Identification of metabolic pathways using pathfinding approaches: a systematic review

Zeyad Abd Algfoor, Mohd Shahrizal Sunar*, Afnizanfaizal Abdullah, and Hoshang Kolivand

*Corresponding author. Mohd Shahrizal Sunar, MaGIC-X (Media and Games Innovation Centre of Excellence), UTM-IRDA Digital Media Centre, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia. E-mails: shahrizal@utm.my

Abstract

Metabolic pathways have become increasingly available for various microorganisms. Such pathways have spurred the devel-opment of a wide array of computational tools, in particular, mathematical pathfinding approaches. This article can facilitate the understanding of computational analysis of metabolic pathways in genomics. Moreover, stoichiometric and pathfinding approaches in metabolic pathway analysis are discussed. Three major types of studies are elaborated: stoichiometric identifi-cation models, pathway-based graph analysis and pathfinding approaches in cellular metabolism. Furthermore, evaluation of the outcomes of the pathways with mathematical benchmarking metrics is provided. This review would lead to better comprehension of metabolism behaviors in living cells, in terms of computed pathfinding approaches.

Key words: pathfinding algorithms; stoichiometric approaches; graph methods; metabolic pathways; identification

Introduction

Metabolism is a biological subsystem in cells, which is respon-sible for extracting the energy. The process of generating energy and necessary materials is considered as a highly complex cel-lular process. Enzymes play a critical role in catalyzing bio-chemical reactions within the cell [1–3]. Cooperatively, the reactions of these cellular machinery components produce en-ergy metabolism. Metabolism integrates all parts of cells, and its study is important to understand the function of the system and furthermore, to understand alterations that occur in dis-ease state, [4] and, hence, for subsequent applications in drug discovery [5]. Therefore, the reconstruction of genome-scale metabolic graph representation from genomics and other mo-lecular or biochemical data is now feasible.

In biological research, one of the most important topics is identifying different metabolic pathways within species, which might be exposed to subtle shifts or malfunctions. Metabolism has many important activities that can lead to finding the drug resistance of pathogenic bacteria. However, identifying a pathway in the laboratory is a complicated process. The process includes difficult subtasks such as metabolic flux analyses [6] and labeling techniques for dynamic metabolism profiling [7]. All these require advanced technologies, which are expensive and time consuming. Another direction is to make use of the quali-fied approach by comparing metabolic networks of related spe-cies. A shorter pathway might be preferred over a longer pathway because of the smaller number of enzymes required. This is both advantageous because the genome capacity is often limited and because the shortest pathway is more likely to be conserved in evolution. Hence, we classified these approaches into stoichiometric approaches and pathfinding approaches. Stoichiometric approaches, which underlie flux balance analysis (FBA) approach, elementary flux modes (EFMs) and extreme pathways (EPs), define a metabolic pathway as a minimal set of

Zeyad Abd Algfoor is a PhD candidate in Computer Science from Universiti Teknologi Malaysia.

Mohd Shahrizal Sunar obtained his PhD from National University of Malaysia. He is currently working as a Associate Professor at the Department of Computer Graphics & Multimedia, Universiti Teknologi Malaysia.

Afnizanfaizal Abdullah received his PhD in Computer Science from Universiti Teknologi Malaysia. He is currently working as senior lecturer at the Department of Software Engineering, Universiti Teknologi Malaysia.

Hoshang Kolivand received his PhD in Computer Science from Universiti Teknologi Malaysia. He is currently working as a lecturer at the Department of Computer Graphics & Multimedia, Universiti Teknologi Malaysia.

biochemical reactions in pseudo steady state. Pathfinding approaches view a metabolic pathway as a set of biochemical re-actions that contain a directed path from a given source com-pound to a given target compound [8–10].

However, pathfinding relies on a metabolism network where compounds and reactions are nodes linked by edges represent-ing substrate/product relationships [11–13]. Various algorithms can be used to the find single shortest path or k-shortest paths between a given pair of start and end nodes. Experimental find-ings on various organisms indicate that (despite the presence of many possible biochemically feasible pathways) organisms often have a single 'preferred' pathway when converting a source com-pound into a target compound [14]. However, this terminology is not unique, and many authors described these pathways using diverse terms, for instance, annotated pathway [15], consensus pathway [8], empirically elucidated pathway [16] and experimen-tally determined pathway [17]. Of course, it is important to note that from the physiological viewpoint, metabolism pathways do not operate in isolation, and many pathways work together to produce an overall global flux distribution in reactions. Much re-search has been proposed to analyze pathways based on graph models for identifying biologically relevant pathways in meta-bolic pathways [18]. Graph-based metabolism pathfinder algo-rithms complement stoichiometric approaches, as they focus on different aspects of modeling and understanding metabolism [19–21]. In addition, they are typically used for modeling specific organisms or metabolism systems [22]. Most models are based on the steady-state assumption and, therefore, require explicit labeling of internal and external compounds [23].

In recent years, there has been an increasing amount of lit-erature on the metabolic pathway model's design. Some model design have problems with large-scale data for cell biology and the applications of graph theory [24]. The authors used compu-tational approaches to reconstruct the genome-scale metabolic models into numerical graphs. In this article, we discuss the main purpose of representing cell biology information in a mathematical graph model form. Moreover, to present biologic pathway data in a useful graph form, the characteristics and mechanisms of the inner workings of the cell must be con-sidered [25]. Although the pathway data sets have been col-lected worldwide through observations and experiments, there is still a lack of satisfying explanations and theories to give them sound biological meaning [26]. The varieties of data sets available have expanded from a handful in the mid-1990s to a several thousand today [27]. These data sets are categorized into main roots such as MetaCyc, KEGG, Reactome, Model SEED and BiGG families [28].

This systematic review focuses on two major points: first, it highlights the most relevant mathematical metabolism path-way models with pathfinding approaches; second, it exports the progress of pathfinding algorithms on metabolic pathways and illustrates the formalization and constraint rules. The rest of the article is organized as follows. Section 2 reviews the relevant previous surveys on metabolism, Section 3 briefly reviews the details of metabolic pathways, Section 4 reviews the pathfind-ing approaches, Section 5 validates the measurements for pathfinding approaches and Section 6 discusses the limitations and gaps in this area. The article is concluded in Section 7.

Stoichiometric approaches

There are two key distinctions between stoichiometric approaches and pathfinding approaches. First, the pathfinding approaches focus on searching a directed path within the metabolic pathway, while stoichiometric approaches seek the complete pathway. Second, pathfinding approaches do not make direct use of reac-tion stoichiometry, while stoichiometric approaches do.

There are two main categories of stoichiometric approaches: flux balance analyses (FBAs) and EFMs. The FBA is a fundamen-tal computational framework applied to metabolic networks, derived from the steady-state assumption and mathematically defined metabolic pathways. The EFMs are based on analyzing metabolic networks from a pathway-oriented perspective.

FBA approach

FBA is a cornerstone in mathematical optimization methods for metabolic networks [29–32]. The main objective of this model is to achieve accuracy in the reaction metabolic network based on some conditions: First, the metabolic network must be complete and must fully cover the metabolic network capabilities to be ready for modeling. Second, the metabolic network should have no gaps or dead ends, and the representation of the metabolic network based on the metabolic phenotype of the organism must be correct. Typically, this objective function is the maxi-mization of the flux through the biomass formation reaction. Most experiments done for this approach use a mixed-integer linear program formulation [33], and with a variety of data sets such as KEGG pathway [34] and MetaCyc [35] as in Figure 1.

The most significant feature of FBA is its ability to make quantitative predictions without any need for detailed kinetic descriptions, needing only the stoichiometry of reactions. The network stoichiometry (metabolic model), a biological objective and the growth and environmental conditions (substrate avail-ability) are the only necessary inputs for FBA. One of the draw-backs of the FBA approach is that it is somewhat limited in its predictive power, unless additional constraints are appended to the optimization problem. This approach can be systematically extended into several techniques, such as regulatory FBA [36], steady-state regulatory FBA [37], gene inactivity moderated by metabolism and expression [38] and probabilistic regulation of metabolism [39], all of which have been developed to integrate regulatory information with the metabolic network.

The reason for using these models is to enable the mathemat-ical representation of the bio-transformations and metabolic processes occurring within the organism. Although a large num-ber of methodologies have been proposed, such as flux variability analysis, linear optimization for identifying the maximum bio-mass [40] and flux coupling analysis [41], constraints imply that not all fluxes in a metabolic network can vary independently.

Elementary flux modes

EFMs is a computational technique to analyze metabolic path-ways in metabolic networks [42–44]. EFMs can be defined as a minimal set of enzymes able to operate at a steady state, with the enzymes weighted by the relative flux they need to carry for the mode to function [9]. One significant aspect of EFM is that every flux distribution can be decomposed into a set of EFMs, and a number of the methods to study flux distributions origin-ate from it [45, 46]. Continuous efforts have been made to im-prove the computational speed and memory demand required to compute EFMs. Nonetheless, the distribution of any particu-lar steady-state flux can be represented as a nonnegative linear combination of elementary modes [47, 48].

In the metabolic network, the internal reversible reactions are decomposed into pairs of opposite irreversible reactions, where the corresponding flux cone is augmented and the run of

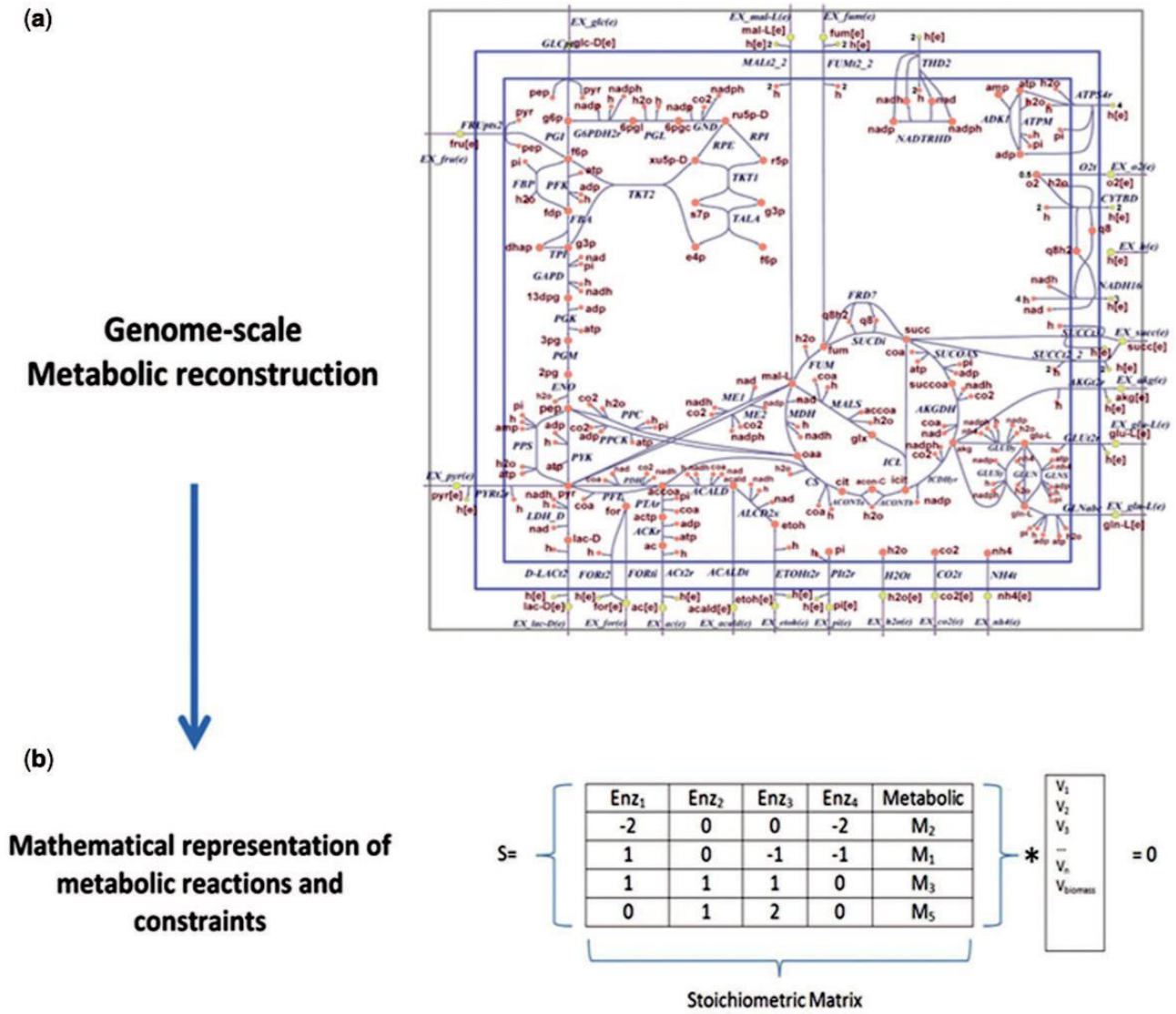

Figure 1. (A) A metabolic network reconstruction consists of a list of stoichiometrically balanced biochemical reactions. (B) This reconstruction is then converted into a mathematical model by forming a matrix (labeled S), in which each row represents a metabolite and each column represents a reaction.

the algorithm for the enumeration of extreme rays produces a minimal generating set whose augmented space is known as EPs [49]. EPs in the original flux cone are not conically independ-ent, unless there are no internal reversible reactions. Furthermore, EPs are used in the analysis of the metabolic capa-bilities of red blood cell metabolism [50–52].

Finally, each of the EFMs and EPs represent a more general and elegant concept for metabolic pathways than paths [9, 49]. Both approaches, in terms of computational complexity, are more expensive in large metabolic networks. Despite several ef-forts that have been made to make EFMs as a practical tool to analyze huge metabolic pathways, these approaches still need more improvements [53–57].

Graph-based metabolic pathfinding

In graph-theoretic metabolic modeling, the main consideration is the connectivity of metabolism network. The basic abstrac-tion level of metabolic networks can be represented as mathematical graphs, using nodes to represent metabolic com-ponents, and edges to represent their various types of inter-actions [58].

Generally, the exploratory analysis in graph theory, where a metabolic network is queried to discover connections between metabolites, is called metabolic pathfinding. Here, we introduce some notations from basic graph theory:

- Let G ¼ (V, E) be a weighted graph; V is a node of Vertex set and E is the edge set, where E V V.
- ð u; vÞ 2 E is the degree of the node.
- W is an edge weight function, where E ! $R^{þ}$.

In the context of a metabolite graph, there are the collections of reactions R and metabolites M for an organism. The number of reactions will be determined by the consumption and pro-duction of the metabolite's processes. Meanwhile, the metabol-ite graph context will be as follows:

- m_i 2 M Set of consumption or production by each reaction r_j 2 R
- d_{ij} 2 R Stoichiometric constants, if $d_{ij} > 0$ r_j produces m_i or if $d_{ij} < 0$, r_j consume m_i

- The substrates S ðri Þ and target products Pð ri Þ of a reaction r_i are defined
- as S r_j ¼ f $m_i$jd$_{ij}$ < 0g and P r_j ¼ f m_i jd$_{ij}$ > 0g, respectively.
- A node u_r 2 V_R for each reaction, a node u_m 2 V_M for each sub-strate or product m 2 Sð rÞ [Pðr Þ of each reaction.

An edge ðu$_r$; u$_m$Þ 2 E for producing metabolite m, and ð u $_m$; u$_r$Þ 2 E for consuming m.

In a metabolic pathway graph context, a node to reaction or metabolite can be related. In reaction cases, reactions are represented as a directed graph G_R, and the edge r_i; r_i 2 E exists whenever P $_{ð}^{r_i}$ Þ \ S r_j 6¼ ø. In the metabolite case, metabolites are represented as a directed graph G_M and the edge m_i; m_j 2 E whenever m_i 2 Sðr Þ; m_j 2 Pð rÞ for some reactions.

Several graph-based methods have been proposed for search-ing and eventually enumerating pathways in metabolic networks, such as Pathfinding [20–22] and the Pathway Hunter Tool [59]. These methods are computing pathways and the shortest path-ways in graphs compared with hyper-graphs. Practically, these methods concern only the main substrates (start compounds), and the main products (target compounds) are considered during construction of the pathways. In addition, these main compounds are unlike the cofactors such as co-substrates and co-products.

Subgraph based-metabolic pathway

A subgraph is represented as a subset of nodes with a specific set of edges connecting them, and the total number of subgraphs exponentially depends on the set of nodes as in Figure 2B. Therefore, efficient and scalable heuristics are developed for de-tecting the given subgraphs and their frequencies in large meta-bolic networks. Many researchers mention that it is equally important to understand the organization of infrequently observed subgraphs [60–62].

In particular, the highest performance of pathfinding identifi-cation in a subgraph is through a sub-pathway mining module [63]. Substantially, the complex structure of metabolic pathways drives the sub-pathway to become a general problem. Generally, the metabolic pathways are represented as directed graphs, and the protein-to-protein interaction pathways are represented as undirected graphs. On the other hand, it is possible to identify the path based on enzymes, and subgraphs will be represented as un-directed graphs. The disadvantage of undirected subgraphs is that pathfinding approaches will not differentiate between reac-tion products and substrates [62, 64].

Bipartite graph-based-metabolic pathway

Recently, many researchers have analyzed the metabolic net-work by transforming the biochemical reactions into a bipartite graph, where the nodes and links take the form of metabolites and enzymatic reactions. Generally, a metabolic network con-sists of enzymes, reactions and metabolites. Based on a bipart-ite model graph, each enzyme is shared by two nodes, and

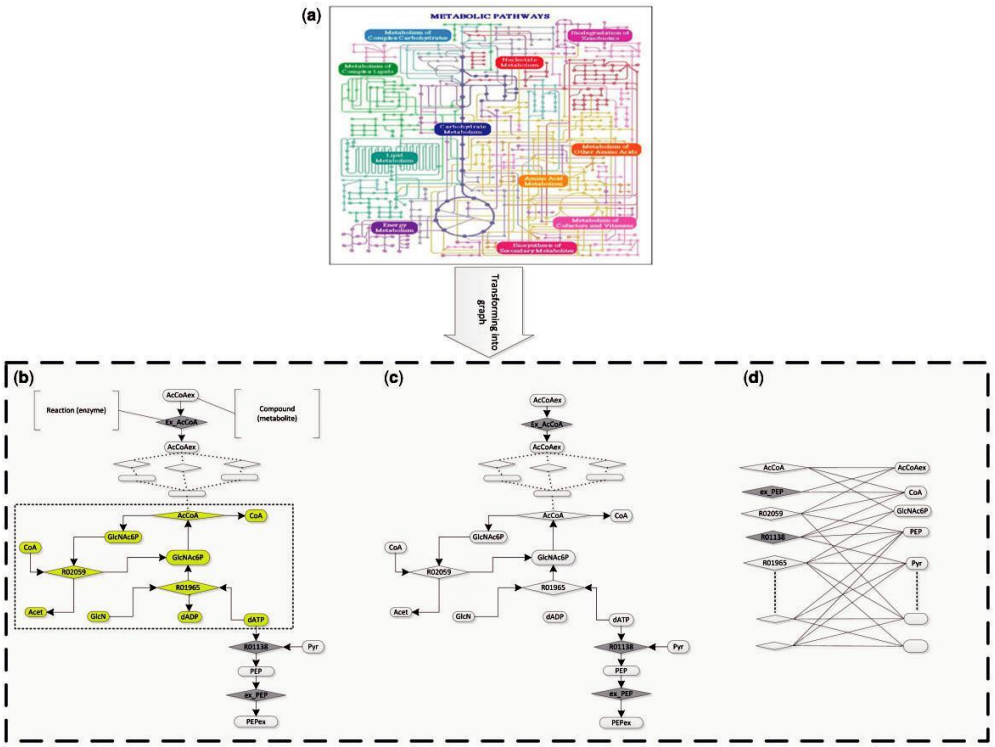

Figure 2. (A) A graph G is metabolic pathway for genome-scale, which consist set of compounds and reactions. (B) The yellow reactions and compounds represent sub-graph. (C) All chemical reactions and compounds in graph G represent the bipartite graph. (D) The hyper-graph, where each compound connects at least more than two reactions.

edges define the biological relationship between a set of metab-olites and enzymes as in Figure 2C. However, this model graph only facilitates drug discovery and ranking of choke points and load points, and both points are used to find enzymes (edges), which uniquely consume or produce a particular metabolite (nodes) [65, 66].

There are online databases available as bipartite graphs, ob-tained from 43 different organisms, which were collected and published in recent studies [12, 67]. The bipartite graph has been used in some works related to pathfinding via the use of the metabolic data set KEGG [68]. RPAIR is a database from KEGG categorized into 'main', 'trans', 'cofac', 'ligase' and 'leave', depending on their roles in a chemical reaction [20].

Hyper-graph based-metabolic pathway

A hyper-graph is a generalization of an ordinary graph where an edge, called a hyper-edge, can connect more than two verti-ces [69], as in Figure 2D. The vertices in the hyper-graph are the compounds, and the hyper-edges are the reactions connecting the compounds. The metabolism of living cells can be repre-sented using a metabolic network in the form of a directed hyper-graph that encodes a set of elementary biochemical reac-tions taking place within the cell. In this hyper-graph, the nodes represent the involved metabolites, and the edges represent the metabolic fluxes or reaction rates.

Identification of the most relevant pathways based on hyper-graphs have been proposed by [70]. The hyper-graph can model the metabolic network where the reactions are repre-sented by hyper-arc. The hyper-arc will be used as an enzyme-catalyzed reaction, which in turn leads to transforming the set of substrates' compounds into product compounds [71]. In a hyper-graph, each hyper-arc connects a set of vertices, corres-ponding to reactants, to a disjoint set of vertices, representing the products. Each hyper-arc corresponds to a reaction that can be catalyzed by an enzyme. The hyper-graph models have al-ready been used to find minimal sets of metabolites sufficient to produce a set of target metabolites [72].

Pathfinding approaches for metabolic pathway identification

An interactive navigation through metabolic networks is pos-sible, given a set of biochemical reactions belonging to a par-ticular organism or cell, by using different algorithms based on artificial intelligence to compute the meaningful metabolic pathways from a source compound to a target compound as in Figure 3. However, searching without information other than the connectivity (for example, two successive reactions are con-nected if they have a metabolite in common) often delivers meaningless results. Several pathfinding algorithms have been proposed, such as A* (A star), best-first search and depth-first-search (DFS; backtracking) [15, 57, 73–76]. Because the algo-rithms do not depend on predefined pathways, they can effi-ciently identify plausible routes using known biochemical transformations.

Metabolic networks in any living organism can be repre-sented as a direct graph. Practically, to formalize the traveling cost in state-space, we need to define the difference between any two compounds as Dx. We denote a compound as x and state-space describe it by a set of chemical descriptors, x_k. Thus, every metabolite m can be placed at a point in hyper-space, which is defined by x ¼ (x_1, x_2, x_3 . . . x_n). The formulation t ¼ Dx, simply state-transition, and jDxj, the distance, that can be calculated using the Euclidean metric or the Manhattan metric are given in Equations (1) and (2), respectively. Both distance metrics (Euclidean or Manhattan) are heuristics functions H that always represent the shortest distance between any two compounds or reactions. However, the Manhattan distance is computationally more efficient than the Euclidean distance in terms of identifying the path in discrete chemical changes [73, 77]. Below are the general formalizations and constraints:

$$jDxj_E \text{ ¼ } \sum_{k¼0}^{k¼N} \text{ð } Dx_k Þ^2 \quad (1)$$

$$jDxj_M \text{ ¼ } \sum_{k¼0}^{k¼N} Dx_k \quad (2)$$

Using metric distance to identify the shortest path leads to evaluation of the functions F, G and H, which are required for a heuristic search [78, 79]. F is the heuristic evaluation function, which can be calculated using different methods, G is the cost if reaching the current state to the initial state and H is the greedy search, which minimizes the cost of reaching the goal state from the current state.

By considering the hypothetical pathway as $p^{0;L}$ ¼ x^0 ! x^1 . . . x^m ! x^L, which begins with the initial state x^0, ends with the final state x^L and has any intermediate state x^m, the total cost can be calculated by G and H at the intermediate state, x^m, as given in Equations (3) and (4), respectively.

$$\text{ð } G\ 0;\ m \text{ Þ ¼ } \sum_{i¼4m}^{x} x^i \quad X^{i\ 1}_{j} \quad (3)$$

$$H\ \text{ð}m;\ L\text{Þ ¼ } jx^m \quad x^L j \quad (4)$$

For pathfinding algorithms, the main target is to identify the path with minimum total cost, F ¼ G þ H, as shown in Equation (5), where G (0, m) is the actual distance through chemical tran-sitions between x^0 to x^m, and H (m, L) is 'estimation' for the shortest path to the target state x^L.

$$F\text{ð}0;\ m;\ L\text{Þ ¼ } G\text{ð}0;\ m\text{Þ þ } H\ \text{ð } m;\ L\text{Þ ¼ } \sum_{i¼1}^{X_{i¼m}} \text{ð}jx^i \text{ þ } x^{i\ 1}j\text{Þ þ } jx^m \quad x^L j \quad (5)$$

The pathfinding approach based on a heuristic search can provide a series of efficient biochemical transformations that transform the start compound into a target compound. The heuristic function will monitor any chemical proximity for an intermediate compound to target compound. Equation (5) is used to evaluate and select the pathway that efficiently con-verts the input compound to the output compound. The effi-ciency of transforming is not based on the length of the path, but is determined by an optimal value of the heuristic function F [73].

Identification-based-weighted pathway

Here, the metabolic network is described as a weighted graph, in which all the compounds are included, and each compound is assigned a weight equal to the number of reactions in which it participates. The simplest one ('unit weight') sets all node weights to one. The complex one ('compound degree weight') penalizes highly connected compounds by assigning each com-pound a weight equal to its degree, while setting each reaction to a weight of one. The result from shortest-pathfinding is also

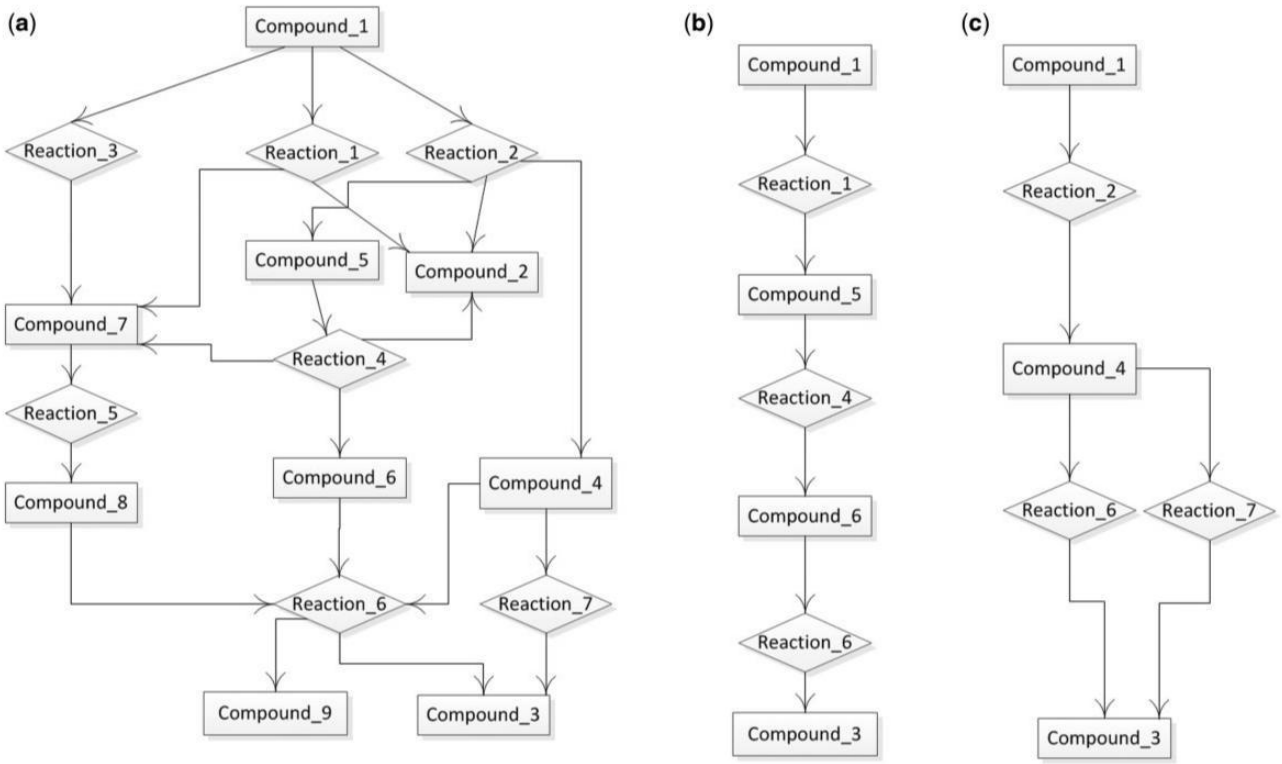

Figure 3. (A) A simple example of metabolic pathway for converting compound_1 'source' into compound_3 'goal'. (B) Shows a possible pathway to convert com-pound_1 into compound_3 and (C) is another possible way to reach the compound_3.

dependent on the weights associated with nodes or edges, which are modeled differently in various approaches [68]. The weights are assigned to the edges based on compound structure similarity [59]. In a degree-weighted graph representing a meta-bolic network [15], each node is assigned a weight equal to its degree. A weighted graph will reduce the probability of finding unfeasible biotransformation paths during the pathfinding pro-cess [80]. Recently, the weighted graph representation has all the compounds included in the graph, whereas the weight (cost) is associated with each compound equaling its connectiv-ity in the entire metabolic network. Pathfinding algorithms will tend to avoid highly connected compounds whenever possible [81, 82].

Several researchers have proposed the use of metabolic pathfinding methods that use weighted atom maps [10, 20, 83– 85]. In metabolic paths, at least one atom is transferred between adjacent metabolites and computed. Additionally, the edges are weighted according to the degree of metabolite nodes, to reduce the number of paths that traverse via high-connectivity metab-olites [86, 87]. To cope with this problem, high-degree metabol-ites can be removed from the network, or edges of a metabolite graph can be weighted according to the degree of metabolite nodes, such that low-degree nodes are preferred [62, 68, 88].

Branched metabolic pathway

Branched pathways are the nonlinear pathways that can arrive at a target compound through combinations of pathways that split compounds into smaller ones, work in parallel with many compounds and join compounds into larger ones. The identifica-tion of branched pathways has a number of important applica-tions in areas that require deeper understanding of metabolism, including metabolic engineering and drug target identification.

Furthermore, branched pathways enable the analysis of meta-bolic processes with a more comprehensive perspective as com-pared with the limited picture provided by linear pathways [89, 90]. However, pathfinding requires specifying a single start and a single end node. It cannot deal with branched pathways or with sets of query reactions. A more challenging question is to predict pathways from multiple seed nodes (e.g. reactions catalyzed by a cluster of co-expressed genes) by extracting the subnetwork that connects them best [20, 57, 62].

Several algorithms have been proposed to handle branched pathways, in a multi-genome scale, and metabolic data such as the KEGG RPAIR data set [91]. BPAT-M is a graph-based algo-rithm for identifying branched metabolic pathways [92]. This al-gorithm is based on two techniques; the first is the rapid search algorithm 'Linear Pathfinding with Atom Tracking', which can track across thousands of reactions and compounds from mul-tiple species; and the second is Branched Pathfinding using Atom Tracking, 'BPATS', which is based on seed pathways [93]. The BPAT-M algorithms will perform better if all branches con-serve at least the given number of atoms and are of similar length. In contrast, the algorithm does not perform well with all pathways such as the 'inosine monophosphate' pathway.

Metabolic pathway tools and systems

Metabolic pathway tools treat a genome as far more than a se-quence and a set of annotations. Moreover, it links the molecu-lar parts list of the cell to the genome, and to a carefully constructed web of functional interactions. The Pathway Tools ontology is defined as an extensive set of object attributes and object relations that allows a rich conceptualization of biology to be represented within a Pathway/Genome database (PGDB),

Table 1. Pathway tools and types of MODs

Tool	Task	Link Accessibility (URL)	Reference
PathFinder	Dynamically represents and provides visualization of biochemical information	http://bibiserv.TechFak.Uni-Bielefeld.DE/pathfinder	[95]
PathMiner	Identifies plausible routes using known biochemical transformations	http://pathminer.uchsc.edu/	[73]
Pathway Hunter Tool (PHT)	Analyzes the shortest paths and calculates the average shortest paths	http://www.pht.uni-koeln.de	[59, 65]
aMAZE	Web interface to the aMAZE relational database	http://www.amaze.ulb.ac.be	[96]
Pathway Prediction System (PPS)	Predicts microbial catabolism of organic compounds	http://umbbd.ahc.umn.edu/predict/	[97]
KEGG genes	Identifies the link between genomic information in the GENES database	http://www.genome.jp/kegg	[98]
MetaRoute	Explores genome-scale metabolic networks	http://www-bs.informatik.uni-tuebingen.de/Services/	[86]
KEGG pathway	Provides reference knowledge for pathway mapping	http://www.genome.jp/kegg/	[99]
Pathway Tools version 13.0	Allows the user to interrogate and explore relationships within the network	http://www.biocyc.org/download.shtml	[94]
Pathway projector	Provides intuitive browser pathway map with the addition of gene and enzyme nodes	http://www.g-language.org/PathwayProjector/	[100]
PathPred	Functions as knowledge-based prediction system	http://www.genome.jp/tools/pathpred/	[101]
Atom tracking system	Enables pathfinding algorithms to avoid unrealistic connections	http://www.kavrakilab.org/atommetanet	[93]
EcoCyc Database	Provides pathway/genome navigator software with the EcoCyc database	http://ecocyc.org/	[102]
MetaCyc Database	Provides a uniquely high-quality resource for metabolic pathways and enzymes	http://metacyc.org/	[103]

which is queried and manipulated by the user [94]. Pathway Tools provides a broad range of functionalities. It can manipu-late genome data, metabolic networks and regulatory networks. For each data type, it provides query, visualization, editing and analysis functions. It also provides model-organism databases (MODs), development capabilities, editors who allow for refine-ment of a PGDB, web publishing and comparative analysis. A family of curated PGDBs has been developed using these tools for modeling important organisms. The software also provides visual tools for analysis of omics data sets, and tools for the analysis of biological networks. Table 1 shows the investigated pathway tools and the common types of MODs.

Evaluating pathfinding approaches

Evaluating pathfinding approaches to metabolic pathways can be hard, even for linear pathways, because there is no standard test set [93]. In other words, the effectiveness of any pathfinding approach is tested by seeing how well it performs with respect to a known metabolic pathway. If the source and goal com-pound and the entire metabolic network are given, how well does a specific pathfinding approach perform at navigating the reactions and compounds involved in a known metabolic path or pathway? Generally, by using benchmarking metrics func-tions commonly used for computed path.

When we use a particular pathfinding approach, it is neces-sary to compare the results with other methods or other works. Thus, we need some measurement to evaluate the accuracy of metabolic pathways. Here, we present some criteria frequently used by researchers for comparison [14, 15, 20, 86]. These crite-ria are used to compare algorithm results with the ground truth labeling, to define the following corresponding values which in-dicate, numerically, correspondence between the computed path and the metabolic path. Here, we introduce some criteria to measure the degree of correspondence between any com-puted path and metabolic path:

- True Positives (TP), in which the total number of metabolic path of reactions and compounds are found (except for source and target nodes whether reaction or compound are not considered).
- False Negatives (FN), signifies the total number of computed path found for number of reactions and compounds.
- False Positives (FP), signifies the total number of reactions and compounds found in the computed path that are not in the metabolic path.

For each pathway, measurements found previously in the literature are used to calculate the

- Sensitivity (Sn) ¼ TP/(TP þ FN), which is the fraction of the reac-tions and compounds in the metabolic path,
- Positive Predictive Value (PPV) ¼ TP/(TP þ FP) and
- Accuracy (Ac) ¼ (Sn þ PPV)/2

The above measurement criteria are used to compare the com-puted path (by pathfinding approach) and the metabolic path or pathway. Next to these criteria, we must take into account the exe-cution time and the memory overhead for pathfinding approaches. Table 2 shows the summary results for pathfinding approaches in different databases. The most common data set that has been used frequently by researchers is the KEGG network. For each of A* (A-star) search, breadth-first search (BFS) and depth-first search (DFS), heuristic search algorithms have been used to predict meta-bolic pathways that are based on a biochemical state-space. The results show that using these algorithms always produces bio-chemical paths that are optimal in terms of the heuristic. Recently, DFS search strategy was used to generate EFMs, where its shows the scalability, low memory overhead, a CPU-limited problem and can incorporate additional flux constraints to generate a full subset of EFMs of interest. Another approach of pathfinding is k-shortest

Table 2. Summarizing results for pathfinding approaches

Algorithm	Operating system	Year	Time		Average (Best)		Accuracy	Data set	Reference
			Best performance	Worst performance	Sn	PPV			
3A*Linux7. operating system.		2003	831. ms	906. ms	——	——	——	KEGG network	[73]
BFS	Sony PCG-C1MW laptop.		6. ms76	19. ms42	——	——	——		
DFS	Transmeta CrusoeTM5800 Processor.		1. ms64	104 ms>	——	——	——		
384 MB of main memory.									
DFS	——	2006	——	——	2%92.	1%95.	Most accurate with EcoCyc 7%pathways(weightedgraph)93.	KEGG and EcoCyc networks	[15]
k-shortest path	——	2009	——	——	100%	100%	Reaction–reaction for 10-k paths 100%	Escherichia coli	[14]
					46%	77%	Compound–compound for 10-k paths 62%		
BPAT-M	XML representations of the KEGG	2011	4 min	33 min	——	——	——	KEGG	[91]
BPAT-S	Single core from a 2. GHz83		200 min>	600 min>					
ReTrace	Intel Xeon E5440 with access		40 min	300 min>					
to 16 GB of RAM									
Branch-and-Bound	18. version of Pathway Tools0	2014	0. s3	3. s6	——	——	——	EcoCyc and MetaCyc version 17.5	[79]
2. GHz computer3									
DFS	Linux OS	2014	310 s	5000 s	——	——	——	Efmtool EFMs	[75]
Intel Core i7 860									
93Processor2. GHz									
RAM 8 GB									
MATLAB (R2011a)									
6 weighting schemes	——	2014	40 min	——	——	——	——	KEGG	[104]

path, where 'shortest path' is interpreted as minimizing the number of arcs (reactions) involved in the path, and k refers to the number of shortest paths; for instance, k ¼1 corresponds to the shortest path, k ¼ 2 corresponds to the second shortest and k ¼ 3 to the third shortest path, etc. Giving weights to a metabolic graph can reflect various kinetic and thermo-dynamic parameters.

Discussion

Pathfinding-based distance metric

Finding shortest paths in metabolic pathways constitutes a considerable advance with respect to stoichiometric approaches. In genome-scale metabolic networks, k-shortest path is a well-known graph theory problem for computa-tional tracking in metabolic networks. The most suitable dis-tance metric will determine the computed k-shortest paths instead of computing all paths because not all computed paths will have biological significance. Because analysis of the computed paths becomes simpler, k is usually restricted to a small number to determine biologically meaningful metabolic paths; thus, the distance metric appears to be the most effective metric presented to date [15, 80]. The found paths are implicitly and completely determined by the dis-tance metric adopted, thus having biological significance.

One major drawback of pathfinding approaches is that they are relatively inflexible in terms of adding additional biologically meaningful constraints. Currently, for instance, extracting stoichiometric information for a metabolic path must be done as a separate stage (e.g. by balancing inter-mediate compounds in the path) once the path has been computed without using stoichiometric information. Being able to add biologically-based constraints (e.g. stoichiometric, regulatory [105], topological or energetic [106]) as an intrinsic part of the pathfinding process would significantly refine the search for biologically meaningful metabolic paths, provided this can be done without excessively complicating the algo-rithmic/computational expense of finding k-shortest paths.

Metabolic pathfinding limitation

A limitation of metabolic pathfinding in the graph-theoretical setting is that all metabolites are by default considered equal in importance. Many currency and cofactor metabolites ap-pear in a large number of reactions, consituting a large frac-tion of high-degree metabolites in a typical metabolic network. Because of their high degree of connectivity, a ran-dom shortest path traverses through such a metabolite with a high probability. However, in many cases, the biological relevance of such a path is low. For instance, almost all func-tionally dissimilar parts of metabolism can be connected with short paths using adenosine triphosphate as an inter-mediate. To cope with this problem, high-degree metabolites can be removed from the network or edges of a metabolite graph can be weighted according to the degree of metabolite nodes, such that low-degree nodes are preferred [80]. However, non-cofactor metabolites with a high degree, such as pyruvate, then become a challenge to metabolic pathfind-ing. Indeed, deciding a suitable set of cofactor metabolites is a nontrivial problem, depending on the modeling task [107].

Several pathfinding algorithms are not efficient, such as A*, BFS and DFS (backtracking). For example, they either have a large overhead, yield far from optimal paths, do not

easily scale upto many cores or are cache unfriendly. Therefore, in recent years, many researchers have proposed several new algorithms to solve these weaknesses of previous algorithms. Algorithms such as Jump point Search [108, 109] and Sub [110] are high-performance algorithms in terms of time and memory overhead complexities, and they can provide optimality paths (shortest paths). Finally, an important topic of metabolic path-finding is further refinement of the approach and integration with other techniques, which is a useful methodology. For bet-ter results in the future, we suggest using these algorithms or new algorithms for biological purposes such as metabolic pathways.

Conclusion and future works

In this article, we have described stoichiometric and pathfinding approaches that are related with metabolic pathways. Notwithstanding that stoichiometric approaches have a theoret-ical basis, FBA, EFMs and EPs face many difficulties from both the computational and analytic points of view when applied to large-scale experimental data for metabolic biology pathways. Contrary to stoichiometric approaches, pathfinding approaches do enable analysis of genome-scale metabolic networks to be performed. We believe the key research challenges are choice of an appropri-ate distance metric and addition of biologically based constraints. We would emphasize here that pathways produced by any math-ematical/computational approach need a clear validation with experimentally determined metabolic pathways from the bio-chemical literature. Being better able to model mathematically metabolic pathways will help in the metabolic engineering of or-ganisms to create (currently) unknown, novel pathways useful for biotechnological or biomedical purposes.

Most importantly, each algorithm should guarantee the opti-mality of the routes found when the cost of the atoms is lost in the routes as a part of the optimality criteria. Future work should take into account the symmetry of compounds for atom mappings, where the atoms lost and conserved in routes are more precisely tracked. In addition, any pathfinding algorithm should consider the cost of side compounds, the toxicity of some of the metabolites produced and the taxonomic range of the new enzymes introduced in routes. Another worthy re-search topic of metabolic pathfinding in the near future is to find optimal routes from multiple start compounds, and to identify more complex multiple target compounds because the possible candidate solutions are no longer linear routes but multiple linear routes merged together.

Key points

- A computational problem in metabolic engineering is finding efficient metabolic routes from a source to a target compound in genome-scale reaction networks.
- There exist a multitude of tools freely available for metabolic pathway analysis.
- The article clearly reviews stoichiometric approaches, graph theory and pathfinding approaches to metabolic pathways.
- Pathfinding approaches use shortest path and k-short-est path concepts.
- Stoichiometric approaches involve FBA, EFMs and EPs.

Funding

This article is supported by ERGS grant vote: R.J130000.7828.4L092 at MaGICX UTM-IRDA Universiti Teknologi Malaysia.

References

1. Fersht A. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. New York, NY: Macmillan, 1999.
2. Cornish-Bowden A. Fundamentals of Enzyme Kinetics. Weinheim, Germany: John Wiley & Sons, 2013.
3. Berg JM, Tymoczko JL, Stryer L. Biochemistry: International Edition. Basingstoke, UK: WH Freeman & Company Limited, 2006.
4. Porth C. Essentials of Pathophysiology: Concepts of Altered Health States. Philadelphia, USA: Lippincott Williams & Wilkins, 2011.
5. Ingelman-Sundberg M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. Trends Pharmacol Sci 2004;25:193–200.
6. Harrigan GG, Goodacre R. Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis: Its Role in Biomarker Discovery and Gene Function Analysis. USA: Springer, 2003.
7. Wiechert W, Mollney€ M, Petersen S, et al. A universal frame-work for 13 C metabolic flux analysis. Metabol Eng 2001;3:265–83.
8. Arita M. Metabolic reconstruction using shortest paths. Simulat Pract Theory 2000;8:109–25.
9. Schuster S, Fell DA, Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 2000;18:326–32.
10. Planes FJ, Beasley JE. An optimization model for metabolic pathways. Bioinformatics 2009;25:2723–9.
11. Kuffner€ R, Zimmer R, Lengauer T. Pathway analysis in meta-bolic databases via differential metabolic display (DMD). Bioinformatics 2000;16:825–36.
12. Jeong H, Tombor B, Albert R, et al. The large-scale organiza-tion of metabolic networks. Nature 2000;407:651–4.
13. Fell DA, Wagner A. The small world of metabolism. Nat Biotechnol 2000;18:1121–2.
14. Planes FJ, Beasley JE. Path finding approaches and metabolic pathways. Discrete Appl Math 2009;157:2244–56.
15. Croes D, Couche F, Wodak SJ, et al. Inferring meaningful pathways in weighted metabolic networks. J Mol Biol 2006;356:222–36.
16. Keseler IM, Collado-Vides J, Gama-Castro S, et al. EcoCyc: a comprehensive database resource for Escherichia coli. Nucleic Acids Res 2005;33:D334–7.
17. Beasley JE, Planes FJ. Recovering metabolic pathways via op-timization. Bioinformatics 2007;23:92–8.
18. Deville Y, Gilbert D, Van Helden J, et al. An overview of data models for the analysis of biochemical pathways. Brief Bioinform 2003;4:246–59.
19. Planes FJ, Beasley JE. A critical examination of stoichiomet-ric and path-finding approaches to metabolic pathways. Brief Bioinform 2008;9:422–36.
20. Faust K, Croes D, van Helden J. Metabolic pathfinding using RPAIR annotation. J Mol Biol 2009;388:390–414.
21. de Figueiredo LF, Schuster S, Kaleta C, et al. Response to comment on ‘Can sugars be produced from fatty acids? A

test case for pathway analysis tools'. Bioinformatics 2009;25:3330–1.

22. Gombert AK, Nielsen J. Mathematical modelling of metabol-ism. Curr Opin Biotechnol 2000;11:180–6.
23. Kauffman KJ, Prakash P, Edwards JS. Advances in flux bal-ance analysis. Curr Opin Biotechnol 2003;14:491–6.
24. Aittokallio T, Schwikowski B. Graph-based methods for analysing networks in cell biology. Brief Bioinform 2006;7:243–55.

25. Baldan P, Cocco N, Marin A, et al. Petri nets for modelling metabolic pathways: a survey. Nat Comput 2010;9:955–89.
26. Kher S, Dickerson J, Rawat N. Biological pathway data inte-gration trends, techniques, issues and challenges: a survey. In: Nature and Biologically Inspired Computing (NaBIC), 2010 Second World Congress on. Fukuoka: IEEE, 2010, 177–82.
27. Karp PD, Caspi R. A survey of metabolic databases empha-sizing the MetaCyc family. Arch Toxicol 2011;85:1015–33.
28. Zomorrodi AR, Suthers PF, Ranganathan S, et al. Mathematical optimization applications in metabolic net-works. Metab Eng 2012;14:672–86.
29. Fell DA, Small JR. Fat synthesis in adipose tissue: an examin-ation of stoichiometric constraints. Biochem J 1986;238:781–6.
30. Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product se-cretion in wild-type Escherichia coli W3110. Appl Environ Microbiol 1994;60:3724–31.
31. Schilling CH, Edwards JS, Palsson BO. Toward metabolic phenomics: analysis of genomic data using flux balances. Biotechnol Prog 1999;15:288–95.
32. Price ND, Reed JL, Palsson BØ. Genome-scale models of mi-crobial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2004;2:886–97.
33. Herrga˚rd MJ, Fong SS, Palsson BØ. Identification of genome-scale metabolic network models using experimentally measured flux profiles. PLoS Comput Biol 2006;2:676–86.
34. Kanehisa M, Goto S, Kawashima S, et al. The KEGG resource for deciphering the genome. Nucleic Acids Res 2004;32: D277–80.

35. Krieger CJ, Zhang P, Mueller LA, et al. MetaCyc: a multiorgan-ism database of metabolic pathways and enzymes. Nucleic Acids Res 2004;32:D438–42.
36. Covert MW, Schilling CH, Palsson B. Regulation of gene ex-pression in flux balance models of metabolism. J Theor Biol 2001;213:73–88.

37. Shlomi T, Eisenberg Y, Sharan R, et al. A genome-scale com-putational study of the interplay between transcriptional regulation and metabolism. Mol Syst Biol 2007;3:101.
38. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol 2008;4: e1000082.

39. Chandrasekaran S, Price ND. Probabilistic integrative model-ing of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. Proc Natl Acad Sci USA 2010;107:17845–50.
40. Mahadevan R, Schilling C. The effects of alternate optimal solutions in constraint-based genome-scale metabolic mod-els. Metab Eng 2003;5:264–76.
41. Burgard AP, Nikolaev EV, Schilling CH, et al. Flux coupling analysis of genome-scale metabolic network reconstruc-tions. Genome Res 2004;14:301–12.
42. Schuster S, Hilgetag C. On elementary flux modes in bio-chemical reaction systems at steady state. J Biol Syst 1994;2:165–82.
43. Schuster S, Dandekar T, Fell DA. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol 1999;17:53–60.

44. Papin JA, Stelling J, Price ND, et al. Comparison of network-based pathway analysis methods. Trends Biotechnol 2004; 22:400–5.

45. Chan SJ. A Decomposition Approach for Computing Elementary Flux Modes in Genome-Scale Metabolic Networks. The Hong Kong Polytechnic University, 2011.
46. Klamt S, Stelling J. Combinatorial complexity of pathway analysis in metabolic networks. Mol Biol Rep 2002;29:233–6.
47. Pfeiffer T, Nu J, Montero F, et al. METATOOL: for studying metabolic networks. Bioinformatics 1999;15:251–7.
48. Klamt S, Stelling J. Two approaches for metabolic pathway analysis? Trends Biotechnol 2003;21:64–9.
49. Schilling CH, Letscher D, Palsson BØ. Theory for the sys-temic definition of metabolic pathways and their use in in-terpreting metabolic function from a pathway-oriented perspective. J Theor Biol 2000;203:229–48.
50. Wiback SJ, Palsson BO. Extreme pathway analysis of human red blood cell metabolism. Biophys J 2002;83:808–18.
51. Papin JA, Price ND, Edwards JS, et al. The genome-scale metabolic extreme pathway structure in haemophilus influ-enza shows significant network redundancy. J Theor Biol 2002;215:67–82.

52. Price ND, Papin JA, Palsson BØ. Determination of redun-dancy and systems properties of the metabolic network of Helicobacter pylori using genome-scale extreme pathway analysis. Genome Res 2002;12:760–9.
53. Rezola A, de Figueiredo LF, Brock M, et al. Exploring meta-bolic pathways in genome-scale networks via generating flux modes. Bioinformatics 2011;27:534–40.
54. De Figueiredo LF, Podhorski A, Rubio A, et al. Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics 2009;25:3158–65.
55. Arita M. The metabolic world of Escherichia coli is not small. Proc Natl Acad Sci USA 2004;101:1543–7.
56. Wagner A, Fell DA. The small world inside large metabolic networks. Proc Biol Sci 2001;268:1803–10.
57. van Helden J, Wernisch L, Gilbert D, et al. Graph-based ana-lysis of metabolic networks. In: Bioinformatics and Genome Analysis. Berlin Heidelberg: Springer, 2002, 245–74.
58. Carter GW. Inferring network interactions within a cell. Brief Bioinform 2005;6:380–9.
59. Rahman SA, Advani P, Schunk R, et al. Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). Bioinformatics 2005;21:1189–93.
60. Przulj N, Corneil DG, Jurisica I. Modeling interactome: scale-free or geometric? Bioinformatics 2004;20:3508–15.
61. Przulj N, Corneil DG, Jurisica I. Efficient estimation of graph-let frequency distributions in protein–protein interaction networks. Bioinformatics 2006;22:974–80.
62. Faust K, Dupont P, Callut J, et al. Pathway discovery in meta-bolic networks by subgraph extraction. Bioinformatics 2010; 26:1211–18.

63. Li C, Li X, Miao Y, et al. SubpathwayMiner: a software pack-age for flexible identification of pathways. Nucleic Acids Res 2009;37:e131.

64. Leung S, Leung HC, Xiang CL, et al. Predicting metabolic pathways from metabolic networks with limited biological knowledge. In: 2010 IEEE International Conference on

Bioinformatics and Biomedicine Workshops (BIBMW), Hong Kong, 2010, 7–12.

65. Rahman SA, Schomburg D. Observing local and global prop-erties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks. Bioinformatics 2006; 22:1767–74.

66. De Figueiredo LF, Schuster S, Kaleta C, et al. Can sugars be produced from fatty acids? A test case for pathway analysis tools. Bioinformatics 2008;24:2615–21.
67. Eom Y-H, Lee S, Jeong H. Exploring local structural organiza-tion of metabolic networks using subgraph patterns. J Theor Biol 2006;241:823–9.
68. Lim K, Wong L. CMPF: class-switching minimized pathfind-ing in metabolic networks. BMC Bioinformatics 2012;13:S17.
69. Yeung M, Thiele I, Palsson BØ. Estimation of the number of extreme pathways for metabolic networks. BMC Bioinformatics 2007;8:363.

70. Nabli F. Approches de programmation par contraintes pour l'ana-lyse des proprietes structurelles des reseaux de Petri et application aux reseaux biochimiques. Universite Paris-Diderot-Paris VII. France: Paris Diderot University, 2013.
71. Carbonell P, Fichera D, Pandit SB, et al. Enumerating meta-bolic pathways for the production of heterologous target chemicals in chassis organisms. BMC Syst Biol 2012;6:10.
72. Cottret L, Milreu PV, Acua V, et al. Enumerating precursor sets of target metabolites in a metabolic network. In: Algorithms in Bioinformatics. Berlin Heidelberg: Springer, 2008, 233–44.

73. McShan DC, Rao S, Shah I. PathMiner: predicting metabolic pathways by heuristic search. Bioinformatics 2003;19: 1692–8.

74. Heinonen M, Lappalainen S, Mielikainen€ T, et al. Computing atom mappings for biochemical reactions without subgraph isomorphism. J Comput Biol 2011;18:43–58.
75. Quek L-E, Nielsen LK. A depth-first search algorithm to com-pute elementary flux modes by linear programming. BMC Syst Biol 2014;8:94.
76. Abd Algfoor Z, Sunar MS, Kolivand H. A comprehensive study on pathfinding techniques for robotics and video games. Int J Comput Games Technol 2015;2015:111.
77. Schuster S, Pfeiffer T, Moldenhauer F, et al. Exploring the pathway structure of metabolism: decomposition into sub-networks and application to Mycoplasma pneumoniae. Bioinformatics 2002;18:351–61.
78. DeJongh M, Formsma K, Boillot P, et al. Toward the auto-mated generation of genome-scale metabolic networks in the SEED. BMC Bioinformatics 2007;8:139.
79. Latendresse M, Krummenacker M, Karp PD. Optimal meta-bolic route search based on atom mappings. Bioinformatics 2014;30:2043–50.

80. Croes D, Couche F, Wodak SJ, et al. Metabolic PathFinding: inferring relevant pathways in biochemical networks. Nucleic Acids Res 2005;33:W326–30.
81. Planes FJ. Metabolic pathway analysis via integer linear program-ming. London, UK: Brunel University, School of Information Systems, Computing and Mathematics, 2008.
82. Faust K, van Helden J. Predicting metabolic pathways by sub-network extraction. In Bacterial Molecular Networks, New York, NY: Springer, 2012, 107–30.
83. Bonchi F, Gionis A, Gullo F, et al. Distance oracles in edge-labeled graphs. In: 17th International Conference on Extending Database Technology (EDBT/ICDT), Athens, Greece. 2014, 547–58.
84. Ingalalli V, Ienco D. Subgraph Matching for Single Large Multigraphs. Doctoral dissertation. France: Montpellier 2 University (UniversitO Montpellier 2), 2014.
85. Jevremovic D. Scalable Computation and Analysis of Elementary Flux Modes in Metabolic Networks. Minnesota, USA: University of Minnesota, 2013.
86. Blum T, Kohlbacher O. MetaRoute: fast search for relevant metabolic routes for interactive network navigation and visualization. Bioinformatics 2008;24:2108–9.
87. Pitkanen€ E, Rousu J, Ukkonen E. Computational methods for metabolic reconstruction. Curr Opin Biotechnol 2010;21:70–7.
88. Pitkanen€ E. Computational Methods for Reconstruction and Analysis of Genome-Scale Metabolic Networks. Doctoral dis-sertation, Doctoral dissertation. Finland: University of Helsinki. 2010.

89. Michal G, Schomburg D. Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology. New York, NY: Wiley, 1999.

90. Pitkanen€ E, Jouhten P, Rousu J. Inferring branching pathways in genome-scale metabolic networks. BMC Syst Bioly 2009;3:103
91. Heath AP, Bennett GN, Kavraki LE. An algorithm for efficient identification of branched metabolic pathways. J Comput Biol 2011;18:1575–97.
92. Heath AP, Bennett GN, Kavraki LE. Identifying branched metabolic pathways by merging linear metabolic pathways. In: Research in Computational Molecular Biology. Springer Berlin Heidelberg, 2011, 70–84.
93. Heath AP, Bennett GN, Kavraki LE. Finding metabolic path-ways using atom tracking. Bioinformatics 2010;26:1548–55.
94. Karp PD, Paley SM, Krummenacker M, et al. Pathway tools version 13.0: integrated software for pathway/genome in-formatics and systems biology. Brief Bioinform 2009;11:40–79.
95. Goesmann A, Haubrock M, Meyer F, et al. PathFinder: recon-struction and dynamic visualization of metabolic pathways. Bioinformatics 2002;18:124–9.
96. Lemer C, Antezana E, Couche F, et al. The aMAZE LightBench: a web interface to a relational database of cellu-lar processes. Nucleic Acids Res 2004;32:D443–8.
97. Ellis LB, Roe D, Wackett LP. The University of Minnesota bio-catalysis/biodegradation database: the first decade. Nucleic Acids Res 2006;34:D517–21.
98. Kanehisa M, Goto S, Hattori M, et al. From genomics to chem-ical genomics: new developments in KEGG. Nucleic Acids Res 2006;34:D354–7.
99. Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res 2008;36: D480–4.

100. Kono N, Arakawa K, Ogawa R, et al. Pathway projector: web-based zoomable pathway browser using KEGG atlas and Google Maps API. PLoS One 2009;4:e7710.
101. Moriya Y, Shigemizu D, Hattori M, et al. PathPred: an en-zyme-catalyzed metabolic pathway prediction server. Nucleic Acids Res 2010;38:W138–43.
102. Karp PD, Weaver D, Paley S, et al. The EcoCyc database. EcoSal Plus 2014;30:56–8.
103. Caspi R, Foerster H, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collec-tion of pathway/genome databases. Nucleic Acids Res 2008;36:D623–31.

104. Ghosh S, Baloni P, Vishveshwara S, et al. Weighting schemes in metabolic graphs for identifying biochemical routes. Syst Synth Biol 2014;8:47–57.
105. Covert MW, Palsson BO. Constraints-based models: regula-tion of gene expression reduces the steady-state solution space. J Theor Biol 2003;221:309–25.
106. Henry CS, Broadbelt LJ, Hatzimanikatis V. Thermodynamics-based metabolic flux analysis. Biophys J 2007;92:1792–805.
107. Huss M, Holme P. Currency and commodity metabol-ites: their identification and relation to the modularity of metabolic networks. IET Syst Biol 2007;1:280–5.
108. Harabor DD, Grastien A. Online graph pruning for pathfind-ing on grid maps. In: Proceedings of the 25th National Conference on Artificial Intelligence (AAAI), San Francisco, CA, 2011.
109. Harabor D, Grastien A. Improving jump point search. In: Proceedings of the 24th International Conference on Automated Planning and Scheduling (ICAPS), Portsmouth, NH, 2014.
110. Uras T, Koenig S, Hernandez C. Subgoal graphs for optimal pathfinding in eight-neighbor grids. In: Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS), Rome, Italy, 2013.