BIOINFORMATICS ORIGINAL PAPER

Vol. 30 no. 22 2014, pages 3232–3239
doi:10.1093/bioinformatics/btu529

Systems biology

Advance Access publication August 6, 2014

Estimating biological elementary flux modes that decompose a flux distribution by the minimal branching property

Siu Hung Joshua Chan[1,*], Christian Solem[1], Peter Ruhdal Jensen[1] and Ping Ji[2,*]

[1]Systems Biotechnology and Biorefining, National Food Institute, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark and [2]Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

Associate Editor: Igor Jurisica

ABSTRACT

Motivation: Elementary flux mode (EFM) is a useful tool in constraint-based modeling of metabolic networks. The property that every flux distribution can be decomposed as a weighted sum of EFMs allows certain applications of EFMs to studying flux distributions. The existence of biologically infeasible EFMs and the non-uniqueness of the decomposition, however, undermine the applicability of such methods. Efforts have been made to find biologically feasible EFMs by incorporating information from transcriptional regulation and thermodynamics. Yet, no attempt has been made to distinguish biologically feasible EFMs by considering their graphical properties. A previous study on the transcriptional regulation of metabolic genes found that distinct branches at a branch point metabolite usually belong to distinct metabolic pathways. This suggests an intuitive property of biologically feasible EFMs, i.e. minimal branching.

Results: We developed the concept of minimal branching EFM and derived the minimal branching decomposition (MBD) to decompose flux distributions. Testing in the core *Escherichia coli* metabolic network indicated that MBD can distinguish branches at branch points and greatly reduced the solution space in which the decomposition is often unique. An experimental flux distribution from a previous study on mouse cardiomyocyte was decomposed using MBD. Comparison with decomposition by a minimum number of EFMs showed that MBD found EFMs more consistent with established biological knowledge, which facilitates interpretation. Comparison of the methods applied to a complex flux distribution in *Lactococcus lactis* similarly showed the advantages of MBD. The minimal branching EFM concept underlying MBD should be useful in other applications.

Contact: sinhu@bio.dtu.dk or p.ji@polyu.edu.hk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on January 7, 2014; revised on July 11, 2014; accepted on July 31, 2014

1 INTRODUCTION

Metabolic networks consisting of biochemical reactions and compounds in a cell have proven to be an extremely useful *in silico* tool to study cellular metabolism (Edwards and Palsson, 1999; Oberhardt *et al*., 2009; Price *et al*., 2004). The structure of a metabolic network is reflected in the stoichiometric matrix **S**, and the state of the network is embodied in the flux distribution (FD) **v**, which is a vector containing all reaction rates in the network. Based on the well-established quasi-steady-state assumption, **v** can be constrained in a polyhedral cone ($\mathbf{Sv}=0, \quad \mathbf{v}_{ub} \geq \mathbf{v} \geq \mathbf{v}_{lb}$) and studied using a series of so-called constraint-based reconstruction and analysis (COBRA) methods (Lewis *et al*., 2012).

One of the approaches of theoretical importance and application value is the elementary flux mode (EFM) analysis (Schuster and Hilgetag, 1994). An EFM is a FD with a minimal number of active reactions that together satisfy the steady-state condition. It is thus a minimal metabolic pathway in a metabolic network. EFMs have been successfully applied to examine network structures, explore new metabolic pathways, suggest rational strain design, etc. (reviewed in Trinh *et al*., 2009; Zanghellini *et al*., 2013).

Another property of EFMs useful for studying FDs is that every FD **v** is a non-negative sum of EFMs:

$$\mathbf{v}=\sum_{k=1}^{K}\alpha_k\mathbf{e}_k \tag{1}$$

where $\mathbf{e}_1 \cdots \mathbf{e}_K$ are all EFMs of the network, and α_k is the corresponding weight coefficient. This decomposition of FDs from EFMs or extreme pathways (another type of pathways with a similar definition) has been investigated and applied to certain biological studies by choosing particular decompositions [reviewed in Chan and Ji (2011)]. The usefulness of these methods, however, is limited for two main reasons. First, the practical computation of the decomposition requires the set of all EFMs a priori, which is usually not available in large-scale networks owing to combinatorial explosion. This computational issue can be overcome in two ways, either by sampling EFMs (Machado *et al*., 2012; Tabe-Bordbar and Marashi, 2013) or by finding EFMs with particular properties to solve for the decomposition without requiring the entire set of EFMs (Chan and Ji, 2011; de Figueiredo *et al*., 2009; Ip *et al*., 2011; Pey and Planes, 2014; Rezola *et al*., 2011). The second limitation is the non-uniqueness of decomposition and its dependence on the choice of algorithms or optimization objective (Zanghellini *et al*., 2013). The ambiguity severely weakens the biological significance of decomposition. Poolman *et al*. (2004) has proposed a method to obtain unique decompositions by combining EFMs with variable weights into compound modes with unique weights. Another method for unique decompositions is the minimization of the sum of squares of weights (Schwartz and Kanehisa, 2005, 2006), but the resulting decomposition can be dependent on how EFMs are scaled.

*To whom correspondence should be addressed.

© The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

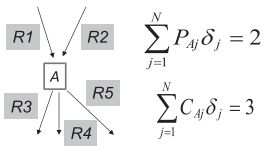

Fig. 1. The graphical meaning of NP_{ik}. A is a metabolite in the network. If EFM 1 contains all five reactions, then $NP_{A1} = 2 \times 3 = 6$, equal to the number of all locally different paths passing through A

The usefulness of EFMs can be enhanced by finding biologically feasible EFMs. Attempts have been made to incorporate transcriptional regulatory information to enumerate biologically feasible EFMs (Jungreuthmayer *et al.*, 2013) and to rule out thermodynamically infeasible EFMs (Jol *et al.*, 2012). These approaches integrating information other than stoichiometry did give promising results. The information, for instance, metabolomics data and transcriptomics data, however, is not always available. Surprisingly, all the proposed methods to find EFMs or to choose decomposition have not made full use of their graphical properties, which should be intuitively important for interpreting EFMs. An earlier study on the co-regulation of metabolic pathways found that transcriptional regulation tends to favor linear metabolic flow by co-expressing only distinct branches at metabolic branch points (Ihmels *et al.*, 2004). This means, in Figure 1 where 'A' is a hypothetical branch point metabolite, the genes for the linear pair of reactions 'R1' and 'R3', or the pair 'R2' and 'R4', are more likely to be co-expressed than 'R3', 'R4' and 'R5', which are distinct branches. When decomposing a FD containing 'R1' to 'R4', two EFMs containing pair 'R1'and 'R3' and pair 'R2' and 'R4' are thus a more preferable choice than one EFM containing all 'R1' to 'R4', as there are probably two distinct pathways. This suggests the rationale of minimizing branching pathways to choose EFMs that are more likely to be biologically feasible. This rationale is also expected to allow clearer interpretation of the roles of the chosen EFMs because they are expected to contain fewer distinct pathways. We believed that such EFMs can have wide applicability. As the first attempt based on this rationale, in this article, a new optimization objective called minimal branching decomposition (MBD) is proposed to identify more biologically relevant EFMs to decompose FDs using stoichiometric information alone.

2 METHODS

2.1 Optimization objective

The goal is to derive an optimization objective for finding $\alpha_1 \cdots \alpha_K$ to compute an MBD composed of EFMs with as few branching pathways as possible. Given the set of all EFM contained in the matrix $\mathbf{E} = [\mathbf{e}_1 \cdots \mathbf{e}_K] = [e_{jk}] \in \mathbf{R}^{n \times K}$ and the stoichiometric matrix $\mathbf{S} = [S_{ij}] \in \mathbf{R}^{m \times n}$ in a metabolic network with m metabolites and n reactions, assume that the FD to be decomposed is $\mathbf{v} = [v_j] \in \mathbf{R}^n$.

First, define the consumption matrix $\mathbf{C} = [C_{ij}]$ and the production matrix $\mathbf{P} = [P_{ij}]$ for metabolites:

$$C_{ij} = \begin{cases} 1 & \text{if} \quad S_{ij} < 0 \\ 0 & \text{if} \quad S_{ij} \geq 0 \end{cases}, \; P_{ij} = \begin{cases} 1 & \text{if} \quad S_{ij} > 0 \\ 0 & \text{if} \quad S_{ij} \leq 0 \end{cases} \tag{2}$$

In other words, if $C_{ij} = 1$, then metabolite i is consumed by reaction j, and if $P_{ij} = 1$, then metabolite i is produced by reaction j. Based on $\mathbf{C}$ and $\mathbf{P}$, compute the number of paths $\mathbf{NP} = [NP_{ik}] \in \mathbf{R}^{m \times K}$ passing through metabolite i in EFM k:

$$NP_{ik} = \left(\sum_{j=1}^{n} C_{ij}\delta_{jk} \right) \left(\sum_{j=1}^{n} P_{ij}\delta_{jk} \right) \tag{3}$$

where $\delta_{jk} = 1$ if $e_{jk} \neq 0$ and $\delta_{jk} = 0$ if $e_{jk} = 0$. NP_{ik} is in fact the number of reactions consuming metabolite i multiplied by the number of reactions producing metabolite i in EFM k. It is thus the number of locally different paths passing through metabolite i. Figure 1 illustrates an example. **NP** is used for distinguishing EFMs that deviate from linear flow through branch point metabolites. Perfectly linear EFMs would have $NP_{ik} = 1$ for all metabolites, and EFMs with more branching flow would have higher NP_{ik}. The minimum use of EFMs with high NP_{ik} therefore defines the optimization objective to obtain an MBD. Hence, the objective coefficient c_k for EFM k is expressed as the following sum:

$$c_k = \sum_{i=1}^{m} w_i NP_{ik} \tag{4}$$

where w_i is the weight for metabolite i. It is manually assigned to account for the relative importance of a metabolite as a branch point metabolite. The simplest way for assigning weights to metabolites is to just assign 1 for all weights meaning that there is no bias toward the branching of any metabolite in an EFM. In practice, biological knowledge is required to choose suitable weights to obtain better decomposition. For example, for pyruvate, which is a true branch point metabolite involved in different metabolic pathways, a decomposition by EFMs with different fates of pyruvate, for instance, one going into the anaerobic fermentation and the other going into the tricarboxylic acid cycle (TCA) cycle, is preferred to a decomposition by all EFMs involving both the fermentation and the tricarboxylic acid cycle (TCA) cycle. In this case, branching of pyruvate within an EFM is strongly undesired so the weight can be set to be much higher than 1. In contrast, cofactors like ATP/ADP are global cellular currencies, and their distinct linear flow may not be biologically relevant. Low or zero weights can be assigned to them. The weighting also depends on the particular purpose of application. For example, if one focuses on comparing the glycolysis and the pentose phosphate pathway, a high weight can then be assigned to glucose 6-phosphate, which is a branch point of the two pathways.

2.2 Minimal branching decomposition

Having derived the objective, an MBD can be obtained by solving the following mixed integer linear programming problem (MILP):

$$\min \sum_{k=1}^{K} \sum_{i=1}^{m} w_i NP_{ik} \beta_k \tag{5}$$

$$\text{subject to} \sum_{k=1}^{K} e_{jk}\alpha_k = v_j \quad \text{for} \quad j = 1, \ldots, N \tag{6}$$

$$0 \leq \alpha_k \leq \beta_k \text{ for } \quad k = 1, \ldots, K \tag{7}$$

$$\beta_k \in \{0, 1\} \text{ for } \quad k = 1, \ldots, K \tag{8}$$

Objective (5) minimizes the use of EFMs with higher numbers of nonlinear pathways being active at branch point metabolites. Equation (6) is the constraint for decomposition. Constraint (7) is the on/off switch for an EFM. If $\beta_k = 0$, EFM k is switched off. If $\beta_k = 1$, EFM k is switched on. In the problem stated here, two assumptions without loss of generality are made on the EFM matrix $\mathbf{E} = [e_{jk}]$ for simpler presentation. First, $\mathbf{E}$ is calculated from a stoichiometric matrix consisting of

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

irreversible reactions only. It can be achieved by dividing each reversible reaction into two irreversible reactions with stoichiometry negative to each other. This avoids negative fluxes and additional constraints on reversible EFMs. Second, all EFMs are assumed to be scaled to have maximum weights not greater than 1 ($\alpha_k \leq 1$). One way for scaling is to multiply each $\mathbf{e}_k$ by $r_k = \min_j\{v_j/e_{jk}|e_{jk}>0\}$. Another method used in Wiback *et al.* (2003) is scaling to an uptake reaction by multiplying each $\mathbf{e}_k$ by $r_k = v_{\text{uptake}}/e_{\text{uptake},k}$.

It is noted that the 'no cancellation' property of decomposition by EFMs (Llaneras and Picó, 2010) ensures that if for a reaction j, $v_j = 0$, then all EFMs with $e_{jk} \neq 0$ must have $\alpha_k = 0$. It facilitates optimization by first selecting EFMs able to have non-zero weights for a FD, called 'contributable EFMs' and defined by $\mathbf{E}_{CONTRI} = \{\mathbf{e}_k | e_{jk} = 0 \quad \text{if} \quad v_j = 0\}$, out of $\mathbf{E}$, which is usually intractable for larger networks (Klamt and Stelling, 2002; Trinh *et al.*, 2009).

2.3 Enumeration of alternative optima

One of the major drawbacks of interpreting decomposition by EFMs is the non-uniqueness. Despite the use of certain optimization objectives, for instance, a minimum number of EFMs as in Wiback *et al.* (2003), the number of alternative optimal solutions can still be large. The performance of MBD regarding the uniqueness of decomposition was examined by enumerating alternative optima. The R-th alternative optimal MBD was calculated by adding the following constraint to the model:

$$\sum_{k=1}^{K} \beta_k^r \beta_k \leq \left(\sum_{k=1}^{K} \beta_k^r \right) - 1 \quad \text{for} \quad r = 1, \ldots, R-1 \tag{9}$$

where $\mathbf{b}_r = [\beta_1^r \quad \cdots \quad \beta_K^r]$ is the binary integer solution of the r-th alternative optimal MBD found. Suboptimal solutions can also be found.

2.4 Other decompositions for comparison

MBD was compared with decomposition by a minimum number of EFMs (MinEFMD) and random decomposition (RD). MinEFMD was calculated by replacing equation (5) with $c_k = 1$ for all k. RD was sampled as follows:

Initialization: $\mathbf{D} = \{\}$.

Step 1: find contributable EFMs $\mathbf{E}_{CONTRI} = \{\mathbf{e}_k | e_{jk} = 0 \quad \text{if} \quad v_j = 0\}$.

Step 2: randomly choose an $\mathbf{e}_k$ from $\mathbf{E}_{CONTRI}$. $\mathbf{D} \leftarrow \mathbf{D} \cup \{\mathbf{e}_k\}$.

Step 3: $\mathbf{v} \leftarrow \mathbf{v} - r\mathbf{e}_k$ where $r = \min_j\{v_j/e_{jk}|e_{jk}>0\}$

Step 4: if $\mathbf{v} = \mathbf{0}$, terminate and $\mathbf{D}$ is a decomposition. Else, go to step 1.

2.5 Implementations

EFMs were calculated using *efmtool* (Terzer and Stelling, 2008), and optimization models were solved by Gurobi 5.5® via the COBRA toolbox in MATLAB® in which all other calculations were performed. The MATLAB script file is available in the Supplementary Material. The desktop computer used had a 3.1 GHz quad-core CPU and 16 GB of RAM.

3 RESULTS

MBD was tested on randomly sampled FDs in the core *Escherichia coli* metabolic network as a validation of the method (Orth *et al.*, 2010). It was then applied to a FD determined by isotopomer distributions reported previously for the cardiomyocyte in perfused mouse heart (Vo and Palsson, 2006) and to a FD estimated from a dataset of metabolite assimilation and excretion in the recently published genome-scale metabolic network of *Lactococcus lactis* (Flahaut *et al.*, 2013). In the latter two cases, FDs were not sampled because they were estimated from experimental data.

3.1 Core *E.coli* metabolic network

The core *E.coli* metabolic network (Orth *et al.*, 2010) with 95 reactions and 72 metabolites simplified from the *i*AF1260 genome-scale network of *E.coli* (Feist *et al.*, 2007) includes glycolysis, TCA cycle, oxidative phosphorylation, pentose phosphate pathway and a biomass reaction, alongside with a set of transcriptional regulatory rules that can be readily implemented. Three different tests were performed to examine (i) the ability to separate fluxes at branch point metabolites; (ii) the performance in a simulated case consistent with regulatory rules and (iii) the issue of multiple optimal decompositions. As the set of all possible FDs has infinitely many solutions, random sampling based on the hit-and-run method (Almaas *et al.*, 2004) was applied to sample FDs for the tests. Table 1 summarizes the results. More details can be found in Supplementary Material.

3.1.1 Separation of fluxes at branch points To check that the proposed method has the desired ability to choose EFMs of distinct branches at branch points, two scenarios were

Table 1. Summary of test results in the core *E.coli* metabolic network

Test	Result	
1. Separation of combined fluxes at branch points. 83 cases in scenario 1 (Fig. 2a), 94 cases in scenario 2 (Fig. 2b)	Correct recovery rate Scenario 1 MBD: 100% RD: 7.5%	 Scenario 2 MBD: 94% RD: 1%
2. Coupling of glyoxylate cycle and acetate uptake implied by the regulatory rules in the presence of fructose and acetate (Fig. 2c).	No EFMs with simultaneous fructose uptake and glyoxylate cycle activity MBD: 100% MinEFMD : 0% RD: 1%	
3. Uniqueness of decompositions. Enumeration of alternative MBDs and MinEFMDs	MBD: 2000 FDs tested. 45% unique; 95% ≤ 8 optima. MinEFMD: 20 FDs tested. 100% ≥ 1000 optima	

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

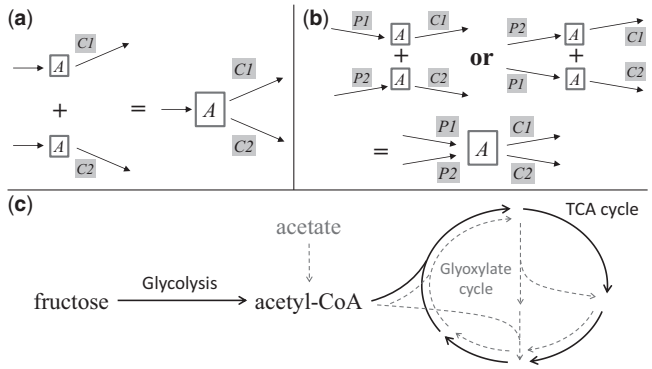

Fig. 2. Testing the ability of MBD to separate fluxes at branch points. (**a**) Scenario 1 in test 1: separation of two individual consuming reactions. (**b**) Scenario 2 in test 1: correct separation of two pairs of coupled producing and consuming reactions. (**c**) Test 2: a simulated case of glyoxylate cycle coupled to acetate uptake implied by regulatory rules

tested (Fig. 2a). First, two of the consuming reactions (*C1*, *C2*) of a branch point were chosen. Two sets of FDs, one with $v_{C1} \neq 0$, $v_{C2} = 0$ and vice versa, were sampled and summed up as the set of FDs being tested. The second scenario is more difficult in which metabolites have two producing reactions (*P1*, *P2*) individually and coupled to *C1* and *C2*, respectively (Fig. 2b). FDs were sampled and summed up as complex FDs for testing. In both scenarios, FDs were decomposed into EFMs to check whether the distinct branches were correctly associated in the chosen EFMs. In this test, the available carbon sources were set to be fructose and acetate to allow more branch points while their simultaneous uptake does not conflict the given regulatory rules. As this test aimed to assess the basic ability of MBD, the weight w_i was set to 1 for the branch point being tested and 0 for the rest of metabolites. All metabolites were balanced during the calculation of EFMs. In all, 19 500 EFMs were calculated.

There are 11 branch points in the first scenario. In each case, 10 FDs were randomly sampled and then subjected to MBD. In all, 1000 sRDs were obtained for each FD as comparison. In all cases, no EFMs in an MBD use two consuming reactions simultaneously, whereas it is true for only 7.5% of the RDs overall. In scenario 2, six branch points with two or more producing and consuming reactions were identified, and 94 combinations were tested. For each combination, 10 FDs for each of the four producing and consuming pairs were sampled and combined as shown in Figure 2b. MBDs for the resultant FDs were then calculated. In 94% of the MBDs for near 2000 FDs, all EFMs involve the correct reaction pairs, whereas only 1% of the RDs is correct. These results confirmed that the proposed MBD is capable of separating fluxes at branch point metabolites.

3.1.2 Coupling of glyoxylate cycle and acetate uptake The given regulatory rules of the core *E.coli* network imply that the glyoxylate cycle is active only if acetate is being consumed. Thus, in the presence of fructose and acetate, the system can consume both substrates with the glyoxylate cycle used primarily for acetate metabolism. Similar experiments and results have been reported in yeast (dos Santos *et al.*, 2003). To test whether MBD can distinguish the acetate-induced glyoxylate cycle from the TCA cycle, two sets of FDs were sampled, one without acetate uptake and glyoxylate cycle activity and the other without fructose uptake (Fig. 2c). One FD from each set was summed to form a complex FD, which was subject to MBD. $w_i = 1$ for all metabolites except the obvious branch point Acetyl-CoA with $w_i = 1000$. MinEFMDs and RDs were also computed for comparison.

In all 200 FDs tested, no EFMs in any MBDs have simultaneous fructose uptake and glyoxylate cycle activity, whereas all MinEFMDs do. Only 1.5% of MBDs use one EFM cometabolizing fructose and acetate, whereas all MinEFMDs do. For RDs, >99% have neither properties. This example demonstrated the unique capability of MBD to distinguish individually regulated pathways in a realistic system setting with underlying regulatory rules.

3.1.3 Uniqueness of MBDs To investigate the uniqueness of MBDs, 2000 FDs with glucose as carbon source were sampled. Alternative optimal MBDs and MinEFMDs were enumerated. To obtain a smaller set of EFMs for easier optimization, several cofactors and small metabolites were unbalanced when calculating EFMs. It is noted that though EFMs resulting from unbalancing cofactors can still properly decompose any FDs, they cannot reflect the subtle interdependence between different metabolic pathways. More discussion can be found in the Supplementary Information.

MBD has a satisfactory performance regarding uniqueness. There are 45% of FDs that have a unique MBD, and 95% of the FDs have no more than eight multiple MBDs. To compare the results with MinEFMDs, we randomly picked 20 FDs and enumerated the alternative MinEFMDs. All FDs turned out to have ≥1000 different MinEFMDs. As the minimum number of EFMs required for decomposition is much smaller than the number of contributable EFMs, e.g. dozens out of hundreds or thousands, there are virtually countless MinEFMDs. The number of multiple MBDs is extremely small in this sense. The results suggested that MBD is effective in identifying a small subset of decompositions by EFMs with fewer branching pathways, which are suitable for understanding the FDs.

3.2 Mouse myocardial metabolic network

To demonstrate that the proposed method indeed yield a more intuitive and biologically relevant decomposition, MBD was applied to a FD of mouse cardiomyocyte estimated from isotopomer labeling experiments performed in perfused mouse hearts (Vo and Palsson, 2006). The model and the FD are available as the Supplementary Information in that publication. The mouse myocardial metabolic network consists of 240 metabolites and 257 reactions compartmentalized into cytoplasm, mitochondrion and extracellular space. There are 99 active reactions in the FD (Fig. 3). Glucose and oleate (a fatty acid) were available as carbon sources. Lactate, pyruvate and the ketone bodies acetoacetate and 3-hydroxybutanoate were excreted. A small amount of citrate and succinate were also produced.

The primary function of cardiomyocyte is energy production for heart contraction. It is represented by the reaction 'DMatp'

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

hydrolyzing ATP in the network. By decomposing the FD into a set of EFMs, different pathways for energy production and other cellular activities can be revealed. All metabolites were balanced for EFM calculation. In all, 1236 EFMs were computed. Optimality could not be proved after 24 h in the initial trial to find an MBD using a uniform weight. We then attempted to put emphasis on the obvious branch point of glucose and fatty acid oxidation, acetyl-CoA, by giving a weight of 1000 to it and 1 to other metabolites. An optimal MBD comprising 14 EFMs was obtained and was proved to be unique. EFMs in the MBD can be divided into four types. The first is three EFMs of fatty acid oxidation that produce CO_2 and ketone bodies. The second is nine EFMs producing CO_2 and different organic acids from glucose. The third is an EFM consuming glucose and oleate and producing citrate, β-hydroxybutyrate (BHB) and CO_2. In this EFM, it is unambiguous that BHB is produced from oleate because pyruvate dehydrogenase, the only reaction in the FD converting pyruvate into acetyl-CoA, is not included in this EFM, and acetyl-CoA is the precursor of BHB. The fourth type is an isolated cycle between citrulline and ornithine. The difference between EFMs lies in the products and the different cellular activities or pathways, including the malate–aspartate shuttle, pseudoketogenesis and an apparent glutamate–glutamine futile cycle.

We examined whether finding decomposition after removing EFMs with simultaneous oleate and glucose uptake is possible. No solution could be found, showing that at least one such EFM is required in any decomposition. To see whether this necessary interdependence of glucose and fatty acid oxidation originated from the balance of cofactors, we tried to find MBDs using sets of EFMs calculated from unbalancing some cofactors and small metabolites. When unbalancing NAD^+ and NADH, an MBD without any EFM using both carbon sources simultaneously was found, whereas unbalancing other metabolites such as ATP, ADP, H^+ or Pi did not give the desired result. This suggested that glucose and fatty acid oxidation interacted to have a balanced use of $NADH/NAD^+$. This kind of comparison between MBDs obtained from balanced and partially balanced EFMs can provide insights into the interdependence of pathways.

For comparison, MinEFMDs and RDs were computed. There were 35 optimal MinEFMDs, each with 13 EFMs. To have a fair comparison, nine suboptimal MBDs were enumerated. Comparing the three types of decomposition can distinguish several properties of an MBD (Table 2). First, only one EFM in each MBD consumes glucose and oleate simultaneously. As one is hexose and the other is fatty acid, their uses lie in different pathways, and thus, a good decomposition should avoid using EFMs consuming both. In all MinEFMDs, however, at least five such EFMs were used.

Second, most EFMs in the MBDs excrete at most one organic acid or ketone body except two EFMs that excrete two, whereas all MinEFMDs use several EFMs producing three or four products simultaneously. Such decompositions are more difficult to interpret because the sole effect of producing a certain compound alone (e.g. role in ATP production) might not be observed by looking at individual EFMs. EFMs excreting more metabolites are also expectably longer, which might be more difficult to visualize. Another relevant advantage of MBD is that because an MBD would in general use EFMs that produce as few compounds as possible, if coproduction is observed in an EFM in an MBD solution, it may suggest the possibility that the

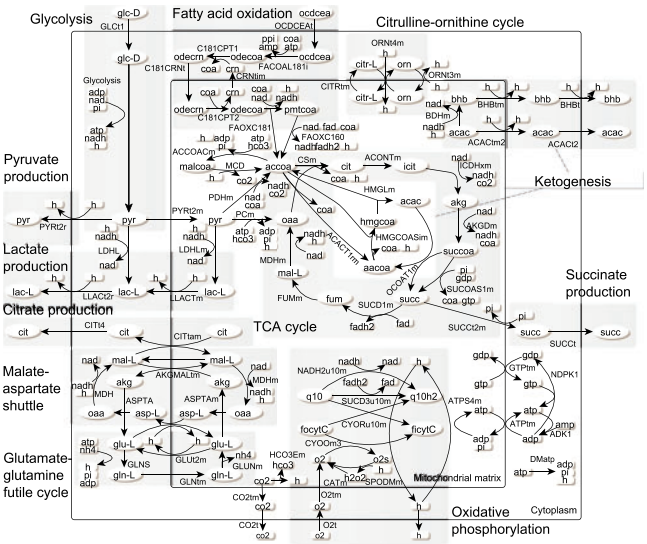

Fig. 3. Reactions with non-zero fluxes in the mouse myocardial metabolic network. Intermediate steps of glycolysis are omitted. Water involved in reactions is not shown. All abbreviations of reactions and metabolites are the same as in the original network. Colored version is available in Supplementary Material

Table 2. Comparison of MBD, MinEFMD and RD for the mouse myocardial flux distribution

Property	MBD	MinEFMD	RD
1. Number of EFMs	14	13	14
2. Number of multiple optima	1 (unique)	35	–
3. Maximum number of products in EFMs	2	3 or 4	$\geq$3 (88.8%)
4. Number of EFMs with simultaneous glucose and oleate uptake	1[a]	$\geq$5	$\geq$5 (99.5%)
5. No EFMs producing ketone bodies from glucose	True	True for 8.6%	True for 6.4%
6. No EFMs involving oleate uptake and malonyl-CoA simultaneously	True	None of the solutions	True for 0.1%

[a]Our results showed that at least one EFM consuming both carbon sources is required for decomposition.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

production of the compounds is coupled under the particular condition where the FD is measured.

Third, the EFMs used in the MBDs appear to be more consistent with the regulation of metabolic pathways. For instance, ketogenesis by which ketone bodies are produced is known to be associated with fatty acid oxidation rather than glycolysis. In MBDs, no EFMs producing ketone bodies from glucose are present, but 19 of 35 MinEFMDs contain at least one of these EFMs. Another example is malonyl-CoA ('malcoa' in Fig. 3). Production of malonyl-CoA is a known mechanism leading to the repression of fatty acid oxidation by glucose by inhibiting the transfer of fatty acyl moieties into mitochondria in cardiac muscle (Depre *et al*., 1999). Use of EFMs that do not couple fatty acid oxidation and the cycle between malonyl-CoA and acetyl-CoA is thus a more rational choice. In the MBD, no EFMs simultaneously use oleate and involve malonyl-CoA whereas all MinEFMDs do. The same comparison was also performed between MBDs and MinEFMDs obtained from EFMs unbalancing NAD^+ and NADH. All properties still hold (see Supplementary Material).

From the comparison, MBD is shown to have several properties of biological relevance that are possessed by only a tiny portion of MinEFMDs. We argue that the more biologically relevant decomposition by MBD does not occur by chance but rather by the minimization of the use of EFMs with branching pathways. The reasons are 2-fold. First, from the consideration of the network structure alone, naturally occurring metabolic pathways tend to prefer linearity to branching at branch point metabolites as suggested in Ihmels *et al*. (2004) based on their observation on data from $\geq$1000 experimental conditions. Second, the flux of a certain reaction is the sum of the fluxes of different pathways containing that particular reaction. As different pathways have different fluxes, minimizing the use of those 'hybrid' EFMs would force the use of EFMs representing the true pathways to fit the flux values rather than a combination of 'hybrid' EFMs, which is also able to fit the values of fluxes with a similar number of EFMs.

3.3 *Lactococcus lactis* genome-scale metabolic network

Lactococcus lactis is a model organism in the group of lactic acid bacteria and has an important role in cheese production. *Lactococcus lactis* is in general auxotrophic for several amino acids (AAs) (Jensen and Hammer, 1993) and grows on complex media. It also produces compounds involved in flavor formation (Ayad *et al*., 1999). From the recently published genome-scale metabolic network of *L.lactis subsp. cremoris* MG1363, which specifies many flavor-forming pathways (Flahaut *et al*., 2013), a FD with a complex profile of substrate consumption and production can be obtained and can thus test the applicability of MBD. The FD was simulated by FBA using the measured consumption/production rates of glucose, lactate, formate, acetate, ethanol and different AAs in a chemostat experiment at dilution rate 0.05 h^{-1} as constraints. The subnetwork spanned by the simulated FD contains 294 metabolites and 280 reactions, including 26 uptake and 13 excretion reactions. In all, 518 EFMs were computed when balancing all the metabolites.

A unique MBD of 11 EFMs was computed. Nine other suboptimal MBDs, 10 MinEFMDs and 10^4 RDs were obtained for

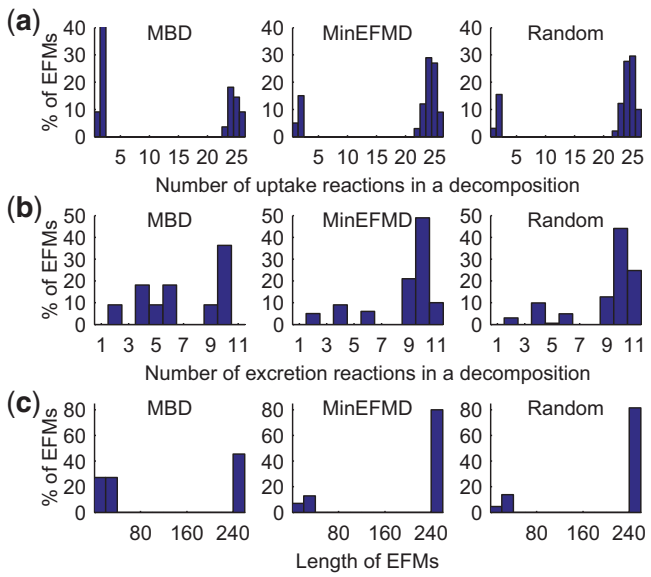

Fig. 4. (**a**) Number of uptake reactions, (**b**) number of excretion reactions and (**c**) length of EFMs in MBD, MinEFMD and RD

comparison. First, we focused on two properties, the number of exchange reactions and the length of the EFMs. They can reflect the interpretability and visualizability of the EFMs. Figure 4 shows their distribution in the three types of decompositions. MBD in general used shorter EFMs with lower numbers of uptake and excretion reactions, whereas MinEFMD and RD performed similarly. This indicates that MBD is easier for interpretation and visualization.

Second, we looked into the organization of individual pathways. An interesting observation pertains to the fermentation mode. *Lactococcus lactis* has been well known for its shift between homolactic fermentation producing only lactate and mixed-acid fermentation that produces formate, acetate and ethanol in addition depending on the glucose uptake rate (Thomas *et al*., 1979). The two modes are regulated differently. The reactions responsible for lactate, formate, acetate and ethanol production are lactate dehydrogenase (LDH), pyruvate-formate lyase (PFL), acetate kinase (ACK) and alcohol dehydrogenase (ADH), respectively. No MBDs contain EFMs involving concurrent use of LDH and PFL, ACK or ADH, while 97% of RDs and all but 1 MinEFMDs do.

The *L.lactis* network also models detailed AA metabolism including pathways for AA interconversion and tRNA biosynthesis. According to a recently reconstructed transcriptional regulatory network of *L.lactis* (Ravcheev *et al*., 2013) available in RegPrecise database (Novichkov *et al*., 2010), they belong to different regulons and coincide at many metabolites. It is relevant to check how MBD performs regarding these branch points. In the current MBD, however, the EFM matrix used to calculate MBD includes the biomass reaction, forcing many pathways to participate in one EFM to produce biomass. This is also the reason for the small size of the EFM set.

To cope with this, in a new round of MBD calculation, the biomass reaction was split into several reactions producing

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

Table 3. AA branch points in the flux distribution in *L.lactis* network

Branch point	Flow from	Flow toward	% with no EFM of >1 branches	
			MBD	RD
Ala	Ala[e], pyruvate	Ala-tRNA[a], D-alanine[a]	100%	13%
Cys	Cys[e]	Cys-tRNA[a], L-cysth, pyruvate, CoA[a]	80%	0
Met	Met[e]	Met-tRNA[a], L-cysth.	100%	1%
Gly	Gly[e]	Gly-tRNA[a], Thr	100%	0.1%
Thr	Thr[e], Gly	Thr-tRNA[a], 2OBUT	100%	0
Ser	Ser[e]	Ser-tRNA[a], pyruvate	0	0.8%
Val	Val[e]	Val-tRNA[a], 2H3MB[e][b]	100%	1.4%
Glu	Glu[e]	Glu-tRNA[a], D-glutamate[a]	100%	14%
Leu	Leu[e]	Leu-tRNA[a], 2HXIC[e][b]	0	0
Lys	Lys[e]	Lys-tRNA[a], uAGL[a]	100%	14%
2OBUT	Thr, L-cysth	2H3MP[e][b]	100%	50%
R5P	ribose, RU5P	PRPP	100%	0.4%
PRPP	R5P	Trp-tRNA[a], NAD^+ [a]	100%	6%

Notes. [e], extracellular; L-cysth, L-cystathionine; 2H3MB, 2-hydroxy-3-methylbutanoate; 2H3MP, 2-hydroxy-3-methylpentanoate; 2HXIC, L-2-hydroxyisocaproate; uAGL, UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-L-lysine; RU5P, ribulose 5-phosphate; PRPP, phosphoribosyl pyrophosphate.
[a]Biomass precursors.
[b]Flavor compounds.

individual biomass components, including DNA, RNA, proteins, lipid, peptidoglycan, etc. This then resulted in 32 962 EFMs. A weight equal to 1000 was assigned to branch points in AA metabolism (Table 3) and 1 to other metabolites. Most branch points are AAs acting as precursors for tRNA, other biomass components or flavor compounds. Cysteine and serine as an exception contribute to the pyruvate pool. Other branch points include 2-oxobutanoate (2OBUT), a precursor for a flavor compound where the fluxes from threonine and cystathionine meet; ribose 5-phosphate (R5P), synthesized from the pentose phosphate pathway or the ribose produced from methionine and phosphoribosyl pyrophosphate (PRPP), synthesized from R5P and used for producing either tryptophan or NAD^+.

Convergence to optimality seemed impractical and five suboptimal MBDs were obtained. The best one has an optimality gap of 22% after 24 h of calculation. For MinEFMD, no solution using less EFMs than RD could be found so only RD was considered.

Comparing the original MBDs with the current ones, differences can be observed. First, the present MBDs consist of eight more EFMs than the original. Second, EFMs are shorter in the present MBDs. These changes are expectable because smaller parts of metabolism are sufficient for producing individual biomass components at steady state while more EFMs are required when each EFM covers a smaller part. We also looked into the ATP turnover; in the original MBDs, ATP consumption is perfectly distributed to growth requirement in five EFMs and non-growth–associated maintenance in the other six EFMs. In the current MBDs, however, only 70% of ATP maintenance is consumed in EFMs independent of synthesis of biomass components. This may be the result of suboptimality. Alternatively, this suggests that simultaneous clear roles in both ATP consumption and AA branch points are impossible. However, as an advantage of the current MBDs, the ATP consumption by synthesizing each biomass components can be reflected. tRNA and nucleotide synthesis consume >80% of the ATP.

Regarding AA branch points, MBD successfully chose separate EFMs representing the production of biomass components and flavor compounds. Table 3 compares the five MBDs and 10^4 RDs. All tRNAs are produced in three EFMs, and other amino-acid–consuming reactions are in other EFMs, except for the cases of serine and leucine. For example, beside tRNA charging in three EFMs, cysteine participates in the reaction of phosphopantothenate-cysteine ligase for biosynthesis of coenzyme A in another EFM, goes into the pyruvate pool by cysteine desulfhydrase in three other EFMs and produces cystathionine (for producing a flavor compound) in one EFM. Taking glycine and threonine as another example, beside tRNA charging of the two AAs in 3 EFMs, threonine is converted into 2OBUT by threonine deaminase for flavor compound production in four other EFMs, of which threonine is taken up from the medium in three of the EFMs and is converted from glycine by threonine aldolase in the other one. In most cases, RD failed to have such properties. More details can be found in Supplementary Material.

For leucine, we found that, in all EFMs, tRNA charging was coupled to the production of a flavor compound, thus inseparable. For serine, if EFMs with only one serine-consuming reaction are used, EFMs branching at either threonine or cysteine must be used. It was observed that all three AAs had one consuming reaction producing NH_4. We thus suspected that branches were not separated due to NH_4 balance. When finding MBD from EFMs with NH_4 unbalanced, indeed the issue was resolved.

From this and the previous examples, an interesting observation is that the FDs can be represented by most EFMs of individual pathways together with a few necessary EFMs coupling

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

the pathways. It may suggest that while a large part of the metabolism can be interpreted as the sum of individual pathways, at least a small part must be ascribed to cooperation between pathways.

4 CONCLUSION

MBD to decompose a FD into a set of EFMs was derived. The EFMs contain as few branching pathways as possible. MBD advances the analysis of FD by decomposition into EFMs in three aspects. First, it originates from the consideration of the branching structure in metabolic network based on experimental evidence, which is a novel attempt. Second, MBD results in more biologically relevant EFMs leading to clearer interpretation. Third, MBD showed good performance regarding uniqueness. It diminishes the ambiguity of decomposition.

The minimal branching EFM concept underlying MBD can be useful in further applications, e.g. identifying metabolic pathways for engineering by acting as an estimate for biologically feasible EFMs. Scaling up the method to genome-scale networks is necessary for practical applicability. In conclusion, MBD is able to bring useful insights when analyzing FDs. Combining information from e.g. transcriptional regulation and thermodynamics, wider application of EFMs can be anticipated.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the helpful comments made by the three anonymous reviewers.

Funding: This work was supported by the Danish research council for Nature and Universe (FNU; 10-085115).

Conflict of interest: none declared.

REFERENCES

Almaas,E. *et al.* (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*, **427**, 839–843.

Ayad,E.H.E. *et al.* (1999) Flavour forming abilities and amino acid requirements of *Lactococcus lactis* strains isolated from artisanal and non-dairy origin. *Int. Dairy J.*, **9**, 725–735.

Chan,S.H.J. and Ji,P. (2011) Decomposing flux distributions into elementary flux modes in genome-scale metabolic networks. *Bioinformatics*, **27**, 2256–2262.

de Figueiredo,L.F. *et al.* (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics*, **25**, 3158–3165.

dos Santos,M.M. *et al.* (2003) Identification of *in vivo* enzyme activities in the cometabolism of glucose and acetate by *Saccharomyces cerevisiae* by using ^{13}C-labeled substrates. *Eukaryot. Cell*, **2**, 599–608.

Depre,C. *et al.* (1999) Glucose for the heart. *Circulation*, **99**, 578–588.

Edwards,J.S. and Palsson,B.Ø. (1999) Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J. Biol. Chem.*, **274**, 17410–17416.

Feist,A.M. *et al.* (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.

Flahaut,N.A. *et al.* (2013) Genome-scale metabolic model for *Lactococcus lactis* MG1363 and its application to the analysis of flavor formation. *Appl. Microbiol. Biotechnol.*, **97**, 8729–8739.

Ihmels,J. *et al.* (2004) Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nat. Biotech.*, **22**, 86–92.

Ip,K. *et al.* (2011) Analysis of complex metabolic behavior through pathway decomposition. *BMC Syst. Biol.*, **5**, 91.

Jensen,P.R. and Hammer,K. (1993) Minimal requirements for exponential growth of *Lactococcus lactis*. *Appl. Environ. Microbiol.*, **59**, 4263–4266.

Jol,S.J. *et al.* (2012) System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes. *PLoS Comput. Biol.*, **8**, e1002415.

Jungreuthmayer,C. *et al.* (2013) regEfmtool: speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic. *Biosystems*, **113**, 37–39.

Klamt,S. and Stelling,J. (2002) Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.*, **29**, 233–236.

Lewis,N.E. *et al.* (2012) Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Micro.*, **10**, 291–305.

Llaneras,F. and Picó,J. (2010) Which metabolic pathways generate and characterize the flux space? A comparison among elementary modes, extreme pathways and minimal generators. *J. Biomed. Biotechnol*, **2010**, 753904.

Machado,D. *et al.* (2012) Random sampling of elementary flux modes in large-scale metabolic networks. *Bioinformatics*, **28**, i515–i521.

Novichkov,P.S. *et al.* (2010) RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res.*, **38**, D111–D118.

Oberhardt,M.A. *et al.* (2009) Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.*, **5**, 320.

Orth,J.D. *et al.* (2010) Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. In: Böck, A. *et al.* (eds) *Escherichia coli and Salmonella: Cellular and Molecular Biology*. 10.2.1. ASM Press, Washington DC.

Pey,J. and Planes,F.J. (2014) Direct calculation of elementary flux modes satisfying several biological constraints in genome-scale metabolic networks. *Bioinformatics*, **30**, 2197–2203.

Poolman,M.G. *et al.* (2004) A method for the determination of flux in elementary modes, and its application to *Lactobacillus rhamnosus*. *Biotechnol. Bioeng.*, **88**, 601–612.

Price,N.D. *et al.* (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.*, **2**, 886–897.

Ravcheev,D.A. *et al.* (2013) Genomic reconstruction of transcriptional regulatory networks in lactic acid bacteria. *BMC Genomics*, **14**, 94.

Rezola,A. *et al.* (2011) Exploring metabolic pathways in genome-scale networks via generating flux modes. *Bioinformatics*, **27**, 534–540.

Schuster,S. and Hilgetag,C. (1994) On the elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst*, **2**, 165–182.

Schwartz,J.M. and Kanehisa,M. (2005) A quadratic programming approach for decomposing steady-state metabolic flux distributions onto elementary modes. *Bioinformatics*, **21** (Suppl. 2), ii204–ii205.

Schwartz,J.M. and Kanehisa,M. (2006) Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. *BMC Bioinformatics*, **7**, 186.

Tabe-Bordbar,S. and Marashi,S.-A. (2013) Finding elementary flux modes in metabolic networks based on flux balance analysis and flux coupling analysis: application to the analysis of *Escherichia coli* metabolism. *Biotechnol. Lett.*, **35**, 2039–2044.

Terzer,M. and Stelling,J. (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, **24**, 2229–2235.

Thomas,T.D. *et al.* (1979) Change from homo- to heterolactic fermentation by Streptococcus lactis resulting from glucose limitation in anaerobic chemostat cultures. *J. Bacteriol.*, **138**, 109–117.

Trinh,C.T. *et al.* (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl. Microbiol. Biotechnol.*, **81**, 813–826.

Vo,T.D. and Palsson,B.O. (2006) Isotopomer analysis of myocardial substrate metabolism: a systems biology approach. *Biotechnol. Bioeng.*, **95**, 972–983.

Wiback,S.J. *et al.* (2003) Reconstructing metabolic flux vectors from extreme pathways: defining the α-spectrum. *J. Theor. Biol.*, **224**, 313–324.

Zanghellini,J. *et al.* (2013) Elementary flux modes in a nutshell: properties, calculation and applications. *Biotechnol. J.*, **8**, 1009–1016.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016