A Markov constraint to uniquely identify elementary flux mode weights in unimolecular metabolic networks

Justin G. Chitpin[1,2] and Theodore J. Perkins[1*,2]

[1]Regenerative Medicine Program, Ottawa Hospital Research Institute

[2]Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON, Canada

[*]To whom correspondence should be addressed

July 25, 2022

Abstract

Motivation: Elementary flux modes (EFMs) are minimal, steady state pathways characterizing a flux network. Fundamentally, all steady state fluxes in a network are decomposable into a linear combination of EFMs. While there is typically no unique set of EFM weights that reconstructs these fluxes, several optimization-based methods have been proposed to constrain the solution space by enforcing some notion of parsimony. However, even these approaches may fail to uniquely identify EFM weights and, therefore, return different feasible solutions across objective functions and solvers.

Results: Using simulated and biological networks, we demonstrate how optimization-based methods may return one of infinitely many EFM weight solutions that reconstruct the network fluxes. For unimolecular flux networks, we propose a natural, Markovian constraint to uniquely identify EFM weights explaining steady state fluxes. We describe an algorithm for computing these weights by reformulating the flux decomposition problem in terms of a certain discrete-time Markov chain. We demonstrate our method with a differential analysis of EFM activity in a lipid metabolic network comparing healthy and Alzheimer's disease patients.

Conclusion: Our method is the first to uniquely decompose steady state fluxes into EFM weights for any unimolecular metabolic network.

Availability: The Julia software is available through https://www.perkinslab.ca/ or on Github at https://github.com/jchitpin/MarkovWeightedEFMs.jl.

Contact: tperkins@ohri.ca

1 Introduction

Cellular metabolism consists of carefully-tuned biochemical reactions operating collectively to maintain metabolic homeostasis. These reactions form highly connected metabolic networks whose dynamics regulate bioenergetic processes [1, 2] and cell-specific functions [3]. Data from large-scale genomics projects has led to comprehensive, genome-wide reconstructions of numerous cell systems including bacterial, yeast, and human metabolic networks [4, 5]. Typically, these large-scale networks are applied in kinetic or stoichiometric models to predict system-wide fluxes [6], optimize metabolite production [7], and identify metabolic alterations associated with diseases [8].

Elementary flux modes (EFMs) are a pathway-based approach to study properties of flux networks derived from kinetic or stoichiometric models of metabolism. The EFM is defined as a non-decomposable, steady state pathway in a metabolic network [9]. EFMs may either traverse the network, connecting input and output fluxes of external metabolites, or encode looped pathways, such as futile cycles, to maintain steady state. The non-decomposable property specifies that pathway flux is lost if any reaction in an EFM is abolished. Thus, EFMs have been described as functional units of flux expressing any steady state metabolic phenotype [10].

EFM analyses have been extensively applied in literature to characterize metabolic activity. For example, the number of EFMs producing a given metabolite is an indicator of network robustness. Considerable work has been devoted to quantitatively assessing the contribution of individual EFMs towards external metabolite production of bioreactors [11] and correlating activity of specific EFMs with metabolic phenotypes [12]. More recently, EFM analyses have been applied to predict metabolic interactions in complex microbial communities [13], estimate cellular processes explaining exponential cell growth [14], and optimize renewable fuel production in bioreactors [15].

A major computational challenge associated with EFMs is their enumeration in large-scale networks. EFM analyses are typically limited to small biological networks (< 100 reactions) or subnetworks such as central carbon metabolism. The number of EFMs explodes combinatorially with network size and number of external metabolites [16, 17]. Considerable

attention has therefore been devoted to scaling EFM enumeration towards genome-wide networks [18–20]. Most recently, advances in parallelized EFM computation have enabled EFM enumeration in medium-sized networks containing as many as 296 internal reactions and 15 uptake reactions, yielding over twelve billion EFMs [21].

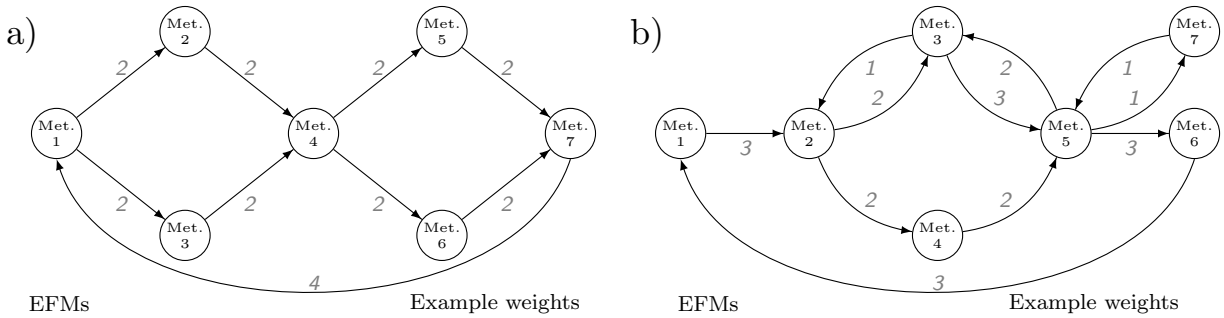
Another key computational problem is to determine how much flux is travelling along each EFM, given the steady state fluxes of each reaction in a metabolic network. This flux decomposition problem commonly arises in differential EFM analyses to identify metabolic subpathways altered across disease-versus-control or growth conditions [22, 23]. The challenge when analyzing biological networks, however, is that decomposing steady state fluxes onto EFMs is generally not unique. In a unimolecular flux network with M metabolites, the steady state phenotype can be described by the N reaction fluxes, of which there are at most M^2. In general, there can be up to $M!$ EFMs through a flux network. While any weighting of the K EFMs corresponds to a unique set of N fluxes, the reverse is not true, as a set of N fluxes can be explained in many different ways by K EFMs, especially when $K \gg N$.

Existing strategies for solving the flux decomposition problem rely on computational convenience, rather than strong biophysical motivation. These optimization-based methods use linear, quadratic, or mixed-integer linear programming (LP/QP/MILP) to assign EFM weights according to a user-specified objective function. Common optimization-based methods minimize or maximize activity through the shortest or fewest EFMs to reconstruct the observed fluxes. However, these methods may return one among many equally good solutions, often without acknowledging the other possibilities.

Figure 1 highlights the flux decomposition problem on two example networks whose fluxes can be perfectly reconstructed from infinitely many different sets of EFM weights. The first example network contains two parallel pathways transforming metabolite 1 to 4 and metabolite 4 to 7 through different intermediates. This branching-converging pattern is observed in biological networks, for example, when the glycolytic metabolite D-glucose 1-phosphate is converted to 2-phospho-D-glycerate through shared intermediates D-glyceraldehyde 3-phosphate and 3-phospho-D-glycerate [24]. The second example network shows an internal cycle formed by an irreversible and reversible set of reaction intermediates between metabolite 1 and 6. Similar network structures are observed in folate metabolism, whereby

5,10-methylene THF converts to THF irreversibly by methylenetetrahydrofolate reductase or reversibly by serine hydroxymethyltransferase [25]. Applying an objective function based on EFM length to the first example network in Figure 1a would fail to constrain the solution space as all EFMs share the same number of reactions. Similarly, maximizing the number of active EFMs would lead to an infinite number of decimal weights, while a minimization function would yield two equally good integer solutions. The same problems arise in Figure 1b, where EFM weights are not uniquely identifiable regardless of the specified objective function. Although the weight for EFM 5 remains constant across all solutions, these uniquely identifiable weights only occur when there exists at least one reaction unique to that EFM. As EFM 5 is the only pathway containing reactions between metabolites 5 and 7, its weight must be equal to these reaction fluxes. Thus, uniquely decomposing fluxes onto EFMs is uncommon in highly connected networks and any downstream analysis of these unidentifiable EFM weights is biologically meaningless.

Despite the well-established use of mathematical optimization, uniquely identifying EFM weights remains a longstanding problem limiting EFM applications. In this article, we show how common objective functions described in literature may fail to unambiguously decompose fluxes onto EFM weights. We further argue these methods are inappropriate for assigning EFM weights, as the metabolites in a network are assumed to follow a predetermined path along each EFM. A metabolite has no memory of the reactions leading to its current state nor the sequence of reactions it will undergo to complete an EFM in the future. We argue that a more natural approach to assigning EFM weights is to model how a single particle converts between metabolites according to the observed, steady state fluxes. This Markovian assumption motivates our discrete-time Markov chain model of a single particle converting between metabolites in a unimolecular flux network. By decomposing the Markov chain trajectories into EFMs, we compute a unique probability of a particle transiting an EFM that is proportional to the EFM weights reconstructing those observed fluxes. We solve the problem of computing the EFM transit probabilities via steady state analysis of a new construction we propose, a cycle-history Markov chain. We demonstrate our method by computing EFM weights for the example flux networks in Figure 1 and a kinetic model of sphingolipid metabolism of healthy versus Alzheimer's disease patients.



a)

EFMs	Example weights				
1. $1 \to 2 \to 4 \to 5 \to 7 \to 1$	2	0	1	1.1	1.5
2. $1 \to 3 \to 4 \to 6 \to 7 \to 1$	2	0	1	0.4	1.5
3. $1 \to 2 \to 4 \to 6 \to 7 \to 1$	0	2	1	0.9	0.5
4. $1 \to 3 \to 4 \to 5 \to 7 \to 1$	0	2	1	1.6	0.5

b)

EFMs				
1. $1 \to 2 \to 3 \to 5 \to 6 \to 1$	1	2	1.1	1.5
2. $1 \to 2 \to 4 \to 5 \to 6 \to 1$	2	1	1.9	1.5
3. $2 \to 3 \to 2$	1	0	0.9	0.5
4. $3 \to 5 \to 3$	2	1	1.1	1.5
5. $5 \to 7 \to 5$	1	1	1	1
6. $2 \to 4 \to 5 \to 3 \to 2$	0	1	0.1	0.5

Figure 1: Example networks demonstrating EFM weight ambiguity. Nodes and edges denote metabolites and fluxes, respectively. EFMs are shown below with example EFM weights that reconstruct the observed fluxes.

2 Methods

2.1 Preliminaries

A closed, connected, unimolecular, steady-state flux network can be represented as a weighted graph $G = (V, E, f)$, where V is the set of nodes corresponding to distinct metabolites, E is the set of directed edges corresponding to chemical transformations of one metabolite into another, and f are positive real-valued weights of those edges that denote the amount of metabolite per unit time being transformed. The graph being connected means that there is a path in the graph from any node i to any other node j. The network being closed means that no material is flowing into or out of the network. The network representing steady state means that the total flux into each node equals the total flux out–for every $i \in V$, $\sum_j f_{ji} = \sum_j f_{ij}$. Otherwise, metabolite i would be increasing or decreasing. This network G can be decomposed into a finite set of K EFMs transiting a series of edges in E. A pathway is considered an EFM if it satisfies steady state and is support-minimal such that all edges are not contained in any other EFM. Equivalently, the EFMs are the set of all simple cycles in the graph. Once enumerated, any set of steady state fluxes in G may be explained as a positive, linear combination of their EFMs [26]. Specifically, the weighted sum of all EFMs contributing towards the same reaction should equal the observed, steady

state flux of that reaction. This flux decomposition problem is usually solved by identifying the set of EFM weights w satisfying $Aw = f$, where A is an $|E|$ by K matrix with one or zero elements denoting whether a unimolecular reaction is active or inactive in each EFM. When the number of EFMs exceeds the number of linearly independent fluxes, w is undetermined. While optimization-based methods are not guaranteed to constrain the solution space of w, the method we present uniquely identifies w and can be trivially extended to some other types of flux networks.

2.2 Single particle, discrete-time Markov chain model

Our approach to uniquely identifying EFM weights depends on imagining a single "particle" transiting the flux network–repeatedly being transformed from one metabolite to another. Specifically, we assume the particle moves stochastically according to a discrete-time Markov chain, with states and transitions corresponding to metabolites and reactions, respectively. Transition probabilities are proportional to the flows out of each state: for all $i, j \in V$, $T_{ij} = f_{ij} / \sum_{j'} f_{ij'}$. Starting from any initial state s_0, the particle will transit a random sequence of states $s_0, s_1, s_2, \ldots$ Because we have assumed the flux network is strongly connected, and by basic theory of Markov chains, with probability 1, that sequence will include all states infinitely many times. Therefore, every time the particle visits a state i, it must eventually return to state i. The sequence of states in between, $s_t, \ldots, s_{t'}$ constitutes a cycle. That need not be a simple cycle–it may comprise multiple nested or interleaved simple cycles. However, it can always be reduced to a unique simple cycle by repeatedly removing the earliest simple cycle found within the sequence. In this way, every revisit to any state can be said to have occurred via a particular simple cycle, which of course corresponds to an EFM. Therefore, we propose finding unique EFM weights by setting them proportional to the (steady state) probabilities of our Markovian particle transiting the corresponding simple cycles.

2.3 Cycle-history Markov chain

Equating EFM weights with cycle probabilities in the Markov chain of section 2.2 creates the problem of computing those cycle probabilities, a problem that has not been previously

addressed in the Markov chain literature. In this section we take the first step in that direction by defining a cycle-history Markov chain (CHMC). In the next section, we show how to use the CHMC to compute the desired probabilities. The CHMC is like the Markov chain in section 2.2, but its notion of state is expanded to include enough of the history of the process to determine each time the particle finishes transiting a simple cycle. The CHMC is constructed as follows. We first choose an arbitrary metabolite s_0 in the flux network. Each state S in the CHMC corresponds to an entire simple path s_0, s_1, ..., s_m of the Markov chain envisaged in section 2.2. If there is a reaction transforming metabolite s_m to s_{m+1}, then there is a corresponding transition in the CHMC. If s_{m+1} is not among s_0, s_1, ..., s_m, then it is simply a transition to the longer simple path–i.e. from the state $S \equiv s_0, s_1, \ldots, s_m$ to a state $S' \equiv s_0, s_1, \ldots, s_m, s_{m+1}$. However, if s_{m+1} is among s_0, s_1, ..., s_m, say $s_{m+1} = s_i$ for $0 \leq i < m$, then this corresponds to closing a simple cycle, and there is a transition from $S \equiv s_0, s_1, \ldots, s_m$ back to the shorter simple path $S' \equiv s_0, s_1, \ldots, s_i$. In either case, the probability of that transition in the CHMC is the same as the transition in the Markov chain: $\mathcal{T}_{S,S'} = T_{s_m,s_{m+1}}$

To make that abstract discussion more concrete, let us first consider the flux network from Figure 1a. Suppose a CHMC for that flux network is initialized at metabolite 1. The chain is "grown" recursively by adding children metabolites 2 and 3, and so forth, until there are four simple paths ending at metabolite 7. These four nodes encoding metabolite 7 transition back to metabolite 1 in the original flux network and are therefore linked together as upstream edges. These edges complete the simple cycle for each EFM starting and ending at the root node. The resulting CHMC is shown in Figure 2a. The CHMC could have been rooted on any of the other metabolites in the original network as the edge probabilities linking metabolites would remain unchanged. The CHMC for the second example network is shown in Figure 2b. Every EFM appears as a loop somewhere in the transformed network. EFMs involving the initial metabolite occur as precisely one path from root to leaf and then back up to root, such as EFMs 1 and 2. EFMs not involving the root metabolite also occur somewhere in the CHMC, such as EFM 3 which cycles between metabolites 2 and 3. However, other EFMs may occur in multiple parts of the tree. For instance, EFM 4, which is a loop between metabolites 3 and 5, is represented by two cycles in the CHMC: from state

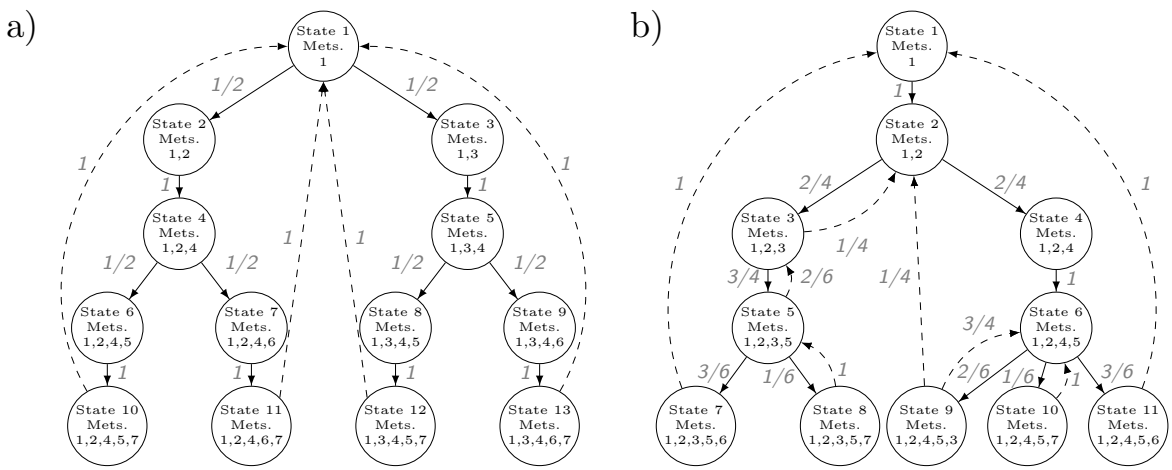


Figure 2: Networks from Figure 1 converted into trees that are arbitrarily rooted on metabolite 1. Nodes and edges denote metabolite sequences and CHMC transition probabilities, respectively. Dashed arrows represent the completion of a simple cycle corresponding to an EFM.

3 to 5 and back to 3, and also from state 6 to 9 and back to 6.

2.4 Computing steady state EFM probabilities

Again, assuming a strongly connected flux network, the corresponding CHMC will also be strongly connected–i.e. every state is reachable from every other. This implies the existence of a unique steady state distribution π. To compute the steady state probability of an EFM k, we identify all transitions in the CHMC that correspond to closing that cycle, E_k. For example, returning to Figure 2b and EFM 4, we have $E_4 = \{(5,3),(9,6)\}$, as those two edges close cycles involving the 3-5-3 metabolic loop. For other EFMs, E_k may contain just a single transition. We define the steady state probability of transiting EFM k as the sum of the steady state probabilities of those transitions occurring, or:

$$P_k = \sum_{(S,S')\in E_k} \pi_S \cdot \mathcal{T}_{S,S'}. \tag{1}$$

In our proposal, the steady state EFM probabilities P_k are proportional to the desired EFM weights w_k and must be scaled appropriately to account for the absolute magnitude of the fluxes. Specifically, if $f_{total} = \sum_{i,j} f_{i,j}$ is the total flux in the network summed across reactions, and if $f_{efm} = \sum_k w_k L_k$ is the total flux obtained by summing EFM weights w_k and their lengths L_k, then our proportionality constant α, where $w_k = \alpha P_k$, should satisfy

$f_{total} = f_{efm} = \alpha \sum_k P_k L_k$, so $\alpha = f_{total} / \sum_k P_k L_k$.

Table 1: Benchmarked optimization-based approaches

Description and code	Objective function	Ref.
Minimize L2 norm (L2)	$\min \sum w_k^2$	[22]
Maximize shortest pathway activity (qSPA)	$\max \sum (L_k w_k)^2$	[27]
Maximize shortest pathway activity (lSPA)	$\max \sum (L_k w_k)$	[28]
Minimize shortest pathway activity (lSPA)	$\min \sum (L_k w_k)$	[29]
Minimize active pathways (milAP)	$\min \sum w_k^0$	[23]

2.5 Comparison to optimization-based methods

We benchmarked the Markov model against five optimization-based methods reported in literature. The LP, QP, and MILP objective functions are described in Table 1. Briefly, these approaches optimize either the zero-, one- or two-norm of the EFM weights, possibly multiplied by the EFM lengths, and subject to constraints that ensure the fluxes are correctly reconstructed. EFM enumeration and optimization programs were implemented in Julia (version 1.6.5) using the commercially-licensed Gurobi solver (version 9.1.2) and all open-source solvers supported in the Julia JuMP package. In total, EFMs weights for the optimization-based methods were computed using eight LP, two QP, and one MILP solver (see Supplemental Information). The full codes to reproduce all analyses can be found at https://github.com/jchitpin/reproduce-efm-paper-2022.

3 Results and discussion

3.1 Markov model uniquely estimates EFM weights in example networks

We first compared the Markov and optimization-based methods on the example networks in Figure 1. In the first example network, the steady state probabilities of a particle occupying metabolic state 7, denoted by the CHMC nodes 10-13 in Figure 2a, were 5% because the path probabilities along each prefix were identical. As the probability of completing each simple cycle was 100% from any of those states, the steady state EFM probabilities under

Table 2: EFM weights from Figure 1a) with solvers shown below.

EFM	Method									
	Markov	Min L2	Max qSPA	Max lSPA			Min lSPA			Min milAP
1	1	1[a]	1[b]	2[c]	0[d]	1[e]	2[f]	0[g]	1[h]	0[i]
2	1	1	1	2	0	1	2	0	1	0
3	1	1	1	0	2	1	0	2	1	2
4	1	1	1	0	2	1	0	2	1	2

[a]COSMO, OSQP; [b]COSMO, OSQP; [c]Gurobi, CDDLib, GLPK; [d]SCIP; [e]OSQP, ECOS, ProxSDP, Tulip; [f]Gurobi, GLPK; [g]SCIP, CDDLib; [h]OSQP, ECOS, ProxSDP, Tulip; [i]Gurobi.

Table 3: EFM weights from Figure 1b) with solvers shown below.

EFM	Method											
	Markov	Min L2	Max qSPA	Max lSPA				Min lSPA				Min milAP
1	1.29	1.8[a]	1.5[b]	2[c]	1[d]	1.5[e]	1.8[f]	2[g]	1[h]	1.5[i]	1.8[j]	1[k]
2	1.71	1.2	1.5	1	2	1.5	1.2	1	2	1.5	1.2	2
3	0.71	0.2	0.5	0	1	0.5	0.2	0	1	0.5	0.2	1
4	1.71	1.2	1.5	1	2	1.5	1.2	1	2	1.5	1.2	2
5	1.00	1.0	1.0	1	1	1.0	1.0	1	1	1.0	1.0	1
6	0.29	0.8	0.5	1	0	0.5	0.8	1	0	0.5	0.8	0

[a]COSMO, OSQP; [b]COSMO, OSQP; [c]Gurobi, GPLK; [d]SCIP, CDDLib; [e]Tulip; [f]OSQP, ECOS, ProxSDP; [g]Gurobi, GLPK; [h]SCIP, CDDLib; [i]Tulip; [j]OSQP, ECOS, ProxSDP; [k](1) Gurobi.

the Markov model were 5%. Scaling the EFM probabilities to weights yielded the unique solution $w = [1111]$. In contrast, the solutions for the five optimization-based methods, presented in Table 2, returned three feasible EFM weight solutions across all eleven solvers. Of the five objective functions used, minimizing the two QP objective functions, across all solvers, returned only the Markovian solution. However, the LP-based programs returned three solutions across all eight solvers, including the Markovian weights. Given all EFMs were of the same length, neither LP objective function constrained the solution space and could have returned an infinite set of decimal-valued EFM weights. The MILP method was the only one that failed to return the Markovian solution and returned one of two feasible solutions that minimized the number of active EFMs.

The EFM probabilities were next computed for the second example network shown in Figure 1b. The EFM probabilities for Figure 2b were $P_k = [0.19, 0.26, 0.11, 0.26, 0.15, 0.04]$

which, when rescaled by proportionality constant α, yielded $w_k = [1.29, 1.71, 0.71, 1.71, 1.00, 0.29]$. This Markovian solution differed greatly from all five optimization-based methods presented in Table 3. Of the four feasible solutions returned across all objective functions, none were the Markovian solution and methods that returned only a single solution were different from each other. Collectively, these results demonstrated that the Markovian constraint identified EFM weights that were unique and distinct from the other optimization-based methods, which returned multiple feasible solutions depending on the objective function and solver.

3.2 Applications to a sphingolipid network

We next compared the Markov model and optimization-based methods to assign EFM weights in an empirically-validated network of sphingolipid metabolism of non-specific human tissue across nine subcellular compartments [30]. Figure 3 shows the model which consists of 69 reactions across 39 lipid species and kinetic parameters for the wildtype and Alzheimer's disease condition. The within-compartment reactions were modelled by Michaelis-Menten kinetics, while transport reactions across compartments were modelled by mass-action kinetics. This sphingolipid network is an ideal unimolecular model because reactions either move a lipid family from one compartment to another or transform one lipid family into another by the addition or removal of an N-acyl fatty acid or phosphocholine head group. Although Figure 3 suggests an open-loop network, source reactions 64-68 and sink reaction 6 contained zero-valued kinetic parameters in both conditions and therefore no flux into or out of the system. The kinetic parameter for source reaction 1 was zero in the wildtype condition and set to zero (from $V_m = 0.04$) in the disease condition to ensure a closed-loop network. Altogether, there were 62 remaining reactions involving 37 lipid species.

The full ordinary differential equation models provided in the supplementary material by Wronowska et al. were reproduced and solved to identify the the steady state lipid concentrations in both conditions. The steady state fluxes in both models were next computed from the kinetic parameters and steady state lipid concentrations. We noticed that all pairs of reversible transport reactions contained forward fluxes that were several orders of magnitude larger than the reverse fluxes. These reactions were simplified into net reactions to avoid enumerating additional EFMs with negligible weights. The Markov model was

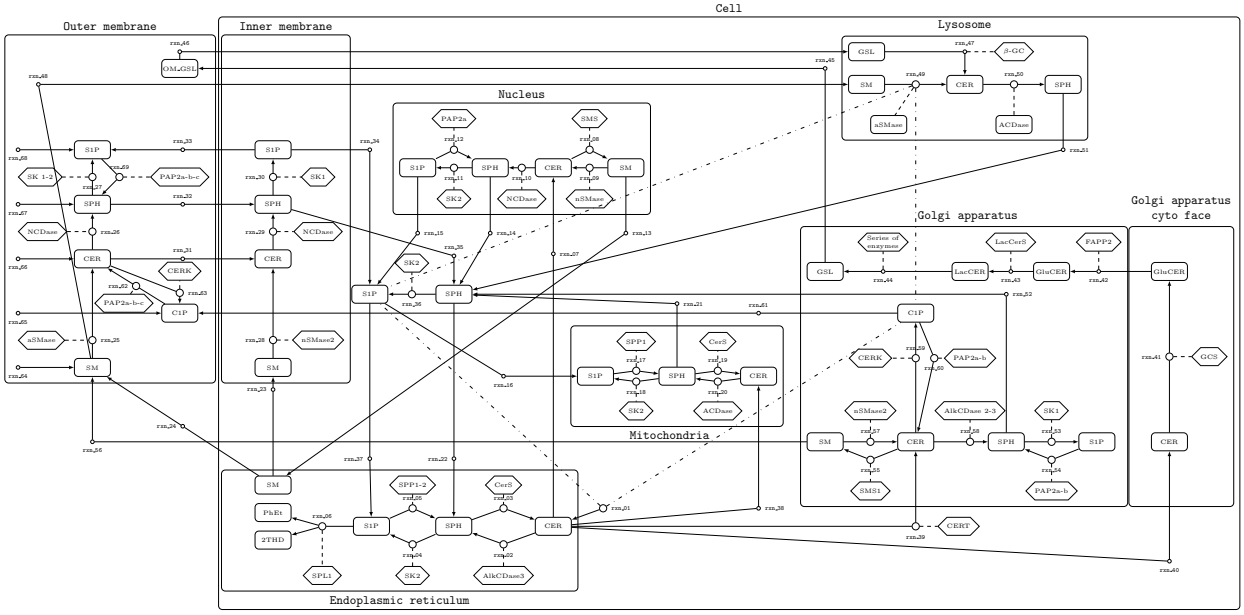


Figure 3: Sphingolipid kinetic model adapted from Wronowska et al. Oval boxes denote lipid families and hexagonal boxes denote enzymes. Small circles denote transport reactions modelled by mass-action kinetics. Bigger circles denote metabolic reactions modelled by Michaelis-Menten kinetics. Solid lines connect metabolites and reactions, dashed lines denote the reaction enzyme, and dash-dotted lines denote inhibition.

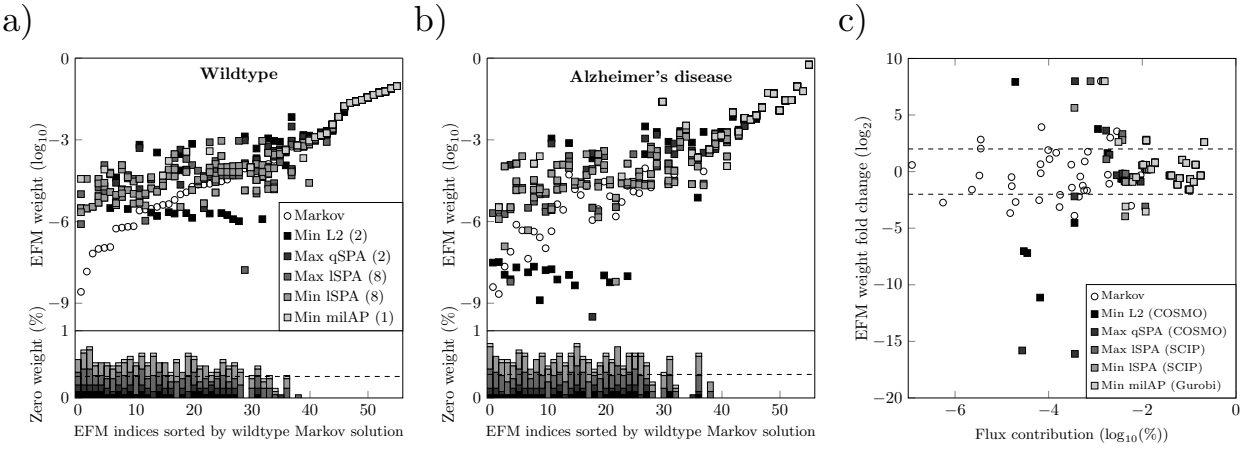


Figure 4: EFM weights under the Markov constraint versus five objective functions from optimization-based methods across all solvers for the a) wildtype, b) disease condition, c) fold change with respect to wildtype with a representative solver from each method The optimization-based method codes are described in Table 1. Numbers in parentheses correspond to the solvers used for each mathematical optimization. Histogram bars show the frequency of zero-valued EFM weights from the optimization-based methods normalized across all 21 objective functions and solvers.

used to both enumerate the 55 network EFMs (see Supplemental Information) and compute their weights. These values were compared against EFM weights calculated from the five optimization-based methods across all solvers.

Figure 4 shows the EFM weights across all tested methods for the wildtype and disease conditions, where all methods, including ours, returned quite different solutions. The greatest agreement in EFM weights, across methods and solvers, was observed for the top 9 EFM weights in the wildtype condition of Figure 4a. These EFMs explained 85.8% of the total network fluxes and consisted nearly exclusively of two-step reversible reactions between lipid families within a compartment. Their weights were consistent across methods because the large fluxes between pairs of reversible reactions can only be explained by these two-step pathways. Beyond these reversible reactions, there was much less agreement for the remaining EFMs consisting of longer reactions which varied by up to three orders of magnitude between solvers and across objective functions. As an increase in one EFM weight must be balanced by a decrease in another to satisfy flux conservation, EFMs assigned very large values must force others towards zero. Across all optimization-based methods and solvers, we observed a 32% probability of assigning a zero-valued weight to each of the 55 EFMs. Different objective functions displayed varying tendencies towards sparse solutions with the mixed-integer and maximizing shortest pathway activity by quadratic programming functions assigning the most zero-valued weights on average (29 and 26.5). Both linear programming methods that either maximized or minimized the shortest pathway activity assigned the same average number of zero-valued weights in the wildtype condition. These observations were consistent in the Alzheimer's disease condition in Figure 4b with high agreement for the same top 9 EFM weights shared in the wildtype condition. These EFMs explained 86.7% of the total network flux with greater disagreement within solvers and across methods for the remaining EFMs. Minimizing the L2 norm, for instance, yielded weights that were up to four magnitudes smaller than the Markov solution or other objective functions. This larger variability corresponded to a 35% probability of any optimization-based method assigning a zero-valued weight. In contrast, our Markov solution is not based on sparsity or parsimony, so all EFMs receive some non-zero weight, regardless of their length or how much flux they explain in the network. All EFM weights assigned by our method are comparable across conditions and consistently identifiable under the Markov constraint as opposed to the objective functions, whose weights are strongly dependent on the chosen solver. We observed that the largest 13 EFM weights explained 95% of the fluxes in both conditions and that 11 of them were

shared between wildtype and Alzheimer's disease (see Supplementary Information). One of the two largest EFMs unique to the wildtype and Alzheimer's disease were two-step reactions cycling sphingosine and sphingosine-1-phosphate in the outer membrane and sphingomyelin to ceramide in the nucleus, respectively. The other largest EFM unique to the wildtype condition was a five reaction EFM carrying flux through the nucleus, cytoplasm, and endoplasmic reticulum, while the largest EFM unique to Alzheimer's disease also consisted of five reactions but through the mitochondria, cytoplasm and endoplasmic reticulum.

Given the majority of network fluxes across both conditions were explained by the same 11 EFMs, we next investigated whether metabolic differences were explainable through other differentially active EFMs. Figure 4c shows the fold change between the Alzheimer's disease and wildtype condition EFM weights as a function of the total explained flux in the wildtype condition. Weights were selected from the objective function solver that best reconstructed the individual and total fluxes across both networks (see Supplemental Information). The top 9 EFM weights explaining the majority of the network fluxes are located on the right-most cluster of Figure 4c where there is high agreement across all methods. Nevertheless, only one of those EFMs showed over four-fold increase in the Alzheimer's disease versus wildtype condition. This two-step EFM pathway corresponded to the remodelling of sphingosine and sphingosine-1-phosphate in the endoplasmic reticulum. These data indicated that the sphingolipidome in Alzheimer's disease is altered through pathways carrying smaller units of steady state flux. According to the Markov solution, we observed extreme fold changes ($|\log_2 FC| \geq 4$) for 17 EFMs that explained between $10^{-6} - 0.1\%$ of the total network fluxes. In contrast, there were fewer, and possibly more misleading, fold changes for the optimization-based methods due to the sparsity of EFM weights for the disease and wildtype conditions. Both linear and mixed-integer programming methods showed between 6-10 extreme fold changes that were limited to EFMs explaining fluxes contributing 0.001% or greater in the wildtype network. The quadratic programming method yielded 8-10 fold changes that were more variable than the Markov method, yet only for EFMs that explained 0.00001% of the total wildtype fluxes.

4 Conclusions

In this article, we explored the longstanding computational problem of uniquely identifying EFM weights from steady state flux networks. Although this problem has mainly been addressed by optimization-based methods, we showed how different objective functions may nonetheless fail to properly constrain the solution space of EFM weights. We then proposed a Markovian constraint under our CHMC model, wherein the probability of a particle transiting through a unimolecular network is proportional to the observed, steady state fluxes and leads to a unique probability of transiting each EFM. In comparison, other optimization-based methods not only disagree on different "parsimony" principles for constraining solutions, but different numerical solvers for optimization-based methods can arbitrarily return different, equally-good solutions. In contrast with the objective functions described in literature, which assumes fluxes transit predetermined paths through the network, Markov chain models are well-established in the stochastic chemical kinetic literature. Furthermore, several simplifying assumptions we made are easily lifted. Our analyses were applied to closed-loop networks, for example, yet open networks may be analyzed similarly. Also, networks with separate components or transient metabolite species can similarly be analyzed.

We applied our Markov method to identify differentially active EFM weights in a model of sphingolipid metabolism in Alzheimer's disease and compared the results against optimization-based methods. While all methods strongly agreed on the top 9 (16%) EFM weights, the Markov and objective functions greatly differed for the remaining EFMs. The Markov method showed large fold changes between EFM weights, regardless of their contribution to the total network flux. In contrast, the objective functions largely returned extreme fold changes between EFMs that explained the greatest network flux. Upon inspection, these EFMs consisted primarily of two-step remodelling pathways between pairs of lipid families, while the EFMs explaining less network flux contained more reactions transiting multiple compartments. Taken together, the Markov results suggest that longer pathways containing less network flux are equally relevant to explaining metabolic differences between wildtype and Alzheimer's disease under the sphingolipid kinetic model.

There exist several avenues of future work to extend our Markov chain model. First,

our method is only applicable to metabolic networks of unimolecular reactions. While these networks do occur in biology, such as the sphingolipid network analyzed, expanding our methodology to higher order reactions would enable one to compute EFM weights for any metabolic network. Currently, one must simplify higher-order reactions into a series of elementary reactions, or remove co-enzymes, co-factors, and other (highly abundant) metabolic byproducts from the reaction equations. Second, algorithmic developments are required to scale our Markov method to genome-wide networks containing thousands to millions of EFMs. The Julia code we provide can only feasibly compute weights for networks containing several thousand EFMs (see Supplemental Information). Third, we anticipate our work will encourage differential analysis of EFM weights to quantify metabolic changes across conditions in the future. Previous analyses have typically computed EFM weights using a variety of objective functions and used a consensus-based approach to identify active EFMs that citeOrmanMehmetA2011Paol,SchwartzJean-Marc2006Qema. As our method uniquely identifies EFM weights, we foresee a deeper analysis of EFM weights and their incorporation in statistical learning models to study differential metabolism.

Funding

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), Discovery grant RGPIN-2019-0660 to T.J.P. J.G.C. was supported by an NSERC CREATE Matrix Metabolomics Scholarship and an NSERC Alexander Graham Bell Canada Graduate Scholarship.

Conflict of Interest: none declared.

References

1. Podrini, C. *et al.* Dissection of metabolic reprogramming in polycystic kidney disease reveals coordinated rewiring of bioenergetic pathways. *Commun. Biol.* **1,** 194–194 (2018).

2. Chen, W. W., Freinkman, E., Wang, T., Birsoy, K. & Sabatini, D. M. Absolute Quantification of Matrix Metabolites Reveals the Dynamics of Mitochondrial Metabolism. *Cell* **166,** 1324–1337.e11 (2016).

3. Lau, A. N. *et al.* Dissecting cell-type-specific metabolism in pancreatic ductal adenocarcinoma. *eLife* **9,** 1– (2020).

4. Brunk, E. E. *et al.* Recon3D: A Resource Enabling A Three-Dimensional View of Gene Variation in Human Metabolism. *Nat. Biotechnol.* **36,** 272–281 (2018).

5. Duarte, N. C. *et al.* Global Reconstruction of the Human Metabolic Network Based on Genomic and Bibliomic Data. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 1777–1782 (2007).

6. De Martino, D., Mc Andersson, A., Bergmiller, T., Guet, C. C. & Tkačik, G. Statistical mechanics for metabolic networks during steady state growth. *Nat. Commun.* **9,** 2988–2988 (2018).

7. Fernandez-de-Cossio-Diaz, J., Leon, K. & Mulet, R. Characterizing steady states of genome-scale metabolic networks in continuous cell cultures. *PLoS Comput. Biol.* **13,** e1005835– (2017).

8. Schwartz, J.-M., Otokuni, H., Akutsu, T. & Nacher, J. C. Probabilistic controllability approach to metabolic fluxes in normal and cancer tissues. *Nat. Commun.* **10,** 2725–2725 (2019).

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

10. Zanghellini, J., Ruckerbauer, D. E., Hanscho, M. & Jungreuthmayer, C. Elementary flux modes in a nutshell: Properties, calculation and applications. *Biotechnol. J.* **8,** 1009–1016 (2013).

11. Carlson, R. & Srienc, F. Fundamental Escherichia coli biochemical pathways for biomass and energy production: Identification of reactions. *Biotechnol. Bioeng.* **85,** 1–19 (2004).

12. Srivastava, M. *et al.* Aspergillus fumigatus Challenged by Human Dendritic Cells: Metabolic and Regulatory Pathway Responses Testify a Tight Battle. *Front. Cell. Infect. Microbiol.* **9,** 168–168 (2019).

13. Hunt, K. A., Jennings, R. M., Inskeep, W. P. & Carlson, R. P. Multiscale analysis of autotroph-heterotroph interactions in a high-temperature microbial community. *PLoS Comput. Biol.* **14,** e1006431–e1006431 (2018).

14. Müller, S., Széliová, D. & Zanghellini, J. Elementary vectors and autocatalytic sets for resource allocation in next-generation models of cellular growth. *PLoS Comput. Biol.* **18,** e1009843–e1009843 (2022).

15. Narisetty, V. *et al.* Valorisation of xylose to renewable fuels and chemicals, an essential step in augmenting the commercial viability of lignocellulosic biorefineries. *Sustain. Energy Fuels* **6,** 29–65 (2021).

16. Klamt, S. & Stelling, J. Combinatorial Complexity of Pathway Analysis in Metabolic Networks. *Mol. Biol. Rep.* **29,** 233–236 (2002).

17. Acuña, V. *et al.* Modes and cuts in metabolic networks: Complexity and algorithms. *Biosystems* **95,** 51–60 (2009).

18. Klamt, S., Stelling, J., Ginkel, M. & Gilles, E. D. FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics* **19,** 261–269 (2003).

19. Hunt, K. A., Folsom, J. P., Taffs, R. L. & Carlson, R. P. Complete enumeration of elementary flux modes through scalable demand-based subnetwork definition. *Bioinformatics* **30,** 1569–1578 (2014).

20. Van Klinken, J. B. & Willems van Dijk, K. FluxModeCalculator: an efficient tool for large-scale flux mode computation. *Bioinformatics* **32,** 1265–1266 (2016).

21. Buchner, B. A. & Jürgen, Z. EFMlrs: a Python package for elementary flux mode enumeration via lexicographic reverse search. *BMC Bioinformatics* **22,** 547–547 (2021).

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

23. Nookaew, I. *et al.* Identification of flux regulation coefficients from elementary flux modes: A systems biology tool for analysis of metabolic networks. *Biotechnol. Bioeng.* **97,** 1535–1549 (2007).

24. Huang, Y., Xie, Y., Zhong, C. & Zhou, F. Finding branched pathways in metabolic network via atom group tracking. *PLoS Comput. Biol.* **17,** e1008676–e1008676 (2021).

25. Coleman, D. N. *et al.* Multifaceted role of one-carbon metabolism on immunometabolic control and growth during pregnancy, lactation and the neonatal period in dairy cattle. *J. Anim. Sci. Biotechnol.* **12,** 27–28 (2021).

26. Schuster, S., Dandekar, T. & Fell, D. A. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* **17,** 53–60 (1999).

27. Orman, M. A., Berthiaume, F., Androulakis, I. P. & Ierapetritou, M. G. Pathway analysis of liver metabolism under stressed condition. *J. Theor. Biol.* **272,** 131–140 (2011).

28. Rügen, M., Bockmayr, A., Legrand, J. & Cogne, G. Network reduction in metabolic pathway analysis: Elucidation of the key pathways involved in the photoautotrophic growth of the green alga Chlamydomonas reinhardtii. *Metab. Eng.* **14,** 458–467 (2012).

29. Ren, L., Sun, X., Zhang, L., Zhao, Q. & Huang, H. Identification of active pathways of Chlorella protothecoides by elementary mode analysis integrated with fluxomic data. *Algal Res.* **45,** 101767– (2020).

30. Wronowska, W., Charzyńska, A., Nienałtowski, K. & Gambin, A. Computational modeling of sphingolipid metabolism. *BMC Syst. Biol.* **9,** 47–47 (2015).