

METHODOLOGY

Open Access



Discovery of Boolean metabolic networks: integer linear programming based approach

Yushan Qiu¹, Hao Jiang^{2*}, Wai-Ki Ching³ and Xiaoqing Cheng⁴

From The Sixteenth Asia Pacific Bioinformatics Conference
Yokohama, Japan. 15-17 January 2018

Abstract

Background: Traditional drug discovery methods focused on the efficacy of drugs rather than their toxicity. However, toxicity and/or lack of efficacy are produced when unintended targets are affected in metabolic networks. Thus, identification of biological targets which can be manipulated to produce the desired effect with minimum side-effects has become an important and challenging topic. Efficient computational methods are required to identify the drug targets while incurring minimal side-effects.

Results: In this paper, we propose a graph-based computational damage model that summarizes the impact of enzymes on compounds in metabolic networks. An efficient method based on Integer Linear Programming formalism is then developed to identify the optimal enzyme-combination so as to minimize the side-effects. The identified target enzymes for known successful drugs are then verified by comparing the results with those in the existing literature.

Conclusions: Side-effects reduction plays a crucial role in the study of drug development. A graph-based computational damage model is proposed and the theoretical analysis states the captured problem is NP-completeness. The proposed approaches can therefore contribute to the discovery of drug targets. Our developed software is available at "<http://hkumath.hku.hk/~wkc/APBC2018-metabolic-network.zip>".

Keywords: Metabolic network, Integer linear programming, Boolean model

Background

One of the most important biological processes in organism is metabolism and its dysregulation can contribute to many human diseases. The manipulation of metabolism has been extensively studied in the field of metabolic engineering as many of metabolic process produce commodity and specialty chemicals. In particular, production of industrially valuable products using a microbial host with recombinant technology becomes one of the most successful applications of metabolic engineering [1–3]. Metabolic network is used to model the behavior of reactions and chemicals [4]. Authors in [5] have proposed a new mathematical model for large metabolic network

and demonstrated that the outcome of network expansion is robust against of single or few reactions. Indeed, for the analysis of metabolic networks, many studies have been conducted so as to develop Boolean models. For example, Smart et al. [6] studied the effect of deletion of each enzyme in metabolic network of a Boolean model, and Lemke et al. [7] considered almost the same problem from the viewpoint of the Boolean aspect. Tamura et al. [8] developed an integer linear programming method for Boolean reaction cut (BRC) problem. Furthermore, in [9], the computational complexity of BRC was analyzed. In the Boolean model of metabolic networks, compounds and reactions can be represented by "OR" and "AND" nodes, respectively. Li et al. [10] have developed methods for finding a set of enzymes whose inhibition stops the production of the target compound with a minimum elimination of non-target compounds. Furthermore, Takemoto et al. [11] and Lee et al. [12] estimated the size distribution

*Correspondence: jiangh@ruc.edu.cn

²Department of Mathematics, School of Information, Renmin University of China, No.59 Zhong Guan Cun Avenue, Hai Dian District, 100872 Beijing, China
Full list of author information is available at the end of the article

of the deletion effect of enzymes using a branching process. Enumerating all the possible enzyme combination is infeasible since the number of combination increases exponentially with the number of enzymes. Lu et al. [13] developed an integer linear programming method for designing synthetic metabolic networks by considering minimum reaction insertion in a Boolean model. The proposed method can appropriately solve minimum reaction insertion in a Boolean network.

In pharmaceuticals, the development of drugs mainly focus on target identification and lead inhibitor identification [14]. Furthermore, many researchers have done a lot of work on the efficacy of drugs regardless of their side-effects (toxicity). However, toxicity or lack of efficacy may occur when the compounds which are not the intended targets in the metabolic network. Thus how to effectively reduce the side-effect of drugs has become an important and challenging issue. The current works aim to identify the biological targets (could be enzyme or protein) for drugs which can be manipulated so as to produce the desired effect (i.e., curing a disease) with minimum disruptive side-effects [15, 16]. Drug targets, in particular enzymes, are chosen to reduce abnormal metabolites by formulating an optimal combination problem in enzyme combination in metabolic networks [10, 17]. It is known that the function of enzymes is to catalyze reactions and produces metabolites (compounds) in metabolic networks. Human diseases, especially metabolic diseases, may be directly caused by the accumulation of certain compounds due to the malfunctions of enzymes. We denote such compounds as *target compounds* while the others as *non-target compounds*. For example, the malfunction of enzyme phenylalanine hydroxylase results in accumulation of the amino acid phenylalanine which causes the disease phenylketonuria [18]. Therefore, there is a great need to identify a set of enzymes which are manipulated by drugs to prevent the buildup of target compounds with minimal damage. Here *damage* corresponds to the number of non-target compounds whose production are forced to stop by the inhibition of that enzyme (or enzyme set).

Then the problem becomes to identify the optimal enzyme set whose inhibition eliminates the target compounds and at the same time, incurs minimum damage based on the given metabolic network and a set of target compounds. We denote such problem as enzyme combination identification (ECI). Sridhar et al. [19] developed a branch-and-bound algorithm to dynamically explore the search space. Furthermore, two filtering strategies are proposed to prune the search space to guarantee an optimal solution. However, their algorithm is complicated and impractical for us to use it. In this paper, we develop an efficient method based on Integer Linear Programming [20–22] which has a wide application in solving NP-

hard problems. Furthermore, all instances of the original problem are needed to convert into integer programming formalization so as to apply the existing free solver called CPLEX [23].

This paper has two main contributions. First, we formulate a new biological problem based on Boolean metabolic network to identify the drug target with minimum damage. Second, we propose an effective and efficient method, Integer Linear Programming, (needs $O(m + n)$ variables) to solve the captured problem and it is easy to implement. By integrating the human metabolic networks, we have shown that the proposed approach can accurately identify the target enzyme set for known drugs in the literature. Furthermore, experiments have shown that our proposed method is extremely efficient which can effectively solve the problem in seconds.

The remainder of the paper is organized as follows. In “**Problem formulation**” section, we formulate the captured problem. “**Methods**” section presents the algorithm ILP-ECI. In “**Results and discussions**” section, we discuss the experimental results and give a theoretical illustration. Finally, conclusions are given in the last section.

Problem formulation

A metabolic network can be represented by a directed graph $G = (V, A)$ which captures the interaction between reactions, compound, and enzymes. Let C, R , and E denote the set of compounds, reactions, and enzymes, respectively. Then V_C, V_R and V_E denote the set of vertices from C, R , and E , where $V_C = \{v_{c_1}, v_{c_2}, \dots, v_{c_m}\}$ denotes a set of *compound nodes*, $V_R = \{v_{r_1}, v_{r_2}, \dots, v_{r_n}\}$ represents a set of *reaction nodes* and $V_E = \{v_{e_1}, v_{e_2}, \dots, v_{e_l}\}$ defines a set of *enzyme nodes*. The subscript m, n, l denote the number of *compound nodes*, *reaction nodes* and *enzyme nodes*, respectively. Here $V = V_C \cup V_R \cup V_E$ and $V_C \cap V_R = \{\}, V_C \cap V_E = \{\}, V_R \cap V_E = \{\}$ hold. Here $\{\}$ is the empty set. Let $V_s \subseteq V_C$ and $V_t \subseteq V_C$ be sets of *source nodes* and *target nodes*, respectively. Source nodes are those nodes having no incoming edges and they are the seed compounds. A metabolic network can be defined as follows:

- For each edge $(u, v) \in A$, either $(u \in V_C) \wedge (v \in V_R)$ or $(u \in V_R) \wedge (v \in V_C | v \in V_E)$ holds and “ \wedge ” represents “AND” function;
- Each source node has no incoming edges;
- For nodes $v \notin V_s$ and $v \in V_C$ have at least one incoming edge.

The main problem **Enzyme Combination Identification (ECI)** in a Boolean network is first described with a simple example and then followed by its mathematical formalization.

A small hypothetical metabolic network is shown in Fig. 1, where circles, rectangles and triangles represent compounds, reactions and enzymes, respectively. For instance, R_1 is a reaction, its substrates are C_1 and E_2 and its product is C_2 . Here C_5 is the target compound in this figure which shall be stopped. To stop the production of C_5 , reaction R_2 must be prevented from taking place. One of the possible solution is to disrupt one of its catalyzing enzymes (E_1 for instance). Another is to stop the production of its reactant compounds (i.e., C_2 or C_3). If C_2 is stopped, then two possible ways can achieve this effect (i.e., enzyme E_2 or E_1). It is noted that C_2, R_2, C_7 and R_1 forms a cycle, the inhibition of E_1 will result in inactivating reaction R_2 which will further makes C_2 in producible. The other way is inhibiting enzyme E_2 which makes R_1 inactive and further stop the production of compound C_2 . Therefore, the inhibition of enzyme E_1 or E_2 can result in stopping the production of the target compound.

We define the **Enzyme Combination Identification** as follows.

- **Input:** A metabolic network and a set of target compounds $T(T \in C)$, C is the set of compounds.
- **Output:** Find a set of enzymes $X(X \in E)$ with minimum damage, whose inhibition stops the production of all the compounds in T .

For simplicity, we assume there are no external inputs to all reactions and all the input compounds related to reactions are shown in the network. We note that different compounds and enzymes may have different levels of importance in the metabolic networks. Here we assume that all the compounds and enzymes are of equal importance. We assign binary value (i.e., 0 or 1) to each node V in the Boolean model. “0” means that

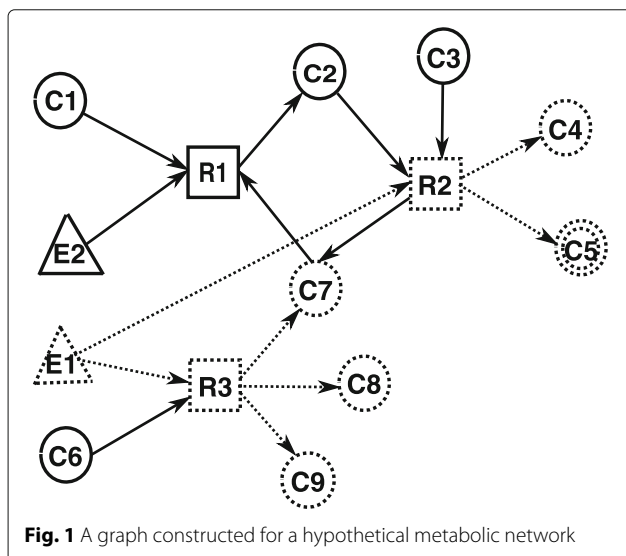


Fig. 1 A graph constructed for a hypothetical metabolic network

the corresponding compound is not producible or the corresponding reaction is inactive while “1” means that the corresponding reaction is active. For the related enzyme catalyzing the reaction, “1” means that it is active and “0” means that it is inhibited. Let G be such an assignment (G is a function from V to $\{0, 1\}$). For each node $v \in V$, we write $v = 1$ (resp., $v = 0$) if 1 (resp., 0) is assigned to v . Then G can be regarded as a *valid assignment* if the following conditions are satisfied:

- For each $v \in V_s, v = 1$.
- For each $v \notin V_s, v = 1$ if and only if there is u such that $(u, v) \in E$ and $u = 1$.
- For each $v \in V_r, v = 1$ if and only if $u = 1$ holds for all u such that $(u, v) \in E$.

Condition (ii) implies that compound nodes correspond to “OR” nodes. Condition (iii) implies that reaction nodes correspond to “AND” nodes, which means that the output is forced to 0 when a node is inactivated.

From Fig. 1, we can see that inhibition of E_1 results in the knock out of compounds C_4, C_7, C_8 and C_9 in addition to the target compound C_5 . Then we denote the number of non-target compounds knocked out as the *damage*, which is caused by the manipulating the enzyme set in the metabolic network. It can be seen that the damage of inhibiting E_2 is 2 (i.e., C_2 and C_4). Compound C_7 is still producible because it can be produced by R_3 even after the inhibition of E_2 . The damage effect of inhibition of E_1 is 4 (i.e., C_4, C_7, C_8 and C_9). Both E_1 and E_2 are potential drug targets since they can achieve the effect of disrupting the target compound C_5 . However, E_2 is a better drug target than E_1 owing to the fact that it causes less damage.

Methods

In this section, we introduce integer programming-based methods for ECI. Integer programming, in particular, Integer Linear Programming (ILP) is set to minimize (or maximize) a linear objective function under linear constraints with all the variables taking integer values. In the following, each variable takes including the binary value (i.e., 0 or 1), representing the Boolean values. We apply ILP to ECI since ILP is widely used for solving NP-hard problems.

The ILP formulation for the network in Fig. 1 is as follows:

ILP-ECI

$$\min \left\{ \sum_{i=1}^m TC_i \right\} \tag{1}$$

subject to

$$\begin{aligned}
 FC5 &= 1 & (2) \\
 TR1 + FC1 + FC7 + FE2 &\geq 1, \\
 FR1 + TC1 &\geq 1, \\
 FR1 + TC7 &\geq 1, \\
 FR1 + TE2 &\geq 1, & (3) \\
 TR2 + FC2 + FC3 + FE1 &\geq 1, \\
 FR2 + TC2 &\geq 1, \\
 FR2 + TC3 &\geq 1, \\
 FR2 + TE1 &\geq 1, & (4) \\
 TR3 + FC6 + FE1 &\geq 1, \\
 FR3 + TC6 &\geq 1, \\
 FR3 + TE1 &\geq 1 & (5) \\
 TC2 &= TR1 & (6) \\
 TC4 &= TR2 & (7) \\
 TC5 &= TR2 & (8) \\
 FC7 + TR2 + TR3 &\geq 1, \\
 TC7 + FR2 &\geq 1, \\
 TC7 + FR3 &\geq 1 & (9) \\
 TC8 &= TR3 & (10) \\
 TC9 &= TR3 & (11) \\
 TC1 = 1, \quad TC3 = 1, \quad TC6 = 1 & & (12) \\
 TX + FX = 1 \text{ for any } X & & (13)
 \end{aligned}$$

We denote the above formalization as **ILP-ECI**. Here all variables including the value of reaction compound and enzyme nodes take either 1 or 0. Thus, v_{r_i} can be either 0 or 1, and v_{c_i} and v_{e_i} also take 0 or 1. In this example, $v_{r_i} = 0$ (resp. $v_{r_i} = 1$) indicating that the value of reaction i takes 0 is represented by $FRi = 1$ (resp. $TRi = 1$) which implies that the reaction is inactivated (otherwise, the $TRi = 1$ implies the reaction is activated). Therefore, $TRi = 0$ (equivalent to $FRi = 1$) means the corresponding value for true reaction takes 0, which implies the reaction is inactive. And $FRi = 0$ (equivalent to $TRi = 1$) indicates the corresponding value for false reaction takes 0, which implies the reaction is active. Thus, $TRi + FRi = 1$ holds for any node i in the network. Similarly, TCi and FCi are used to represent the values of compound nodes. For instance, $TC2 = 1$ means that $v_{c_2} = 1$ and in other words, $FC2 = 0$ since $TCi + FCi = 1$. Furthermore, v_{r_i} corresponds to “AND” node which implies that if $v_{e_i} = 0$ will inactivate v_{r_i} .

The objective function (1) means that the damage should be minimized. $FCi = 1$ (or $TCi = 0$) means that a compound v_{c_i} is not producible. Equation (2) means that the target compound v_{c_5} should be 0 after the 0-1 assignment converges. Equation (3) represents the Boolean relation $v_{r_1} = v_{c_1} \wedge v_{c_7} \wedge v_{e_2}$. Note that the Boolean relations such as “ \vee ” or “ \wedge ” cannot be used in ILP formulation,

we need to convert them into linear equations and/or inequations. Actually, “ \vee ” indicates “AND” function and “ \wedge ” represents “OR” function. Since $x_1 = x_2 \wedge x_3 \wedge \dots \wedge x_n$ can be represented by

$$(x_1 \vee \overline{x_2} \vee \dots \vee \overline{x_n}) \wedge (\overline{x_1} \vee x_2) \wedge (\overline{x_1} \vee x_3) \wedge \dots \wedge (\overline{x_1} \vee x_n) = 1,$$

the constraint $v_{r_1} = v_{c_1} \wedge v_{c_7} \wedge v_{e_2}$ can be converted into

$$(v_{r_1} \vee \overline{v_{c_1}} \vee \overline{v_{c_7}} \vee \overline{v_{e_2}}) \wedge (\overline{v_{r_1}} \vee v_{c_1}) \wedge (\overline{v_{r_1}} \vee v_{c_7}) \wedge (\overline{v_{r_1}} \vee v_{e_2}) = 1.$$

Thus Eq. (3) is obtained. Similarly, Eqs. (4)-(5) represent the constraints of v_{r_2} and v_{r_3} , respectively.

For a compound node with indegree is 1 which indicates the node has only one incoming edge, the value of the predecessor is just copied. For instance, since v_{c_2} has only one predecessor v_{r_1} , v_{c_2} is just copied from v_{r_1} as shown in Eq. (6). Similarly, v_{c_4} is just copied from v_{r_2} which is shown in Eq. (7).

However, for a compound node with indegree more than 1, it is necessary to convert the “ \vee ” relation into linear equation or inequations. Equation (9) represents the Boolean relation $v_{c_7} = v_{r_2} \vee v_{r_3}$. Since $x_1 = x_2 \vee x_3 \vee \dots \vee x_n$ can be represented by

$$(\overline{x_1} \vee x_2 \vee \dots \vee x_n) \wedge (x_1 \vee \overline{x_2}) \wedge (x_1 \vee \overline{x_3}) \wedge \dots \wedge (x_1 \vee \overline{x_n}) = 1,$$

$v_{c_7} = v_{r_2} \vee v_{r_3}$ can be turned into $(\overline{v_{c_7}} \vee v_{r_2} \vee v_{r_3}) \wedge (v_{c_7} \vee \overline{v_{r_2}}) \wedge (v_{c_7} \vee \overline{v_{r_3}})$. Thus, Eq. (9) can be obtained.

Equation (6) - Eq. (12) represent the constraints of v_{c_1}, \dots, v_{c_9} respectively. Equation (12) means that v_{c_1}, v_{c_3} and v_{c_6} are 1 since their indegrees is 0.

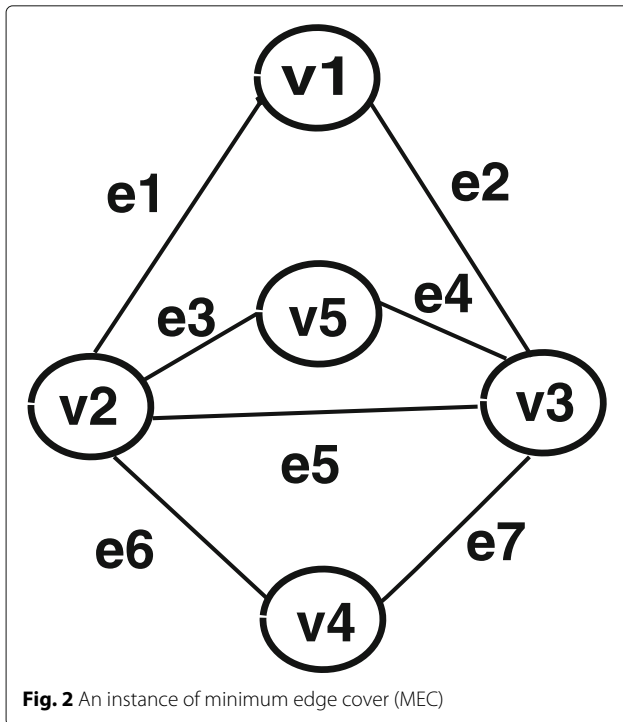
Equation (13) means that “T” and “F” correspond to “true (1)” and “false (0)”, respectively, and complement each other. X in Eq. (13) means any component or reaction in the metabolic network.

The above formalization can clearly solve ECI and obtain the correct solution $\{E_2\}$. Besides, the number of variables is $O(m + n)$ in the above formalization where m and n are the number of compounds and reactions, respectively.

It is noted that solving ILP is NP-complete, however, a problem that can be formalized as ILP is not always NP-complete. Thus in the following, we prove that ECI is NP-complete.

Theorem *ECI is NP-complete problem with the maximum indegree and outdegree being bounded by 2.*

Proof Obviously, the problem is in NP, it suffices to show that it is NP-hardness. The proof is by a polynomial time reduction from minimum edge cover (MEC), which is a problem for a given graph to find the minimum number of edges so that each node is incident to at least one of the selected edges. For instance, $E_1 = \{e_2, e_3, e_6\}$ is one of optimal solutions of MEC for graph shown in Fig. 2. Let $G = (V, E)$ be an instance of MEC,

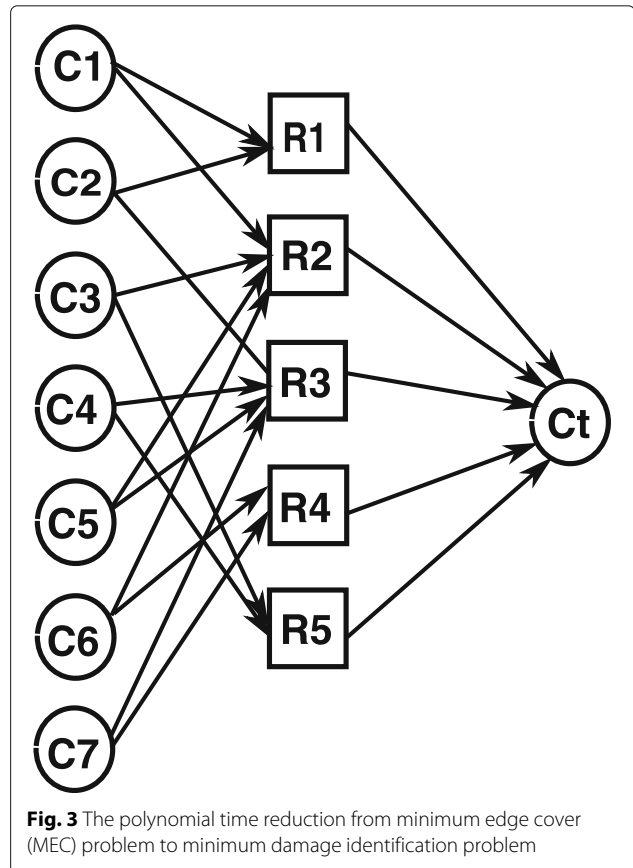


where $V = \{v_1, v_2, \dots, v_n\}$ and $E = \{e_1, e_2, \dots, e_m\}$. We then construct the corresponding ECI as below. The metabolite network $G = (V_c \cup V_r \cup V_e, E)$ is given by

$$V_c = \{c_1, c_2, \dots, c_m\} \cup c_t, \quad V_r = \{r_1, r_2, \dots, r_n\},$$

$$E = \{\{c_i, r_j\} \mid i = 1, \dots, m, j = 1, \dots, n, \text{ if } v_j \text{ is an end point of } e_i.\} \cup \{\{r_j, c_t\} \mid j = 1, \dots, n\}$$

It is noted that the minimum *damage* is determined uniquely by the inhibition of enzyme set. Furthermore, our objective is to minimize the “damage” (side-effects). Then V_e can be regarded as virtual nodes and denoted as an empty set in this case. The ECI problem can be converted into the problem of identifying the minimum set of non-target compounds. Thus the graph for MEC shown in Fig. 2 is converted into ECI shown in Fig. 3. It is clear that this conversion can be done in polynomial time. Then we show that MEC for G has a solution of size z if and only if ECI has a minimum damage of size z . To guarantee that the target compound c_t is stopped (i.e., $c_t = 0$), it implies that all r_j ($j = 1, \dots, n$) takes the value 0. If G has an edge cover of size z , then it follows that the minimum number of c_i taking 0 should be z . On the other hand, if the minimum damage for ECI is z , then each r_j must be 0 so as to satisfy $c_t = 0$, we have at least predecessor of each r_j must be included in the minimum damage set. Since there is an edge between c_i and r_j if and only if v_j is incident to



e_i . Thus $\{c_i \mid c_i \in \text{minimum damage set}\}$ is an edge cover of size z . □

Results and discussions

In this section, we verify the biological validity of our proposed method by using known drugs. Besides, the performance of the ILP-ECI algorithm is evaluated by using the execution time which indicates the total time taken by the method. For the experimental data, we extracted the metabolic network information from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [24]. KEGG is database which provides known drug molecules along the the enzymes they inhibit and their therapeutic category. Then we use drugs at this database as our benchmarks and we report two of them due to the space limitation. And the value in parenthesis that starts with letter “C”, “D”, and “E” (e.g., E1.13.11.34) is the unique identifier assigned to the corresponding compound, drug, and enzyme respectively in KEGG.

The drug we used in this paper is *Benoxaprofen* (D03080) which inhibits arachidonate 5-lipoxygenase (E1.13.11.34) [25]. This enzyme appears in several

networks including arachidonic acid metabolism network (hsa00590). In Pharmacology, the inhibition of 5-lipoxygenase will decrease the biosynthesis of LTB4 (C02165), cysteinylcontaining leukotrienes LTC4 (C02166), LTD4 (C05951) and LTE4 (C05952) (see, for instance, Fig. 4) [26]. According to our graph model, the removal of E1.13.11.34 will eliminate four compounds (LTB4, LTC4, LTD4 and LTE4) in arachidonic acid metabolism network. Furthermore, these four compounds play an important role in the mechanisms of toxic brain damage in acute methanol poisoning in humans [27]. Thus, we chose them as the target compounds. Apart from that, inhibiting this enzyme also eliminates four more compounds (i.e., 5(S)-HPETE(C05356), 5-HETE(C04805), LTA4(C00909), and 20-OH-LTB4(C04853)) [28]. These compounds can be regarded as “damage” in our model. Furthermore, when applying ILP-ECI with LTB4, LTC4, LTD4 and LTE4 as the target compounds, we find LTA4H (E3.3.2.6) and LTC4 synthase (E4.4.1.20) as the optimal enzyme set since the inhibition of these two enzymes eliminates only one non-target compound namely 20-OH-LTB4 (C04853). Thus, it is shown that ILP-ECI potentially finds a better solution in this experiment than the existing drugs since the same target compounds are eliminated by the existing drug in addition to other four compounds. Indeed, recent research validated our model since the anti-inflammatory

effect of the levels of LTA4H [29] and LTC4 [30] has been observed. The computational time in this experiment takes only 5.19 s.

Another experiment we conducted is the histidine metabolism network (hsa00340). The enzyme amine oxidase (E.1.4.3.4) appeared in hsa00340 can be inhibited by drug (Rasagiline (D02562)). According to our graph model, the removal of enzyme E.1.4.3.4 will result in eliminating two compounds Methylimidazoleacetic acid (C05828) and Methylimidazole acetaldehyde (C05827). It should be noted that the level of pros-methylimidazoleacetic acid is closely related to severity of Parkinson disease in patients [31, 32]. Running ILP-ECI with Methylimidazoleacetic acid and Methylimidazole acetaldehyde as the target compounds finds amine oxidase as the optimal enzyme. It takes only 3.26 s to run ILP-ECI. This experiment verifies that Rasgiline targets the optimal enzyme. An important advantage of our Boolean model is its capability of detecting the lack of substrates where the connectivity-based methods fail to handle this. Another advantage of this model is its capability of handling branches and cycles in a pathway from the source compound to the target compound. However, there are still have some limitations in this method. One of the major drawbacks is the assumption that all the compounds and enzymes are of equal importance. However, different compounds and enzymes

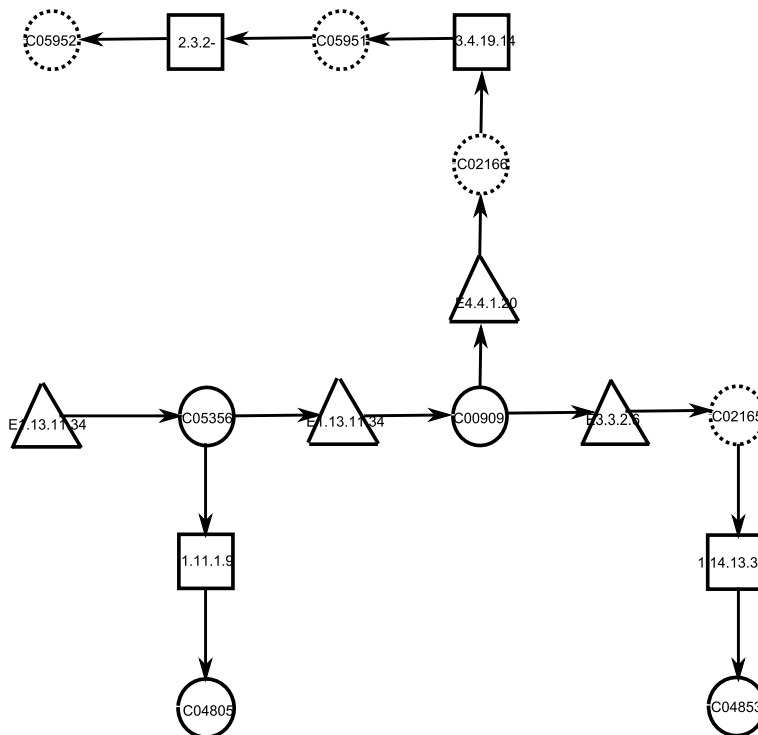


Fig. 4 Subgraph of arachidonic acid metabolism

may have different levels of importance in the metabolic networks. Our future work will focus on developing other models which include the weight of different nodes.

Conclusions

In this paper, we formulate the optimal enzyme combination identification (ECI) problem as an optimization problem in Boolean metabolic networks. We have proven that ECI in the Boolean model is NP-complete and the target enzyme set is uniquely determined when the target compounds are given. Furthermore, considering that an exhaustive search cannot be used to solve ECI when the network is large, we developed an ILP-based algorithm for ECI. Considering that the computational time of IP-based method is exponential to the number of variables, to improve the scalability of the developed method, it is vital to reduce the number of variables appearing in IP formalization. And our proposed IP-based method needs $O(m + n)$ variables.

The efficiency and effectiveness are validated by the computational experiments in which datasets were downloaded from the KEGG database. The results demonstrate that the proposed model can accurately identify the target enzymes for known successful drugs in the literature. Specifically, ILP-ECI has found a different enzyme set for the target compounds of Benoxaprofen which indicates that our method has a great potential to be better than Benoxaprofen. The reason is that the solution of our algorithm damages only one non-target compound while Benoxaprofen damages 4 non-target compounds including the compound damaged by ILP-ECI's solution. Besides, ILP-ECI has found that the same target enzyme like Rasagiline when its target compounds are given in advance. It is to be noted that our problem needs only $O(m + n)$ variables in the IP formalization. The experiments also show that ILP-ECI can solve the problem in a short time which confirms the efficiency of our algorithm.

Acknowledgements

The authors would like to thank anonymous reviewers for your helpful and constructive comments.

Funding

This research is supported in part of Natural Science Foundation of SZU (Grant No. 2017058) and National Natural Science Foundation of China NSFC (Grant Nos. 91730301, 11626229 and 11671158). The publication costs are funded by National Natural Science Foundation of China NSFC (Grant Nos. 91730301).

Availability of data and materials

All the data sets are publicly available and can be accessed from the KEGG databases.

About this supplement

This article has been published as part of *BMC Systems Biology* Volume 12 Supplement 1, 2018: Selected articles from the 16th Asia Pacific Bioinformatics Conference (APBC 2018): systems biology. The full contents of the supplement are available online at <https://bmcsystbiol.biomedcentral.com/articles/supplements/volume-12-supplement-1>.

Authors' contributions

YQ designed the research. YQ and WKC proposed the methods and did theoretical analysis. YQ and HJ collected the data. YQ, HJ and XC conducted the experiments and analyze the results. YQ, HJ, WC and XC wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Mathematics and Statistics, Shenzhen University, Nanhai Avenue 3688, 518060 Shenzhen, China. ²Department of Mathematics, School of Information, Renmin University of China, No.59 Zhong Guan Cun Avenue, Hai Dian District, 100872 Beijing, China. ³Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong, Hong Kong. ⁴School of Mathematics and Statistics, Xi'an Jiaotong University, No.28 West Xianning Road, 710049 Xi'an, China.

Published: 11 April 2018

References

1. Bro C, Regenber B, Förster J, Nielsen J. In silico aided metabolic engineering of *saccharomyces cerevisiae* for improved bioethanol production. *Metab Eng.* 2006;8(2):102–11.
2. Lee SK, Chou H, Ham TS, Lee TS, Keasling JD. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr Opin Biotechnol.* 2008;19:556–63.
3. Alper H, Jin YS, Moxley JF, Stephanopoulos G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *escherichia coli*. *Metab Eng.* 2005;7:155–64.
4. Soh KC, Hatzimanikatis V. Dreams of metabolism. *Trends Biotechnol.* 2010;28:501–8.
5. Handorf T, Ebenhöf O, Heinrich R. Expanding metabolic networks: scopes of compounds, robustness, and evolution. *J Mol Evol.* 2005;61:498–512.
6. Smart AG, Amaral LA, Ottino JM. Cascading failure and robustness in metabolic networks. *Proc Natl Acad Sci.* 2008;105:13223–8.
7. Lemke N, Herédia F, Barcellos CK, Dos Reis AN, Mombach JC. Essentiality and damage in metabolic networks. *Bioinformatics.* 2004;20:115–9.
8. Tamura T, Takemoto K, Akutsu T. Finding minimum reaction cuts of metabolic networks under a boolean model using integer programming and feedback vertex sets. *Int J Knowl Discov Bioinformatics (IJKDB).* 2010;1:14–31.
9. Tamura T, Akutsu T. Exact algorithms for finding a minimum reaction cut under a boolean model of metabolic networks. *IEICE Trans Fundam Electron.* 2010;93:1497–1507.
10. Li Z, Wang RS, Zhang XS, Chen L. Detecting drug targets with minimum side effects in metabolic networks. *IET Syst Biol.* 2009;3:523–33.
11. Takemoto K, Tamura T, Akutsu T. Theoretical estimation of metabolic network robustness against multiple reaction knockouts using branching process approximation. *Physica A Stat Mech Appl.* 2003;392:5525–535.
12. Lee D, Goh KI, Kahng B. Branching process approach for boolean bipartite networks of metabolic reactions. *Phys Rev E.* 2012;86:027101.
13. Lu W, Tamura T, Song J, Akutsu T. Integer programming-based method for designing synthetic metabolic networks by minimum reaction insertion in a boolean model. *Plos ONE.* 2014;9:92637.
14. Drews J. Drug discovery: A historical perspective. *Science.* 2000;287:1960–4.
15. Smith C. Hitting the target. *Nature.* 2003;422:341–7.
16. Takenaka T. Classical vs reverse pharmacology in drug discovery. *BJU Int.* 2001;88:7–10.
17. Li Z, Wang RS, Zhang XS. Two-stage flux balance analysis of metabolic networks for drug target identification. *BMC Syst Biol.* 2011;5:11.
18. Surtees R, Blau N. The neurochemistry of phenylketonuria. *Eur J Pediatr.* 2000;159:109–13.
19. Sridhar P, Song B, Kahveci T, Ranka S. Mining metabolic networks for optimal drug targets. *Pac Symp Biocomput.* 2008;13:291–302.

20. Schrijver A. Theory of Linear and Integer Programming. New York: John Wiley & Sons, Inc; 1986.
21. Zhenping L, Zhang S, Wang Y, Zhang XS, Chen L. Alignment of molecular networks by integer quadratic programming. *Bioinformatics*. 2007;23:1631–9.
22. Cheng X, Qiu Y, Hou W, Ching WK. Integer programming-based method for observability of singleton attractors in boolean networks. *IET Syst Biol*. 2017;11:30–5.
23. IBM 2010, IBM ILOG CPLEX Optimizer. <http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html>. Accessed 2 Mar 2017.
24. Kanehisa M, Goto S. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
25. Fogh K, Herlin T, Kragballe K. In vitro inhibition of leukotriene b4 formation by exogeneous 5-lipoxygenase inhibitors is associated with enhanced generation of 15-hydroxy-eicosatetraenoic acid (15-hete) by human neutrophils. *Arch Dermatol Res*. 1988;280:430–6.
26. Lötzer K, Funk CD, Habenicht AJ. The 5-lipoxygenase pathway in arterial wall biology and atherosclerosis. *Biochim Biophys Acta*. 2005;1736:30–7.
27. Zakharov S, Kotikova K, Nurieva O, Hlusicka J, Kacer P, Urban P, Vaneckova M, Seidl Z, Diblík P, Kuthan P, Navrátil T, Pelcova D. Leukotriene-mediated neuroinflammation, toxic brain damage, and neurodegeneration in acute methanol poisoning. *Int J Toxicol (Phila)*. 2017;55:249–59.
28. Sweeney FJ, Eskra JD, Carty TJ. Development of a system for evaluating 5-lipoxygenase inhibitors using human whole blood. *Prostaglandins Leukot Med*. 1987;28:73–93.
29. Rao NL, Dunford PJ, Xue X, Jiang X, Lundeen KA, Coles F, Riley JP, Williams KN, Grice CA, Edwards JP, Karlsson L, Fourie AM. Anti-inflammatory activity of a potent, selective leukotriene a4 hydrolase inhibitor in comparison with the 5-lipoxygenase inhibitor zileuton. *J Pharmacol Exp Ther*. 2007;321:1154–60.
30. Torres-Galván MJ1, Ortega N, Sánchez-García F, Blanco C, Carrillo T, Quiralte J. Ltc4-synthase a-444c polymorphism: lack of association with nsaid-induced isolated periorbital angioedema in a spanish population. *Ann Allergy Asthma Immunol*. 2001;87:506–10.
31. Blandina P, Cherici G, Moroni F, Prell GD, Green JP. Release of glutamate from striatum of freely moving rats by pros-methylimidazoleacetic acid. *J Neurochem*. 1995;64:788–93.
32. Prell GD, Khandelwal JK, Burns RS, Blandina P, Morrishow AM, Green JP. Levels of pros-methylimidazoleacetic acid: Correlation with severity of parkinson's disease in csf of patients and with the depletion of striatal dopamine and its metabolites in mptp-treated mice. *J Neural Transm*. 1991;3:1435–63.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

