3
Toward Genome-Scale Metabolic Pathway Analysis

Jürgen Zanghellini, Matthias P. Gerstl, Michael Hanscho, Govind Nair, Georg Regensburger, Stefan Müller, and Christian Jungreuthmayer

3.1
Introduction

Constraint-based methods (CBMs) are extensively used to study cellular metabolism. These methods mainly rely on the stoichiometry of the biochemical reactions that form a metabolic network – data that can be fairly reliably obtained from annotated genome sequences and metabolic pathway databases. On the basis of this information, genome-scale metabolic networks can be reconstructed [1]. Mathematically, the topology of this network is captured by the (total) stoichiometric matrix that systematically collects the (signed) stoichiometric coefficients of all metabolites in the participating reactions. It is common practice to partition the set of metabolites into external and internal metabolites by defining a systems boundary. This systems boundary may correspond to a physical boundary, like the cell envelope, or to a virtual boundary. Reactions that connect external metabolites with internal ones are known as *exchange reactions* and account for the communication of a network with its environment. Of particular interest is the internal stoichiometric matrix $\boldsymbol{S} \in \mathbb{R}^{m \times r}$ that consists of the (signed) stoichiometric coefficients of the *m internal* metabolites in the *r* participating reactions (including exchange reactions).

At steady state, owing to mass conservation and thermodynamic feasibility, the metabolic flux vector $\boldsymbol{v} \in \mathbb{R}^r$ fulfills the equations/inequalities

$$\boldsymbol{S}\boldsymbol{v} = \boldsymbol{0}, \tag{3.1a}$$

$$v_i \geq 0 \quad \text{for } i \in I_{\text{irrev}}. \tag{3.1b}$$

The latter inequalities guarantee that all reactions i in the set of irreversible reactions I_{irrev} carry a (steady-state) flux v_i in the thermodynamically feasible forward direction. Typically, as $r > m$, the equations and inequalities (3.1a) and (3.1b) define a set of solutions, rather than a single solution. Characterizing the set of feasible solutions and finding biologically relevant solutions is at the heart of all CBMs.

Elementary flux mode analysis (EFMA) has emerged as a powerful tool in the family of CBMs [2, 3]. The particular power of an EFMA is based on the ability to

Industrial Biotechnology: Microorganisms, First Edition. Edited by Christoph Wittmann and James C. Liao.

unbiasedly decompose a metabolic network into irreducible functional building blocks, called elementary flux modes (EFMs). Three conditions uniquely define the set of EFMs in a metabolic network. An EFM $\boldsymbol{e} \in \mathbb{R}^r$ (i) operates at steady state, $\boldsymbol{Se} = \boldsymbol{0}$, (ii) uses all irreversible reactions in the appropriate direction, $e_i \geq 0$ for $i \in I_{\text{irrev}}$, and (iii) involves a minimal set of active reactions [4]. The first two features are obvious requirements from the equations and inequalities (3.1a) and (3.1b) and apply to all feasible flux distributions, while the characteristic property of an EFM is its support minimality. Recall that the support of a vector is the set of nonzero components. Support minimality of an EFM means that there is no feasible flux distribution having a smaller support, that is, less active reactions. Similar to basis vectors of linear subspaces, EFMs can be seen as generating vectors of a metabolic network as every feasible flux distribution can be represented as a non-negative linear combination [1)]

$$\boldsymbol{v} = \sum_{i=1}^{n} \lambda_i \boldsymbol{e}_i \quad \text{with } \lambda_i \in \mathbb{R} \text{ and } \lambda_i \geq 0, \tag{3.2}$$

where n denotes the number of EFMs in the network. This property allows for numerous applications in basic science and biotechnology. In fact, EFMA has been identified as a useful tool for metabolic engineering [6].

Suppose we are interested in turning a host organism into a growth-coupled cell factory for the production of some (bio)chemical commodity. Is this even possible for a given host and product of interest? The answer is provided by EFMA [7]: growth-coupled (and even partially growth-coupled) production is feasible if and only if there exists at least one growth-coupled EFM. [2)] For an illustration see Figure 3.1, where we plotted the product yield as function of the biomass yield for each EFM in a growth and nongrowth coupled (toy) network. *Escherichia coli* for instance, is in principle capable of producing all central carbon metabolites in a growth-coupled manner even if additional maintenance requirements are taken into account [7]. Thus EFMA provides extremely powerful means of analysis.

Yet, the enumeration of EFMs is computationally hard, which significantly limits the applicability of EFMA. In fact, a complete EFMA is currently limited to medium-scale (metabolic) models. The computational challenge for larger networks is associated with two main problems: (i) the number of EFMs explodes with the size of the metabolic network [9]. This makes a full EFMA of large systems virtually impossible, as even a core metabolic model of *E. coli* may already have several hundred million EFMs [10]. (ii) Even if a full EFMA in large networks was possible, then gazillions of EFMs would need to be stored, processed and further analyzed to gain biological knowledge, which would by far exceed current computational capabilities. For instance, on the basis of a massively parallelized approach, the largest, full EFMA reported to date found

1) Strictly speaking, a biochemically meaningful (thermodynamically feasible) decomposition is a non-negative linear combination of EFMs without cancellations [5].
2) For simplicity, we only discuss the so-called homogeneous case, which is applicable to all systems of the form (3.1). If additionally inhomogeneous constraints, such as minimal maintenance requirements or maximal flux capacities, are considered, the mathematical analysis based on the concept of "elementary flux vectors" yields similar, but more complex decompositions [5, 7, 8].

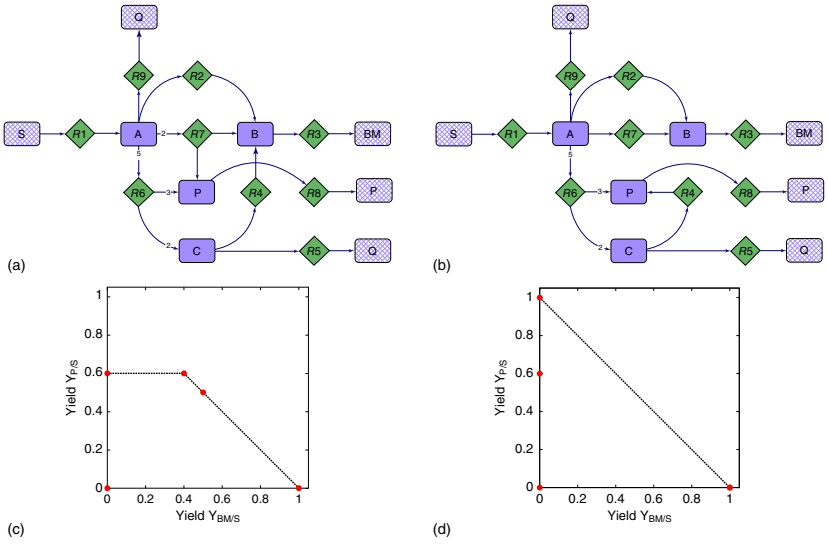

Figure 3.1 Growth- and non-growth-coupled toy metabolic networks (panels a and b, respectively) along with their associated phenotypic yield spaces (panels c and d, respectively). Both networks consist of nine irreversible reactions (diamonds, $R1-R9$), four internal metabolites (full rectangles), four external metabolites (checkered rectangles: BM, biomass; P, product of interest; Q by-product; S, substrate; note that the metabolite Q is the product of two reactions–$R5$ and $R9$) and five EFMs. In the phenotypic yield space EFMs are represented by full circles. Note that the point (1/0) in the phenotypic yield space of network B represents two EFMs with identical yields. The feasible yield space is bounded by the two axes and the dashed line. Growth-coupled production of P is achievable only in network A, but not in network B.

2 billion EFMs in a metabolic model of *Phaeodactylum tricornutum* with 318 reactions [11]. These data take up 1.2 TB of storage. Just reading all EFMs one by one takes up to an hour on our standard high-performance workstations. These numbers indicate that it is essential to find ways that identify and select "relevant" EFMs.

In the following, we briefly review currently available methods and then move on to present a novel method for the calculation of "short" EFMs.

One way to overcome the computational problems associated with an EFMA of large-scale networks is to compute only small subsets of EFMs. Subsets may be selected randomly [12, 13] or based on support information [14] or subject to additional constraints [15, 16]. Characterizing a subset of EFMs is also sufficient to address the feasibility problem of growth-coupled production outlined above (see Figure 3.1) as one has to find only one growth-coupled EFM. This is conveniently achieved by two methods developed by David and Bockmayr [17] and Pey and Planes [18], respectively. The latter method has the advantages that additional minimal yield requirements can be considered as well and that the method even remains feasible in genome-scale models. However, even if the analysis is restricted to a particular subset of EFMs (e.g., the subset of growth-coupled EFMs), their computation remains challenging. In fact, it has been shown that a single EFM can be computed in polynomial time, yet computing the complete subset is NP-complete [19].

One very fast EFM enumeration method uses the binary nullspace implementation [20, 21] of the double description (DD) method [22]. This method has proven particularly useful for the complete enumeration of EFMs in metabolic networks [22].

In the following, we will detail the DD method and then outline a new algorithm for the calculation of "short" EFMs that will make the DD method fit for the analysis of genome-scale metabolic models.

3.2 DD Method

We consider a metabolic network at steady state given by the equations and inequalities (3.1a) and (3.1b). After splitting every reversible reaction into two irreversible reactions (with opposite directions), we can assume that all reactions are irreversible. The equations and inequalities (3.1a) and (3.1b) define the so-called flux cone

$$F = \{\boldsymbol{v} \in \mathbb{R}^r \mid \boldsymbol{S}\boldsymbol{v} = \boldsymbol{0} \text{ and } \boldsymbol{v} \geq \boldsymbol{0}\}. \tag{3.3}$$

All thermodynamically feasible steady-state solutions, in particular all EFMs, are elements of this cone. Since we assume that all reactions are irreversible, the flux cone F is pointed and the EFMs coincide with its extreme rays (the "edges" of the flux cone) [23]. The DD method is widely used to enumerate these extreme rays.

The DD method relies on the fact that, due to the Minkowski–Weyl theorem, a polyhedral cone C can be described implicitly by inequalities (by the intersection

of half spaces),

$$C = C(\mathbf{A}) = \{\boldsymbol{\nu} \in \mathbb{R}^r \mid \mathbf{A}\boldsymbol{\nu} \geq \mathbf{0}\}, \tag{3.4}$$

or explicitly by generators,

$$C = \text{cone}(\mathbf{R}) = \{\boldsymbol{\nu} \in \mathbb{R}^r \mid \boldsymbol{\nu} = \mathbf{R}\lambda,\ \lambda \geq \mathbf{0}\}, \tag{3.5}$$

for some matrices $\mathbf{A}$ and $\mathbf{R}$. A pair $(\mathbf{A}, \mathbf{R})$ of corresponding matrices is called a *DD pair*. If the cone C is pointed and the columns of $\mathbf{R}$ are its extreme rays, then $(\mathbf{A}, \mathbf{R})$ is called a *minimal DD pair*.

The DD method iteratively processes the inequalities given by the rows of $\mathbf{A}$. Let $\mathbf{A}_k$ be the submatrix of $\mathbf{A}$ that contains the first k rows of $\mathbf{A}$ and suppose that $(\mathbf{A}_k, \mathbf{R}_k)$ is the corresponding minimal DD pair. The DD method considers the next row of $\mathbf{A}$ and constructs a new minimal DD pair $(\mathbf{A}_{k+1}, \mathbf{R}_{k+1})$.

In the following, we describe one iteration step of the DD method in more detail. (For simplicity, we omit the iteration index.) Given the minimal DD pair $(\mathbf{A}, \mathbf{R})$ and an additional inequality represented by a vector $\boldsymbol{a} \in \mathbb{R}^r$ (i.e., the next unprocessed row of the full matrix), we consider the matrix $\mathbf{A}' = \begin{pmatrix} \mathbf{A} \\ \boldsymbol{a}^t \end{pmatrix}$ defining the cone $C(\mathbf{A}') = C(\mathbf{A}) \cap C(\boldsymbol{a}^t)$. We determine the corresponding matrix $\mathbf{R}'$ as follows: The additional inequality divides the set of column indices J of $\mathbf{R}$ into three sets

$$J^+ = \{j \in J \mid \boldsymbol{a}^t \boldsymbol{r}_j > 0\}, \tag{3.6a}$$

$$J^0 = \{j \in J \mid \boldsymbol{a}^t \boldsymbol{r}_j = 0\}, \tag{3.6b}$$

$$J^- = \{j \in J \mid \boldsymbol{a}^t \boldsymbol{r}_j < 0\}. \tag{3.6c}$$

The first two sets contain the columns of $\mathbf{R}$ that lie in the intersection $C(\mathbf{A}) \cap C(\boldsymbol{a}^t)$, and the third set contains the columns of $\mathbf{R}$ that lie outside. We keep the extreme rays that fulfill the additional inequality and form positive linear combinations of (so-called adjacent) rays such that the new rays lie on the cutting hyperplane $\{\boldsymbol{\nu} \in \mathbb{R}^r \mid \boldsymbol{a}^t \boldsymbol{\nu} = 0\}$. Hence, we obtain a matrix $\mathbf{R}' \in \mathbb{R}^{r \times J'}$, where

$$J' = J^+ \cup J^0 \cup J^{\text{adj}}, \tag{3.7a}$$

$$J^{\text{adj}} = \{(j^+, j^-) \in J^+ \times J^- \mid \boldsymbol{r}^{j^+}, \boldsymbol{r}^{j^-} \text{ adjacent}\}, \tag{3.7b}$$

$$\boldsymbol{r}'_j = \boldsymbol{r}_j \text{ for } j \in J^+ \cup J^0, \tag{3.7c}$$

$$\boldsymbol{r}'_{(j^+, j^-)} = (\boldsymbol{a}^t \boldsymbol{r}_{j^+}) \boldsymbol{r}_{j^-} - (\boldsymbol{a}^t \boldsymbol{r}_{j^-}) \boldsymbol{r}_{j^+} \quad \text{for } (j^+, j^-) \in J^{\text{adj}}, \tag{3.7d}$$

that is, for adjacent rays $\boldsymbol{r}^{j^+}, \boldsymbol{r}^{j^-}$. Then, $(\mathbf{A}', \mathbf{R}')$ is a minimal DD pair that satisfies also the additional inequality.

A cone $C(\mathbf{A})$ as in (3.4) is defined by inequalities $\mathbf{A}\boldsymbol{\nu} \geq \mathbf{0}$. A representation of the flux cone F as in (3.3) by inequalities is given by $F = C(\mathbf{A})$ with

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{S} \\ -\boldsymbol{S} \\ \boldsymbol{I}_r \end{pmatrix}, \tag{3.8}$$

where $\boldsymbol{I}_r \in \mathbb{R}^{r \times r}$ denotes the identity matrix.

Of course, first we need to find an initial minimal DD pair. Following the null space approach [23, 24], we compute a basis of the kernel of the stoichiometric matrix $\boldsymbol{S}$. More specifically, we compute a column-reduced echelon form of the basis and (after a permutation of rows) obtain

$$\mathbf{R}_0 = \begin{pmatrix} \boldsymbol{I}_p \\ \boldsymbol{K} \end{pmatrix}, \tag{3.9}$$

where p is the dimension of the kernel of $\boldsymbol{S}$. By construction, the columns of $\mathbf{R}_0$ satisfy the inequalities coming from the stoichiometric matrix $\boldsymbol{S}$ and the first p non-negativity constraints. The remaining $r-p$ non-negativity constraints are processed consecutively. Explicitly, we consider

$$\mathbf{A}_k = \begin{pmatrix} \boldsymbol{S} \\ -\boldsymbol{S} \\ \boldsymbol{I}_{p+k} \ \mathbf{0} \end{pmatrix}, \quad k = 0, \ldots, r-p. \tag{3.10}$$

Clearly, the first $p+k$ rows of the corresponding matrices $\mathbf{R}_k$ are non-negative. This fact is the basis for the binary null space approach [23, 24] for which efficient implementations are available [20, 21]. Still, the DD suffers from the combinatorial explosion of intermediate extreme rays resulting from the combination of adjacent rays $\boldsymbol{r}^{j^+}, \boldsymbol{r}^{j^-}$. Thus, a full EFMA is only applicable to medium-scale metabolic networks.

We conclude with the following key observation: at each iteration step $k = 1, \cdots, r-p$, we either keep an extreme ray or compute a positive linear combination of two adjacent rays. In any case, *the number of positive entries within the first $p+k$ components of the resulting rays either remains the same or increases.* In particular, the parents of an extreme ray, that is, the intermediate rays that are combined, never have a larger support than their offspring. (Recall that the support of a vector is the set of nonzero components.) This observation can be exploited computationally, as explained below.

3.3 Calculating Short EFMs in Genome-Scale Metabolic Networks

Suppose we are interested in calculating all EFMs with a maximal cardinality $c_{\max}$ of their support. In this case, we can omit an intermediate extreme ray from the analysis as soon as the number of positive entries in the first $p+k$ components exceeds the maximal cardinality. By omitting intermediate extreme rays, the combinatorial explosion of extreme rays is curbed and the method becomes applicable to larger networks, without compromising the efficiency and speed of the method.

We implemented the maximum cardinality feature as an extension to the open source program *Efmtool* [25]. The software is freely available at the author's webpage [26]. The principal workflow of the algorithm is illustrated in Figure 3.2. This extended version was used throughout the remainder of the study and compared

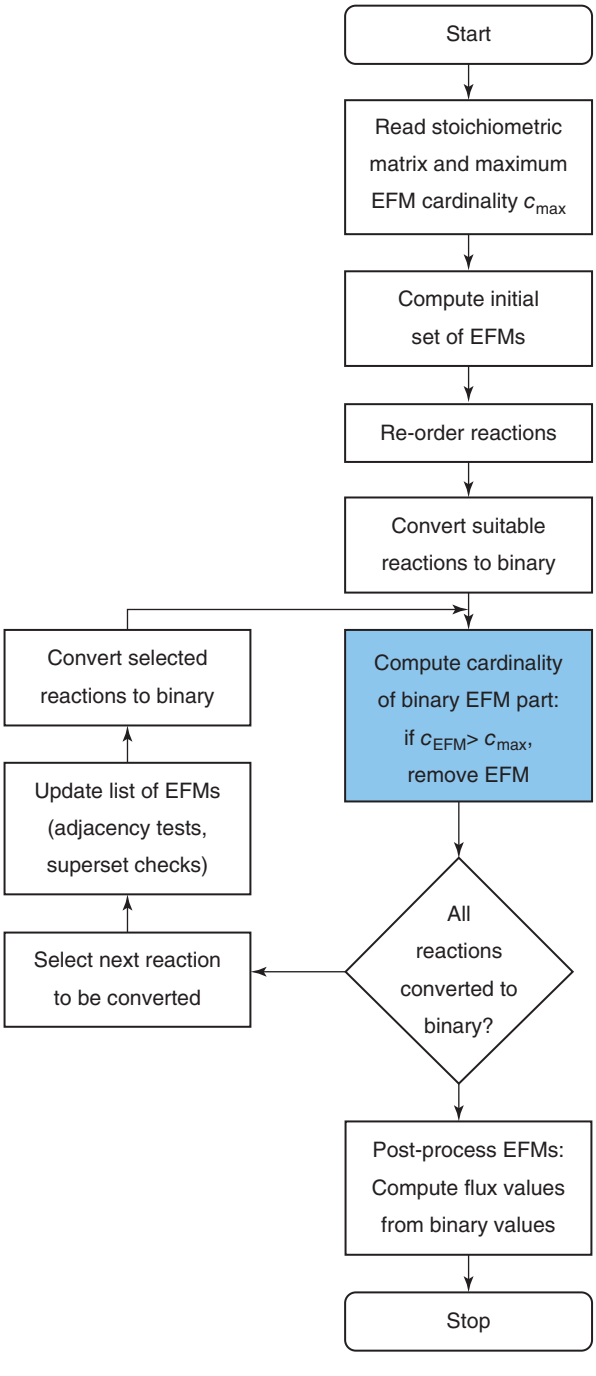

Figure 3.2 Computing the shortest EFMs with the DD method. The box with blue background color highlights the main extension of the maximum cardinality feature over common implementations of the DD method.

to a standard approach of calculating the k-shortest EFMs using a mixed-integer linear program (MILP) [16].

The MILP-based approach was realized by a hybrid code. Performance uncritical parts of the program (e.g., setting up the linear program, input/output handling) were implemented in Perl. C-libraries of the IBM ILOG CPLEX Optimization Studio were used for the computationally expensive task of solving the linear program. CPLEX is a commercial optimization software package for which free academic licenses are available.

We calculated all EFMs up to a cardinality of $c_{\max}$ in six metabolic models of various sizes ranging from small-scale to genome-scale. The key properties of these models are listed in Table 3.1. All networks were compressed [23] before the EFMA. We used a computer with two Intel Xeon CPUs (each with six cores, 2.67 GHz) running Ubuntu 14.04. Both programs were allowed to use up to eight parallel threads during the execution.

Figure 3.3 illustrates the runtimes (primary y-axis) for both methods, and the number of enumerated EFMs (secondary y-axis) as function of the EFMs' length for the metabolic networks listed in Table 3.1. Note that both y-axes were scaled logarithmically and that the depicted runtimes did not include pre- or postprocessing steps such as network compression and decompression, which took up to an hour in genome-scale metabolic networks.

Naturally, we found that both, runtimes and the number of EFMs increased with increasing cardinality. Moreover, execution times for both approaches strongly corresponded with the number of EFMs. However, for a fixed number of computed EFMs the execution time varied widely across the models. For instance, in the liver cancer model it took almost 15 h to compute all 307 444 EFMs up to a cardinality of 7, while it took less than 2 h to compute all 990 797 EFMs up to a cardinality of 100 in the *P. tricornutum* model. This indicates not only that the relation between runtime, EFM cardinality, and number of EFM is nontrivial but also that the number of EFMs had a larger effect on the runtime than the cardinalities of the EFMs.

Table 3.1 Main topological properties of the six metabolic models used in this study.

Model	Internal metabolites	External metabolites	Total reactions	Irreversible reactions	EFMs
E. coli core I	69	17	82	35	5 011
E. coli core II	53	15	71	20	429 276
E. coli core III	72	21	95	59	226 269 020
P. tricornutum	327	11	318	103	1 934 729 551
Blattibacteriacae cuenoti Bge	306	43	350	45	Unknown
Liver cancer	1 754	195	2 423	454	Unknown

Note however, that the number of metabolites and the number of reactions has no known functional correlation to the number of EFMs. In fact, previous attempts to find such a correlation failed [27] and the question remains open.

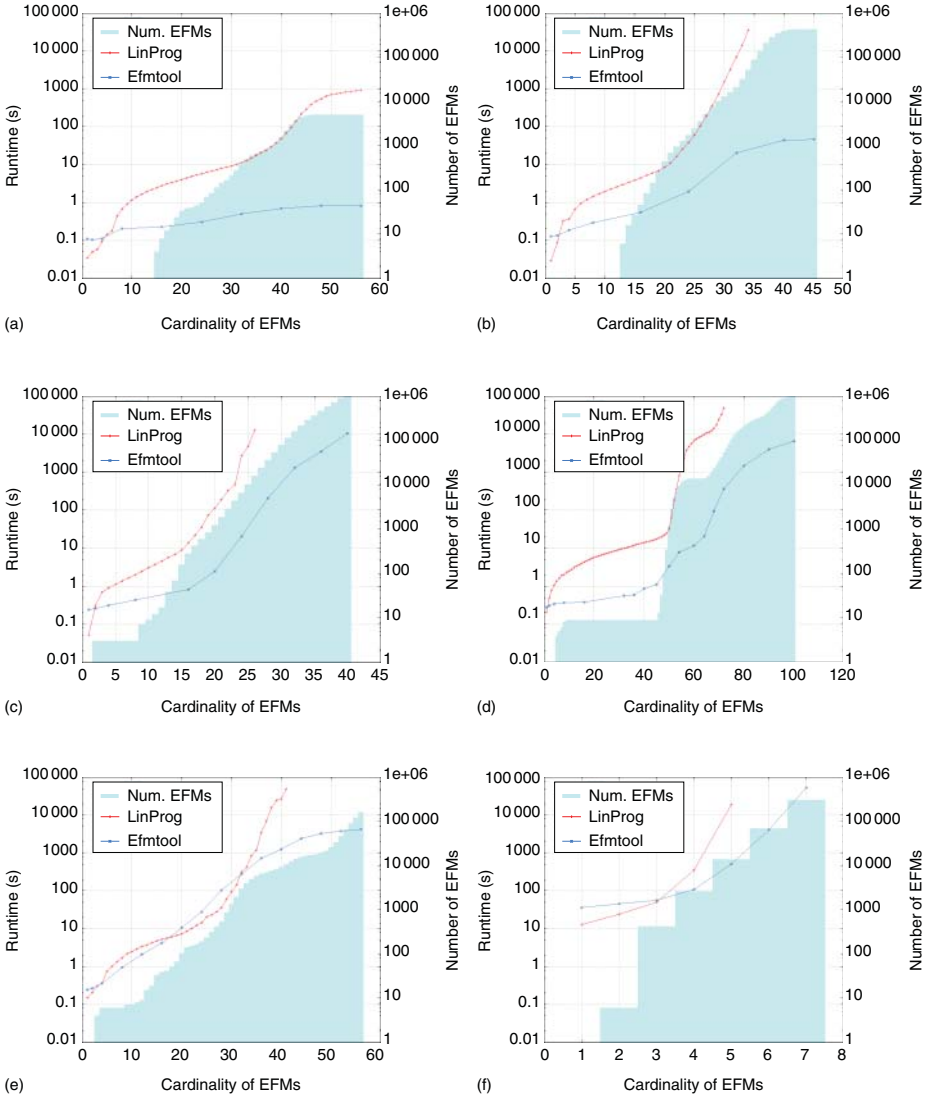

Figure 3.3 Comparison of execution times. Runtime comparison between maximum cardinality EFMA and the MILP based approach in various models (see Table 3.1). (a) *E. coli* core I [28], (b) *E. coli* core II, (c) *E. coli* core III, (d) *P. tricornutum* [11], (e) *Blattibacteriacae cuenoti* Bge [29], and (f) liver cancer [30].

Figure 3.3 shows that for low cardinalities and, therefore, for very low runtimes, maximum cardinality EFMA and the MILP approach exhibit a similar performance. However, Figure 3.3 also clearly demonstrates that for larger cardinalities, maximum cardinality EFMA is much faster than the MILP approach. Importantly, the performance gain grows with increasing EFM cardinalities.

Moreover, in the models (b) to (f) the runtime time requirements for the MILP approach became quickly prohibitively large, while maximum cardinality EFMA successfully continued to output EFMs.

3.4 Conclusions

Here we introduced a maximum cardinality EFMA that efficiently enumerates short EFMs in (large) metabolic networks. Maximum cardinality EFMA is based on the DD method and exploits the fact that new (intermediate) EFMs are always constructed by combining two appropriately selected shorter EFMs. Thus the cardinality of any new (intermediate) EFM is always larger than the cardinality of both of its parent EFMs. An EFM exceeding the user-specified cardinality threshold can therefore safely be removed from further analysis.

Maximum cardinality EFMA was implemented as a minimal invasive extension to *Efmtool* and fully utilizes the computational advantages of the binary null space implementation of the DD method. Both factors, the maximum cardinality EFMA strategy and the binary implementation, result in a major speedup that outperforms other, MILP-based approaches by orders of magnitude.

Maximum cardinality EFMA requires a specific, user-defined maximum cardinality threshold, $c_{\max}$, as input. However, the number of EFMs that will be calculated for the specified threshold is not known *a priori*. In contrast to MILP-based approaches, maximum cardinality EFMA is therefore not able to terminate after it has found a predefined number of EFMs. In practice, this drawback had little effect on the usability of our program. Beginning with low values of $c_{\max}$, we repeatedly executed maximum cardinality EFMA while increasing $c_{\max}$ until the desired number of EFMs were calculated. Even if we cumulated all execution times, still maximum cardinality EFMA outperformed the MILP approach in calculating the 10 000 shortest EFMs in all tested models. Conversely, for a given runtime, many more EFMs can be computed by maximum cardinality EFMA than the MILP approach, which in turn results in larger networks that can be analyzed and studied more thoroughly with maximum cardinality EFMA.

In principle, maximum cardinality EFMA is able to fully enumerate all EFMs of a network, as for $c_{\max} \to r$ the ordinary DD method is retrieved. The DD method suffers from a combinatorial explosion in the number of (intermediate) EFMs. The same problem applies to maximum cardinality EFMA. Thus $c_{\max}$ has to be sufficiently small in order to remain computationally feasible. Nevertheless, our analysis revealed that in all investigated models maximum cardinality EFMA was able to calculate the 70 000 shortest EFMs within about 1 h on standard workstation computers. Comparable runtimes are not achievable with currently available alternative methods.

Note that the maximum cardinality feature of the DD method as observed above is a very general feature that can be exploited for the integration of -omics data. Suppose an (intermediate) EFM contains two active reactions that together are infeasible. For instance, in *E. coli* the glyoxylate shunt is downregulated under

high-glucose conditions [31]. In this case, the glucose uptake and the glyoxylate shunt are not simultaneously active. Any EFM that contains both reactions is therefore infeasible. On the basis of our observation, these EFMs can be discarded without impact on the further analysis. In fact, recent methodological developments exploited this fact and identified thermodynamically feasible [32, 33] or transcriptionally regulated [34, 35] EFMs. However, the underlying principle has never been clearly formulated. Thus by extensively including -omics data based on the principle introduced here, an EFMA of a genome-scale metabolic model seems feasible.

Acknowledgments

CJ, MPG, MH, GMN, and JZ acknowledge support by the Federal Ministry of Science, Research and Economy (BMWFW), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol, the Government of Lower Austria and ZIT – Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG.

SM and GR were supported by the Austrian Science Fund (FWF), project P28406.

References

1. Thiele, I. and Palsson, B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.*, **5** (1), 93 – 121, doi: 10.1038/nprot.2009.203.
2. Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A., and Palsson, B.O. (2003) Metabolic pathways in the post-genome era. *Trends Biochem. Sci.*, **28** (5), 250 – 258.
3. Zanghellini, J., Ruckerbauer, D.E., Hanscho, M., and Jungreuthmayer, C. (2013) Elementary flux modes in a nutshell: properties, calculation and applications. *Biotechnol. J.*, **8** (9), 1009 – 1016.
4. Schuster, S., Fell, D.A., and Dandekar, T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.*, **18** (3), 326 – 332, doi: 10.1038/73786.
5. Müller, S. and Regensburger, G. (2016) Elementary vectors and conformal sums in polyhedral geometry and applications in metabolic pathway analysis, *Front. Genet.*, **7**, 90, doi: 10.3389/fgene.2016.00090.
6. Trinh, C.T. and Thompson, R.A. (2012) Elementary mode analysis: a useful metabolic pathway analysis tool for reprograming microbial metabolic pathways, in *Reprogramming Microbial Metabolic Pathways*, Vol. **64** (eds X. Wang, J. Chen, and P. Quinn), Springer Netherlands, Dordrecht, pp. 21 – 42.
7. Klamt, S. and Mahadevan, R. (2015) On the feasibility of growth-coupled product synthesis in microbial strains. *Metab. Eng.*, **30**, 166 – 178, doi: 10.1016/j.ymben.2015.05.006.
8. Urbanczik, R. (2007) Enumerating constrained elementary flux vectors of metabolic networks. *IET Syst. Biol.*, **1** (5), 274 – 279.
9. Klamt, S. and Stelling, J. (2002) Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.*, **29** (1-2), 233 – 236, doi: 10.1023/A:1020390132244.

10. Jungreuthmayer, C., Nair, G., Klamt, S., and Zanghellini, J. (2013) Comparison and improvement of algorithms for computing minimal cut sets. *BMC Bioinformatics*, **14** (1), 318, doi: 10.1186/1471-2105-14-318.
11. Hunt, K.A., Folsom, J.P., Taffs, R.L., and Carlson, R.P. (2014) Complete enumeration of elementary flux modes through scalable demand-based subnetwork definition. *Bioinformatics*, **30** (11), 1569–1578.
12. Machado, D., Soons, Z., Patil, K.R., Ferreira, E.C., and Rocha, I. (2012) Random sampling of elementary flux modes in large-scale metabolic networks. *Bioinformatics*, **28** (18), i515–i521, doi: 10.1093/bioinformatics/bts401.
13. Kaleta, C., De Figueiredo, L.F., Behre, J., and Schuster, S. (2009) EFMEvolver: computing elementary flux modes in genome-scale metabolic networks, in Proceedings *of the German Conference on Bioinformatics*, *Lecture Notes in Informatics (LNI) P-157*, Gesellschaft für Informatik, Bonn, pp. 179–190.
14. Pey, J., Villar, J.A., Tobalina, L., Rezola, A., García, J.M., Beasley, J.E., and Planes, F.J. (2015) TreeEFM: calculating elementary flux modes using linear optimization in a tree-based algorithm. *Bioinformatics*, **31** (6), 897–904.
15. Kelk, S.M., Olivier, B.G., Stougie, L., and Bruggeman, F.J. (2012) Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Sci. Rep.*, **2**, 580, doi: 10.1038/srep00580.
16. De Figueiredo, L.F., Podhorski, A., Rubio, A., Kaleta, C., Beasley, J.E., Schuster, S., and Planes, F.J. (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics*, **25** (23), 3158–3165, doi: 10.1093/bioinformatics/btp564.
17. David, L. and Bockmayr, A. (2014) Computing elementary flux modes involving a set of target reactions. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **11** (6), 1099–1107, doi: 10.1109/TCBB.2014.2343964.
18. Pey, J. and Planes, F.J. (2014) Direct calculation of elementary flux modes satisfying several biological constraints in genome-scale metabolic networks. *Bioinformatics*, **30** (15), 2197–2203, doi: 10.1093/bioinformatics/btu193.
19. Acuna, V., Chierichetti, F., Lacroix, V., Marchetti-Spaccamela, A., Sagot, M.F., and Stougie, L. (2009) Modes and cuts in metabolic networks: Complexity and algorithms. *Biosystems*, **95** (1), 51–60.
20. van Klinken, J.B. and van Dijk, K.W. (2016) FluxModeCalculator: an efficient tool for large-scale flux mode computation. *Bioinformatics*, **32** (8), 1265–1266, doi: 10.1093/bioinformatics/btv742.
21. Terzer, M. and Stelling, J. (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, **24** (19), 2229–2235, doi: 10.1093/bioinformatics/btn401.
22. Fukuda, K. and Prodon, A. (1996) Double description method revisited, in *Combinatorics and Computer Science*, *Lecture Notes in Computer Science*, Vol. **1120**, Springer-Verlag, Berlin Heidelberg, pp. 91–111.
23. Gagneur, J. and Klamt, S. (2004) Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, **5**, 175.
24. Wagner, C. (2004) Nullspace approach to determine the elementary modes of chemical reaction systems. *J. Phys. Chem. B*, **108** (7), 2425–2431, doi: 10.1021/jp034523f.
25. efmtool - elementary flux mode tool, http://www.csb.ethz.ch/tools/software/efmtool.html (accessed 22 June 2016).
26. Regulatory elementary flux mode tool, regEfmtool, http://www.biotec.boku.ac.at/en/arbeitsgruppenresearch-groups/research-group-mattanovich/staff/associated-group-metabolic-modelling/regefmtool/ (accessed 22 June 2016).
27. Yeung, M., Thiele, I., and Palsson, B. (2007) Estimation of the number of extreme pathways for metabolic networks. *BMC Bioinformatics*, **8** (1), 363, doi: 10.1186/1471-2105-8-363.
28. Trinh, C.T., Unrean, P., and Srienc, F. (2008) Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl. Environ. Microbiol.*, **74** (12), 3634–3643.
29. González-Domenech, C.M., Belda, E., Patiño-Navarrete, R., Moya, A., Peretó,

J., and Latorre, A. (2012) Metabolic stasis in an ancient symbiosis: genome-scale metabolic networks from two Blattabacterium cuenoti strains, primary endosymbionts of cockroaches. *BMC Microbiol.*, **12** (Suppl. 1), S5, doi: 10.1186/1471-2180-12-S1-S5.

30. Human Metabolic Atlas, INIT models of cancer cell, http://www.metabolicatlas.com/downloads/initcan (accessed 22 June 2016).
31. Kornberg, H.L. (1966) The role and control of the glyoxylate cycle in *Escherichia coli*. *Biochem. J.*, **99** (1), 1–11.
32. Gerstl, M.P., Ruckerbauer, D.E., Mattanovich, D., Jungreuthmayer, C., and Zanghellini, J. (2015) Metabolomics integrated elementary flux mode analysis in large metabolic networks. *Sci. Rep.*, **5**, 8930.
33. Gerstl, M.P., Jungreuthmayer, C., and Zanghellini, J. (2015) tEFMA: computing thermodynamically feasible elementary flux modes in metabolic networks. *Bioinformatics*, **31** (13), 2232–2234, doi: 10.1093/bioinformatics/btv111.
34. Jungreuthmayer, C., Ruckerbauer, D.E., Hanscho, M., and Zanghellini, J. (2015) Avoiding the enumeration of infeasible elementary flux modes by including transcriptional regulatory rules in the enumeration process saves computational costs. *PLoS ONE*, **10** (6), e0129840.
35. Jungreuthmayer, C., Ruckerbauer, D.E., and Zanghellini, J. (2013) regEfmtool: speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of a three-state logic. *Biosystems*, **113** (1), 37–39.