



Universität für Bodenkultur Wien
University of Natural Resources
and Life Sciences, Vienna

Department of Biotechnology

Head of Department:

Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Florian Rümer

Advisor:

Univ.Prof. Dipl.-Ing. Dr. Diethard Mattanovich

THERMODYNAMIC METABOLIC PATHWAY ANALYSES
EFFICIENT CALCULATION OF BIOLOGICALLY FEASIBLE ELEMENTARY FLUX MODES

Dissertation

for obtaining a doctorate degree

at the University of Natural Resources and Applied Life
Sciences Vienna

Submitted by

Matthias P. Gerstl

Vienna, May 2015

Abstract

Elementary flux modes (EFMs) are indivisible steady-state pathways through a metabolic network. They can be used to unbiasedly analyse the network, provide a better insight into an organism or optimize organisms for using them in biotechnological processes. The unbiased view has the price of high hardware requirements during the EFM enumeration process as the number of EFMs increases dramatically with the network's size. In the last years, some progress has been made to improve the calculation. For example, networks were split and calculated separately on a cluster or only a subset of all EFMs was taken. Nevertheless, the calculation of EFMs in large genome-scale networks is still out of reach. The present work was motivated by the fact that not all topologically possible EFMs are biologically meaningful. Here, tEFMA (thermodynamical EFM analysis), which is an extension of a previous enumeration tool called *efmtool* is presented. The new tool uses metabolite concentrations to enumerate only those EFMs that are thermodynamically consistent with the given metabolome. tEFMA is published as open source software and available at <https://github.com/mpgerstl/tEFMA>. An *Escherichia coli* model was analysed to show that tEFMA uses less memory and has a decreased runtime for enumerating EFMs. Furthermore, infeasible reaction patterns that are reported by tEFMA were analysed. Besides expected infeasible pathways, like gluconeogenesis when *E. coli* is grown on glucose, it is shown that glutamate dehydrogenase is inactive when grown on glucose whereas it is active when grown on glycerol or acetate, which was confirmed by previously published experiments. This work presents a large step forward to the goal of calculating the full set of biologically feasible EFMs in large genome-scale networks.

Keywords: metabolic network, elementary flux mode, thermodynamic, metabolic pathway

Kurzfassung

Elementare Flussmoden (EFMs) sind unteilbare Stoffwechselwege in einem metabolischen Netzwerk, die sich in einem stabilen Zustand befinden. Sie können für die Analyse des Netzwerks, zum besseren Verständnis oder auch zur Optimierung eines Organismus für biotechnologische Anwendungen verwendet werden. EFMs liefern einen unverzerrten Blick auf das Netzwerk. Dieser Blick hat aber den Preis von hohen Hardwareanforderungen während der Berechnung, da die Anzahl der EFMs mit zunehmender Netzwerkgröße dramatisch ansteigt. Verschiedene Anstrengungen wurden in den letzten Jahren unternommen, um die Berechnung zu verbessern. Zum Beispiel wurden Netzwerke aufgespalten und separat auf einem Cluster berechnet oder es wurde nur ein Teil aller EFMs gewählt. Dennoch ist die EFM-Berechnung in Netzwerken, die vollständige, große Genome widerspiegeln noch immer außer Reichweite. Die Motivation zur vorliegenden Arbeit lag darin, dass nicht alle topologisch möglichen Stoffwechselwege biologisch sinnvoll sind. In dieser Arbeit wird tEFMA (thermodynamische EFM Analyse), eine Erweiterung eines früheren Berechnungsprogramms mit dem Namen *efmtool* präsentiert. tEFMA verwendet gemessene Metabolitkonzentrationen und berechnet nur jene EFMs, die mit den gegebenen Metabolitdaten konsistent sind. Unter <https://github.com/mpgerstl/tEFMA> ist der Quellcode frei verfügbar. Die Vorteile im Speicherbedarf als auch in der verringerten Laufzeit der EFM Berechnung werden am Beispiel eines *Escherichia coli* Modells gezeigt. Weiters werden die von tEFMA zurückgegebenen, undurchführbaren Reaktionskombinationen analysiert. Neben den zu erwartenden inaktiven Stoffwechselwegen, wie Gluconeogenese, wenn *E. coli* auf Glukose wächst, wird gezeigt, dass Glutamate Dehydrogenase inaktiv ist, wenn *E. coli* auf Glukose wächst, aber aktiv ist bei einem Wachstum auf Glycerol oder Acetat, was durch frühere Experimente bestätigt wird. Diese Arbeit stellt einen großen Schritt dar, um in der Zukunft auch in großen Netzwerken alle EFMs, die biologisch sinnvoll sind, berechnen zu können.

Schlagwörter: Metabolisches Netzwerk, Elementare Flussmoden, Thermodynamik, Stoffwechselweg

Acknowledgments

First of all I want to thank all my colleagues and supervisors, who helped me a lot to finish my studies. I want to thank Diethard Mattanovich for his support the last years. Furthermore, I want to thank all people in our office of the metabolic modeling group for fruitful discussions and a lot of fun. Special thanks belong to Christian Jungreuthmayer, who took always the time to teach me programming. Whatever programming language I used and whatever problem I had with it, you knew the solution and shared it with me, thank you! The last sentence at this place is designed for our group leader Jürgen Zanghellini: I do not complain!

Furthermore, I want to thank the members of the Grillari LABS. You have not forgotten about me after I had changed to Jürgen's group and you have still invited me to contribute to very interesting projects and publications.

Ich möchte hier auch die Gelegenheit nützen und mich bei meinen Verwandten und Freunden dafür entschuldigen, dass ich die letzten Jahre zu wenig Zeit hatte. Ein großer Dank gebührt den Musikerinnen und Musikern unseres Musikvereins. Ihr habt es mir immer verziehen, wenn ich einmal nicht richtig vorbereitet zur Probe erschienen bin. Besonderen Dank möchte ich dabei unserem Obmann Ferdinand und meinen stellvertretenden Kapellmeistern Martin und Josef aussprechen. Ihr habt mich die letzten Jahre immer unterstützt und mir viel Arbeit abgenommen. Danke!

Zum Schluss möchte ich jenen danken, die den größten Dank verdienen. Conni, du warst die letzten Jahre immer für mich da und hast mich immer unterstützt. Auch als ich dir nach fünf anstrengenden und zeitraubenden Jahren des berufsbegleitenden Studiums sagte, dass ich noch ein Doktoratsstudium beginnen möchte hast du dich nicht darüber beschwert. Emilie, dir danke ich dafür, dass du mich so erfolgreich von meiner Arbeit und meinem Studium abgelenkt hast. Deine großen leuchtenden Augen, dein Lachen, deine schelmische Art und deine besonderen Begrüßungen, wenn ich am Abend nach Hause kam, haben im letzten Jahr jeden Tag zu einem perfekten Tag werden lassen.

Contents

Abstract	i
Kurzfassung	iii
Acknowledgments	v
Contents	vii
1 Introduction	1
2 Background	5
2.1 Linear program (LP)	5
2.2 Metabolic networks	8
2.3 Elementary Flux modes (EFMs)	10
2.4 Enumeration of EFMs	11
2.5 Internal loops in metabolic networks	13
2.6 Network embedded thermodynamic (NET) analysis	14
3 Results	17
3.1 Detection of internal loops	17
3.2 Integration of NET analysis into EFM enumeration	19
3.3 Functionality and performance of tEFMA	20
3.4 Stability of tEFMA	20
3.5 Biological correctness of tEFMA	21
3.6 Comparison of methods	23
4 Conclusion and Outlook	25
5 tEFMA: computing thermodynamically feasible elementary flux modes in metabolic networks	27
6 Metabolomics integrated elementary flux mode analysis in large metabolic networks	33

Contents

Bibliography	71
Annex	79
List of Publications	79
Curriculum Vitae	81

A common approach in the field of systems biology is to reconstruct and then analyse biochemical networks. These networks may represent for instance regulatory, transcriptional, signaling or metabolic states of organisms [1]. Many reconstructions, starting from small bacterial networks [2] to huge and more detailed reconstructions, like a genome scale model of human metabolism [3], were published in the last years. The present work focuses on metabolic networks. Metabolic networks are mathematically described by a stoichiometric matrix, which lists the reactants and products of all reactions in an ordered fashion. They are used for metabolic engineering, biological discovery, determining phenotypic behavior or analysing the networks [4]. Common approaches to analyse fluxes through these networks are based on flux balance analysis (FBA) [5], which finds fluxes in networks by an optimization procedure [6], like optimizing for biomass or energy. In contrast to them, elementary flux modes (EFMs) do not rely on an optimization procedure and provide an unbiased view of the network [7]. EFMs are steady-state fluxes through networks and fulfill the minimality constraint. The minimality condition states that an EFM cannot be further decomposed into other pathways that would still fulfill the steady-state condition [8]. The usefulness of those EFMs can be illustrated when they are described geometrically. Provided that all reactions in the network are irreversible, EFMs are the rays of a polyhedral cone, spanning the space of the metabolic capabilities of the organism. Every possible flux through the network sits within this space and can be described as a non-negative linear combination of these rays [9]. In this work all reactions in a network are considered to be irreversible, as every reversible reaction can be split into an irreversible forward and an irreversible backward reaction.

Besides the identification of pathways [10], defining the flexibility of networks [11] and determining the importance of reactions [12], EFMs provide means to predict genotype-phenotype relationships [9] and to improve the production in biotechnological processes. For instance, EFMs of an *E. coli* network were successfully analysed to genetically modify the organism, so that growth [13] was improved or the production of ethanol [14] or diapolycopendioic acid [15] was enhanced.

However, the unbiased view of the network has its price, as the number of EFMs increases dramatically with the network's size. Even more, the total number of EFMs in a network and the computational complexity of their calculation is not known until they are fully enumerated [16]. A pessimistic upper bound of expected EFMs in a

network can be calculated as described by Klamt and Stelling [17]. Different concepts exist for the calculation of EFMs. On the one hand they can be calculated one after another, e.g. by starting with the shortest EFM, followed by the next shortest and so on [18] or by using a sampling method [19, 20]. Such methods can return identified EFMs immediately. On the other hand tools exist that are based on the double description method. For this iterative method the nullspace of the stoichiometric matrix is created and the algorithm is started with an initial solution consisting of intermediate EFMs. At each iteration step an additional reaction is added to the set of intermediate EFMs that separates valid from invalid rays of the cone. Invalid rays are combined with adjacent valid rays. Only those rays are kept that are valid and fulfill the minimality constraint. In contrast to the former described methods, EFMs can only be returned after the last iteration phase [12, 21]. To the best of my knowledge, *efmtool* [22], which is based on the double description method, is currently the most efficient tool to calculate the total set of EFMs in a network [23]. The efficiency of *efmtool* is based on the usage of binary pattern trees to speed up the adjacency and superset tests. Nevertheless, this tool still requires huge amount of memory.

Hunt et al. [24] suggested to split the network into subnetworks and calculate the EFMs of these subnetworks with *efmtool* on a cluster. With this approach they were able to enumerate ~ 2 billion EFMs in a genome-scale metabolic network of *Phaeodactylum tricornutum*. However, Schwartz and Kanehisa [25] argued that only few EFMs are biologically relevant. Recently, Jungreuthmayer et al. [26] extended *efmtool* to follow regulatory rules. They showed, that their tool calculates only those subset of EFMs, that fulfill the given rules. Based on a similar idea, the present work describes a way to further reduce the set of EFMs considering their thermodynamical properties.

In 2006, Kümmel et al. [27] developed network embedded thermodynamic (NET) analysis, which is based on first principles of thermodynamics, as described by Alberty [28]. The basic idea of NET analysis is the second law of thermodynamics, which states that a reaction in a closed system at constant pressure can only occur spontaneously in the direction of negative Gibbs energy. The Gibbs energy of a reaction depends on the pH value, temperature and ionic strength of the environment as well as on the value of Gibbs free energy of formation and on the concentration of the participating species of the reaction. The formulation of the NET analysis ensures that not only a specific reaction, but all reactions that support a thermodynamic feasible flux have a negative Gibbs energy. As concentration values of the participating species are not fixed values the NET analysis is formulated as a linear problem, which can be solved by various solvers, like the Matlab Optimization Toolbox, GLPK or *cplex*.

Jol et al. [29] applied a NET analysis to the full set of calculated EFMs of a reduced *Saccharomyces cerevisiae* model and found that approximately half of all topologically possible EFMs are thermodynamically infeasible. The authors did not tackle the hardware and runtime problem of the double description algorithm as the total set of EFMs was calculated before the thermodynamical analysis was applied, but it was shown that it is worth the effort to find a way to include the NET analysis into the double description algorithm. The present work describes such a way. Furthermore, it will be shown that including the NET analysis into the enumeration algorithm improves the efficiency of the EFM calculation and simplifies the analysis of infeasible pathways.

In the next chapter background information on metabolic networks, elementary flux modes and their enumeration is given. Furthermore, linear programming is introduced and theory on internal loops in metabolic networks as well as on network embedded

thermodynamic analysis is provided.

In chapter 3 the main results of this thesis are summarized, followed by the conclusion and an outlook in chapter 4.

In chapter 5 tEFMA, which is an extension of *efmtool*, is presented. *efmtool* is published under an open source license, which allowed me to modify and extend the source code. The now extended version, tEFMA, avoids the enumeration of thermodynamically infeasible EFMs by an integrated NET analysis. It will be shown that tEFMA requires weaker hardware and saves runtime compared to an ordinary EFM analysis. The goal of the present work was not only to enumerate those EFMs that are thermodynamically consistent with a given metabolome, but also to understand what causes the infeasibility of pathways. To this purpose additional code was implemented that returns those reaction patterns that are identified to cause the infeasibility.

In chapter 6 an analysis of an *E. coli* core network, which includes the pathways of glycolysis, tricarboxylic acid (TCA) cycle and the pentose phosphate pathway (PPP) [2], is presented. Published metabolite concentrations [30] are applied to tEFMA and the organism is analysed under different growth conditions. It is shown that up to 80% of all topologically feasible EFMs are thermodynamically infeasible. Moreover, it can be seen, that all these results are obtained by an approach, which is free of data fitting and relies only on first principles of thermodynamics. Furthermore, the results of the analyses of identified infeasible reaction patterns are presented. For instance, it is shown that glutamate dehydrogenase (GDH) is inactive when *E. coli* is grown on glucose, while it is active when grown on glycerol or acetate. This result was confirmed by previous publications [31, 32]. To demonstrate that tEFMA is stable against perturbations in concentration and thermodynamics data, additional results are presented. To show the scalability of tEFMA, thermodynamic consistent EFMs of a larger and more detailed *E. coli* model are calculated.

2.1 Linear program (LP)

In the successive chapters LPs are used to identify thermodynamically infeasible network states. That is why the essentials of LPs are briefly reviewed below. Many practically relevant problems are LPs, e.g. optimizing production processes or solving scheduling problems [33]. In the field of systems biology LPs are used to find optimal fluxes in networks, as it is done by FBA approaches, or to enumerate EFMs by solving LPs iteratively (see Section 2.4).

A linear problem has the following standard form:

$$\text{minimize } \mathbf{c}' \cdot \mathbf{x} \quad (2.1a)$$

$$\text{subject to } \mathbf{A} \cdot \mathbf{x} = \mathbf{b} \quad (2.1b)$$

$$\mathbf{x} \geq \mathbf{0} \quad (2.1c)$$

where \mathbf{x} is the vector of decision variables, \mathbf{A} is a $m \times n$ matrix of coefficients, \mathbf{b} is the solution vector and \mathbf{c}' is the cost vector. Eqn. (2.1a) is the objective function of the problem. If the problem is feasible the minimized cost vector will be calculated. Eqs (2.1b – 2.1c) define the constraints of the system. Note, that the objective function can be maximized by changing the sign of the cost vector.

Simple LPs can be solved graphically as shown in Figure 2.1. As the LP of this example consists only of two variables a two dimensional plot can be produced to find the solution. The problem is constraint by four different inequations. Every constraint is converted to a line in the plot. The area, which is surrounded by these lines is the feasible set of the LP. To find the optimal solution the sum of x_1 and x_2 is maximized. For this purpose the objective function is also drawn and then shifted to its maximum value, which is a vertex of the feasible set.

More complicated examples need to be solved by a computer. A common method to solve LPs is to use the simplex method. The simplex method is based on the fact that the solution of the problem is a vertex of the feasible set and is never a point inside the feasible set. In short, this method calculates first a basic feasible solution (BFS). A

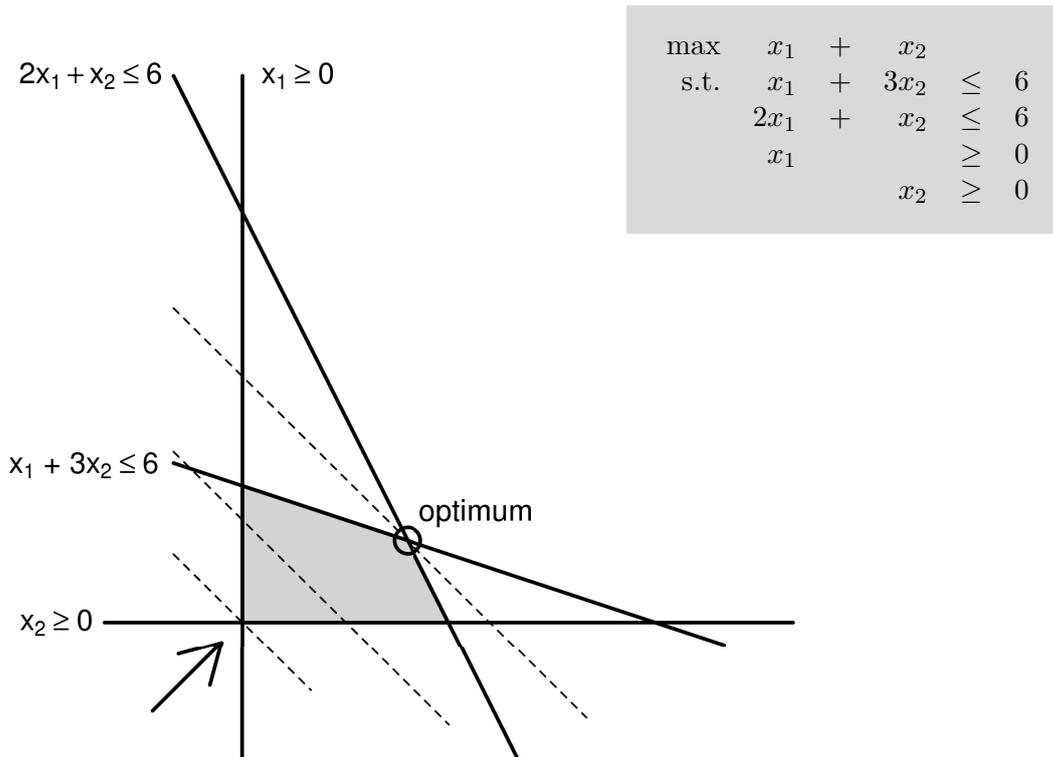


Figure 2.1: Solving a LP by a graphical approach: the four constraints of the problem are plotted by thick lines. The gray area, which is bound by these constraints, shows the feasible set of the LP. Dashed lines show the objective function. By moving the dashed line in the direction indicated by the arrow the objective is maximized and the optimum solution is found in the point (2.4/1.2) marked by a circle.

BFS is a solution of the problem that fulfills all given constraints, but does not need to be the optimum solution. In iterative steps the algorithm moves along the edges of the feasible set to the next BFSs, as long as the cost vector can be optimized and terminates when the optimum is found (see Figure 2.2).

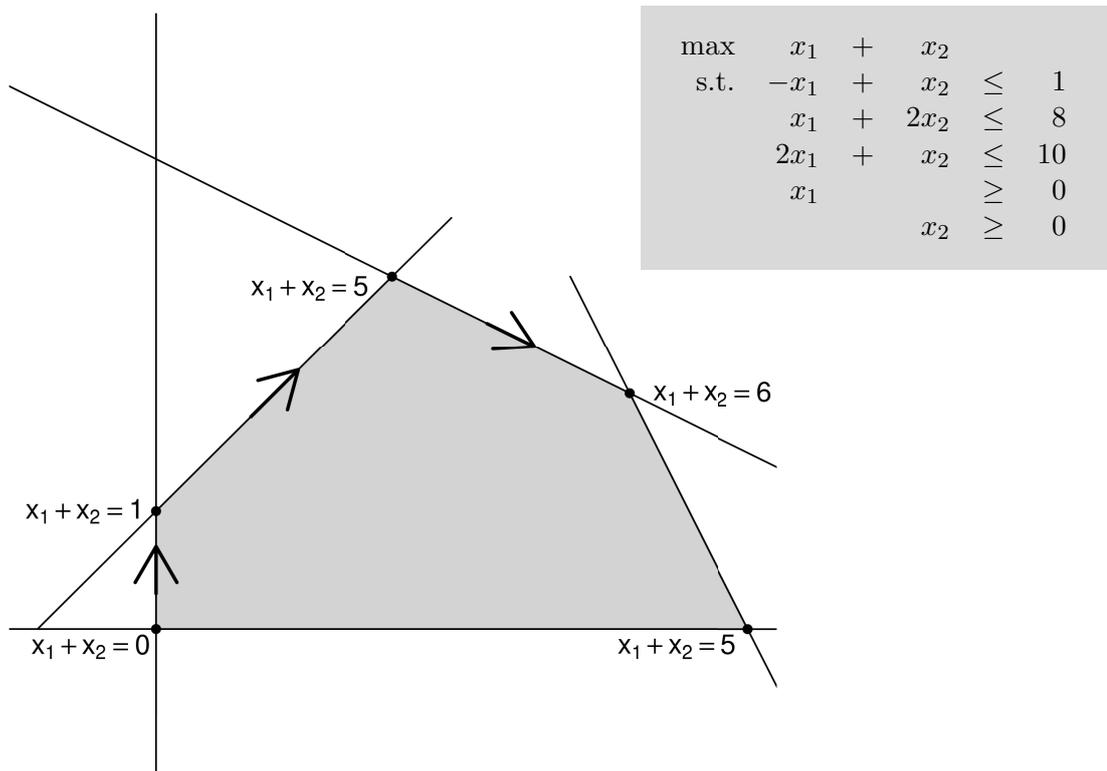


Figure 2.2: Solving a LP by the simplex method: the gray area, which is bound by five given constraints, shows the feasible set of the LP. The simplex method searches the maximum of $x_1 + x_2$ and starts at the first BFS ($x_1 = 0; x_2 = 0$) with an objective value of 0. In the next steps it moves to the next BFSs ($x_1 = 0; x_2 = 1; \text{obj} = 1$), ($x_1 = 2; x_2 = 3; \text{obj} = 5$) and ($x_1 = 4; x_2 = 2$) where it finds the optimum value 6 and terminates.

2.2 Metabolic networks

In this thesis metabolic networks and the feasibility of fluxes through these networks are analysed. In the following a brief summary on the reconstruction of metabolic networks as well as their preparation for the EFM enumeration is given. A detailed protocol to generate a metabolic reconstruction was previously published [34]. This protocol shows that the generation of a model is a very time consuming process. Biochemical knowledge need to be collected from different data sources and experiments are necessary to determine the substrate and cofactor usage as well as the biomass composition. Additionally, lower and upper bounds of reaction fluxes are defined. The model is built in repeating steps, which require manual curations to achieve high quality. The final model consists of different reactions, of which each converts its reactants to its products. The mass balance states that for every metabolite the following differential equation holds:

$$\frac{d[\mathbf{x}]}{dt} = \mathbf{S} \cdot \mathbf{v}([\mathbf{x}]) \quad (2.2)$$

where $[\mathbf{x}]$ is the concentration of the metabolites, t is the time, \mathbf{S} is the stoichiometric matrix of the model and $\mathbf{v}([\mathbf{x}])$ is the vector of the reaction rates. Reactions in the model can be reversible and proceed in both directions or irreversible if they are unidirectional (see Figure 2.3). For a comprehensive computational analysis of a metabolic network kinetic data and rate laws need to be considered. However, when analysing metabolic networks with common tools, like flux balance analysis (FBA) or elementary flux mode analysis (EFMA), it is assumed that (internal) metabolite concentrations do not change over time and therefore the fluxes \mathbf{v} are in steady state and the following equation holds:

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0} \quad (2.3)$$

As a result all external metabolites, which are not in steady state and act as sources and sinks, are omitted when creating the steady state stoichiometric matrix. An example of a (steady state) stoichiometric matrix is shown in Table 2.1. The rows of the stoichiometric matrix show the metabolites whereas the columns show the reactions of the model. The values of the matrix represent the stoichiometric coefficients of the participating metabolites in the reactions. The sign of the coefficients identifies reactants (negative values) and products (positive values).

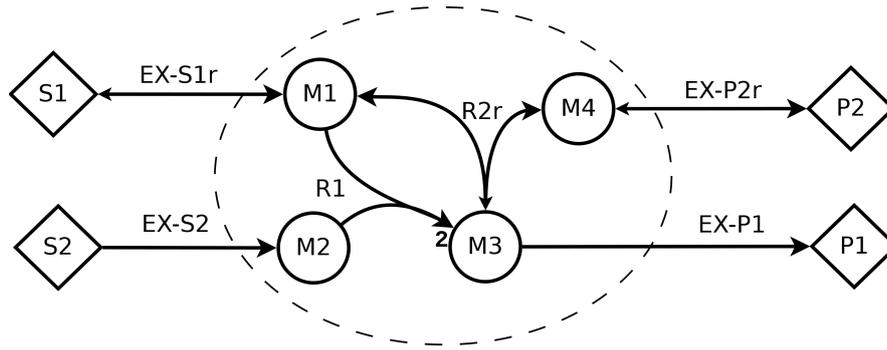


Figure 2.3: Toy network. The dashed line defines the system border of the network. Substrates (S1 and S2) and products (P1 and P2) are external metabolites, which act as sources and sinks and are not in steady state, whereas M1, M2, M3 and M4 are internal metabolites. External metabolites are connected to internal metabolites by exchange reactions (EX-S1r, EX-S2, EX-P1r and EX-P2r). Reactions can be irreversible (EX-S2, R1 and EX-P1) or reversible (EX-S1r, R2r, EX-P2r). Large arrowheads show the forward direction of the reaction and small arrowheads the reverse direction. As reaction R1 produces 2 M3 out of M1 and M2 the stoichiometric coefficient is added to this reaction.

		EX-S1r	EX-S2	R1	R2r	EX-P1	EX-P2r
steady state	M1	1	0	-1	1	0	0
	M2	0	1	-1	0	0	0
	M3	0	0	2	-1	-1	0
	M4	0	0	0	1	0	-1
not in steady state	S1	-1	0	0	0	0	0
	S2	0	-1	0	0	0	0
	P1	0	0	0	0	1	0
	P2	0	0	0	0	0	1

Table 2.1: Stoichiometric matrix of the toymodel (see Figure 2.3). Rows show the metabolites and columns the reactions of the model. The values describe the stoichiometric coefficients of the participating metabolites in the corresponding reaction. Zero indicates that the metabolite is not part of the corresponding reaction, whereas a negative value indicates that the metabolite acts as an reactant and a positive value defines a product. Substrates (S1 and S2) and products (P1 and P2) are not in steady state and therefore not part of the steady state stoichiometric matrix \mathbf{S} .

2.3 Elementary Flux modes (EFMs)

In the present work metabolic networks are analysed with (thermodynamic consistent) EFMs. The following section briefly introduces them. An EFM is a flux through the network that meets several properties. The first property of an EFM is that the flux vector must not be zero ($\mathbf{v} \neq \mathbf{0}$). Furthermore, an EFM fulfills the steady state condition ($\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$) ensuring that the flow rate to a metabolite is the same as from the metabolite. The next property states that an EFM meets all reversibility constraints of the network, ensuring that the flux of a reaction flows in the allowed direction. Furthermore, EFMs need to be minimal, which states that a flux vector that is a superset of another flux vector in the network is not an EFM. Figure 2.5 shows examples that do not fulfill these requirements. Four different EFMs (see upper panel in Figure 2.4), which fulfill all required properties can be found in the former presented toy model (see Figure 2.3). The flux values of the EFMs are summarized in Figure 2.4. Note, that all flux values of an EFM can be multiplied by any positive value ($\mathbf{v} \cdot \mathbf{a} \quad a > 0$) and the result is still an EFM.

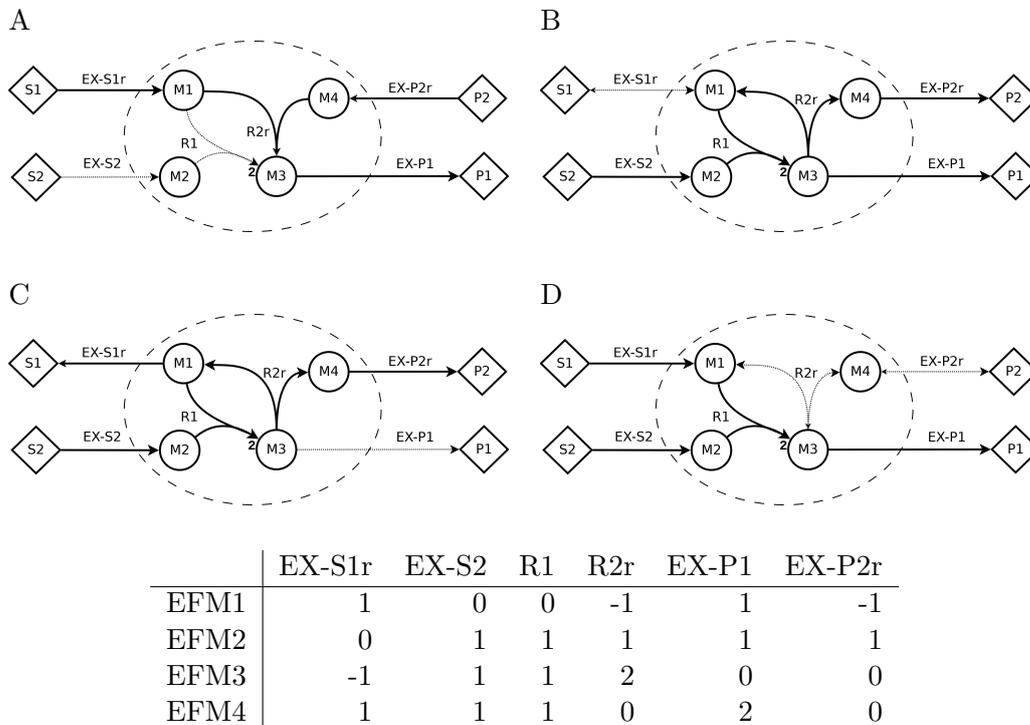


Figure 2.4: Elementary flux modes of the toy network (see Figure 2.3). The upper panel shows all four EFMs that can be found in the network and the lower panel shows the corresponding flux values of these EFMs. Thick lines represent those reactions that support the EFM and thin dotted lines represent reactions that are not part of the EFM. Positive values in the matrix represent reactions that have a flux in the forward direction, whereas a negative value represents a flux in the backward direction.

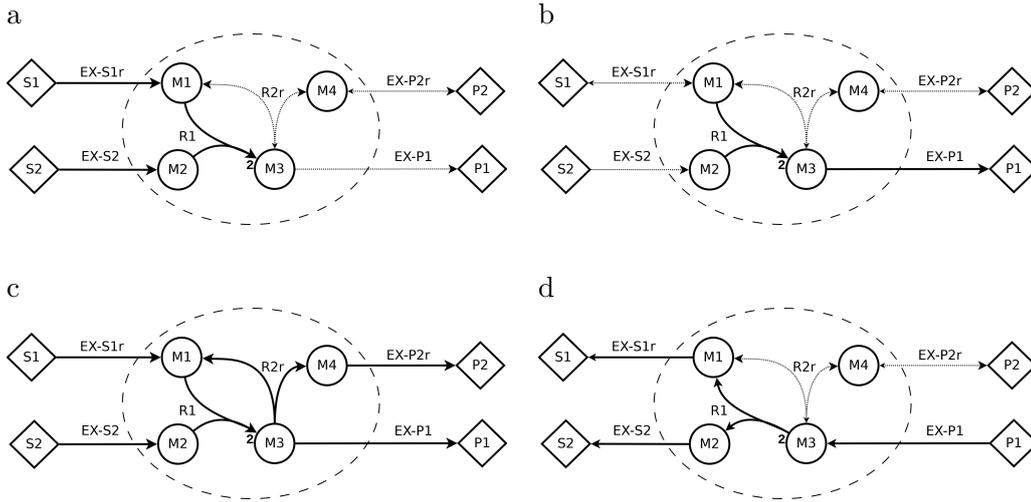


Figure 2.5: Examples showing constructs that do not fulfill the properties of EFMs. (a) and (b) do not fulfill the steady state condition ($\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$) as in (a) M3 would accumulate in the system and in (b) M1 and M2 are missing their sources. Though the next example (c) shows a flux that fulfill the steady state condition it is not minimal as it is a superset of the EFMs shown in Figure 2.4 (B and D). (d) is not an EFM as it contradicts the reversibility constraint of the reactions EX-S2, R1 and EX-P1, as the fluxes flow in the backward direction which is not allowed by the network model (see Figure 2.3).

2.4 Enumeration of EFMs

Several tools exist to enumerate EFMs in metabolic networks. The present work is based on *efmtool* [22], which belongs to the group of tools that use the double description method [12, 35]. *efmtool* uses a binary approach of this algorithm, which is briefly reviewed below.

In a first step, without loss of generality, all reversible reactions are split into an irreversible forward and an irreversible backward reaction. The double description method starts with an initial set of intermediate EFMs. This set is found by calculating the kernel matrix of \mathbf{S} , which is a basis for all vectors \mathbf{v} , that are true for $\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$. The kernel matrix is calculated by the equation

$$\mathbf{S} \cdot \mathbf{K} = \mathbf{0} \quad (2.4)$$

where \mathbf{S} is the stoichiometric matrix of the model and \mathbf{K} is the kernel matrix. For the next steps the kernel matrix is considered to be the mode matrix, where the columns are the (intermediate) EFMs and the rows are the reactions. In the binary approach of the double description method the mode matrix is split into a binary part and a numerical part. In iterative steps each line of the matrix is converted from numerical values into its binary values. Rows that contain only zero and positive values can be easily converted into its binary form. *efmtool* uses an inverse logic and considers zeros, which are reactions without a flux to be true and positive values, which are

2 Background

reactions containing a flux to be false. For rows with negative values additional steps are necessary. Intermediate EFMs with a negative value in the corresponding row are combined with adjacent intermediate EFMs that have a positive value in this row, such that the numerical part is combined by applying following formula to all left rows:

$$v_{comb_r} = \frac{v_1^+ v_r^- - v_1^- v_r^+}{v_1^+ - v_1^-} \quad (2.5)$$

where v_r are the values of the negative (-) and positive (+) column of the corresponding row r and the index 1 is the first row containing numerical values. The binary part of the two intermediate EFMs are combined by *efmtool* using the bitwise *AND* function. A negative side effect of this method is that it can create supersets of already found intermediate EFMs. As EFMs are minimal (see section 2.3), only those combined columns are added to the mode matrix that do not turn out to be supersets of other columns. After combining the negative column with all adjacent positive columns, the negative column is removed from the mode matrix.

This procedure is repeated until the last row is converted to its binary form. The resulting matrix contains all EFMs in binary form. As all reversible reactions of the network were split into irreversible forward and irreversible backward reactions futile-2-cycle modes are created during the enumeration, which are removed from the EFM matrix.

In the last step the numerical flux values of the EFMs are calculated. The calculation of the numerical values is performed by solving a homogeneous linear system for each single EFM. Therefore, a subset \mathbf{K}_{sub} of the kernel matrix \mathbf{K} is built. \mathbf{K}_{sub} contains only those reactions that support the EFM, which can be easily found in the binary representation of the EFM. In the next step the linear system is solved by following equation:

$$\mathbf{K}_{sub} \cdot \mathbf{v} = \mathbf{0} \quad \mathbf{v} \in \mathbb{R}_+^n \quad 0 \quad (2.6)$$

where \mathbf{v} is the flux vector of the EFM in numerical form.

The advantage of this implementation of the double description method is that it uses a binary form to store (intermediate) EFMs. This leads to reduced memory usage and allows using bit operations to combine intermediate EFMs. Additionally, *efmtool* uses bit pattern trees to increase the performance of adjacency and superset tests, as these tests are computationally expensive operations. Therefore, at each iteration step intermediate EFMs are partitioned into three trees T^0, T^+ and T^- based on the numerical value of the particular reaction of the current iteration step. The tree implementation leads to an improvement of the runtime, as not all EFM combinations need to be checked. If it turns out while traversing the trees that intermediate EFMs are not adjacent or that all leaves of the (sub)tree do not contain any supersets the test can be stopped. To further improve the performance of the EFM enumeration Terzer and Stelling [22] added different compression methods to *efmtool*.

However, the double description method has also disadvantages. As the algorithm uses only a single matrix that grows dramatically for large networks during the enumeration it needs huge amount of memory for such large systems. To tackle this problem Hunt et al. [24] suggested to split the network into subnetworks and calculate the EFMs of the subnetworks on a cluster. Another disadvantage is that the true complexity of

the algorithm is not known [16], which makes it difficult to predict the runtime and the memory usage of the overall algorithm. Moreover, it is known that the order of the reactions in the kernel matrix influences the performance of the algorithm [21, 36]. However, at the moment no method exists that can predict the best order.

Alternative methods to calculate EFMs use LPs (see Section 2.1). For instance, EFMs with an increasing number of supported reactions can be calculated with LPs as suggested by De Figueiredo et al. [18], whereas Pey and Planes [37] calculated EFMs that satisfy several biological constraints. Another approach uses LPs to decompose fluxes into EFMs [38]. LPs can also be combined with the power of genetic algorithm to find EFMs by knocking out reactions and enumerate only those EFMs that are not supported by the knocked out reactions [19].

2.5 Internal loops in metabolic networks

The goal of the present work was to identify thermodynamically infeasible EFMs. One possible way to determine infeasible EFMs is to identify internal loops, which are introduced in the following. Internal loops in metabolic networks can be seen analogous to the second law for electrical circuits as described by Kirchhoff [39]. The change of free energy across a circle flux over a closed loop is zero. That means that the energy level is a state function and independent of the path (see Figure 2.6). Considering first principles of thermodynamics a flux through a closed loop is not feasible, as the second law of thermodynamics states that a reaction only occurs in the direction of a negative Gibbs energy ($\Delta_r G < 0$) [28]. Beard et al. [40] described constraints for the detection of internal loops in metabolic networks and Schellenberger et al. [41] published a loopless constraint-based reconstruction and analysis (COBRA) method which removes internal loops from the solution space when calculating optimal fluxes. However, fluxes through internal loops of metabolic networks are EFMs, as they fulfill the minimality and the

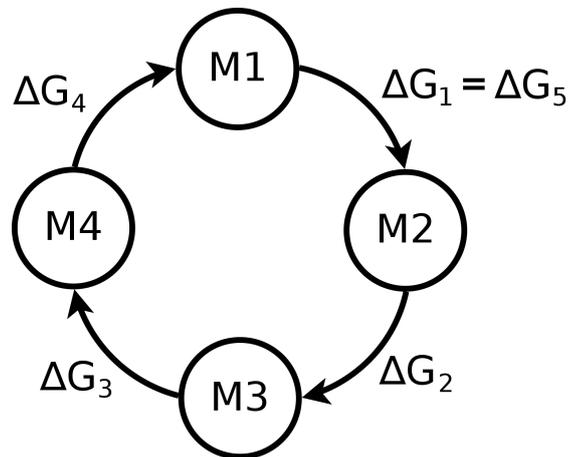


Figure 2.6: Loop in metabolic networks. The loop law states that the change of free energy across a closed loop is zero. The figure shows that a conversion from M1 to M1 over M2, M3 and M4 proceeds without a change of free energy ($\Delta G_1 = \Delta G_2 = \Delta G_3 = \Delta G_4 = \Delta G_5$).

2 Background

steady state condition. Therefore, they are enumerated by EFM calculation tools. As internal loops are in steady state they can be detected by following equation:

$$\mathbf{S}_{red} \cdot \mathbf{v} = \mathbf{0} \quad \mathbf{v} \in \mathbb{R}_+^n, \quad \mathbf{v} > \mathbf{0} \quad (2.7)$$

where \mathbf{v} is the flux vector of the loop and \mathbf{S}_{red} is the reduced internal matrix of the metabolic model. \mathbf{S}_{red} is defined as the steady state matrix of the network without those reactions that provide connections to sources and sinks of the network.

2.6 NET analysis

A different way to determine the thermodynamic state of an EFM can be done by network embedded thermodynamic (NET) analysis [27], which is briefly reviewed below. The NET analysis follows first principles of thermodynamics and is based on the fact that a biochemical reaction can only proceed spontaneously under constant pressure in the direction of negative Gibbs energy of reaction ($\Delta_r G_i$) [28]. NET analysis consists of variable metabolite concentrations and equations that define free Gibbs energy levels and can therefore be formulated as a LP.

To use NET analysis Gibbs free energy of formation ($\Delta_f G$) values for all metabolites need to be adapted to the environmental conditions as described by Alberty [28] using following equations

$$\Delta_f G^{0'} = -RT \ln \left\{ \sum_i \exp \left[-\Delta_f G_i^0(pH, I)/RT \right] \right\} \quad (2.8a)$$

$$\Delta_f G_i^0(pH, I) = \Delta_f G_i^0 - \Delta_f G_i^0(pH) - \Delta_f G_i^0(I) \quad (2.8b)$$

$$\Delta_f G_i^0(pH) = N_{H_i} RT \ln (10^{-pH}) \quad (2.8c)$$

$$\Delta_f G_i^0(I) = \frac{\bar{I}A z_i^2 - N_{H_i}}{1 + \bar{I}B} \quad (2.8d)$$

$$A = 2.91482 \text{ kJ mol}^{-1} \text{ M}^{-0.5}, \quad B = 1.6 \text{ M}^{-0.5} \quad (2.8e)$$

where $\Delta_f G^{0'}$ is the standard transformed Gibbs energy of formation, R is the molar gas constant ($R = 8.314 \text{ J/mol K}$), T is the temperature, pH is the pH value and I is the ionic strength of the environment. Furthermore, each charge state i of the metabolite is considered with its standard Gibbs free energy of formation $\Delta_f G_i^0$, charge z_i and number of H atoms N_{H_i} .

In order to show that an EFM is thermodynamically feasible, each of the reaction that supports this EFM needs to have a negative Gibbs energy of reaction, which is ensured if following LP can be solved [27].

$$\min 0 \quad (2.9a)$$

$$\text{s.t. } \Delta_r G_i \leq 0, \quad i \in \text{supp}(\mathbf{e}_k) \quad (2.9b)$$

$$\Delta_r G_i = \sum_{j=1}^m S_{ji} \Delta_f G'_j \quad (2.9c)$$

$$\Delta_f G'_j = \Delta_f G_j^0 + RT \ln(c_j/c_0), \quad c_0 = 1 \text{ M} \quad (2.9d)$$

$$\ln(c_j^{\min}/c_0) \leq \ln(c_j/c_0) \leq \ln(c_j^{\max}/c_0). \quad (2.9e)$$

The constraints of the LP ensure that every reaction i that supports the EFM \mathbf{e}_k has a negative Gibbs free energy of reaction $\Delta_r G_i$ (Eqn. (2.9b)). $\Delta_r G_i$ depends on the transformed Gibbs free energy of formation ($\Delta_f G'_j$) of all contributing metabolites according to their stoichiometric coefficients (Eqn. (2.9c)). To calculate $\Delta_f G'_j$ values the former calculated standard transformed Gibbs free energy of the metabolite $\Delta_f G_j^0$ as well as the molar gas constant R , the temperature T and the logarithm of the concentrations of the metabolites are used (Eqn. (2.9d)). Concentrations of the metabolites are constrained by lower (c_j^{\min}) and upper (c_j^{\max}) bounds. To keep the problem linear the logarithm of the concentrations can already be used to define the bounds (Eqn. (2.9e)).

In contrast to the LP examples in section 2.1 and to previously published works on NET analyses [27, 29] no optimum value will be searched in the NET analyses of the present work (Eqn. (2.9a)). In this case the LP solver only returns the feasibility of the model. The reason to use this approach is that the optimum value is not necessary to define the feasibility state of an EFM. If the EFM is thermodynamically feasible there also exists an optimal solution and if an optimal solution exists there also exists a basic feasible solution. On the other hand if the EFM is not feasible the algorithm can neither find an optimal solution nor a basic feasible solution. From this follows that if the algorithm returns a basic feasible solution we know that the EFM is thermodynamically feasible. In contrast to an approach that searches the optimum solution, which can cause high computational burden, an approach that relies only on the first basic feasible solution uses less computation time, which leads to an increase of the performance of the overall NET analysis.

The present work is based on the fact that an infeasible EFM can be safely removed during the enumeration phase of the double description method. In detail, the double description algorithm combines two intermediate EFMs in such a way, which ensures that each active reaction of the single EFMs stays active in the combined EFM. In other words, if the combination of reactions of an intermediate EFM turns out to be infeasible all offsprings of this intermediate EFM will also be infeasible. A comprehensive proof is shown on page 65. Following this idea every infeasible EFM can be removed immediately as soon as it is detected. In the present work the infeasibility was determined by considering the thermodynamical state of EFMs. In particular, two different methods were developed to detect thermodynamically infeasible EFMs. The first method finds internal loops by matrix vector multiplications, whereas the second method detects thermodynamically infeasible pathways with the help of NET analyses.

3.1 Detection of internal loops

According to the loop law no net flux can exist across a closed network cycle at steady state [41]. The following equation describes an internal loop in a network:

$$\mathbf{S}_{red} \cdot \mathbf{v} = \mathbf{0} \quad \mathbf{v} \in \mathbb{R}_+^n, \quad \mathbf{v} > 0 \quad (3.1)$$

where \mathbf{S}_{red} is the reduced internal matrix and \mathbf{v} is the flux vector of the internal loop [42]. To detect and remove these inner loops from the set of EFMs during the iteration phase of *efmtool* the following steps were added to the program. In the first step the reduced internal matrix \mathbf{S}_{red} of the network is built. Therefore, a matrix is created from all reactions in the original stoichiometric matrix that are not connected to sources or sinks of the network (Figure 3.1). The next step is to find and remove loops during the iteration phase of the double description method. Therefore, every newly created intermediate EFM is checked by Eqn. (3.1) and removed if the equation holds. However, in special cases the internal loop is still needed by the double description method to prohibit the enumeration of supersets of the loop. If at least two metabolites of an internal loop are involved in a second loop (Figure 3.2), I observed that a superset of these two loops can be generated by the double description method if the single loops were removed before. The found superset is not an EFM, as it does not fulfill the minimality condition. Even worse, the found superset is not in steady state as the

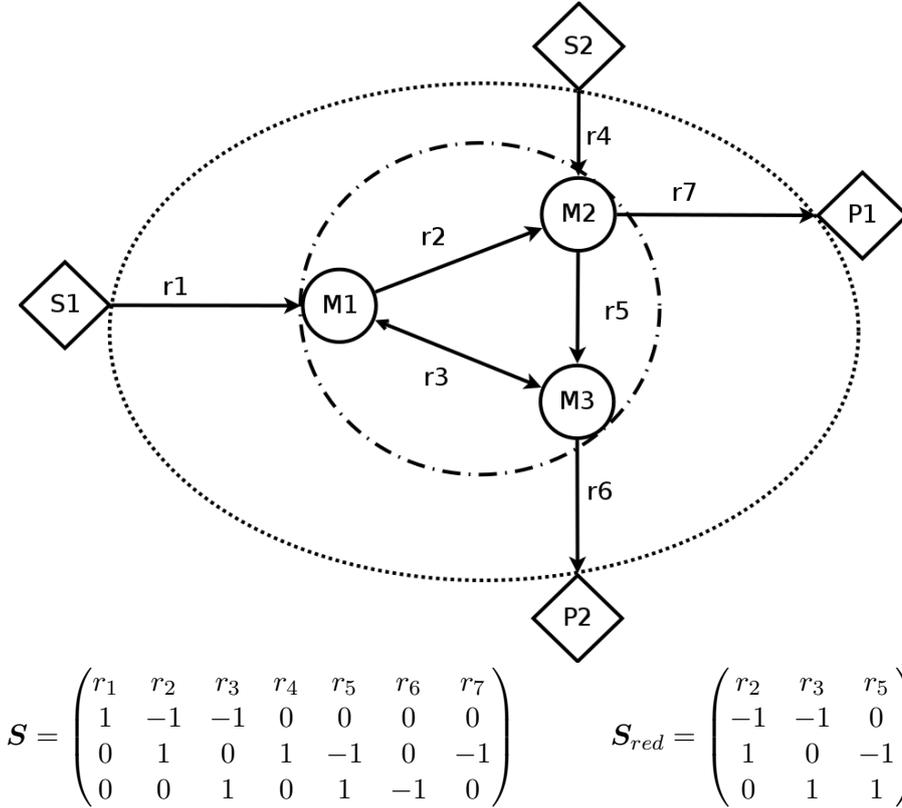


Figure 3.1: Search of the reduced internal stoichiometric matrix: substrates (S_1 and S_2) and products (P_1 and P_2) are not part of the steady state stoichiometric matrix \mathbf{S} . The reactions r_1 and r_4 are connected to the sources and r_6 and r_7 to the sinks of the network. Removing these reactions from the stoichiometric matrix leads to the reduced internal matrix \mathbf{S}_{red} . The dotted line visualizes the matrix \mathbf{S} and the dashed dotted line the reduced internal matrix \mathbf{S}_{red} with its reactions r_2, r_3 and r_5 .

equations $\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$ and $\mathbf{S}_{red} \cdot \mathbf{v} = \mathbf{0}$ do not hold anymore. Therefore, such a superset is not detected by this approach itself. Consequently, if the loop is removed during the iteration phase of the double description method it needs to be kept in memory. Every new found EFM needs to be checked against all former detected loops and is removed from the set of EFMs if it turns out that it is a superset of any of the loops.

In the tree implementation of the double description method internal loops are located on tree T^0 . During the iteration phase of the double description method intermediate EFMs of T^+ and T^- are checked for adjacency and combined if adjacent. Therefore, intermediate EFMs on T^0 will not decrease the runtime of the adjacency search. Even worse, the opposite is the case. Because of the additional superset checks of new found EFMs against already detected internal loops the runtime of the overall enumeration algorithm increases. In summary, using this approach to detect internal loops of the network already during the enumeration of EFMs by the double description method does not increase the performance of the algorithm. In practice, I could observe that a consecutive approach by enumerating the EFMs first and removing the internal loops

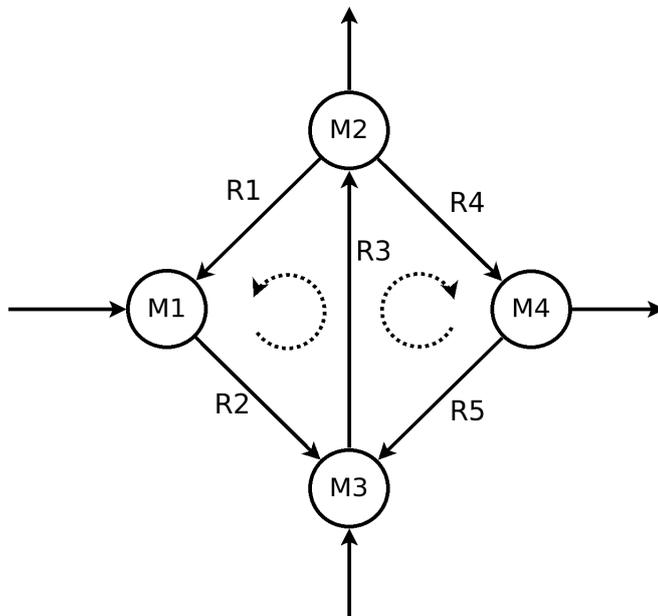


Figure 3.2: Removing loops during the iteration phase of the double description method can lead to wrong results. In this network two EFMs that are internal loops (R_1, R_2, R_3) and (R_3, R_4, R_5) are found by *efmtool*. By removing these internal loops during the iteration phase the superset of the loops (R_1, R_2, R_3, R_4, R_5) are obtained as a result. This result is not an elementary flux mode, as it is not minimal. Moreover, it is not a flux, as it does not fulfill the steady state condition $S \cdot v = \mathbf{0}$.

in a second step is faster than the former described approach.

3.2 Integration of NET analysis into EFM enumeration

To find thermodynamically infeasible EFMs in metabolic networks, NET analysis (see Section 2.6) was integrated into *efmtool* (see also Chapter 5). Therefore, a Java package with several classes was written and new options were added to the program. The program was renamed to tEFMA and the source code was uploaded to GitHub (<https://github.com/mpgerstl/tEFMA>). To perform NET analysis in tEFMA several steps are necessary. The first step is to read the additional input files and calculate the standard transformed Gibbs energy of formation $\Delta_f G_j^{\prime 0}$ for each metabolite according to the pH value and ionic strength of the environment (see Eqs. 2.8). To keep the problem of the NET analysis linear the logarithms of the minimum and maximum concentrations of each metabolite are calculated. Moreover, to increase the performance a template LP for the NET analysis (see Eqs. 2.9) is created that considers equations for the transformed Gibbs energy $\Delta_f G_j^{\prime}$ of all metabolites. Furthermore, the equations for the Gibbs energy of reaction $\Delta_r G$ of all characterized reactions are prepared.

Before new trees are built the thermodynamic state of each intermediate EFM that is designated for the trees T^+ or T^- is determined with *cplex*, which is a commercial solver by IBM for which academic licenses are available. Therefore, equations for $\Delta_r G$

Table 3.1: Comparison between tEFMA and an ordinary EFM analysis (EFMA). Numbers in brackets list the percentage compared to EFMA.

Method	Carbon source	Number of EFMs		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%)

of all thermodynamically characterized reactions that support a particular EFM are added to the LP template. In the next step *cplex* is called with the newly created LP problem. To increase the performance and reduce the runtime of the NET analysis only the feasibility of the problem is determined and in contrast to standard LPs an optimal value is not searched. If the solver returns that the problem is infeasible the intermediate EFM is immediately removed from the mode matrix. The same procedure is executed in the postprocessing step, where all calculated EFMs are scanned for their feasibility and removed if necessary.

Additionally, those intermediate EFMs that turned out to be thermodynamically infeasible are further analysed by tEFMA. To this end the functionality of *cplex* is used to calculate conflicts for unsolvable problems. These conflicts are then converted by tEFMA to infeasible reaction patterns.

3.3 Functionality and performance of tEFMA

The first step was to ensure that tEFMA still calculates correct elementary flux modes (see also Chapter 5 and 6). For this purpose an *E. coli* core model with 155 irreversible reactions and 53 intracellular metabolites was used to find (thermodynamically consistent) EFMs with different metabolite concentration sets. Required concentration values were taken from the literature [30] and Gibbs energy values were taken from the eQuilibrator homepage [43]. Thermodynamically consistent EFMs of the models were calculated on the one hand with the new approach and on the other hand by the original *efmtool* followed by a NET analysis step. This comparison showed that the set of EFMs found by the new approach was a subset of the total set of EFMs, which confirmed that no additional (wrong) EFMs were calculated by tEFMA. Moreover, the set calculated by tEFMA was the same set as calculated by the consecutive approach, which confirmed that all thermodynamically inconsistent EFMs were detected by the new approach. In the next step the computational costs of the new tool were measured. Again, tEFMA was compared to the original *efmtool*. As shown in Table 3.1 tEFMA increases the performance in respect to the runtime as well as in respect to the needed memory.

3.4 Stability of tEFMA

To study the influence of perturbations in metabolite concentrations and Gibbs energy of formation values repeated runs of tEFMA with different concentration and Gibbs energy values were performed (see also Chapter 6). For this purpose the former described

E. coli core model was used as a basis to create three condition specific models. The new models were adapted to grow only on the given carbon sources glucose, glycerol or acetate. In the first step the upper and lower concentrations of the metabolites were randomly perturbed by $\pm 5\%$ and thermodynamic consistent EFMs were calculated 100 times. The procedure was repeated with additional perturbed values of $\pm 10\%$, $\pm 15\%$ and $\pm 20\%$. Figure 3.3 shows that tEFMA is very stable against perturbation of metabolite concentrations. The calculated EFMs were compared to reference EFMs enumerated by tEFMA with not perturbed values. For *E. coli* grown on glycerol and acetate only outliers were found that differs in the number of EFMs compared to the reference. For glucose, the majority of the runs returned the same EFMs compared to the reference run, although some runs using higher perturbed values found additional EFMs to be thermodynamically feasible. In a second step Gibbs free energy of formation values were randomly perturbed by ± 0.3 kJ, ± 1 kJ, ± 3 kJ and ± 9 kJ. Again tEFMA was performed 100 times for each perturbation value and each model. Figure 3.4 shows that up to ± 1 kJ tEFMA returns stable results, whereas for high perturbations of ± 9 kJ many different thermodynamic feasible EFMs were found compared to the reference EFMs.

3.5 Biological correctness of tEFMA

The former described *E. coli* models grown on glucose, acetate and glycerol were used to show that tEFMA returns results that are consistent with known phenotypical properties of the cell. Therefore, infeasible reaction patterns, that were returned by tEFMA, were

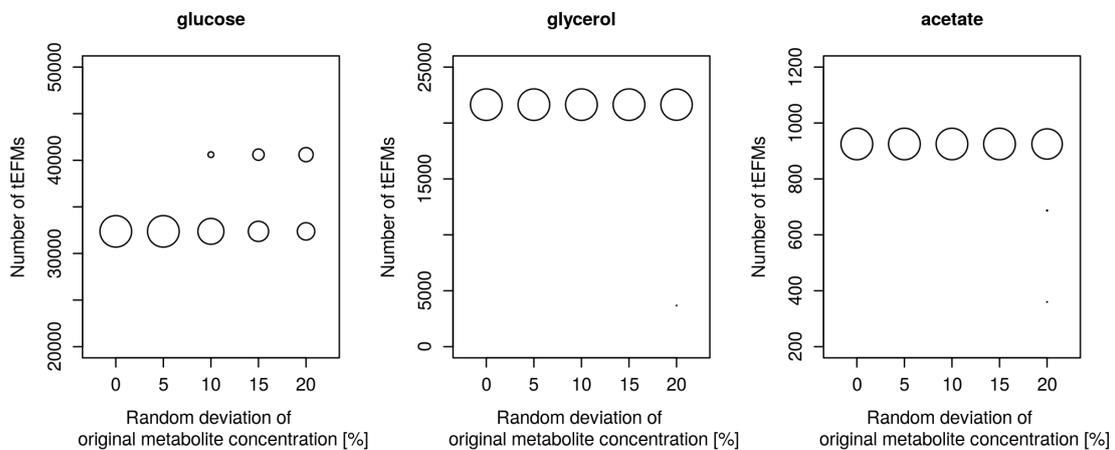


Figure 3.3: Number of calculated EFMs as a function of perturbed metabolite concentrations. Upper and lower metabolite concentration limits for glucose, glycerol and acetate were randomly perturbed by $\pm 0\%$, $\pm 5\%$, $\pm 10\%$, $\pm 15\%$ and $\pm 20\%$ and EFMs were calculated. The size of the circles show the number of runs containing the same number of EFMs. In total 100 runs were performed for each perturbation value. For glycerol and acetate only outliers can be found that differ from not perturbed runs. Even for glucose the majority of cases found the same EFMs in the different enumeration runs.

3 Results

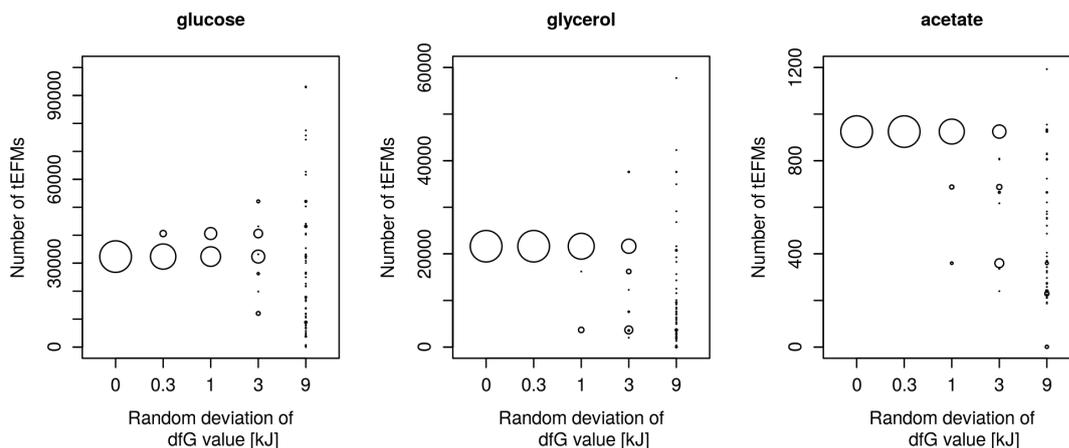


Figure 3.4: Number of calculated EFMs as a function of perturbed Gibbs energies of formation. $\Delta_f G$ values were randomly perturbed by $\pm 0, \pm 0.3, \pm 1, \pm 3$ and ± 9 kJ and EFMs were calculated. The size of the circles show the number of runs containing the same number of EFMs. In total 100 runs were performed for each perturbation value. Up to a perturbation of 1 kJ for all $\Delta_f G$ values the number of calculated EFMs keeps stable, whereas a perturbation of 9 kJ leads to different results in all cases.

further analysed and compared with the literature (see also Chapter 6).

In the first step infeasible reaction patterns, which are related to growth under aerobic conditions, were analysed. In contrast to an ordinary EFMA where all topologically feasible pathways are enumerated tEFMA correctly reported that malate dehydrogenase, which is part of the tricarboxylic acid cycle and oxidizes malate to generate oxaloacetate as well as acetaldehyde dehydrogenase, which reduces acetyl coenzyme A to acetaldehyde were thermodynamically infeasible when grown under aerobic conditions. The result reflects the well known inactivity of both enzymes under the described condition [44, 45].

In the next step infeasible reaction patterns related to different carbon sources were analysed. Here, tEFMA correctly identified that gluconeogenesis is thermodynamically not feasible when *E. coli* is grown on glucose, while such an infeasible reaction pattern was not reported by tEFMA when grown on acetate. Moreover, tEFMA distinguished correctly between glutamate dehydrogenase (GDH) and the glutamine oxoglutarate aminotransferase pathway (GOGAT) for the production of glutamate. It was previously published that *E. coli* uses GDH to form glutamate out of α -ketoglutarate. However, under energy rich conditions the alternative GOGAT pathway, which uses ATP to produce glutamine out of glutamate and forms in a second step two glutamate out of the previously produced glutamine, is active [32, 31]. tEFMA reported the infeasibility of the GDH reaction when *E. coli* is grown under the energy rich carbon source glucose while the infeasibility did not show up when analysing growth on glycerol or acetate confirming the published results.

3.6 Comparison of methods

Both methods, detecting internal loops by a matrix operation and determining infeasible EFMs with LPs by the consideration of metabolite concentrations, pursued the goal of removing thermodynamically infeasible EFMs during the EFM enumeration. However, detecting and removing internal loops by a simple matrix vector multiplication led to artifacts, which could not be detected anymore by the algorithm itself. Nevertheless, the basic idea, which states that all offsprings of an infeasible EFM are also infeasible, still held. Yet, the detection of the artifacts needed additional comparisons, which were computationally expensive. On the other hand, artifacts are not a problem when using NET analysis. If such artifacts appear during the enumeration of EFMs, they would be detected by the NET analysis itself. Using NET analysis to detect thermodynamically infeasible EFMs has another advantage as this method is not limited to detect only loops but can determine the thermodynamic state of every topologically possible EFM.

Conclusion and Outlook

In this work ways were presented to enumerate and analyse thermodynamically feasible EFMs. In a first step internal loops were analysed. It was shown that it is possible to detect and remove internal loops during the iteration phase of the double description method. However, special care must be taken as simply dropping internal loops from the analysis causes inconsistencies. Additional checks were necessary to avoid them, which was found to be computationally expensive. It was also carried out that a consecutive approach where all EFMs are enumerated in a first step followed by a second step that finds and removes internal loops is computationally less expensive than the described single step approach.

Additionally, tEFMA, a freely available open source tool that uses metabolite concentration data to calculate thermodynamically feasible elementary flux modes in metabolic networks, was presented. A previously published method [29] calculated thermodynamically feasible EFMs in a consecutive approach, where in the first step all EFMs of the network are enumerated and in a second step the thermodynamical state of these EFMs are determined. In the present work the thermodynamic feasibility was already determined during the enumeration of the EFMs, which leads to several advantages. The new method results in significantly decreased hardware requirements and time savings. Moreover, tEFMA reports those reaction patterns that cause the infeasibility of pathways. The analysis of these few patterns is much easier than an analysis of thousands or millions of infeasible EFMs.

It was also pointed out that in contrast to the internal loop detection method, tEFMA prevents the calculation of artifacts without the need of additional superset checks. Additionally, tEFMA can be applied to define the thermodynamic state of any topological possible EFM.

To show the functionality of tEFMA, thermodynamically consistent EFMs of an *E. coli* metabolic network were calculated. The final set of EFMs returned by tEFMA was compared to the set of EFMs calculated by a consecutive procedure, which fully enumerates all EFMs of the network by an ordinary EFMA and checks the thermodynamical state of the calculated EFMs in a second step. It was shown that both sets were identical. Additionally, it was shown that tEFMA requires less memory than an ordinary EFMA. As tEFMA relies on additional input data their influence on the stability of the results was analysed. Therefore, the metabolite concentrations as well as Gibbs free energy of formation values were perturbed and EFMs were enumerated. It

was shown that tEFMA returns stable results, as it was insensitive against fluctuation in the metabolite concentration data, but sensible to $\Delta_f G$ data. However, the desired accuracy is achievable with current computational methods [46].

It was also checked if the results returned by the new tool are biologically correct. For this purpose, the reported infeasible reaction patterns were analysed. It was shown that all reported infeasible reaction patterns were identified to be known infeasible pathways. Moreover, it was shown that feasible pathways were not wrongly detected to be infeasible.

EFMA provides an unbiased view on the network. However, the enumeration of all EFMs is computationally challenging. Furthermore, not all topologically feasible EFMs are biologically relevant. The present work presented a large step forward to the goal of calculating the full set of biologically meaningful EFMs in large genome-scale networks. Nevertheless, additional effort is still necessary to reach this goal. While it would not be a high effort to integrate additional constraints for transport reactions into different compartments [47] to the code of tEFMA, it will need a lot of progress in the process of measuring metabolites in eukaryotic cells for their analysis. For an unbiased thermodynamic analysis of eukaryotic cells compartment specific metabolite concentrations would be necessary.

Disregarding the difficulties of the analysis of eukaryotic cells, additional steps are still necessary for the analysis of prokaryotic cells. It was shown in the present work that a combination of thermodynamically feasible reactions can lead to infeasible pathways. Now, we can go a step further. A combination of thermodynamically feasible EFMs need not necessarily lead to a thermodynamically feasible flux distribution, e.g. two thermodynamically consistent EFMs that are supported by a flux of the same reaction, but in different directions cannot form together a thermodynamically feasible flux distribution. Additional work is necessary to separate the feasible flux distributions from the infeasible ones.

In the last years a lot of effort was made to improve the calculation of EFMs. Additional work is also necessary to combine tEFMA with the described methods and ideas. For instance, using gene regulatory information [26] together with the new tool could gain the advantages of both ideas to further decrease the set of calculated and biologically meaningful EFMs.

tEFMA: computing thermodynamically feasible elementary flux modes in metabolic networks

This chapter was published by Matthias P. Gerstl, Christian Jungreuthmayer & Jürgen Zanghellini in *Bioinformatics*, February 2015. doi: 10.1093/bioinformatics. †

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 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 to 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: <https://github.com/mpgerstl/tEFMA>

Introduction

An *elementary flux mode (EFM)* is a steady-state pathway, that consists of an indivisible set of reactions [8]. EFMs permit a system wide analysis of metabolism and can be used to identify engineering targets [7]. Different approaches were published to calculate EFMs in large metabolic models. For instance, Hunt et al. [24] split a metabolic network into subnetworks and calculated the EFMs for each subnetwork in parallel. Others only enumerated the set of the shortest EFMs [18] or randomly sampled the full set of EFMs [20]. A complete EFM analysis for large genome-scale models is still out of reach as

† Author contributions: M.P.G., C.J. and J.Z. conceived and designed the study; M.P.G. and C.J. developed the software; M.P.G. run the experiments and performed the analysis; M.P.G., C.J. and J.Z. wrote the manuscript.

the number of EFMs increases dramatically with the size of the metabolic network [17] 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 [25]. Therefore tools are necessary to find those subsets of (biologically) relevant EFMs. Jungreuthmayer et al. [26] enumerated only those EFMs that were consistent with a given transcriptional regulatory network. Here, we use a network-embedded, thermodynamic (NET) analysis [27] 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 [28]. 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 [29]. 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.

Implementation

We extended the open source program "efmtool" [22] and implemented the usage of metabolomics data. efmtool uses the double description method [36] 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:

$$\min 0 \quad (5.1a)$$

$$\text{s.t. } \Delta_r G_i \leq 0, \quad i \in \text{supp}(\mathbf{e}_k) \quad (5.1b)$$

$$\Delta_r G_i = \sum_{j=1}^m S_{ji} \Delta_f G'_j \quad (5.1c)$$

$$\Delta_f G'_j = \Delta_f G_j^{\prime 0} + RT \ln(c_j/c_0), \quad c_0 = 1 \text{ M} \quad (5.1d)$$

$$c_j^{\min} \leq c_j \leq c_j^{\max}. \quad (5.1e)$$

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, \mathbf{e}_k [Eqn. (5.1b)]. $\Delta_r G$ is calculated using the transformed Gibbs energies, $\Delta_f G'_j$, of the participating metabolites in respect to their stoichiometric coefficient, S_{ji} [Eqn. (5.1c)]. $\Delta_f G'_j$ values dependent on the standard transformed Gibbs energy of formation, $\Delta_f G_j^{\prime 0}$, the molar gas constant, R , the temperature, T , and the concentration of the metabolite, c_j , [Eqn. (5.1d)], which are allowed to vary within a given concentration range [Eqn. (5.1e)]. As any solution

	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%)

Table 5.1: Comparison between tEFMA and an ordinary EFM analysis (EFMA). Numbers in brackets list the percentage compared to EFMA.

showing the thermodynamic feasibility of an EFM is fine, we do not need to search for an optimum [Eqn. (5.1a)]. 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`.

Applications

We tested our tool on an *E. coli* core model [2], consisting of 53 intracellular metabolites and 155 irreversible reactions. We used published metabolite concentrations for *E. coli* grown on glucose, glycerol and acetate [30]. $\Delta_f G^0$ values were retrieved from eQuilibrator [43]. Measured concentration ranges were available for 28 metabolites. For 23 metabolites no measured values were available and default ranges (10^{-7} M to 1M) were applied. For two metabolites no $\Delta_f G$ 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 5.1 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 [29], 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 Table 5.2). *tEFMA* did not falsely identify any feasible pathway to be infeasible.

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.

Supplementary Material

carbon source	infeasible pattern	resulting no of infeasible EFMs	biological meaning
acetate	-GAPD, PGK, PTAr	32,911,177	gluconeogenesis while acetate is produced
	-GAPD, -TPI, PTAr	12,733,908	gluconeogenesis while acetate is produced
	-GAPD, PGK, -SUCOAS	32,977,644	gluconeogenesis via succinyl coenzyme A synthetase
	-GAPD, -TPI, -SUCOAS	12,694,485	gluconeogenesis via succinyl coenzyme A synthetase
glucose	-FBA, -GAPD, -TPI, PGK	25,868,962	gluconeogenesis
	-GAPD, PGK, PTAr	32,911,177	gluconeogenesis
	-GAPD, -SUCOAS	32,977,644	gluconeogenesis via succinyl coenzyme A synthetase
	-ENO, -GLUDy, PGM	37,377,361	inactivity of glutamate dehydrogenase
	-GLUDy, ACONTb	43,727,496	inactivity of glutamate dehydrogenase
	-GLUDy, FUM	57,318,042	inactivity of glutamate dehydrogenase
	-PTAr, -SUCOAS	22,097,625	uptake of acetate while active succinyl CoA synthetase
glycerol	-ENO, -GLUDy, PGM	37,377,361	gluconeogenesis while glutamate dehydrogenase is active
	-FBA, -GAPD, PTAr	12,733,908	gluconeogenesis while acetate is produced
	-GAPD, PGK, PTAr	32,911,177	gluconeogenesis while acetate is produced
	-FBA, -GAPD, -SUCOAS	12,694,485	gluconeogenesis via succinyl coenzyme A synthetase
	-GAPD, -SUCOAS, PGK	32,977,644	gluconeogenesis via succinyl coenzyme A synthetase
	-GAPD, -SUCOAS, -TPI	12,694,485	gluconeogenesis via succinyl coenzyme A synthetase

Table 5.2: Thermodynamically infeasible pattern in *E. coli* core model [2], by applying metabolite concentration data [30]. Each single pattern results in a high number of thermodynamically infeasible EFM. Reaction names starting with a minus indicate the reverse direction.

Metabolomics integrated elementary flux mode analysis in large metabolic networks

This chapter was published by Matthias P. Gerstl, David E. Ruckerbauer, Diethard Mattanovich, Christian Jungreuthmayer & Jürgen Zanghellini in *Sci Rep*, 5:8930, 2015. doi: 10.1038/srep08930. [†]

Elementary flux modes (EFMs) are non-decomposable steady-state pathways in metabolic networks. They characterize phenotypes, quantify robustness or identify engineering targets. An EFM analysis (EFMA) is currently restricted to medium-scale models, as the number of EFMs explodes with the network's size. However, many topologically feasible EFMs are biologically irrelevant. We present thermodynamic EFMA (tEFMA), which calculates only the small(er) subset of thermodynamically feasible EFMs.

We integrate network embedded thermodynamics into EFMA and show that we can use the metabolome to identify and remove thermodynamically infeasible EFMs during an EFMA without losing biologically relevant EFMs. Calculating only the thermodynamically feasible EFMs strongly reduces memory consumption and program runtime, allowing the analysis of larger networks. We apply tEFMA to study the central carbon metabolism of *E. coli* and find that up to 80% of its EFMs are thermodynamically infeasible. Moreover, we identify glutamate dehydrogenase as a bottleneck, when *E. coli* is grown on glucose and explain its inactivity as a consequence of network embedded thermodynamics. We implemented tEFMA as a Java package which is available for download at <https://github.com/mpgerstl/tEFMA>

Introduction

Constraint-based reconstruction and analysis methods have been proven to be valuable tools in gaining system wide understanding of cellular metabolism [48, 49, 50]. These

[†]Author contributions: M.P.G., C.J. and J.Z. conceived and designed the study; M.P.G. and C.J. developed the software; M.P.G. and J.Z. developed the models; M.P.G. run the experiments; M.P.G. and D.E.R. contributed analysis tools and performed the analysis; M.P.G., D.M., C.J. and J.Z. wrote the manuscript.

methods use mathematical reconstructions of metabolism together with (physiochemical, thermodynamical, environmental, *etc.*) constraints to derive their predictions. Based on a steady-state analysis of a stoichiometric matrix (i.e. an ordered collection of the stoichiometric coefficients of all contributing biochemical reactions) these methods allow for phenotypic predictions from genotype data [51]. Here we focus on a method called elementary flux mode (EFM) analysis (EFMA).

EFMA decomposes the stoichiometric matrix into non-decomposable, non-zero steady-state pathways, called EFMs [52]. EFMs are an important structural concept as any metabolic steady-state can be expressed as a non-negative, linear superposition of EFMs. Thus, the complete set of EFMs fully characterizes the available metabolic space. This comes at the price of a dramatically increased computational effort which goes beyond current capabilities for large, genome scale metabolic models [7]. A pessimistic upper bound for the number of EFMs in a network was derived [17], but the exact computational complexity is not yet known [16].

Regardless of the theoretical challenges, several software tools are available and allow the calculation of the full set of EFMs at least in small or medium scale (metabolic) models [7]. Very recently, a massively parallelized approach to completely enumerate EFMs in large-scale networks was presented [24]. For large genome-scale networks particular EFMs, but not all can be calculated. Various strategies ranging from calculating the shortest EFMs [18] to different sampling approaches [20, 19] have been proposed. Recently, Pey and Planes [37] identified a small subset of biologically interesting EFMs in a genome-scale model. Similarly, Kelk et al. [53] search for all EFMs, which span the optimal solution space as defined by a flux balance analysis. Despite all these advances a full enumeration of EFMs in large genome-scale models is as yet out of reach.

EFMA utilizes stoichiometric information only. Yet, many of the topologically feasible EFMs are infeasible *in vivo* as they are in opposition to other constraints that have not been accounted for, like known regulatory mechanisms [26, 54] or thermodynamic properties of biochemical reactions [55]. Incorporating thermodynamic constraints allows us to draw conclusions on the directionality and feasibility of reactions and whole pathways. A single biochemical reaction occurs spontaneously only if its change in Gibbs energy is negative. To derive thermodynamic constraints for the whole network, metabolite data are particularly useful as they determine the Gibbs energy surface.

Here we present a novel computational tool – thermodynamic EFMA (tEFMA) – which integrates the cellular metabolome into the EFMA. This allows us to verify the thermodynamic feasibility of EFMs already during the runtime of the EFMA and curbs the explosion of the number of EFMs without losing any biologically relevant EFMs. Computationally, our new approach successfully tackles the major bottleneck of double description based EFMA by strongly reducing computational costs, both in terms of runtime and resource consumption. Biologically, tEFMA allows the identification of infeasible pathways based on an unbiased analysis derived from first principles. More specifically, tEFMA correctly predicts the inactivity of the glutamate dehydrogenase (GDH) in *E. coli* under glucose saturated conditions.

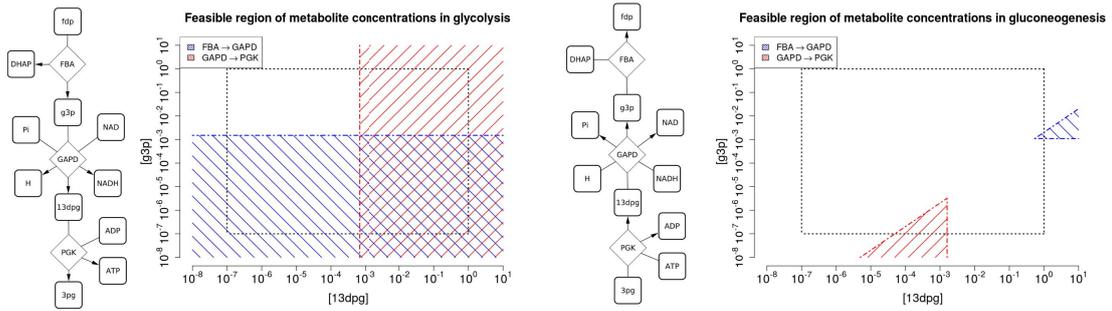


Figure 6.1: Thermodynamically feasible concentration regions for 1,3-bisphosphoglycerate (13dpg) and glyceraldehyde-3-phosphate (g3p) at glycolysis (left) or gluconeogenesis (right) for *E. coli* when growing on minimal media with glucose. Dashed lines indicate the concentration bounds of the metabolites and chain dotted lines the bound of negative Gibbs energy, i.e. the line where $\Delta_r G_{GAPD} = 0$. Blue areas show regions of negative Gibbs energy for the combination of FBA (fructose-bisphosphate aldolase) and GAPD (glyceraldehyde-3-phosphate dehydrogenase) and red areas for the combination of GAPD and PGK (phosphoglycerate kinase). At glycolysis all three reactions are simultaneously thermodynamically feasible indicated by the overlapping red and blue area. At gluconeogenesis such an overlap within the error bounds of the metabolites cannot be found. To find the feasible regions we analyzed the admissible concentrations of the shared metabolites 13dpg and g3p. The minimum and maximum concentration of g3p as function of 13dpg was calculated so that $\Delta_r G_i \leq 0$ held for all reactions.

Methods

Theory

The stoichiometry of a metabolic network with m (internal) metabolites and r reactions can be represented by an $m \times r$ matrix, \mathbf{S} . At steady-state all flux distributions, \mathbf{v} , obey $\mathbf{S}\mathbf{v} = \mathbf{0}$ and $\mathbf{v}_{\text{irrev}} \geq \mathbf{0}$, where $\mathbf{v}_{\text{irrev}}$ is a sub-vector of \mathbf{v} containing only irreversible reactions. We assume that the network contains only irreversible reactions, as any reversible reaction can be split into an irreversible forward reaction and an irreversible backward reaction. Of particular interest are so called EFMs, \mathbf{e}_i [52]. These are steady-state flux distributions of minimal support fulfilling all irreversibility constraints. Minimal support means that if any of the contributing, i.e. supporting reactions ($v_i > 0$) is omitted, the remaining reactions can no longer carry a steady-state flux. Geometrically, the EFMs (in a network of irreversible reactions) can be regarded as extreme rays, i.e. edges, in a convex polyhedral cone [56]. Several EFM-enumeration strategies are known [7]. Here we utilized the binary null-space algorithm [12], which we will briefly outline below.

The binary (null-space) approach represents EFMs as binary bit vectors of the supporting reactions. These bit patterns are generated iteratively. Starting from an initial solution matrix (typically the kernel of \mathbf{S}) each row of this matrix is processed and converted to binary form. For each row (i.e. reaction) intermediate EFMs (that are the columns of the matrix) are combined such that their fluxes are nonnegative and therefore convertible to a bit representation and added to the matrix. New intermediates are added to the quickly growing list of intermediate EFMs if they are not a superset of any other intermediate EFMs. The iteration stops if all reactions are processed and the intermediate EFMs are fully converted into binary format. The remaining intermediate EFMs are then, in fact the EFMs. A step by step example can be found in the supplementary material, section “Proof of safe removal of thermodynamically infeasible EFMs” on page 65.

An important feature of the binary approach is the inheritance of flux activity. When a reaction is converted to binary form and found to be active in an intermediate EFM, all progenies of this EFM will have an active flux in that reaction too [12]. This property is key to our approach. Based on metabolomics data we identify thermodynamically infeasible flux patterns and drop the associated modes from the list of intermediate EFMs, as all their possible offspring will be supersets of these infeasible flux patterns, and therefore will remain infeasible too. Thus, removing thermodynamically infeasible modes has no impact on any feasible (intermediate) EFM. Here, network embedded thermodynamic (NET) analysis [27] is used to identify thermodynamically infeasible EFMs. NET analysis is briefly reviewed below.

The second law of thermodynamics states that at constant pressure any biochemical reaction, i , proceeds spontaneously only in the direction of the negative Gibbs free energy of reaction $\Delta_r G_i$. As our network contains only irreversible reactions this translates into

$$\Delta_r G_i \leq 0 \quad v_i \geq 0, \quad \text{and} \quad \Delta_r G_i \geq 0 \quad v_i = 0. \quad (6.1)$$

$\Delta_r G_i$ can be estimated from the Gibbs free energy of formation, $\Delta_f G_j$, of the contribut-

ing reactants, j :

$$\Delta_r G_i = \sum_{j=1}^m S_{ji} \Delta_f G'_j, \quad (6.2a)$$

$$\Delta_f G'_j = \Delta_f G_j'^0 + RT \ln(c_j/c_0), \quad c_0 = 1 \text{ M}, \quad (6.2b)$$

where S_{ji} represents the stoichiometric coefficient of metabolite j in reaction i and $\Delta_f G'_j$ is used to denote the transformed Gibbs free energy of formation for metabolite j , corrected for its actual, non-standard metabolite concentration, c_j . R is the molar gas constant, and T the absolute temperature. $\Delta_f G_j'^0$ represents the transformed standard Gibbs free energy of formation, which we corrected for ionic strength and pH [28]. See the supplementary materials, section ‘‘Calculation of the transformed standard Gibbs free energy of formation’’ on page 70 for details and the supplementary materials, file 2 for actual $\Delta_f G_j'^0$ -values.

Eqs. (6.1) and (6.2) identify isolated, thermodynamically infeasible reactions based on (measured) metabolite concentrations. However, NET analysis does not only study a reaction in isolation, but rather considers a reaction’s feasibility in the context of pathways. NET analysis utilizes the thermodynamic interdependencies between reactions and verifies if a given network structure is consistent with a (measured) metabolome. To this end NET analysis is solved by the linear program (LP) given by Jol et al. [29]

$$\min 0 \quad (6.3a)$$

$$\text{s.t. } \Delta_r G_i \leq 0, \quad i \in \text{supp}(\mathbf{e}_k) \quad (6.3b)$$

$$\Delta_r G_i = \sum_{j=1}^m S_{ji} \Delta_f G'_j \quad (6.3c)$$

$$\Delta_f G'_j = \Delta_f G_j'^0 + RT \ln(c_j/c_0), \quad c_0 = 1 \text{ M} \quad (6.3d)$$

$$\ln(c_j^{\min}/c_0) \leq \ln(c_j/c_0) \leq \ln(c_j^{\max}/c_0). \quad (6.3e)$$

The program above is linear in $\ln(c_j/c_0)$. That is why the limits in Eq. (6.3e) were expressed in terms of logarithms. The LP checks whether all reactions contributing to an EFM, \mathbf{e}_k , are simultaneously feasible [Eq. (6.3b)] and consistent with a metabolome within the given error bounds c_j^{\min} and c_j^{\max} , respectively [Eq. (6.3e)]. The remaining equations [Eqs. (6.3c) and (6.3d)] account for mapping the metabolome to the Gibbs free energy surface. Note that in the original NET analysis [27] Eqs. (6.3a – 6.3e) are optimized for $\Delta_r G_i$, while we are only interested in the feasibility of Eqs. (6.3a – 6.3e). Therefore, any (non-optimal) solution suffices, which poses a computationally less challenging problem.

The basic feature of NET analysis is illustrated in Figure 6.1. In isolation, each reaction (FBA, GAPD, and PGK) is feasible in both directions. Also the reaction pairs (FBA, GAPD and GAPD, PGK) are feasible in both directions. However, if the reaction triple (FBA, GAPD, PGK) is considered, we find only the forward direction to be consistent with the metabolite concentrations.

In tEFMA every intermediate EFM is checked at the beginning of each iteration against a given metabolome according to Eqs. (6.3a – 6.3e) and immediately removed if infeasible. Figure 6.2 illustrates the basic work flow. For example, in iteration i we may find that 18, 41, and 12 intermediate EFMs have positive, zero, and negative flux

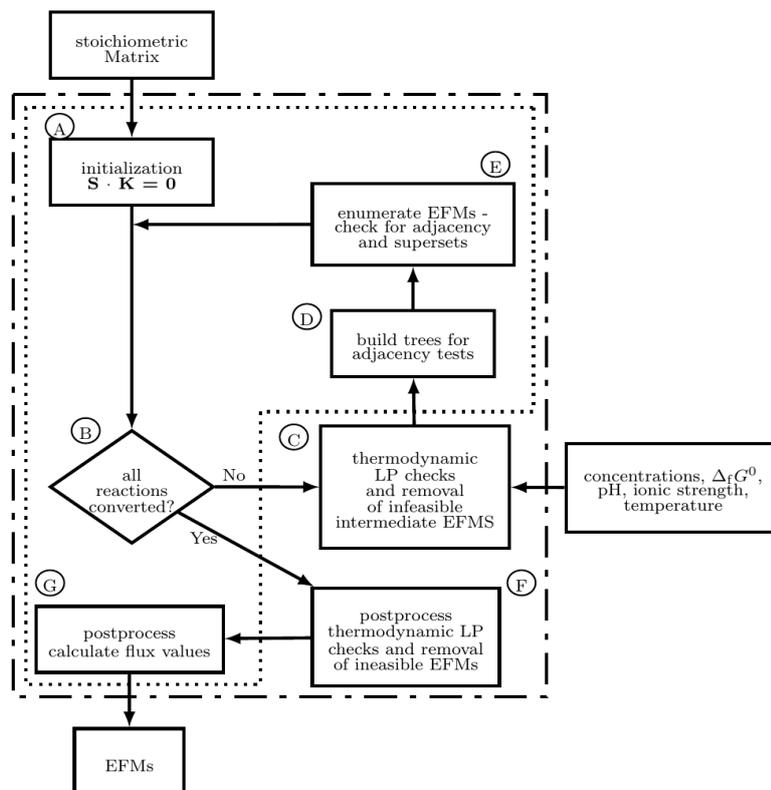


Figure 6.2: Basic work flow of tEFMA. Dashed lines mark the original *efmtool* and chain dotted lines the integration of NET analysis into tEFMA. A) In the initialization phase the stoichiometric matrix is compressed and the kernel matrix created. B) As long as a reaction is not converted from numeric to binary a new iteration is started. C) Intermediate EFMs with positive or negative values on next numeric position are checked for thermodynamic feasibility, based on given input values. Infeasible EFMs are removed here. D) Adjacency trees are built with EFM intermediates. E) New intermediate EFMs are created by combining adjacent EFMs from positive and negative trees. They are added to the list of intermediate EFMs unless they are supersets of other intermediates. F) In the post-processing phase calculated EFMs are finally checked to be thermodynamic feasible. In the last step G) *efmtool* removes futile-2-cycles, decompresses EFMs and calculates the flux values resulting in the enumerated set of EFMs. For an example see the supplementary material, section “Proof of safe removal of thermodynamically infeasible EFMs” on page 65.

values in reaction i . This gives rise to $18 \times 12 = 216$ potentially new intermediate EFMs of which only 22 EFMs are actually added to the list of new intermediates as only these pass the (tree-based) adjacency and superset testing of the EFM enumeration procedure. In tEFMA we check the feasibility of the original $18 + 12$ intermediate EFMs and remove infeasible EFMs there. Suppose that 8 out of the 18 positive intermediate EFMs are infeasible and can be removed instantly. Rather than 216 potentially new intermediates we now just get $10 \times 12 = 120$ potentially new intermediates of which only 17 EFMs are actually added to the list of new intermediate EFMs as these pass the (tree-based) adjacency and superset testing. (The numbers for this example were taken from line $i = 10$ in Table 6.5.) Note that the combination of two intermediate EFMs may create new infeasibilities. If these new intermediates have non-zero flux values in any of the so far unprocessed reactions, they will be checked in a later step of the iteration procedure. In case a new intermediate EFM has only zero flux in the remaining reactions, it will be detected at the end of the iteration phase, where we run a final feasibility check on all remaining EFMs (see Figure 6.2).

The efficiency of this approach is illustrated in Table 6.5, Table 6.6, and Table 6.7, where we show that the total number of LPs is always smaller than the total number of newly generated intermediate EFMs in the non-thermodynamic EFMA. We found heuristically that it is more efficient to check the feasibility of intermediate EFMs first and then do the tree-based adjacency and superset testing, rather than the other way round (data not shown).

In the remainder we assume that $T = 310.15$ K (37°C).

Implementation

We implemented tEFMA as an extension of the open source software *efmtool*, which was originally developed by Terzer and Stelling [22]. We added three new Java packages with 21 new Java classes to *efmtool*. The new classes are responsible for reading the additional information, call CPLEX (a powerful commercial solver by IBM, for which academic licenses are available on request) and handle infeasible EFMs. To invoke the new functionality we modified two already existing Java classes and the XML file that handles command line arguments (see the README-file in the accompanying software package [57] for details). The extended version was compiled by JDK 1.7.11.

Metabolic reconstructions

We used the *E. coli* core model published by Orth et al. [2]. We refer to it as model M1. M1 contained 73 metabolites and 155 irreversible reactions (after splitting each of the 59 reversible reactions into two irreversible forward and backward reactions). The core reconstruction, M1, does not model glycerol uptake, so we added the glycerol uptake pathway from the *E. coli* model *iJR904* [58]. This augmented model is referred as model M2. Specifically, we included glycerol kinase (R_GLYK), glycerol-3-phosphate dehydrogenase (R_G3PD2, R_G3PD5), glycerol transport (R_GLYCt) and glycerol exchange (R_EX_glyc.e). The resulting stoichiometric matrix consisted of 76 metabolites and 163 reactions (62 of them were initially reversible). The rank of this matrix was 71. We used M2 to derive three condition specific sub-models, M2-glc, M2-glyc, and M2-ac, to model growth on minimal medium (containing ammonia, oxygen, phosphate, protons, and water) with glucose, glycerol or acetate as the sole carbon source, respec-

tively. In these models all uptake reactions for nutrients which were not included in the growth media, were removed. If a nutrient transport was reversible we only disabled the nutrient’s uptake but not its secretion into the extracellular environment.

Except for glutamate and glutamine, neither M1 nor M2 model the biosynthesis of the other amino acids. Thus we augmented M2 by adding the amino acid pathways extracted from the *E. coli* model *iJE660a* [59]. This model is referred to as M3. Its stoichiometric matrix consisted of 178 metabolites and 303 irreversible reactions (94 of them were initially reversible). The rank of this matrix was 171. SBML files for M2 and M3 are available in the supplementary materials. M1 can easily be obtained by removing R_GLYK, R_G3PD2, R_G3PD5, R_GLYCt, and R_EX_glyc.e from reconstruction M2.

We summarized the main topological properties of all models in the supplementary material, Table 6.4.

Functionality test

We tested tEFMA for specificity, sensitivity and performance. For the thermodynamic feasibility checks we used previously published metabolite concentration data for *E. coli* when grown on glucose, glycerol or acetate [30]. In comparison to published concentration ranges [60], we used very conservative minimum ($c_j^{\min} = 10^{-7}$ M) and maximum ($c_j^{\max} = 1$ M) default values for unmeasured metabolites to avoid false identification of infeasible EFMs. The necessary $\Delta_f G^0$ data were taken from the online version of eEquilibrator [43]. Independently, we performed a conventional EFMA on the same model using *efmtool* and separately tested each EFM for thermodynamic feasibility using NET analysis.

Stability analysis

We tested the stability of tEFMA against perturbations in the metabolome and the thermodynamic data. We randomly changed all concentrations up to $\pm 5\%$, $\pm 10\%$, $\pm 15\%$, and $\pm 20\%$. This change was on top of the error bounds given by Bennett et al. [30]. That is, all lower and upper bounds (c_j^{\min} and c_j^{\max} , respectively) were independently changed within the intervals given above. Additionally we required that $c_j^{\max} - c_j^{\min} \geq 0.05c_j$, where we used c_j to denote the mean concentration of metabolite j . The perturbed concentrations were then used in the tEFMA. The whole procedure was repeated 100 times. Similarly, all $\Delta_f G'_j$ -values were perturbed by randomly and independently changing each value by up to ± 0.3 kJ/mol, ± 1 kJ/mol, ± 3 kJ/mol, and ± 9 kJ/mol. Again, this procedure was repeated 100 times.

Results

We calculated thermodynamically feasible EFMs in medium scale metabolic models of *E. coli* (models M1 to M3) based on experimental metabolite concentrations measured by Bennett et al. [30]. In the smallest reconstruction (model M1), the experimental data accounted for 56% of the model’s internal metabolites. All unmeasured metabolites were assumed to be within conservative concentration bounds (see method section for details). 15 out of 155 irreversible reactions in M1 were thermodynamically fully characterized by measurements. 56 reactions were at least partially characterized by experimental

data. The overlap between the model M1 and the experimental data is illustrated in the supplementary material, Figure 6.5.

Computational, tEFMA identifies thermodynamically feasible EFMs accurately and economically

tEFMA removes all infeasible EFMs We compared tEFMA against an ordinary EFMA followed by NET analysis. The sets of thermodynamically feasible EFMs were identical in both analyses. Figure 6.3 illustrates a comparison between an EFMA and a tEFMA.

For growth on glucose about one third of all EFMs were thermodynamically feasible. The reduction in the number of feasible EFMs is highly condition specific as on glycerol and acetate the numbers of feasible EFMs are roughly cut in half. These comparisons were based on the full metabolic model, M1, without any other adaptations, i.e. also unused uptake reactions were subject to the analysis. If all unused uptake reactions were removed from the models, then the changes in the number of feasible EFMs was even more pronounced.

On glucose minimal medium only 19% of the EFMs were thermodynamically feasible, while on acetate minimal media 76% were feasible (using model M2). However, in the case of glucose 19% corresponded to more than 30,000 feasible EFMs, while on acetate roughly 900 EFMs remained feasible. Thus, growth on glucose still opened dramatically more metabolic possibilities (counted by the number of feasible EFMs) than growth on any other carbon source.

tEFMA is stable against fluctuations in the metabolome and the thermodynamic data We verified the stability of the feasible EFMs by randomly perturbing the metabolite concentrations (see methods section for details). For all tested perturbations (0, ..., $\pm 20\%$) the median number of feasible EFMs remained constant (see the supplementary material, Figure 6.7). Moreover, all EFMs identified to be feasible without perturbations were re-identified to be feasible in the perturbed runs as well (except for statistical outliers in the case of glycerol and acetate growth at $\pm 20\%$). Note, that the perturbations were added on top of the experimental error (see the supplementary material, Figure 6.8 for details.)

We repeated the analysis to also check tEFMA against variations in $\Delta_f G'$ -values (see methods section for details). Up to a perturbation magnitude of 1kJ/mol our results stayed essentially constant (see the supplementary material, Figure 6.6 and supplementary file 3), i.e. in all these cases we found the same set of EFMs to be thermodynamically feasible. For stronger perturbations large deviations were found.

tEFMA strongly reduces runtime and memory usage Our novel software extended *efmtool* originally developed by Terzer and Stelling [22], which uses a variant of the double description method (DDM) [36] to enumerate EFMs. The method requires to repeatedly solve intermediate EFM enumeration problems, which gives rise to a huge number of intermediate EFMs. Although most of these intermediate EFMs will eventually be rejected, they have to be readily available throughout the calculation. This places high demands on a computer's storage capacity [specifically on the size of the random access memory (RAM)]. Figure 6.3 illustrates the decrease in the number of feasible EFMs for an ordinary EFMA and a tEFMA. The decrease is even stronger in the total number of adjacency candidates, which relaxed the hardware requirements

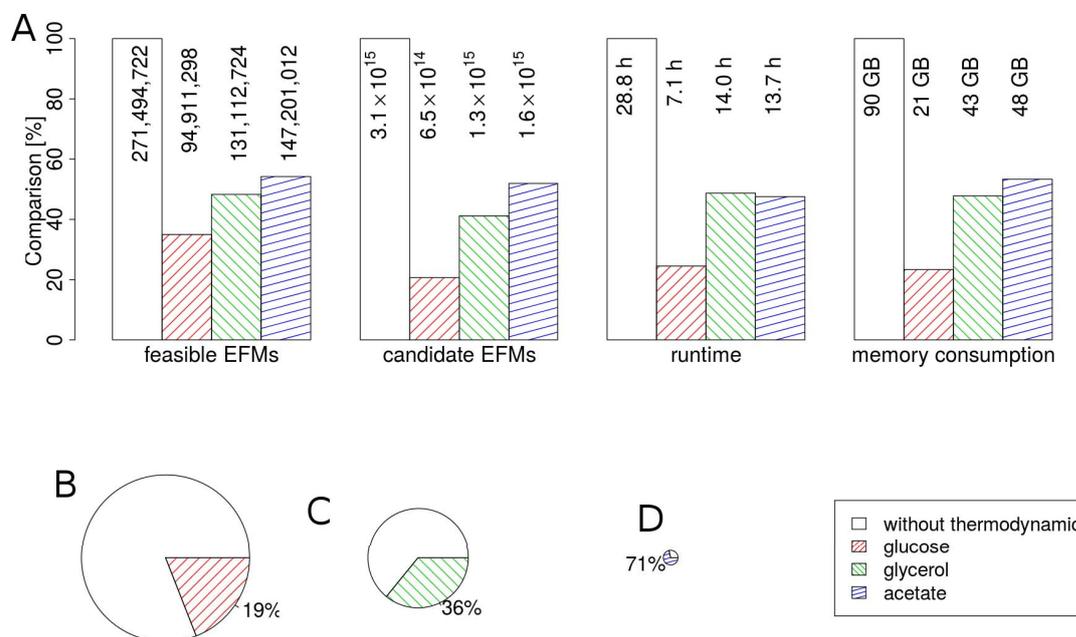


Figure 6.3: Performance analysis of tEFMA with and without thermodynamic feasibility checks using three different metabolomes. Results of an ordinary EFMA (none) were compared against tEFMA using metabolome data [30] for growth on minimal medium (MM, contained ammonia, oxygen, phosphate, protons, and water) and glucose (glc + MM), glycerol (glyc + MM), and acetate (ac + MM). The analysis was performed (A) on the *E. coli* model M1 and (B–D) on condition specific model M2, where all inactive uptake reactions were removed. Using glucose (B) 32,374 EFMs out of 169,916 are feasible, whereas for glycerol (C) 21,642 out of 60,495 and for acetate (D) 925 out of 1,299 EFMs are thermodynamically feasible. In panel A numbers on the top indicate the absolute values. In panel B to C the circle areas are scaled as to represent the total number of topological feasible EFMs in the models.

for tEFMA. In fact, RAM consumption in tEFMA is at least cut in half but can in optimal cases shrink to only 25% compared to the RAM consumption of an ordinary EFMA. Similarly, tEFMA also reduces the runtime of the algorithm and needs only 25% in the best case and 49% in the worst case as compared to an ordinary EFMA (see Figure 6.3).

Biological, tEFMA identifies known infeasible pathways

For the following biological interpretation of the calculated EFMs and infeasible pattern we used the model M2.

It is textbook knowledge that under aerobic conditions malate dehydrogenase (Mdh) oxidizes malate to generate oxaloacetate as part of the tricarboxylic acid cycle. tEFMA correctly identified the reverse reaction to be infeasible. Similarly acetaldehyde dehydrogenase (AdhE), which catalyzes the reduction of acetyl coenzyme A to acetaldehyde, was identified to be infeasible under the three tested growth conditions in accordance with well established knowledge. Both conclusions could have been made without the help of tEFMA, as evaluating Eqs. (6.2a and 6.2b) for Mdh and AdhE unambiguously identified the reactions' directions without considering the network structure of metabolism.

tEFMA correctly distinguished between glycolysis and gluconeogenesis tEFMA correctly classified gluconeogenesis to be infeasible in *E. coli* grown on glucose. The latter could not have been concluded without a NET analysis or tEFMA. For example, within the error bounds of the measured metabolite concentrations the reactions phosphoglycerate kinase (Pgc), glyceraldehyde 3-phosphate dehydrogenase (Gapd), and fructose biphosphate aldolase (Fba) were found to be reversible if analyzed individually. Only together tEFMA identified them to be infeasible in direction of gluconeogenesis (see Figure 6.1). The lower part of gluconeogenesis (from pyruvate to glyceraldehyde-3-phosphate) was also predicted to be infeasible for growth on glycerol while feasible for growth on acetate. Interestingly, gluconeogenesis via succinyl coenzyme A synthetase (SucCD) was inaccessible in the latter (see below and Table 6.3 for further details). Note that Pgc, Gapd, and Fba build a linear, consecutive chain of reactions. In general, however, tEFMA is able to identify thermodynamic inconsistencies between non-consecutive reactions, too (see Table 6.1 to Table 6.3).

tEFMA correctly predicted the inactivity of glutamate dehydrogenase (GDH) during growth on glucose Two pathways for glutamate synthesis are known in *E. coli*. GDH catalyzes the reductive amination of α -ketoglutarate to form glutamate. Alternatively the glutamine oxoglutarate aminotransferase (GOGAT) pathway produces glutamate in two steps: (i) glutamate is used to produce glutamine by the energy dependent glutamine synthase and (ii) the amide group is then transferred reductively from glutamine to α -ketoglutarate to form glutamate. Both pathways were identified in an ordinary EFMA and produce 1 mole of glutamate net. For growth on glucose, however, tEFMA identified inconsistencies between GDH and the lower part of the glycolysis as well as between reactions GDH and aconitate hydratase (ACONTb). We found that on glucose no thermodynamically feasible EFM was supported by an active GDH (see Figure 6.4). This is consistent with experimental evidence that under glucose saturated conditions the alternative GOGAT pathway is active, and not GDH [32]. On the other hand, we identified thermodynamically feasible, GDH supported EFMs when *E. coli* was grown

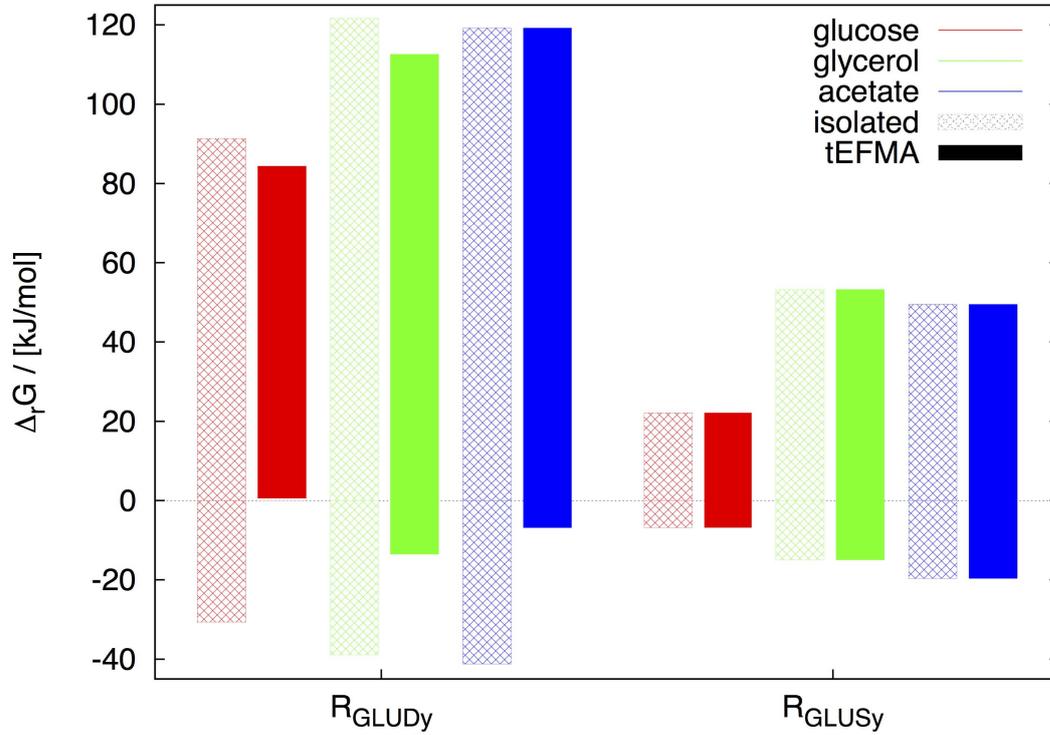


Figure 6.4: Minimum lower and maximum upper bounds of $\Delta_r G$ for the reactions glutamate dehydrogenase (R_{GLUDy}) and glutamate synthase (R_{GLUSy}) for various conditions in model M2. For each single EFM, which was enumerated by *efmtool* (without the *tEFMA* extension), the minimum and maximum $\Delta_r G$ of both reactions were calculated in isolation (open pattern) and within a NET analysis (solid pattern). Note, that only negative $\Delta_r G$ ranges are thermodynamically feasible. Therefore R_{GLUDy} is never feasible, when grown on glucose and analysed by a NET method (red solid pattern).

on acetate or glycerol. Again, this is consistent with experiments, as GDH, but not GOGAT, is energy neutral and therefore favored under energy-stressed conditions [31]. Our analysis revealed that under glucose saturated conditions both reactions are potential thermodynamic bottlenecks as they operate close to $\Delta_r G = 0$ [60]. However, GDH was found to be more sensitive than glutamate synthase (see Figure 6.4). Note that in this analysis it is essential to consider the network structure of metabolism. Within tEFMs GDH is inactive, but by analyzing GDH and glutamate synthase in isolation the inactivity of GDH cannot be determined. In fact a naive interpretation might lead to the erroneous assumption that glutamate synthase rather than GDH is a thermodynamic bottleneck (see Figure 6.4 for an illustration).

tEFMA did not predict false positives

For a given metabolome tEFMA found combinations of reactions that could not operate simultaneously (see Table 6.1 – Table 6.3). We were able to understand all of these combinations of reactions in terms of the (infeasible) pathways described above. In the three test cases tEFMA did not erroneously identify a thermodynamically feasible pathway to be infeasible.

tEFMA is scaleable to larger systems

We repeated a tEFMA using the same experimental data as above together with a more detailed *E. coli* reconstruction, M3. This model did not only contain the core carbon metabolism but was augmented with biosynthesis routes for amino acid production. Compared to its parent model, M3 contained roughly twice as many reactions and also twice as many internal metabolites. The overlap between this model and the experimental data is shown in Figure 6.5. In this model tEFMA identified 1,197,839 thermodynamically feasible EFMs, 37 times more feasible EFMs than in the smaller parent model M1.

In addition, tEFMA identified 15 infeasible flux patterns, i.e. reactions which together must not carry flux (see the supplementary material, Table 6.9 for a listing). The six infeasible flux patterns detected earlier, in the smaller parent model M2, were also found now in the larger reconstruction. The remaining infeasible patterns could not have been detected in the smaller parent model M2, as they all contained reactions which were unique to the larger M3-model.

Discussion

We developed and applied tEFMA to study the metabolic capabilities of *E. coli*. tEFMA integrates experimentally determined metabolomes into an ordinary EFMA to avoid the calculation of thermodynamically infeasible EFMs. Recently this strategy was successfully applied to analyze the metabolic capabilities in yeast grown on glucose [29]. The authors first constrained the metabolic network as tightly as possible and then performed an ordinary EFMA followed by a NET analysis on the EFMs. In contrast to this sequential approach, tEFMA efficiently performs both analyses simultaneously, yielding in huge computational savings. Harvesting these savings is the major achievement of this work.

We exploited the fact that any combination of infeasible EFMs with other (in)feasible EFMs is again infeasible [29] and can be removed from the analysis without impacting biologically relevant EFMs. By doing so, we tackled the major bottleneck in the DDM [36], i.e., the exploding number of (intermediate) EFMs during the calculation [26].

Currently DDM is the most common approach for calculating EFMs [24, 21]. It solves the enumeration problem iteratively by adding one constraint at a time and (re-)enumerating the problem. This is done by a pairwise combination of positive and negative intermediate EFMs. Of the huge number of potential candidates only those intermediate EFMs are used to generate offspring, if they are adjacent. Each newly created intermediate EFM undergoes a superset test which prevents further processing of a new intermediate EFM if it is a superset of any already existing intermediate EFM. Performing the adjacent and superset test, as well, as creating and maintaining this large list of intermediate EFMs is computationally expensive. While Terzer and Stelling [22] efficiently perform adjacency and superset checks using binary bit pattern trees, we also shorten the overall length of intermediate EFMs. By running a NET analysis at every iteration on all (positive and negative) intermediate EFMs we identify infeasible ones and remove them at the moment of birth even before the bit pattern trees are created and adjacency tests are performed. Therefore, infeasible EFMs were unable to proliferate and to inflate the list of (intermediate) EFMs with irrelevant offspring. This dramatically reduced the memory requirements. In fact, if we only used the measured glucose metabolome and the M1-model for tEFMA, a current, high-end personal computer (typically 32GB RAM) would suffice to perform the analysis in a single working day and eliminate the need for a dedicated high performance computing environment. Conversely, tEFMA allowed us to analyse larger systems, which were inaccessible to an ordinary EFMA on our computer infrastructure.

To curb the explosion of the number of (intermediate) EFMs, we solved many LPs to determine their feasibility. In our application LPs are uncritical in terms of memory consumption. Overall we saved memory by removing infeasible EFMs at the price of an increased computational load to evaluate the LPs. Fewer (intermediate) EFMs meant a shorter list of (intermediate) EFMs, too. This reduced the time to perform the adjacency and superset tests on the EFMs. In the tested cases, the overall runtime decreased at least by 50%. Note that the scaling and efficiency of the DDM critically depends on the order in which constraints are processed [21, 36]. This remains also true for tEFMA (data not shown).

It is known that out of all EFMs in large networks few are physiologically significant [25]. Ideally only those will also be calculated. tEFMA (partly) reaches this aim. By adding constraints derived from metabolomics data we reduced the solution space, leading to a substantial reduction in the number of EFMs without losing any biologically relevant EFMs. However, tEFMA only identifies thermodynamically infeasible EFMs. For instance, during growth under high glucose conditions the glyoxylate shunt is inactive due to regulatory interactions. This is not detected by tEFMA. Therefore tEFMA alone does not allow for an EFMA of a (large) genome scale model. In fact, we were unable to complete a tEFMA on a current genome scale model of *E. coli* on our computer infrastructure. More (omics-)data, like regulatory constraints [26], need to be included to tighten the solution space and get rid of irrelevant EFMs. Recently, gene expression data was used to calculate a small subset of characteristic EFMs [61] in genome-scale networks. In contrast to their method, however, tEFMA is comprehensive and builds on first principles, rather than statistical heuristics. Nevertheless a combina-

tion of their method with tEFMA is required to fully enumerate EFMs in genome-scale models, which is the scope of further work.

Although tEFMA utilizes an optimization principle to fit the metabolic profile, it still retains the ability to unbiasedly characterize all metabolic capabilities of an organism. However, tEFMA cannot predict individual metabolic fluxes. In fact, even the combination of two thermodynamically feasible EFMs might result in an infeasible flux distribution [29]. This is in contrast to thermodynamic-based metabolic flux analysis, where an optimization principle is used to determine a particular thermodynamically feasible flux distribution [62, 63, 60]. Predicting intracellular flux distribution from an EFM-spectrum is an active field of research [64]. In fact, metabolite data have increasingly been utilized together with EFMA in order to gain more reliable flux estimates [65, 66, 67, 68]. However, in all these studies an EFMA was carried out first (on a small-scale metabolic model), while the thermodynamic feasibility was only checked *a posteriori*. tEFMA will aid such studies in providing better computational performance and allowing larger systems to be analyzed.

The success of tEFMA is dependent on the availability of a measured metabolome. Measurement errors in the concentrations were taken into account, and tEFMA was found to be robust against further perturbations. More critical for tEFMA is the requirement for accurate data on the Gibbs free energy of formation, $\Delta_f G$, for each metabolite. Our analysis showed that an error in $\Delta_f G$ of up to 1kJ/mol did not cause alterations. Such accuracy is achievable with current (reactant contribution) methods for the estimation of the Gibbs energy [46]. However, these data cover less than one tenth of the reactions in a typical genome scale model. Yet they are sufficient for the kind of medium-scale models accessible to tEFMA. Thus even if only a small fraction of the metabolome were available, tEFMA will still provide a computational advantage. Moreover, missing data do not lead to the identification of false positives. Uncharacterized reactions can simply be omitted in NET analysis. Consequently some thermodynamically infeasible EFMs will not be detected and the overall efficiency of the algorithm will be reduced.

tEFMA is inherently condition specific and in principle has to be repeated upon any change in the environment. In practice, however, that might not be necessary as Ishii et al. [69] observed that metabolite levels were remarkably stable against perturbations.

tEFMA's condition specificity is in strong contrast to the approach taken by Hunt et al. [24]. Those authors pinned their approach on massive parallelization by recursively splitting the network in appropriately selected subnetworks and performing an EFMA there. As the authors did not utilize any additional information, their enumeration is complete and has to be run only once. However, they found close to two billion EFMs in a large-scale model of *P. tricornutum* [24]. The sheer scale makes an interpretation of the EFMs difficult and computationally challenging. Extrapolating our results onto their model, we expect that many EFMs will be infeasible and therefore biologically irrelevant. This could be easily checked by running a NET analysis on their set of EFMs, if experimental data were available. As both approaches are DDM based, it should be possible to integrate tEFMA into the approach of Hunt et al. [24].

tEFMA retains the ability to allow for a fully unbiased analysis of metabolism. In fact, the predicted inactivity of GDH under growth on glucose was completely derived from first principles. This allows to draw very general statements of biological relevance without relying on optimality criteria or particular flux distributions. The inactivity of GDH for instance, allows glutamate synthesis only via the ATP consuming GOGAT

pathway. The increased energy demand for glutamate synthesis might cause problems during recombinant protein production, which induces additional energy requirements in the host. Thus by activating GDH rather than GOGAT the metabolic burden is reduced.

Currently an assumption-free tEFMA can only be performed on prokaryotes. tEFMA on eukaryotes would require compartment specific concentration data. A theory to describe the thermodynamics of inter-compartmental transport is available [47], yet current experimental methods do not allow for a compartment specific resolution of the metabolome. In order to apply tEFMA also to compartmentalised organisms *ad hoc* assumptions are required to estimate the missing compartment specific concentration data [29].

In summary, we developed tEFMA, a tool that presents an important step forward to the analysis of genome-scale metabolic networks. tEFMA integrates NET analysis into EFMA and succeeds in calculating only EFMs, that are thermodynamically consistent with a given metabolome. By doing so, it dramatically reduces the hardware requirements for such an analysis to be carried out and paves the way to enumerate EFMs in large-scale metabolic networks. This is possible as the calculated set of EFMs is reduced by the large number of thermodynamically infeasible EFMs. To show the accuracy of the tool we presented the correct identification of several infeasible pathways without making wrong predictions. Furthermore, we pointed out that tEFMA correctly distinguishes between the GDH and GOGAT pathways to produce glutamate. Additionally, we verified that the patterns, and therefore pathways, which were found to be infeasible in the smaller model remained infeasible in the larger model.

Supplementary Material

6.1–6.3 show infeasible patterns for different cultivation media. Diagrams were created with Cytoscape [70] according to the *E. coli* core model [2] with additional glycerol exchange (model M2). Metabolites are drawn as squares and reactions as diamonds. Reactions of the infeasible pattern are marked red. Directions of the reactions are drawn according to the model, except for those that are part of a pattern, where only the infeasible direction is plotted. Infeasibility patterns are also listed in Table 6.9.

Table 6.1: Minimal infeasible patterns in *E. coli* aerobically growing on glucose.

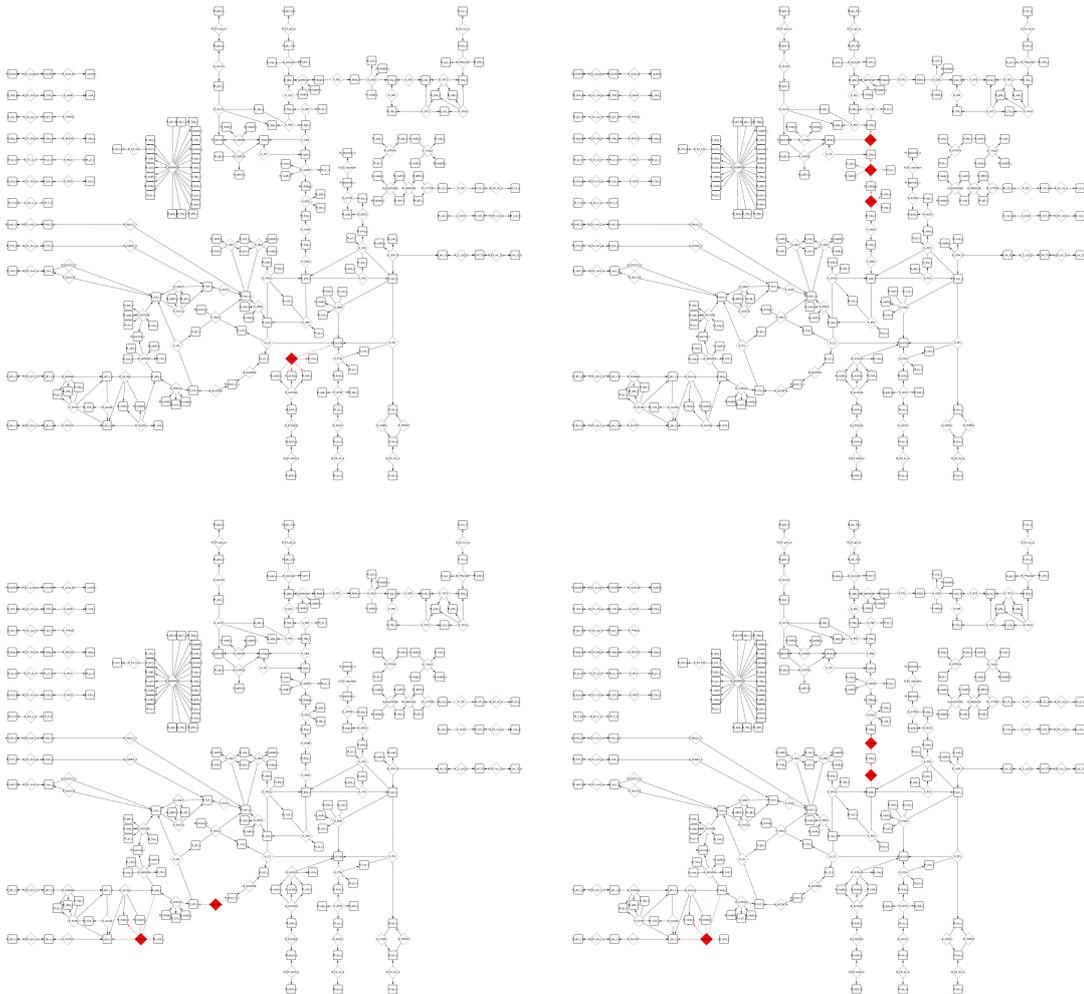


Table 6.1 continues on next page.

Table 6.1 continues from previous page.

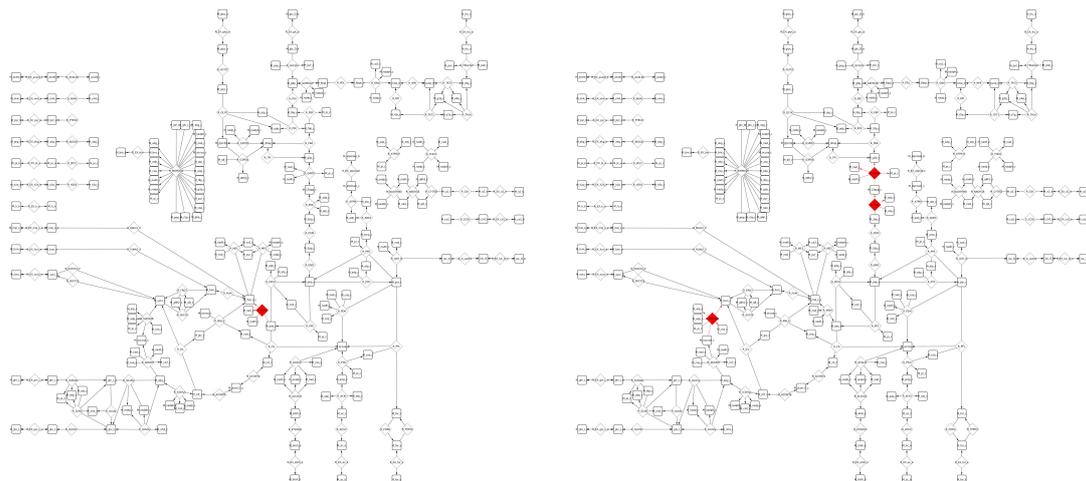


Table 6.1 ends.

Table 6.2: Minimal infeasible patterns in *E. coli* aerobically growing on glycerol.

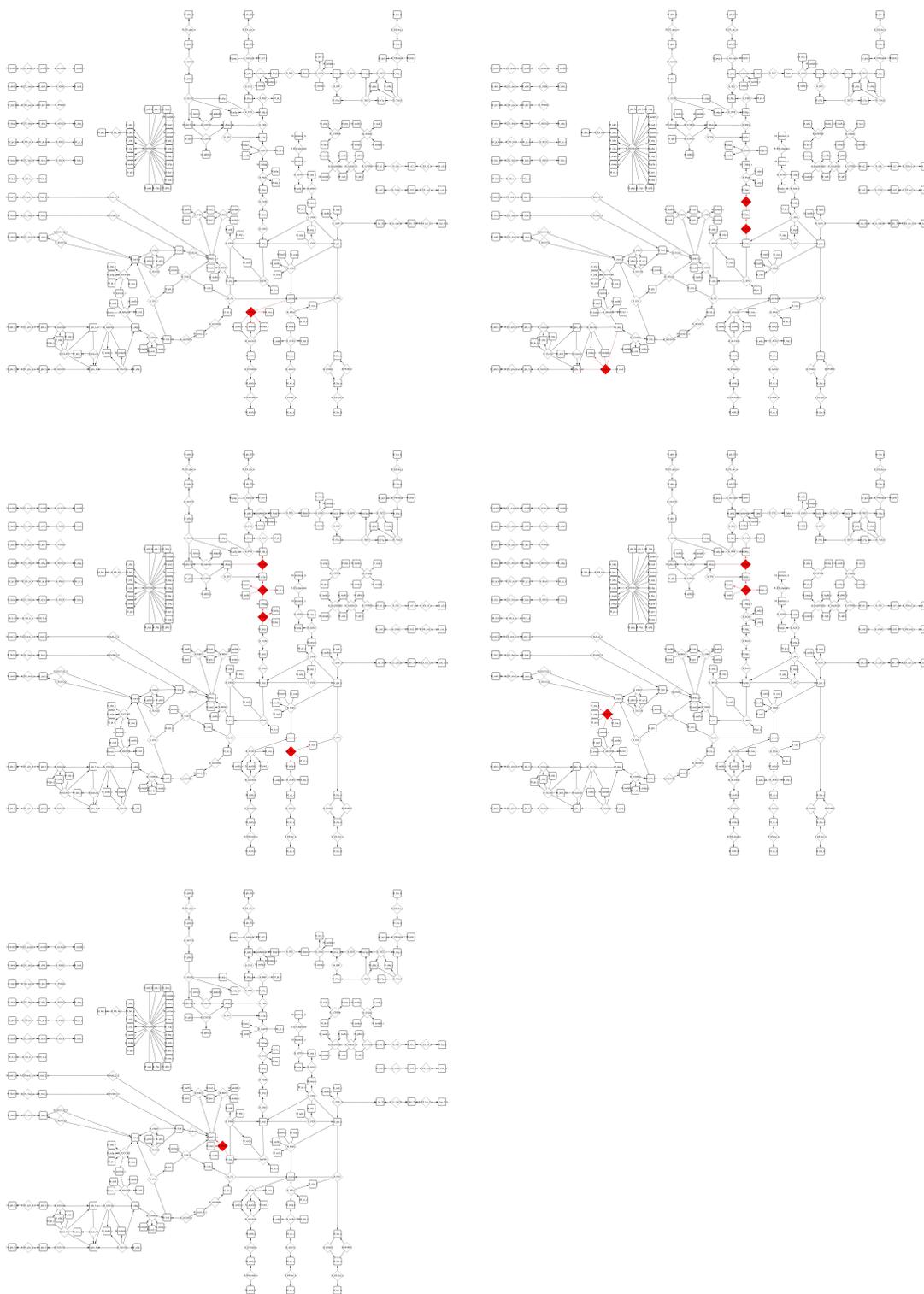


Table 6.2 ends.

Table 6.3: Minimal infeasible patterns in *E. coli* aerobically growing on acetate.

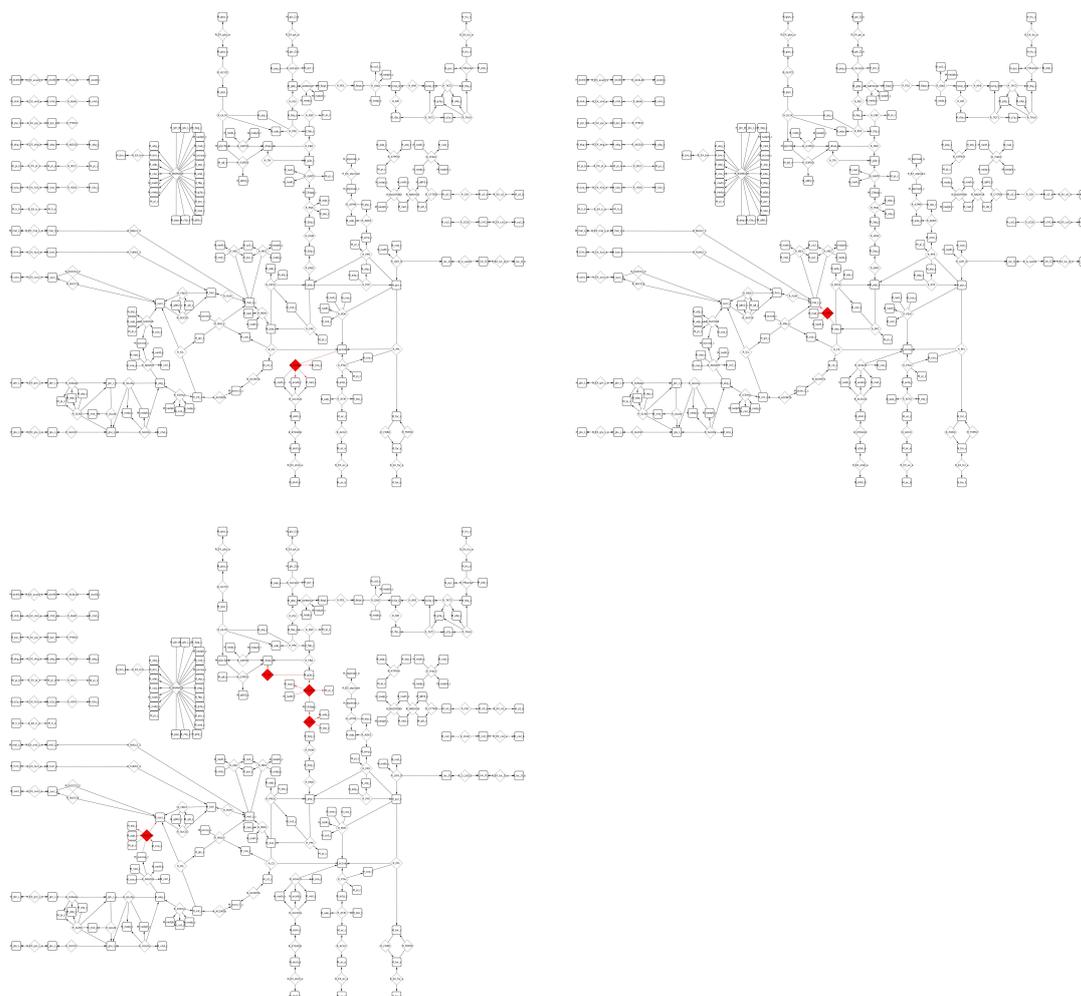


Table 6.3 ends.

Table 6.4: Topological properties of used *E. coli* models.

model	reactions		metabolites	rank
	irreversible	reversible		
M1	37	59	73	68
M2	39	62	76	71
M2-glc	53	48	76	71
M2-glyc	53	48	76	71
M2-ac	53	48	76	71
M3	115	94	178	171

Table 6.5: Performance comparison of model M2 grown on glucose. +, 0, - show the number of (feasible) intermediate EFMs (*i*EFM), having a positive, zero or negative values at reaction (iteration step) *i*. tEFMA checks the feasibility of all + and - *i*EFMs. (The total number of checks per iterations is listed in column “LP”). During post-processing (PP) all enumerated (*i*)EFMs are checked in tEFMA. The columns “cand.” and “new” *i*EFM, list the number of potentially new *i*EFMs and the actual number of new *i*EFMs found in iteration *i*. The column “diff” lists the difference in the number of new *i*EFMs for both methods. Bold values in the table indicate infeasible *i*EFMs were detected.

<i>i</i>	<i>i</i> EFM			LP	tEFMA			<i>i</i> EFM		EFMA			diff		
	+	0	-		feasible <i>i</i> EFM			cand.	new	<i>i</i> EFM					
					+	0	-			+	0	-		cand.	new
0	1	39	2	3	1	39	2	2	2	1	39	2	2	2	0
1	3	39	0	3	3	39	0	0	0	3	39	0	0	0	0
2	2	39	1	3	2	39	1	2	2	2	39	1	2	2	0
3	2	40	1	3	2	40	1	2	2	2	40	1	2	2	0
4	3	40	1	4	3	40	1	3	2	3	40	1	3	2	0
5	4	40	1	5	4	40	1	4	2	4	40	1	4	2	0
6	2	39	5	7	2	39	5	10	10	2	39	5	10	10	0
7	10	40	1	11	10	40	1	10	10	10	40	1	10	10	0
8	1	39	20	21	1	39	20	20	20	1	39	20	20	20	0
9	3	40	17	20	3	40	17	51	28	3	40	17	51	28	0
10	18	41	12	30	10	41	12	120	17	18	41	12	216	22	5
11	14	48	6	20	14	48	6	84	6	18	53	10	180	10	4
12	16	40	12	28	16	40	7	112	13	22	40	19	418	34	21
13	29	39	1	30	29	39	1	29	28	55	40	1	55	54	26
14	2	66	28	30	2	66	28	56	2	2	93	54	108	2	0
15	21	37	12	33	17	37	10	170	82	46	37	14	644	212	130
16	5	34	97	102	5	34	86	430	77	5	34	256	1,280	139	62
17	1	35	80	81	1	35	80	80	80	1	36	141	141	141	61
18	40	26	50	90	40	26	50	2,000	427	64	26	88	5,632	746	319
19	101	211	181	282	101	211	181	18,281	239	135	311	390	52,650	464	225
20	451	36	64	515	451	36	64	28,864	869	767	36	107	82,069	1,524	655
21	93	130	1,133	1,226	90	130	693	62,370	970	100	131	2,096	209,600	1,778	808
22	1,139	49	2	1,141	1,052	49	1	1,052	3	1,906	58	45	85,770	710	707
23	315	65	724	1,039	309	65	723	223,407	1,603	702	135	1,837	1,289,574	6,498	4,895
24	456	617	904	1,360	456	617	904	412,224	788	3,377	1,828	2,130	7,193,010	3,678	2,890
25	1,527	265	69	1,596	1,527	265	69	105,363	564	8,260	502	121	999,460	1,550	986
26	1,085	282	989	2,074	1,085	282	983	1,066,555	5,188	4,193	441	5,678	23,807,854	26,025	20,837
27	2,833	124	3,598	6,431	2,833	124	3,598	10,193,134	29,433	10,949	168	19,542	213,965,358	159,271	129,838
PP	-	-	-	32,390	-	-	-	-	-	-	-	-	-	-	-
Σ				48,578				12,114,435	40,467				247,694,123	202,936	162,469

Table 6.6: Performance comparison of model M2 grown on glycerol. +, 0, - show the number of (feasible) intermediate EFMs (*i*EFM), having a positive, zero or negative values at reaction (iteration step) *i*. tEFMA checks the feasibility of all + and - *i*EFMs. (The total number of checks per iterations is listed in column “LP”). During post-processing (PP) all enumerated (*i*)EFMs are checked in tEFMA. The columns “cand.” and “new” *i*EFM, list the number of potentially new *i*EFMs and the actual number of new *i*EFMs found in iteration *i*. The column “diff” lists the difference in the number of new *i*EFMs for both methods. Bold values in the table indicate infeasible *i*EFMs were detected.

<i>i</i>	<i>i</i> EFM			LP	tEFMA feasible <i>i</i> EFM			<i>i</i> EFM		EFMA					diff
	+	0	-		+	0	-	cand.	new	+	0	-	cand.	new	
0	2	40	1	3	2	40	1	2	2	2	40	1	2	2	0
1	1	41	2	3	1	41	2	2	2	1	41	2	2	2	0
2	1	41	2	3	1	41	2	2	2	1	41	2	2	2	0
3	3	41	0	3	3	41	0	0	0	3	41	0	0	0	0
4	2	41	1	3	2	41	1	2	2	2	41	1	2	2	0
5	2	41	2	4	2	41	2	4	3	2	41	2	4	3	0
6	3	41	2	5	3	41	2	6	3	3	41	2	6	3	0
7	1	41	5	6	1	41	5	5	5	1	41	5	5	5	0
8	6	40	1	7	6	40	1	6	6	6	40	1	6	6	0
9	1	39	12	13	1	39	12	12	12	1	39	12	12	12	0
10	13	38	1	14	13	38	1	13	12	13	38	1	13	12	0
11	2	49	12	14	2	49	12	24	2	2	49	12	24	2	0
12	1	34	18	19	1	34	18	18	18	1	34	18	18	18	0
13	10	38	5	15	10	38	5	50	42	10	38	5	50	42	0
14	55	34	1	56	55	34	1	55	16	55	34	1	55	16	0
15	55	41	9	64	55	41	9	495	17	55	41	9	495	17	0
16	18	41	54	72	18	41	54	972	62	18	41	54	972	62	0
17	55	41	25	80	55	41	25	1,375	33	55	41	25	1,375	33	0
18	46	34	49	95	33	34	49	1,617	157	46	34	49	2,254	218	61
19	156	46	22	178	156	46	22	3,432	98	217	47	34	7,378	122	24
20	21	43	236	257	21	43	236	4,956	1,137	22	50	314	6,908	1,343	206
21	420	28	753	1,173	420	28	708	297,360	5,346	517	30	868	448,756	6,604	1,258
22	1,333	1,917	2,544	3,877	1,333	1,917	2,534	3,377,822	3,074	1,641	2,198	3,312	5,434,992	3,954	880
23	5,706	348	270	5,976	5,706	348	270	1,540,620	4,707	6,822	620	351	2,394,522	5,732	1,025
24	4,971	342	5,448	10,419	4,971	342	5,448	27,082,008	29,120	5,665	419	7,090	40,164,850	36,386	7,266
25	33,531	161	741	34,272	33,530	161	741	24,845,730	14,225	41,313	192	965	39,867,045	17,733	3,508
26	2,216	605	45,095	47,311	1,880	605	11,946	22,458,480	13,538	2,523	710	56,005	141,300,615	22,038	8,500
27	7,157	714	8,152	15,309	6,191	714	7,295	45,163,345	21,729	11,634	1,250	12,387	144,110,358	42,902	21,173
28	4,022	5,002	19,610	23,632	4,022	5,002	19,570	78,710,540	12,697	15,754	8,594	31,438	495,274,252	36,165	23,468
pp	-	-	-	21,721	-	-	-	-	-	-	-	-	-	-	-
Σ				164,594				203,488,953	106,067				869,014,973	173,436	67,369

Table 6.7: Performance comparison of model M2 grown on acetate. +, 0, - show the number of (feasible) intermediate EFMs (*i*EFM), having a positive, zero or negative values at reaction (iteration step) *i*. tEFMA checks the feasibility of all + and - *i*EFMs. (The total number of checks per iterations is listed in column “LP”). During post-processing (PP) all enumerated (*i*)EFMs are checked in tEFMA. The columns “cand.” and “new” *i*EFM, list the number of potentially new *i*EFMs and the actual number of new *i*EFMs found in iteration *i*. The column “diff” lists the difference in the number of new *i*EFMs for both methods. Bold values in the table indicate infeasible *i*EFMs were detected.

<i>i</i>	tEFMA									EFMA					diff
	<i>i</i> EFM			LP	feasible <i>i</i> EFM			<i>i</i> EFM		<i>i</i> EFM			cand.	new	
	+	0	-		+	0	-	cand.	new	+	0	-			
0	2	38	1	3	2	38	1	2	2	2	38	1	2	2	0
1	1	39	2	3	1	39	2	2	2	1	39	2	2	2	0
2	1	39	2	3	1	39	2	2	2	1	39	2	2	2	0
3	1	39	2	3	1	39	2	2	2	1	39	2	2	2	0
4	2	39	1	3	2	39	1	2	2	2	39	1	2	2	0
5	2	39	2	4	2	39	2	4	3	2	39	2	4	3	0
6	5	39	0	5	5	39	0	0	0	5	39	0	0	0	0
7	3	39	2	5	3	39	2	6	3	3	39	2	6	3	0
8	1	39	5	6	1	39	5	5	5	1	39	5	5	5	0
9	6	38	1	7	6	38	1	6	5	6	38	1	6	6	1
10	12	37	1	13	11	37	1	11	10	12	37	1	12	12	2
11	20	38	1	21	20	38	1	20	10	21	39	1	21	10	0
12	2	38	28	30	2	38	28	56	30	2	40	28	56	30	0
13	2	64	4	6	2	64	4	8	8	2	66	4	8	8	0
14	2	68	4	6	2	68	4	8	4	2	70	4	8	4	0
15	38	34	2	40	38	34	2	76	48	38	36	2	76	76	28
16	74	57	17	91	50	57	13	650	206	74	59	17	1,258	340	134
17	127	40	170	297	127	40	170	21,590	474	151	46	276	41,676	1,348	874
18	2	39	600	602	2	39	600	1,200	97	4	45	1,496	5,984	244	147
19	107	28	4	111	107	28	3	321	197	202	36	55	11,110	812	615
20	4	52	278	282	4	52	276	1,104	446	8	92	950	7,600	1,483	1,037
21	308	63	131	439	308	63	131	40,348	604	1,139	115	329	374,731	1,931	1,327
22	947	28	0	947	947	28	0	0	0	3,147	38	0	0	0	0
23	256	48	671	927	256	48	671	171,776	188	304	56	2,825	858,800	702	514
24	477	21	151	628	397	21	74	29,378	1,190	830	24	208	172,640	4,299	3,109
25	31	27	1,678	1,709	31	27	1,630	50,530	917	33	28	5,092	168,036	1,318	401
26	29	353	611	640	29	353	593	17,197	536	30	431	918	27,540	855	319
PP	-	-	-	945	-	-	-	-	-	-	-	-	-	-	-
Σ				7,776				351,501	4,991				1,697,127	13,499	8,508

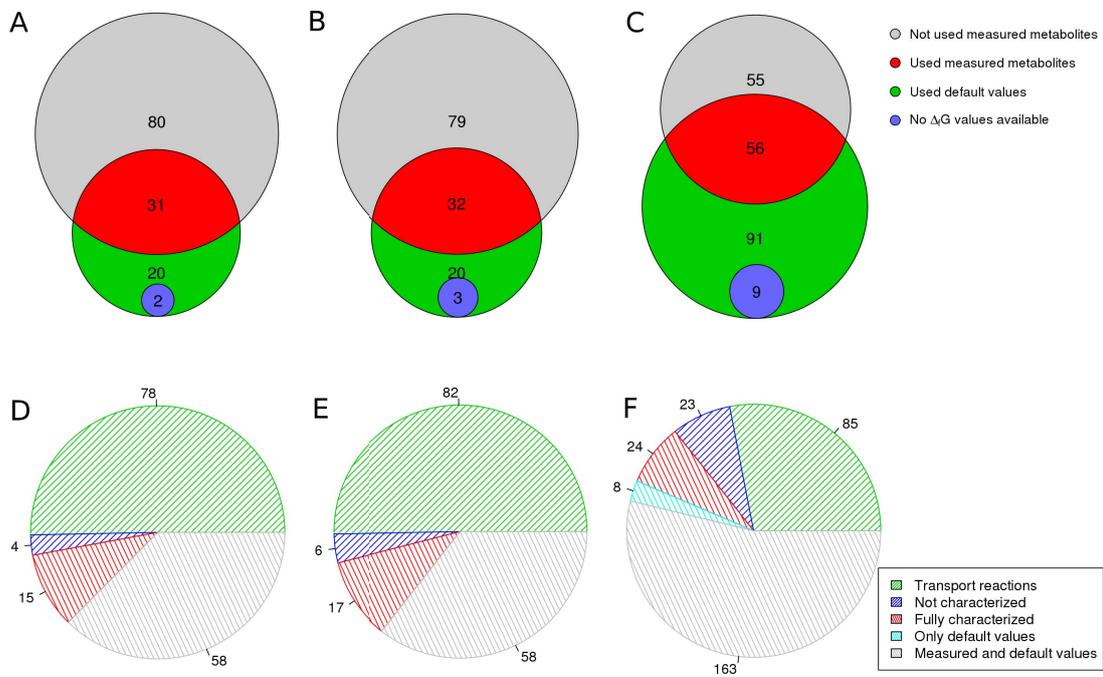


Figure 6.5: Characterization of usable reactions for thermodynamic checks. (A) Out of 111 measured metabolites [30] 31 could be used in the *E. coli* core model M1. For 20 additional metabolites default values were used. Two metabolites, including the biomass, could not be used because of lack of available $\Delta_f G$ value. (B) shows the data for model M2 and (C) for M3. (D) 73 irreversible reactions of the model M1 were used for thermodynamic checks, whereof 15 were fully characterized as for all involved metabolites measured data were available and 58 were partly characterized. 78 transport reactions were not used for the checks as well as 4 uncharacterized reactions, because of missing $\Delta_f G$ information. (E) shows the data for model M2 and (F) for model M3.

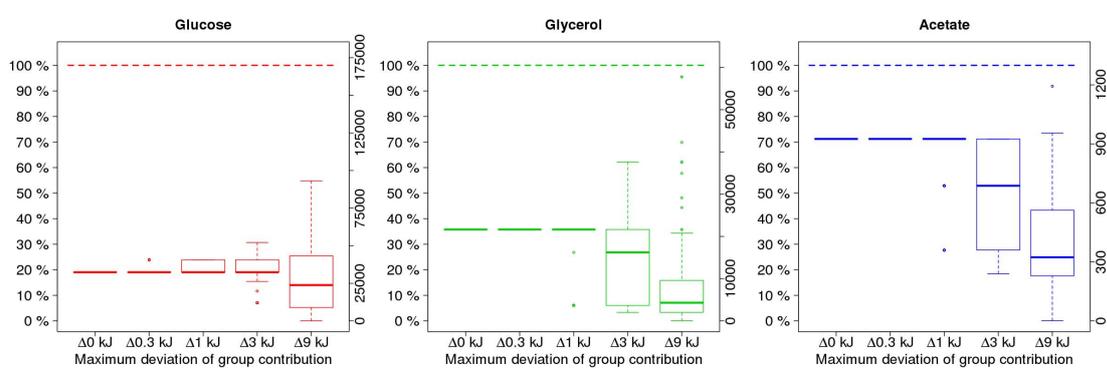


Figure 6.6: The box-plots show the number of calculated EFMs as a function of perturbed $\Delta_f G$ values compared to not perturbed results. Therefore $\Delta_f G$ values for glucose, glycerol and acetate were randomly perturbed by 0, 0.3, 1, 3 and 9 kJ and EFMs of model M2 were calculated with tEFMA. Each box-plot represents 100 runs.

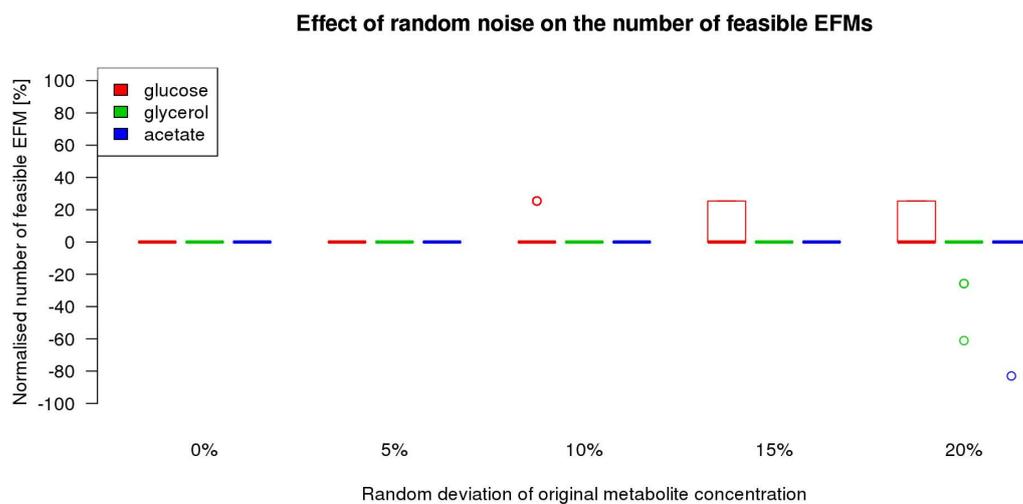


Figure 6.7: Upper and lower metabolite concentration limits for glucose, glycerol and acetate were randomly perturbed by $\pm 5\%$, $\pm 10\%$, $\pm 15\%$ and $\pm 20\%$ and EFMs were calculated. The box-plots show the number of calculated EFMs normalized to not perturbed results averaged over 100 runs each. Note that up to $\pm 15\%$ box-plots are highly degenerated and except for outliers (open circles) only the median value is visible.

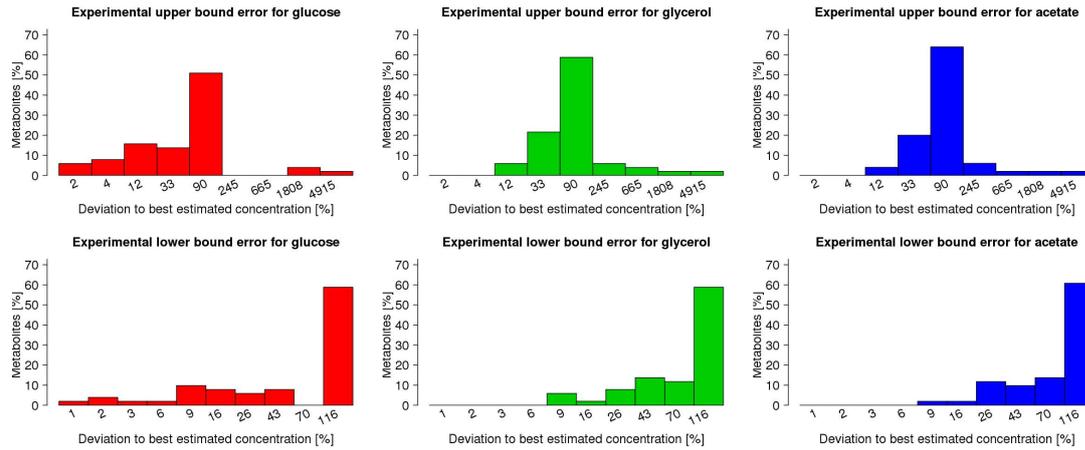


Figure 6.8: Deviation of the upper and lower bound for glucose (left panels), glycerol (middle panels) and acetate (right panels) from the best estimated concentrations as published by Bennett et al. [30]. The deviation was calculated by following formula:

$$\text{deviation of metabolite} = \frac{\text{upper or lower concentration}}{\text{best estimated concentration}}$$

The histogram shows the relative number of metabolites as function of the deviation. Note, that the x axis is in logarithmic scale. For all three metabolites, the peak for upper bound concentration is 90% above the best estimated value and below 116% for lower bound concentrations.

Table 6.8: Changes in the upper and lower concentration bounds on glucose (red), glycerol (green), and acetate (blue) upon random perturbations. The full lines illustrate the non-perturbed values illustrated in Figure 6.8.

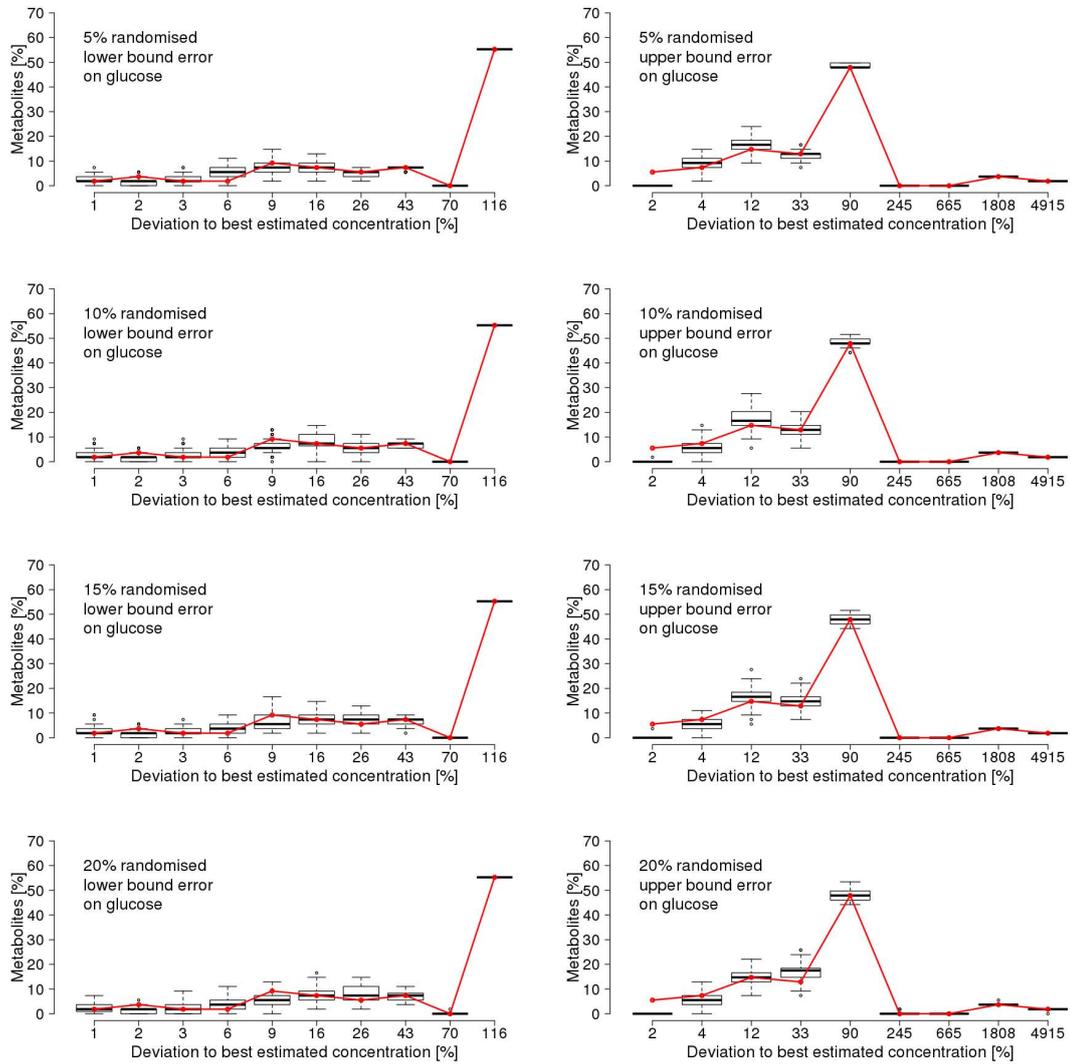


Table 6.8 continues on next page.

Table 6.8 continues from previous page.

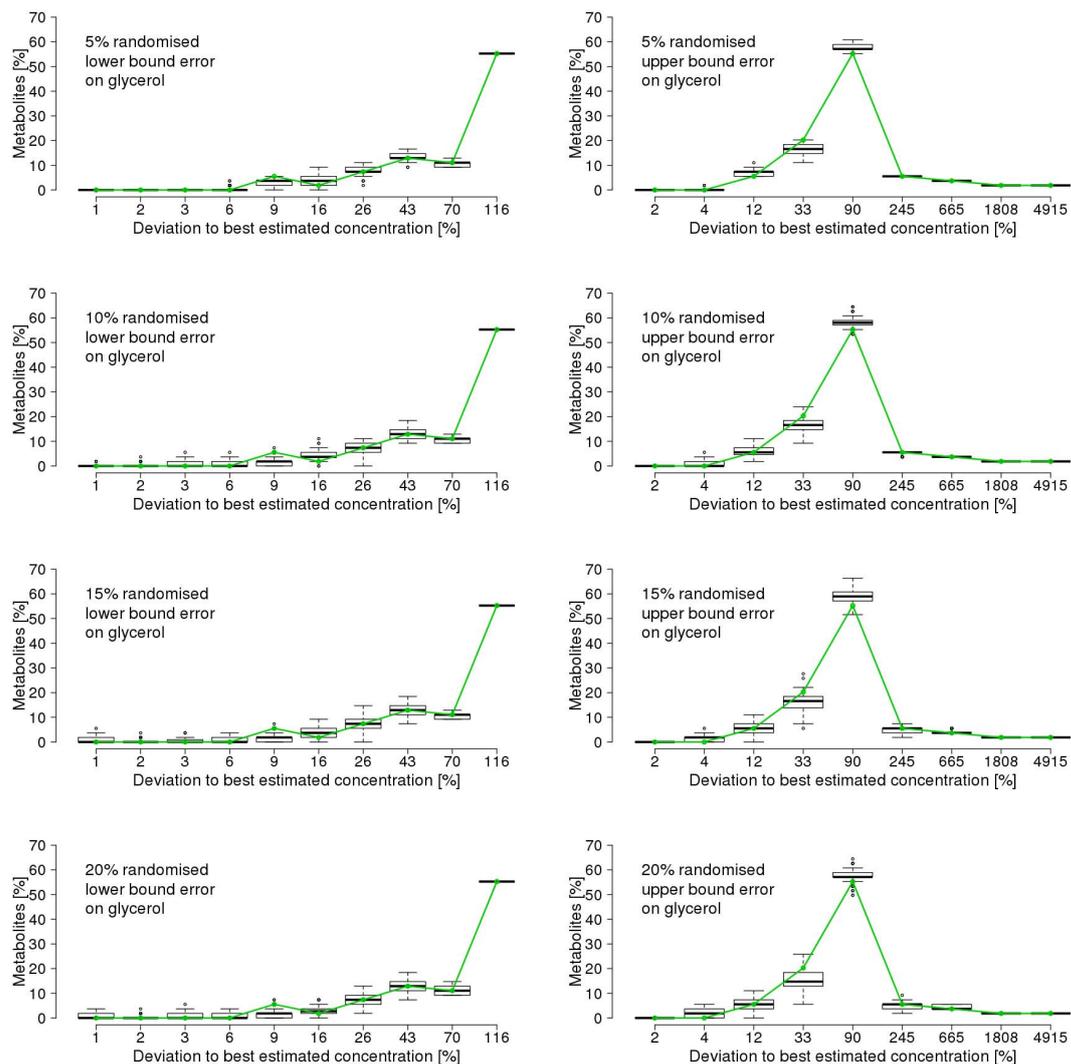


Table 6.8 continues on next page.

Table 6.8 continues from previous page.

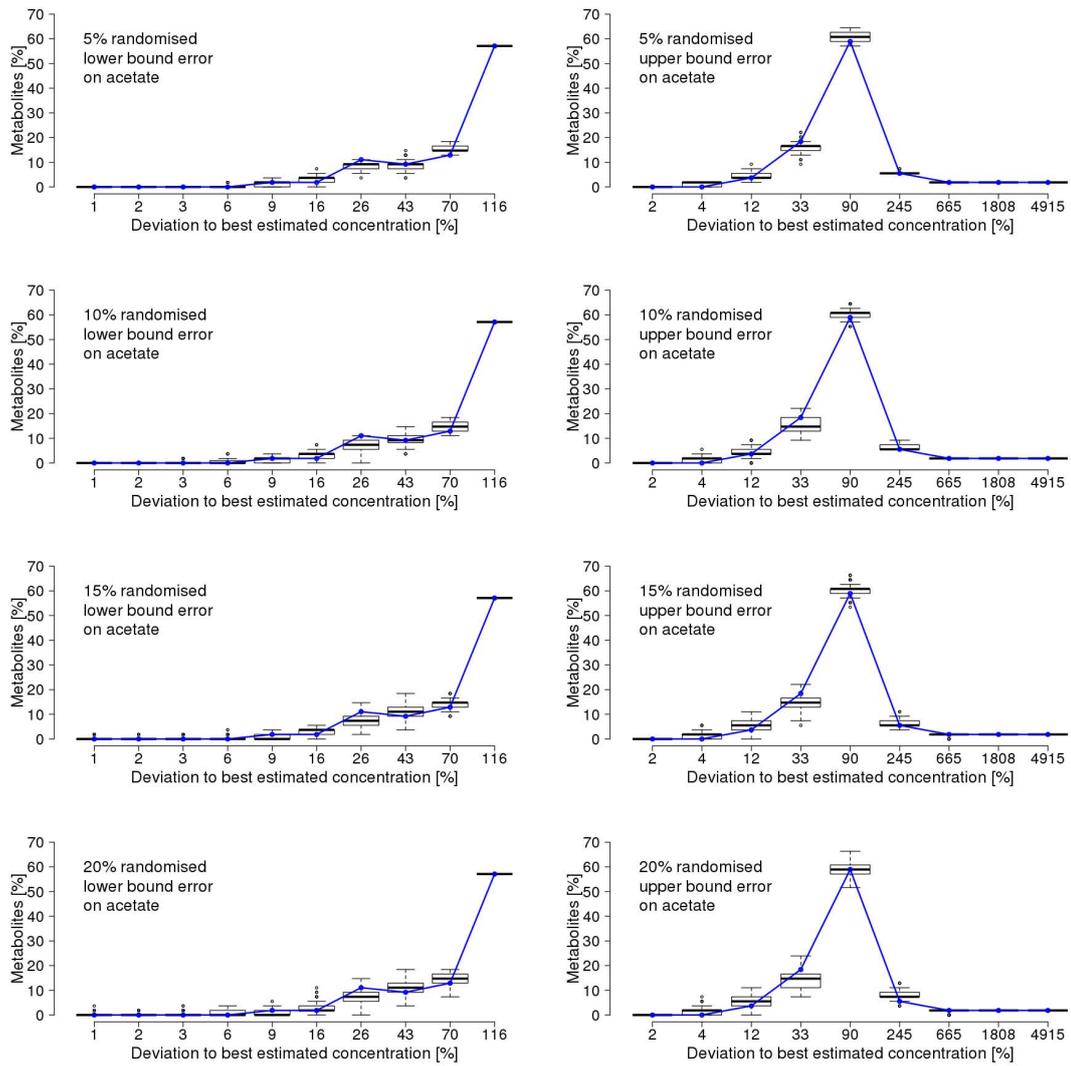


Table 6.8 ends.

Table 6.9: Infeasible patterns found in model M2 and M3, when *E. coli* is grown on glucose, acetate or glycerol. The minus in front of the reaction denotes the reverse direction.

M2 on acetate	-ACALD						
	-MDH						
	-TPI	-SUACOAS	PGK		-GAPD		
M2 on glycerol	-ACALD						
	-MDH						
	-ENO	-GLUD _y	PGM				
	-FBA	-SUACOAS	-GAPD				
	-FBA	PTAr	PGK		-GAPD		
M2 on glucose	-ACALD						
	-MDH						
	-GLUD _y	ACONT _b					
	-ENO	-GLUD _y	PGM				
	-FBA	-GAPD	PGK				
	-GAPD	-SUACOAS	PGK				
M3 on glucose	-ACALD						
	-MDH						
	-GLUD _y	ACONT _b					
	-ENO	-GLUD _y	PGM				
	-FBA	-GAPD	PGK				
	-GAPD	-SUACOAS	PGK				
	-GLUD _y	FUM					
	-MTHFC	FUM					
	-MTHFC	ACONT _b					
	-PTAr	-SUACOAS					
	-GAPD	-TPI	PGK				
	-ENO	G5SAD _s	P5CR	PGM			
	PTAr	ACONT _b	G5SAD _s	G5SD	GLU5K	P5CR	
	-SUACOAS	ACONT _b	G5SAD _s	G5SD	GLU5K	P5CR	
	-SUACOAS	FUM	G5SAD _s	G5SD	GLU5K	P5CR	

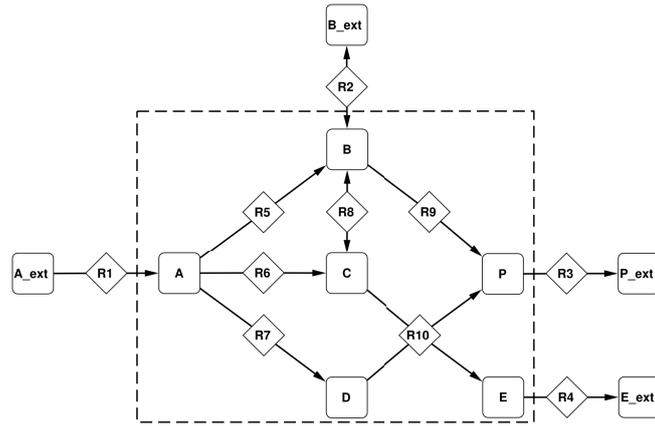


Figure 6.9: Toy network: This network has six internal metabolites (A,B,C,D,E,P) and 10 reactions that are used by *efmtool*. Reactions R_2 and R_8 are reversible.

Proof of safe removal of thermodynamically infeasible EFMs

EFM enumeration by *efmtool*

First, we want to show how EFMs are enumerated by *efmtool*. For this purpose we use the same example as the one in the supplementary material of Terzer and Stelling [22]. Since we want to emphasize on the thermodynamics of EFMs we do not consider the compression methods, adjacency tests and the tree method of the *efmtool* here.

Model

The example is built on the model of Klamt and Stelling [71] (Fig. 6.9) and has the following stoichiometric matrix:

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}
A	1	0	0	0	-1	-1	-1	0	0	0
B	0	1	0	0	1	0	0	-1	-1	0
C	0	0	0	0	0	1	0	1	0	-1
D	0	0	0	0	0	0	1	0	0	-1
E	0	0	0	-1	0	0	0	0	0	1
P	0	0	-1	0	0	0	0	0	1	1

The rows of the stoichiometric matrix represent the internal metabolites (A,B,C,D,E,P) and the columns the reactions ($R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8, R_9, R_{10}$).

Initialization phase

The algorithm starts with the initialization phase and splits the reversible reactions (R_2 and R_8) into two irreversible reactions each.

	R_1	R_2	R_{2r}	R_3	R_4	R_5	R_6	R_7	R_8	R_{8r}	R_9	R_{10}
A	1	0	0	0	0	-1	-1	-1	0	0	0	0
B	0	1	-1	0	0	1	0	0	-1	1	-1	0
C	0	0	0	0	0	0	1	0	1	-1	0	-1
D	0	0	0	0	0	0	0	1	0	0	0	-1
E	0	0	0	0	-1	0	0	0	0	0	0	1
P	0	0	0	-1	0	0	0	0	0	0	1	1

In the next step the kernel matrix is computed and then the rows are permuted

R_1	0	1	-1	1	0	2		R_{2r}	1	0	0	0	0	0
R_2	1	-1	1	-1	1	0		R_5	0	1	0	0	0	0
R_{2r}	1	0	0	0	0	0		R_8	0	0	1	0	0	0
R_3	0	0	0	0	1	1		R_{8r}	0	0	0	1	0	0
R_4	0	0	0	0	0	1		R_9	0	0	0	0	1	0
R_5	0	1	0	0	0	0		R_4	0	0	0	0	0	1
R_6	0	0	-1	1	0	1		R_{10}	0	0	0	0	0	1
R_7	0	0	0	0	0	1		R_7	0	0	0	0	0	1
R_8	0	0	1	0	0	0		R_3	0	0	0	0	1	1
R_{8r}	0	0	0	1	0	0		R_6	0	0	-1	1	0	1
R_9	0	0	0	0	1	0		R_1	0	1	-1	1	0	2
R_{10}	0	0	0	0	0	1		R_2	1	-1	1	-1	1	0

Iterations and thermodynamic feasibility checks

In the iteration phase all EFMs are enumerated. The EFMs are represented by the columns of the generated matrix. Here, in the first step the first nine lines are easily converted to binary values as they contain no negative values. We use \bullet for reactions carrying a flux and \star for non active reactions in the binary notation (note that Terzer and Stelling [22] use the inverse logic in their publication).

R_{2r}	\bullet	\star	\star	\star	\star	\star
R_5	\star	\bullet	\star	\star	\star	\star
R_8	\star	\star	\bullet	\star	\star	\star
R_{8r}	\star	\star	\star	\bullet	\star	\star
R_9	\star	\star	\star	\star	\bullet	\star
R_4	\star	\star	\star	\star	\star	\bullet
R_{10}	\star	\star	\star	\star	\star	\bullet
R_7	\star	\star	\star	\star	\star	\bullet
R_3	\star	\star	\star	\star	\bullet	\bullet
R_6	0	0	-1	1	0	1
R_1	0	1	-1	1	0	2
R_2	1	-1	1	-1	1	0

At the start of each iteration the binary part of modes with positive or negative values on the next numeric position (R_6) is checked for thermodynamic feasibility. For this purpose a linear problem is created containing constraints for all reactions in a column carrying a flux (denoted with \bullet). In this example we assume that it is thermodynamically not feasible that R_8 and R_{10} are active together in the same EFM. In other words, it is not allowed that both reactions are labeled with \bullet in the same column. As none of the modes are infeasible so far we proceed with the next iteration.

As the next row has one negative value in column three the corresponding mode (column) is removed and combined with all modes containing a positive value (column 4 and 6). The combination is performed by a logic OR which is described by the following truth table:

Table 6.10: Truth table for the combination in *efmtool*

column A	column B	new column
\bullet	\bullet	\bullet
\bullet	\star	\bullet
\star	\bullet	\bullet
\star	\star	\star

Without thermodynamic check

R_{2r}	\bullet	\star	\star	\star	\star	\star	\star
R_5	\star	\bullet	\star	\star	\star	\star	\star
R_8	\star	\star	\star	\star	\star	\bullet	\bullet
R_{8r}	\star	\star	\bullet	\star	\star	\bullet	\star
R_9	\star	\star	\star	\bullet	\star	\star	\star
R_4	\star	\star	\star	\star	\bullet	\star	\bullet
R_{10}	\star	\star	\star	\star	\bullet	\star	\bullet
R_7	\star	\star	\star	\star	\bullet	\star	\bullet
R_3	\star	\star	\star	\bullet	\bullet	\star	\bullet
R_6	\star	\star	\bullet	\star	\bullet	\star	\star
R_1	0	1	1	0	2	0	1
R_2	1	-1	-1	1	0	0	1

With thermodynamic check

R_{2r}	\bullet	\star	\star	\star	\star	\star	\star
R_5	\star	\bullet	\star	\star	\star	\star	\star
R_8	\star	\star	\star	\star	\star	\star	\bullet
R_{8r}	\star	\star	\bullet	\star	\star	\star	\bullet
R_9	\star	\star	\star	\bullet	\star	\star	\star
R_4	\star	\star	\star	\star	\bullet	\star	\star
R_{10}	\star	\star	\star	\star	\bullet	\star	\star
R_7	\star	\star	\star	\star	\bullet	\star	\star
R_3	\star	\star	\star	\bullet	\bullet	\star	\star
R_6	\star	\star	\bullet	\star	\bullet	\star	\star
R_1	0	1	1	0	2	0	0
R_2	1	-1	-1	1	0	0	0

In the left table all modes after this iteration step are shown. In the last column, R_8 and R_{10} are both marked with \bullet . Therefore this mode is thermodynamically infeasible. As this mode has a positive value in row R_1 , it is checked at the start of the next iteration and removed immediately (right table).

The next step is again a simple conversion from numeric to binary values as there are no negative values in row R_1 .

Without thermodynamic check		With thermodynamic check
R_{2r}	●	★ ★ ★ ★ ★ ★ ★
R_5	★	● ★ ★ ★ ★ ★ ★
R_8	★	★ ★ ★ ★ ★ ● ●
R_{8r}	★	★ ★ ● ★ ★ ● ★
R_9	★	★ ★ ★ ● ★ ★ ★
R_4	★	★ ★ ★ ★ ● ★ ●
R_{10}	★	★ ★ ★ ★ ● ★ ●
R_7	★	★ ★ ★ ★ ● ★ ●
R_3	★	★ ★ ★ ● ● ★ ●
R_6	★	★ ★ ● ★ ● ★ ★
R_1	★	● ● ● ★ ● ★ ●
R_2	1	-1 -1 1 0 0 1

In the next step two columns with negative values (2 and 3) have to be combined with positive columns (1,4,7 in left table and 1,4 in right table).

Without thermodynamic check		With thermodynamic check
R_{2r}	●	★ ★ ★ ★ ● ● ★ ★
R_5	★	★ ★ ★ ★ ★ ● ★ ● ★
R_8	★	★ ★ ★ ● ● ★ ★ ● ★ ★
R_{8r}	★	★ ★ ★ ● ★ ★ ● ★ ★ ●
R_9	★	● ★ ★ ★ ★ ★ ★ ● ●
R_4	★	★ ★ ● ★ ● ★ ★ ● ★ ★
R_{10}	★	★ ★ ● ★ ● ★ ★ ● ★ ★
R_7	★	★ ★ ● ★ ● ★ ★ ● ★ ★
R_3	★	● ● ● ★ ★ ★ ● ●
R_6	★	★ ★ ● ★ ★ ★ ● ★ ●
R_1	★	★ ★ ● ★ ● ● ● ●
R_2	●	● ● ★ ★ ★ ★ ★ ★

In the left table two new thermodynamically infeasible EFMs which do not appear in the right table are created.

Postprocessing phase

In the postprocessing step of *efmtool* the rows of the matrix are reordered to restore the original order of the reactions:

Without thermodynamic check										With thermodynamic check									
R_1	*	*	•	*	•	•	•	•	•	R_1	*	*	•	*	•	•	•	•	•
R_2	•	•	*	*	•	*	*	*	*	R_2	•	•	*	*	*	*	*	*	*
R_{2r}	•	*	*	*	*	•	•	*	*	R_{2r}	•	*	*	*	•	•	*	*	*
R_3	*	•	•	*	•	*	*	•	•	R_3	*	•	•	*	*	*	•	•	•
R_4	*	*	•	*	•	*	*	•	*	R_4	*	*	•	*	*	*	*	*	*
R_5	*	*	*	*	*	•	*	•	•	R_5	*	*	*	*	•	*	•	*	*
R_6	*	*	•	*	*	*	•	*	*	R_6	*	*	•	*	*	•	*	•	•
R_7	*	*	•	*	•	*	*	•	*	R_7	*	*	•	*	*	*	*	*	*
R_8	*	*	*	•	•	*	*	•	*	R_8	*	*	*	•	*	*	*	*	*
R_{8r}	*	*	*	•	*	*	•	*	*	R_{8r}	*	*	*	•	*	•	*	•	•
R_9	*	•	*	*	*	*	*	*	•	R_9	*	•	*	*	*	*	•	•	•
R_{10}	*	*	•	*	•	*	*	•	*	R_{10}	*	*	•	*	*	*	*	*	*

In the next step the binary values are converted back to numeric values:

Without thermodynamic check										With thermodynamic check									
R_1	0	0	2	0	1	1	1	2	1	1	R_1	0	0	2	0	1	1	1	1
R_2	1	1	0	0	1	0	0	0	0	0	R_2	1	1	0	0	0	0	0	0
R_{2r}	1	0	0	0	0	1	1	0	0	0	R_{2r}	1	0	0	0	1	1	0	0
R_3	0	1	1	0	1	0	0	1	1	1	R_3	0	1	1	0	0	0	1	1
R_4	0	0	1	0	1	0	0	1	0	0	R_4	0	0	1	0	0	0	0	0
R_5	0	0	0	0	0	1	0	1	1	0	R_5	0	0	0	0	1	0	1	0
R_6	0	0	1	0	0	0	1	0	0	1	R_6	0	0	1	0	0	1	0	1
R_7	0	0	1	0	1	0	0	1	0	0	R_7	0	0	1	0	0	0	0	0
R_8	0	0	0	1	1	0	0	1	0	0	R_8	0	0	0	1	0	0	0	0
R_{8r}	0	0	0	1	0	0	1	0	0	1	R_{8r}	0	0	0	1	0	1	0	1
R_9	0	1	0	0	0	0	0	0	1	1	R_9	0	1	0	0	0	0	1	1
R_{10}	0	0	1	0	1	0	0	1	0	0	R_{10}	0	0	1	0	0	0	0	0

Rows representing reversible reactions are combined and modes with futile cycles (1. and 4. column) are removed.

Without thermodynamic check									With thermodynamic check					
	EM_1	EM_2	EM_3	EM_4	EM_5	EM_6	EM_7	EM_8	EM_1	EM_2	EM_4	EM_5	EM_7	EM_8
R_1	0	2	1	1	1	2	1	1	0	2	1	1	1	1
R_2	1	0	1	-1	-1	0	0	0	1	0	-1	-1	0	0
R_3	1	1	1	0	0	1	1	1	1	1	0	0	1	1
R_4	0	1	1	0	0	1	0	0	0	1	0	0	0	0
R_5	0	0	0	1	0	1	1	0	0	0	1	0	1	0
R_6	0	1	0	0	1	0	0	1	0	1	0	1	0	1
R_7	0	1	1	0	0	1	0	0	0	1	0	0	0	0
R_8	0	0	1	0	-1	1	0	-1	0	0	0	-1	0	-1
R_9	1	0	0	0	0	0	1	1	1	0	0	0	1	1
R_{10}	0	1	1	0	0	1	0	0	0	1	0	0	0	0

The toy network results in 8 EFMs(left table), whereas in this example EFM_3 and EFM_6 are thermodynamically infeasible (right table).

Feasible EFMs are never removed

As shown before, new EFMs are generated by a combination of previously calculated EFMs. Here we show that thermodynamically infeasible modes can never contribute to new thermodynamically feasible modes and can safely be removed during the iteration phase.

- In the main part of this paper we have shown that the constraints of the linear program for checking the feasibility of an EFM are defined by all reactions carrying a flux in this mode.
- Here, a reaction carrying a flux is marked by •
- It is shown in the truth table (Table 6.10) that an active reaction will always stay active by the combination of adjacent candidates, as it is done by *efmtool*.
- Hence, a combined mode always contains all active reactions of the two single contributing modes.
- It is not possible to turn an infeasible flux into a feasible flux by adding additional reactions.
- Thus, a combination of an infeasible EFM (e.g. defined by NET analysis) with any other EFM will always result in an infeasible EFM and can therefore be removed without loss of feasible EFMs.

Calculation of the transformed standard Gibbs free energy of formation

$$\Delta_f G'^0 = -RT \ln \left\{ \sum_i \exp \left[-\Delta_f G_i^0(pH, I)/RT \right] \right\} \quad (6.4a)$$

$$\Delta_f G_i^0(pH, I) = \Delta_f G_i^0 - \Delta_f G_i^0(pH) - \Delta_f G_i^0(I) \quad (6.4b)$$

$$\Delta_f G_i^0(pH) = N_{H_i} RT \ln (10^{-pH}) \quad (6.4c)$$

$$\Delta_f G_i^0(I) = \bar{I}A \frac{z_i^2 - N_{H_i}}{1 + \bar{I}B} \quad (6.4d)$$

$$A = 2.91482 \text{ kJ mol}^{-1} \text{ M}^{-0.5}, \quad B = 1.6 \text{ M}^{-0.5} \quad (6.4e)$$

where the summation over i has to be carried out over all net charges, z_i , of a metabolite. $\Delta_f G_i^0(pH, I)$ denotes the standard transformed Gibbs free energy of formation for the metabolite corrected for ionic strength, I [$\Delta_f G_i^0(I)$], and pH [$\Delta_f G_i^0(pH)$] at charge state z_i . The standard Gibbs free energy of formation, $\Delta_f G_i^0$, was estimated using the online version of eQuilibrator [43]. Finally, N_{H_i} denotes the number of H atoms while A and B are constants [28]. We set

$$I = 0.15 \text{ M}, \quad pH = 7, \quad \text{and} \quad T = 310.15 \text{ K} (37^\circ\text{C}). \quad (6.5)$$

Values for the metabolite's $\Delta_f G'^0$ are listed in the supplementary material, file 2.

Bibliography

- [1] Bernhard Ø. Palsson. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, 1 edition, January 2006. ISBN 0521859034.
- [2] Jeffrey D Orth, Ronan M. T. Fleming, and Bernhard Ø. Palsson. Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide. *EcoSal*, 2010. doi: 10.1128/ecosal.10.2.1.
- [3] Ines Thiele, Neil Swainston, Ronan M T Fleming, Andreas Hoppe, Swagatika Sahoo, Maike K Aurich, Hulda Haraldsdóttir, Monica L Mo, Ottar Rolfsson, Miranda D Stobbe, Stefan G Thorleifsson, Rasmus Agren, Christian Bölling, Sergio Bordel, Arvind K Chavali, Paul Dobson, Warwick B Dunn, Lukas Endler, David Hala, Michael Hucka, Duncan Hull, Daniel Jameson, Neema Jamshidi, Jon J Jons-son, Nick Juty, Sarah Keating, Intawat Nookaew, Nicolas Le Novère, Naglis Malys, Alexander Mazein, Jason A Papin, Nathan D Price, Evgeni Selkov, Martin I Sigurdsson, Evangelos Simeonidis, Nikolaus Sonnenschein, Kieran Smallbone, Anatoly Sorokin, Johannes H G M van Beek, Dieter Weichart, Igor Goryanin, Jens Nielsen, Hans V Westerhoff, Douglas B Kell, Pedro Mendes, and Bernhard Ø Palsson. A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.*, 31(5):41925, May 2013. doi: 10.1038/nbt.2488.
- [4] Adam M Feist and Bernhard Ø Palsson. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nat. Biotechnol.*, 26(6):65967, June 2008. doi: 10.1038/nbt1401.
- [5] Thomas Dandekar, Astrid Fieselmann, Saman Majeed, and Zeeshan Ahmed. Software applications toward quantitative metabolic flux analysis and modeling. *Brief. Bioinformatics*, 15(1):91107, January 2014. doi: 10.1093/bib/bbs065.
- [6] Jörg Stelling, Steffen Klamt, Katja Bettenbrock, Stefan Schuster, and Ernst Dieter Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):1903, November 2002. doi: 10.1038/nature01166.
- [7] Jürgen Zanghellini, David E. Ruckerbauer, Michael Hanscho, and Christian Jungreuthmayer. Elementary flux modes in a nutshell: Properties, calculation and applications. *Biotechnology Journal*, 8:1009–1016, 2013. doi: 10.1002/biot.201200269.

- [8] Stefan Schuster, Thomas Dandekar, and David A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology*, 17:53–60, 1999. doi: 10.1016/S0167-7799(98)01290-6.
- [9] Clemens Wagner and Robert Urbanczik. The geometry of the flux cone of a metabolic network. *Biophys. J.*, 89(6):3837–45, December 2005. doi: 10.1529/biophysj.104.055129.
- [10] Stefan Schuster, LuisF. deFigueiredo, and Christoph Kaleta. Predicting novel pathways in genome-scale metabolic networks. *Biochemical Society Transactions*, 38(5):1202, October 2010. ISSN 0300-5127, 1470-8752. doi: 10.1042/BST0381202. URL <http://www.biochemsoctrans.org/bst/038/1202/bst0381202.htm>. 00004.
- [11] Jörn Behre, Thomas Wilhelm, Axel von Kamp, Eytan Ruppin, and Stefan Schuster. Structural robustness of metabolic networks with respect to multiple knock-outs. *Journal of Theoretical Biology*, 252(3):433–441, June 2008. ISSN 0022-5193. doi: 10.1016/j.jtbi.2007.09.043. URL <http://www.sciencedirect.com/science/article/pii/S0022519307004729>.
- [12] Julien Gagneur and Steffen Klamt. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, 5:175, 2004. doi: 10.1186/1471-2105-5-175.
- [13] Cong T Trinh, Ross Carlson, Aaron Wlaschin, and Friedrich Srienc. Design, construction and performance of the most efficient biomass producing E. coli bacterium. *Metab. Eng.*, 8(6):62838, November 2006. doi: 10.1016/j.ymben.2006.07.006.
- [14] Cong T Trinh, Pornkamol Unrean, and Friedrich Srienc. Minimal Escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses. *Appl. Environ. Microbiol.*, 74(12):363443, June 2008. doi: 10.1128/AEM.02708-07.
- [15] Pornkamol Unrean, Cong T Trinh, and Friedrich Srienc. Rational design and construction of an efficient E. coli for production of diapolycopendioic acid. *Metab. Eng.*, 12(2):11222, March 2010. doi: 10.1016/j.ymben.2009.11.002.
- [16] Vicente Acuña, Flavio Chierichetti, Vincent Lacroix, Alberto Marchetti-Spaccamela, Marie-France Sagot, and Leen Stougie. Modes and cuts in metabolic networks: Complexity and algorithms. *Biosystems*, 95:51–60, 2009. doi: 10.1016/j.biosystems.2008.06.015.
- [17] Steffen Klamt and Jörg Stelling. Combinatorial Complexity of Pathway Analysis in Metabolic Networks. *Molecular Biology Reports*, 29:233–236, 2002. doi: 10.1023/A:1020390132244.
- [18] Luis F De Figueiredo, Adam Podhorski, Angel Rubio, Christoph Kaleta, John E Beasley, Stefan Schuster, and Francisco J Planes. Computing the Shortest Elementary Flux Modes in Genome-Scale Metabolic Networks. *Bioinformatics*, 25:3158–3165, 2009. doi: 10.1093/bioinformatics.

- [19] Christoph Kaleta, Luis F De Figueiredo, Jörn Behre, and Stefan Schuster. Efmevolver: computing elementary flux modes in genome-scale metabolic networks. In *Lecture Notes in Informatics (LNI) P-157 - Proceedings of the German Conference on Bioinformatics*, pages 179–190, Bonn, 2009. Gesellschaft für Informatik.
- [20] Daniel Machado, Zita Soons, Kiran Raosaheb Patil, Eugénio C. Ferreira, and Isabel Rocha. Random sampling of elementary flux modes in large-scale metabolic networks. *Bioinformatics*, 28:i515–i521, 2012. doi: 10.1093/bioinformatics.
- [21] Marco Terzer. *Large scale methods to enumerate extreme rays and elementary modes*. PhD thesis, ETH Zurich, Zurich, 2009.
- [22] Marco Terzer and Jörg Stelling. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, 24:2229–2235, 2008. doi: 10.1093/bioinformatics.
- [23] Cong T. Trinh and R. Adam Thompson. Elementary Mode Analysis: A Useful Metabolic Pathway Analysis Tool for Reprogramming Microbial Metabolic Pathways. In Xiaoyuan Wang, Jian Chen, and Peter Quinn, editors, *Reprogramming Microbial Metabolic Pathways*, volume 64, pages 21–42. Springer Netherlands, Dordrecht, 2012. ISBN 978-94-007-5054-8, 978-94-007-5055-5. URL <http://www.springerlink.com/content/1j5n826254v14022/fulltext.html>.
- [24] Kristopher A. Hunt, James P. Folsom, Reed L. Taffs, and Ross P. Carlson. Complete enumeration of elementary flux modes through scalable demand-based subnetwork definition. *Bioinformatics*, page btu021, 2014. doi: 10.1093/bioinformatics.
- [25] Jean-Marc Schwartz and Minoru Kanehisa. Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. *BMC Bioinformatics*, 7: 1–20, 2006. ISSN 1471-2105. doi: 10.1186/1471-2105-7-186. URL <http://www.springerlink.com/content/142w18125521q118/abstract/>.
- [26] Christian Jungreuthmayer, David E. Ruckerbauer, and Jürgen Zanghellini. regEfm-tool: Speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic. *Biosystems*, 113:37–39, 2013. doi: 10.1016/j.biosystems.2013.04.002.
- [27] Anne Kümmel, Sven Panke, and Matthias Heinemann. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Molecular Systems Biology*, 2, 2006. doi: 10.1038/msb4100074.
- [28] Robert A. Alberty. *Thermodynamics of Biochemical Reactions*. John Wiley & Sons, Inc., New Jersey, 2003. ISBN 978-0-471-22851-6.
- [29] Stefan J. Jol, Anne Kümmel, Marco Terzer, Jörg Stelling, and Matthias Heinemann. System-Level Insights into Yeast Metabolism by Thermodynamic Analysis of Elementary Flux Modes. *PLoS Comput Biol*, 8:e1002415, 2012. doi: 10.1371/journal.pcbi.1002415.
- [30] Bryson D. Bennett, Elizabeth H. Kimball, Melissa Gao, Robin Osterhout, Stephen J. Van Dien, and Joshua D. Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology*, 5:593–599, 2009. doi: 10.1038/nchembio.186.

Bibliography

- [31] Robert B. Helling. Pathway Choice in Glutamate Synthesis in *Escherichia coli*. *Journal of Bacteriology*, 180:4571–4575, 1998.
- [32] R. B. Helling. Why does *Escherichia coli* have two primary pathways for synthesis of glutamate? *Journal of Bacteriology*, 176:4664–4668, 1994.
- [33] Dimitris Bertsimas and John N Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, February 1997.
- [34] Ines Thiele and Bernhard Ø Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, 5(1):93121, January 2010. doi: 10.1038/nprot.2009.203.
- [35] Axel von Kamp and Stefan Schuster. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, 22(15):19301, August 2006. doi: 10.1093/bioinformatics.
- [36] Komei Fukuda and Alain Prodon. Double description method revisited. In Michel Deza, Reinhardt Euler, and Ioannis Manoussakis, editors, *Combinatorics and Computer Science*, volume 1120 of *Lecture Notes in Computer Science*, pages 91–111. Springer Berlin, 1996. ISBN 978-3-540-61576-7.
- [37] Jon Pey and Francisco J Planes. Direct calculation of Elementary Flux Modes satisfying several biological constraints in genome-scale metabolic networks. *Bioinformatics*, April 2014. doi: 10.1093/bioinformatics.
- [38] Kuhn Ip, Caroline Colijn, and Desmond S Lun. Analysis of complex metabolic behavior through pathway decomposition. *BMC Syst Biol*, 5:91, 2011. doi: 10.1186/1752-0509-5-91.
- [39] Nathan D Price, Iman Famili, Daniel A Beard, and Bernhard Ø Palsson. Extreme pathways and Kirchhoff’s second law. *Biophys. J.*, 83(5):287982, November 2002. doi: 10.1016/S0006-3495(02)75297-1.
- [40] Daniel A Beard, Eric Babson, Edward Curtis, and Hong Qian. Thermodynamic constraints for biochemical networks. *J. Theor. Biol.*, 228(3):32733, June 2004. doi: 10.1016/j.jtbi.2004.01.008.
- [41] Jan Schellenberger, Nathan E Lewis, and Bernhard Ø Palsson. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J.*, 100(3):54453, February 2011. doi: 10.1016/j.bpj.2010.12.3707.
- [42] Elad Noor, Nathan E Lewis, and Ron Milo. A proof for loop-law constraints in stoichiometric metabolic networks. *BMC Syst Biol*, 6:140, 2012. doi: 10.1186/1752-0509-6-140.
- [43] Avi Flamholz, Elad Noor, Arren Bar-Even, and Ron Milo. eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Res.*, 40:D770–5, January 2012. doi: 10.1093/nar.
- [44] Hugh G. Nimmo. *The Tricarboxylic Acid Cycle and Anaplerotic Reactions*, volume 1, chapter 14, page 156–169. American Society for Microbiology, Washington D.C., 1987.

- [45] D Clark and J E Cronan. Escherichia coli mutants with altered control of alcohol dehydrogenase and nitrate reductase. *J. Bacteriol.*, 141(1):17783, January 1980.
- [46] Elad Noor, Hulda S. Haraldsdóttir, Ron Milo, and Ronan M. T. Fleming. Consistent estimation of gibbs energy using component contributions. *PLoS Comput Biol*, 9:e1003098, July 2013. doi: 10.1371/journal.pcbi.1003098. URL <http://dx.doi.org/10.1371/journal.pcbi.1003098>. 00002.
- [47] Stefan J. Jol, Anne Kümmel, Vassily Hatzimanikatis, Daniel A. Beard, and Matthias Heinemann. Thermodynamic calculations for biochemical transport and reaction processes in metabolic networks. *Biophysical Journal*, 99:3139–3144, November 2010. ISSN 0006-3495. doi: 10.1016/j.bpj.2010.09.043. URL <http://www.cell.com/article/S0006349510011938/abstract>. 00006 PMID: 21081060.
- [48] Tobias Österlund, Intawat Nookaew, and Jens Nielsen. Fifteen years of large scale metabolic modeling of yeast: Developments and impacts. *Biotechnology Advances*, 30:979–988, 2012. doi: 10.1016/j.biotechadv.2011.07.021.
- [49] Erwin P. Gianchandani, Arvind K. Chavali, and Jason A. Papin. The application of flux balance analysis in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2:372–382, 2010. doi: 10.1002/wsbm.60.
- [50] Matthew A. Oberhardt, Bernhard Ø Palsson, and Jason A. Papin. Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology*, 5, 2009. doi: 10.1038/msb.2009.77.
- [51] Nathan E. Lewis, Harish Nagarajan, and Bernhard Ø. Palsson. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10:291–305, 2012. doi: 10.1038/nrmicro2737.
- [52] Stefan Schuster, David A. Fell, and Thomas Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotech*, 18:326–332, 2000. doi: 10.1038/73786.
- [53] Steven M Kelk, Brett G Olivier, Leen Stougie, and Frank J Bruggeman. Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Sci. Rep.*, 2:580; DOI:10.1038/srep00580, 2012. doi: 10.1038/srep00580.
- [54] M. W. Covert and Bernhard Ø Palsson. Transcriptional Regulation in Constraints-based Metabolic Models of Escherichia coli. *Journal of Biological Chemistry*, 277: 28058–28064, 2002. doi: 10.1074/jbc.M201691200.
- [55] Adam M. Feist, Christopher S. Henry, Jennifer L. Reed, Markus Krummenacker, Andrew R. Joyce, Peter D. Karp, Linda J. Broadbelt, Vassily Hatzimanikatis, and Bernhard Ø Palsson. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3, 2007. doi: 10.1038/msb4100155.
- [56] Francisco Llaneras and Jesús Picó. Which Metabolic Pathways Generate and Characterize the Flux Space? A Comparison among Elementary Modes, Extreme Pathways and Minimal Generators. *Journal of Biomedicine and Biotechnology*, 2010: 1–14, 2010. doi: 10.1155/2010.

- [57] Matthias P Gerstl, Jungreuthmayer Christian, and Jürgen Zanghellini. tEFMA: computing thermodynamically feasible elementary flux modes in metabolic networks. *Bioinformatics*, 2015. doi: 10.1093/bioinformatics/btv111.
- [58] Jennifer L. Reed, Thuy D. Vo, Christophe H. Schilling, and Bernhard Ø. Palsson. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biology*, 4:R54, 2003. doi: 10.1186/gb-2003-4-9-r54.
- [59] J. S. Edwards and B. Ø. Palsson. The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, 97:5528–5533, 2000. doi: 10.1073/pnas.97.10.5528.
- [60] Christopher S. Henry, Linda J. Broadbelt, and Vassily Hatzimanikatis. Thermodynamics-Based Metabolic Flux Analysis. *Biophysical Journal*, 92:1792–1805, 2007. doi: 10.1529/biophysj.106.093138.
- [61] Alberto Rezola, Jon Pey, Luis F. de Figueiredo, Adam Podhorski, Stefan Schuster, Angel Rubio, and Francisco J. Planes. Selection of human tissue-specific elementary flux modes using gene expression data. *Bioinformatics*, 29:2009–2016, August 2013. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btt328. URL <http://bioinformatics.oxfordjournals.org/content/29/16/2009>. 00005 PMID: 23742984.
- [62] Joshua J. Hamilton, Vivek Dwivedi, and Jennifer L. Reed. Quantitative Assessment of Thermodynamic Constraints on the Solution Space of Genome-Scale Metabolic Models. *Biophysical Journal*, 105:512–522, 2013. doi: 10.1016/j.bpj.2013.06.011.
- [63] Andreas Hoppe, Sabrina Hoffmann, and Hermann-Georg Holzhütter. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Systems Biology*, 1:23, 2007. doi: 10.1186/1752-0509-1-23.
- [64] Jean-Marc Schwartz and Peter Neal Taylor. *In silico* prediction of elementary mode fluxes. In *Proceedings of the 2nd International Work-Conference on Bioinformatics and Biomedical Engineering*, 2014.
- [65] Mehmet A. Orman, Ioannis P. Androulakis, Francois Berthiaume, and Marianthi G. Ierapetritou. Metabolic network analysis of perfused livers under fed and fasted states: Incorporating thermodynamic and futile-cycle-associated regulatory constraints. *Journal of Theoretical Biology*, 293:101–110, 2012. doi: 10.1016/j.jtbi.2011.10.019.
- [66] Hong Yang, Charles M. Roth, and Marianthi G. Ierapetritou. Analysis of Amino Acid Supplementation Effects on Hepatocyte Cultures Using Flux Balance Analysis. *OMICS: A Journal of Integrative Biology*, 15:449–460, 2011. doi: 10.1089/omi.2010.0070.
- [67] Brett A Boghigian, Hai Shi, Kyongbum Lee, and Blaine A Pfeifer. Utilizing elementary mode analysis, pathway thermodynamics, and a genetic algorithm for metabolic flux determination and optimal metabolic network design. *BMC Systems Biology*, 4:49, 2010. doi: 10.1186/1752-0509-4-49.

- [68] Vidya V. Iyer, Hong Yang, Marianthi G. Ierapetritou, and Charles M. Roth. Effects of glucose and insulin on HepG2-C3A cell metabolism. *Biotechnology and Bioengineering*, 107:347–356, 2010. doi: 10.1002/bit.22799.
- [69] Nobuyoshi Ishii, Kenji Nakahigashi, Tomoya Baba, Martin Robert, Tomoyoshi Soga, Akio Kanai, Takashi Hirasawa, Miki Naba, Kenta Hirai, Aminul Hoque, Pei Yee Ho, Yuji Kakazu, Kaori Sugawara, Saori Igarashi, Satoshi Harada, Takeshi Masuda, Naoyuki Sugiyama, Takashi Togashi, Miki Hasegawa, Yuki Takai, Katsuyuki Yugi, Kazuharu Arakawa, Nayuta Iwata, Yoshihiro Toya, Yoichi Nakayama, Takaaki Nishioka, Kazuyuki Shimizu, Hirotada Mori, and Masaru Tomita. Multiple High-Throughput Analyses Monitor the Response of *E. coli* to Perturbations. *Science*, 316:593–597, 2007. doi: 10.1126/science.1132067.
- [70] Michael E Smoot, Keiichiro Ono, Johannes Ruscheinski, Peng-Liang Wang, and Trey Ideker. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, 27:431–2, February 2011. doi: 10.1093/bioinformatics.
- [71] Steffen Klamt and Jörg Stelling. *Stoichiometric and constraint-based modeling*, pages 73–96. MIT Press (Cambridge / MA), 2006. ISBN 9780262195485.

List of Publications

M. P. Gerstl, D. E. Ruckerbauer, D. Mattanovich, C. Jungreuthmayer, and J. Zanghellini, “Metabolomics integrated elementary flux mode analysis in large metabolic networks,” *Sci. Rep.*, 5:8930, 2015. doi: 10.1038/srep08930.

M. P. Gerstl, C. Jungreuthmayer, and J. Zanghellini, “tEFMA: computing thermodynamically feasible elementary flux modes in metabolic networks.,” *Bioinformatics*, February 2015. doi: 10.1093/bioinformatics.

M. Schosserer, N. Minois, T. B. Angerer, M. Amring, H. Dellago, E. Harreither, A. Calle-Perez, A. Pircher, M. P. Gerstl, S. Pfeifenberger, C. Brandl, M. Sonntagbauer, A. Kriegner, A. Linder, A. Weinhäusel, T. Mohr, M. Steiger, D. Mattanovich, M. Rinnerthaler, T. Karl, S. Sharma, K.-D. Entian, M. Kos, M. Breitenbach, I. B. H. Wilson, N. Polacek, R. Grillari-Voglauer, L. Breitenbach-Koller, and J. Grillari, “Methylation of ribosomal RNA by NSUN5 is a conserved mechanism modulating organismal lifespan.,” *Nat Commun*, vol. 6, no. 6158, 2015.

L. Fliedl, M. Wieser, G. Manhart, M. P. Gerstl, A. Khan, J. Grillari, and R. Grillari-Voglauer, “Controversial role of gamma-glutamyl transferase activity in cisplatin nephrotoxicity.,” *ALTEX*, vol. 31, no. 3, pp. 269–78, 2014.

M. P. Gerstl, M. Hackl, A. B. Graf, N. Borth, and J. Grillari, “Prediction of transcribed PIWI-interacting RNAs from CHO RNAseq data.,” *J. Biotechnol.*, vol. 166, pp. 51–7, June 2013.

Matthias P. Gerstl

- since 2013 PhD student at the group of Jürgen Zanghellini at ACIB under the supervision of Diethard Mattanovich
- 2012 Master Thesis under the supervision of Johannes Grillari at the University of Natural Resources and Applied Life Sciences Vienna (BOKU)
- 2011 – 2012 Technician in the lab of Regina Grillari at the Austrian Centre of Industrial Biotechnology (ACIB)
- 2010 – 2012 Master studies of Bioinformatics at FH Campus Wien degree: Master of Science in Engineering
- 2007 – 2010 Bachelor studies of Bioengineering at FH Campus Wien degree: Bachelor of Science in Engineering
- 1999 – 2010 Police officer in Vienna
- 1998 – 1999 Military service
- 1994 – 1998 Bundesoberstufenrealgymnasium St. Pölten, degree: Matura
- 1990 – 1994 Hauptschule in Hürm
- 1986 – 1990 Volksschule in Hürm