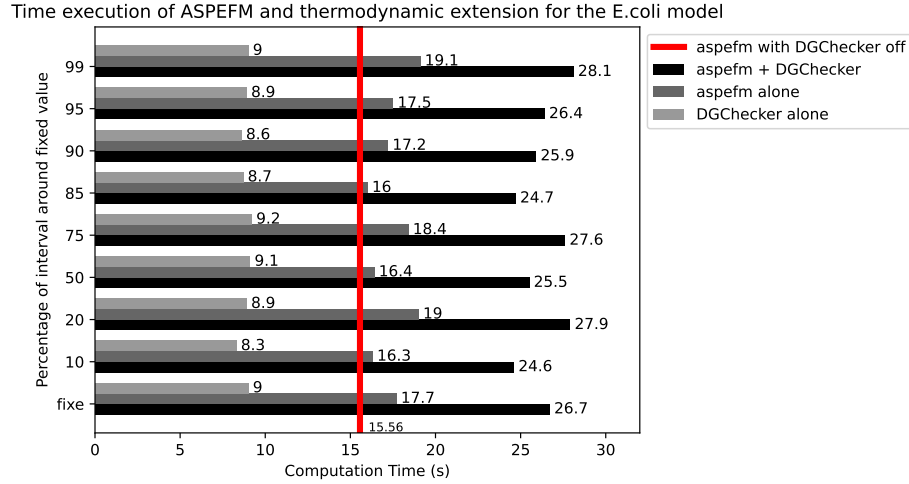


Fig. 2. Evaluation of *aspefm* computation times for *E. coli* depending on percentages around the selected intervals; 1) *aspefm* with DGChecker off: *aspefm* times for computing *E.coli* model EFMs without using extension *DeltaGChecker*; 2) *aspefm* + *DGChecker*: *aspefm* times for computing *E. coli* model EFMs using the *DeltaGChecker* extension, for each percentage around the fixed concentration value; 3) *aspefm* alone: for each percentage, portion of time spent computing *aspefm* solutions, without checking thermodynamics; 4) *DGChecker* alone: for each percentage, portion of time spent checking thermodynamics in the *DeltaGChecker* extension.

This is a promising result, and we hope that computation times for our filtering procedure will be able to scale to larger networks.

References

1. Atkins, P.: The Laws of Thermodynamics: A Very Short Introduction. OUP Oxford (2010).
2. Beber, M.E., Gollub, M.G., Mozaffari, D., Shebek, K.M., Flamholz, A.I., Milo, R., Noor, E.: eQuilibrator 3.0: a database solution for thermodynamic constant estimation. *Nucleic acids research*. 50, D603–D609 (2022).
3. Carlson, R., Sreenc, F.: Fundamental Escherichia coli biochemical pathways for biomass and energy production: identification of reactions. *Biotechnol. Bioeng.* 85, 1–19 (2004).
4. CPLEX Users Manual: Ibm ilog cplex optimization studio. Version. 12, 1 (1987).
5. Flamholz, A., Noor, E., Bar-Even, A., Milo, R.: eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Research*. 40, D770–D775.
6. Gebser, M., Kaminski, R., Kaufmann, B., Ostrowski, M., Schaub, T., Wanko, P.: Theory Solving Made Easy with Clingo 5. In: Technical Communications of the 32nd International Conference on Logic Programming (ICLP 2016). Schloss-Dagstuhl - Leibniz Zentrum für Informatik (2016).

7. Gerstl, M.P., Jungreuthmayer, C., Zanghellini, J.: tEFMA: Computing thermodynamically feasible elementary flux modes in metabolic networks. *Bioinformatics*. 31, 2232–2234 (2015).
8. Gerstl, M.P., Ruckerbauer, D.E., Mattanovich, D., Jungreuthmayer, C., Zanghellini, J.: Metabolomics integrated elementary flux mode analysis in large metabolic networks. *Scientific Reports*. 5, 8930 (2015).
9. Jungreuthmayer, C., Ruckerbauer, D.E., Zanghellini, J.: regEfmtool: Speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic. *Biosystems*. 113, 37–39 (2013).
10. Mahout, M., Carlson, R.P., Peres, S.: Answer Set Programming for Computing Constraints-Based Elementary Flux Modes: Application to Escherichia coli Core Metabolism. *Processes*. 8, 1649 (2020).
11. Morterol, M., Dague, P., Peres, S., Simon, L.: Minimality of metabolic flux modes under boolean regulation constraints. In: *Workshop on constraint-based methods for bioinformatics (WCB)* (2016).
12. Motzkin, T.S., Raiffa, H., Thompson, G.L., Thrall, R.M.: The double description method. In: Kuhn, H.W. and Tucker, A.W. (eds.) *Contributions to theory of games*, Vol. 2. Princeton University Press (1953).
13. Peres, S., Morterol, M., Simon, L.: SAT-Based Metabolics Pathways Analysis without Compilation. In: P. Mendes, J.O.D. and Smallbone, K. (eds.) *Lecture Note in Bioinformatics*. pp. 20–31. Springer International Publishing (2014).
14. Peres, S., Schuster, S., Dague, P.: Thermodynamic constraints for identifying the elementary flux modes. *Biochemical Society Transactions*. 46, 641–647 (2018).
15. Peres, S., Jolicœur, M., Moulin, C., Dague, P., Schuster, S.: How important is thermodynamics for identifying elementary flux modes? *PLOS ONE*. 12, 1–20 (2017).
16. Schuster, S., Hilgetag, C.: On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* 2, 165–182 (1994).
17. Terzer, M., Stelling, J.: Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*. 24, 2229–2235 (2008).