Computational and Structural Biotechnology Journal 13 (2015) 376–381

ELSEVIER

CrossMark

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Mini Review

Computational Methods for Modification of Metabolic Networks

Takeyuki Tamura, Wei Lu, Tatsuya Akutsu *

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 6110011, Japan,

ARTICLE INFO

Article history:
Received 26 March 2015
Received in revised form 20 May 2015
Accepted 24 May 2015
Available online 28 May 2015

Keywords:
Metabolic network
Constraint-based programming
Flux balance analysis
Elementary mode
Boolean model
Overfitting

ABSTRACT

In metabolic engineering, modification of metabolic networks is an important biotechnology and a challenging computational task. In the metabolic network modification, we should modify metabolic networks by newly adding enzymes or/and knocking-out genes to maximize the biomass production with minimum side-effect. In this mini-review, we briefly review constraint-based formalizations for Minimum Reaction Cut (MRC) problem where the minimum set of reactions is deleted so that the target compound becomes non-producible from the view point of the flux balance analysis (FBA), elementary mode (EM), and Boolean models. Minimum Reaction Insertion (MRI) problem where the minimum set of reactions is added so that the target compound newly becomes producible is also explained with a similar formalization approach. The relation between the accuracy of the models and the risk of overfitting is also discussed.

Contents

1. Introduction

A metabolic network represents relations between chemical reactions and compounds in a cell of organisms. Although much knowledge about metabolic networks is available in public databases and references, we often have to modify metabolic networks in various situations. For example, in metabolic engineering, we should modify metabolic networks by newly adding enzymes or/and knocking-out genes to maximize the biomass production with minimum side-effect. The former and latter correspond to adding and deleting chemical reactions, respectively, in a metabolic network. For another example, when we reconstruct a genome-scale metabolic network from a newly determined DNA sequence, the reconstructed metabolic network may need some modification to be consistent with the existing knowledge. Thus, in metabolic network modification, we often add or/and delete reactions so that specified constraints are satisfied.

Although there may exist various modification problems, we focus on the following two major problems in this mini-review: (i) *Minimum Reaction Cut* (MRC) problem: delete the minimum set of reactions so that the target compound becomes non-producible, and (ii) *Minimum Reaction Insertion* (MRI) problem: add the minimum set of reactions so that the target compound newly becomes producible. It should be noted that, for most cases, a target compound can be replaced by a set of target compounds in a straight-forward manner. In order to solve

* Corresponding author. Tel.: +81 774383015.
E-mail addresses: tamura@kuicr.kyoto-u.ac.jp (T. Tamura), Rogi@kuicr.kyoto-u.ac.jp (W. Lu), takutsu@kuicr.kyoto-u.ac.jp (T. Akutsu).

http://dx.doi.org/10.1016/j.csbj.2015.05.004

these problems, three mathematical models have been utilized: flux balance analysis (FBA) model, elementary mode (EM) model, and Boolean model. In this mini-review, we explain these three models in the context of MRC, and briefly review MRI.

Before explaining details of each model, we briefly explain MRC. Suppose that a metabolic network of Fig. 1 is given. Rectangles and circles represent reactions and compounds, respectively. $\{c_1,\ldots,c_{13}\}$ is a set of compounds, and $\{r_1,\ldots,r_5\}$ is a set of reactions. For example, reaction r_1 has the substrates (reactants) $\{c_1, c_2\}$, and products $\{c_6, c_7\}$. If either indegree (the number of input nodes) or outdegree (the number of output nodes) of a compound node is 0, it is called an *external node*. $\{c_1, c_2, c_3, c_4, c_5, c_6, c_9, c_{10}, c_{13}\}$ is a set of external nodes in Fig. 1, and the external nodes consist of *source nodes* and *sink nodes*. Compound nodes with indegree 0 are called source nodes and are assumed to be supplied by the external environment. $\{c_1, c_2, c_3, c_4, c_5\}$ are source nodes. On the other hand, compound nodes with outdegree 0 are called sink nodes. $\{c_6, c_9, c_{10}, c_{13}\}$ are sink nodes. *Target nodes* are chosen from sink nodes. In Fig. 1. $\{c_9\}$ is chosen as a target node.

For example, in MRC, the solution in the Boolean model is deleting $\{r_2, r_3\}$ because c_9 is produced only from r_2 or r_2. However, in the EM and FBA models, if there is a chemical reaction "$A + B \rightarrow C + D$", C and D should also exist for the reaction to take place, in addition to A and B, because steady states are assumed (see latter sections for details of EM and FBA models). Therefore, deletion of any single reaction is the solution of MRC in the FBA and EM models since all reactions must take place at a time if some reaction takes place.

2. Flux Balance Analysis-based Method

Flux balance analysis (FBA) is a constraint-based mathematical framework using the stoichiometry of a given metabolic network. In many cases, FBA is used to optimize a biologically relevant objective function to identify optimal flux distributions [19,32]. In FBA, the state of the whole metabolic network is represented by fluxes for all reactions, and the sum of incoming fluxes must be equal to the sum of outgoing fluxes for each compound, where fluxes may be weighted according to the stoichiometry coefficients.

For MRC and MRI in the FBA model, in addition to the objective function in the standard FBA, the number of added or deleted reactions should also be taken into account. Furthermore, we may need to consider two objectives: cellular objective and bioengineering objective.

In order to identify gene knockout strategies for microbial strain optimization under such a complex situation, a bilevel programming framework was introduced in [2] in which there are outer and inner optimization problems as shown in Fig. 2. The outer problem optimizes the bioengineering objective, whereas the inner problem optimizes the cellular objective.

Here, we consider MRC under the bilevel programming framework. Let v_{target} denote the flux of the reaction that produces the target compound. Our purpose is to find the minimum number of reactions

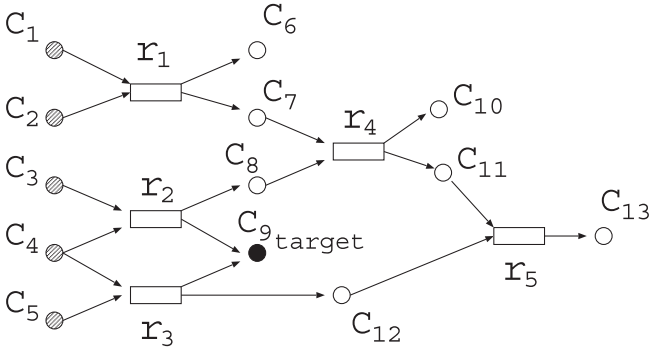

Fig. 1. An example of a metabolic network. Rectangles and circles represent chemical reactions and compounds, respectively.

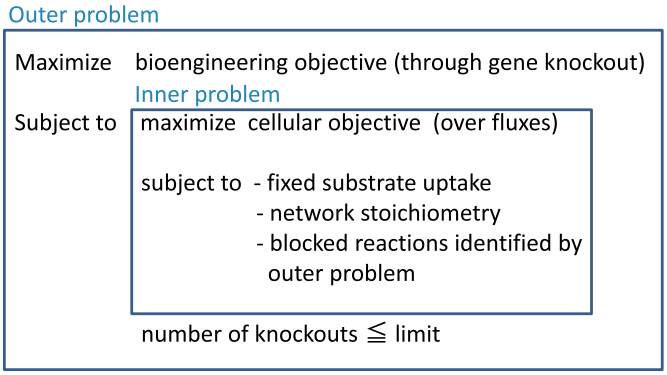

Fig. 2. A framework of the bilevel programming [2].

deletion of which always makes $v_{target} = 0$. Then, MRC in the FBA model can be formalized as follows by starting with $K = 0$, and increment K by 1 until $v_{target} = 0$ is obtained, where K is the upper limit of the number of deleted reactions.

$$\underset{s_j}{Maximize} \quad -v_{target}$$

subject to

$$\underset{v_j}{Maximize} \quad v_{target}$$

subject to

$$\begin{aligned}
&\sum_j S_{ij} \cdot v_j = 0, \forall i \in I,\\
&LB_j \cdot s_j \le v_j \le UB_j \cdot s_j, \forall j \in J,\\
&s_j \in \{0,1\}, \forall j \in J,\\
&\sum_{j \in J} (1 - s_j) \le K,
\end{aligned}$$

where s_j is a 0–1 variable, s_{ij} is a stoichiometry matrix for the ith compound and jth reaction, v_j $(j = 1,\ldots, n)$ is a flux vector, I is a set of compounds, J is a set of reactions, and LB_j and UB_j are the lower and upper bounds of v_j $(j = 1,\ldots, n)$, respectively. s_j represents whether jth reaction is knocked-out, where $s_j = 0$ indicates that jth reaction is knocked-out since v_j is forced to be 0.

In the above, we used the same function (but with different signs) as the objective functions in outer and inner optimization problems. However, there are various versions of the problem setting based on objective functions for the inner problem and the outer problem.

For example, the minimization of metabolic adjustment method (MOMA) minimizes the difference between the wild and the knocked-out flows [25]. In the flux variability analysis (FVA), both the maximum and minimum values of the objective function are calculated, and the range of them is accounted for [26]. OptKnock maximizes the bioengineering objective in the outer problem, and the cellular objective in the inner problem [2], where the upper bound of the number of removed reactions is given as in the above. On the other hand, RobustKnock maximizes the minimal possible rate of the bioengineering objective in the outer max–min problem, while the cellular objective is maximized in the inner min–max structure [30].

3. Elementary Mode-based Method

An elementary mode (EM) represents a feasible and balanced (steady-state) flux distribution of the network [24,23]. It must be minimal with respect to utilized reactions (enzymes). Suppose that a metabolic network of Fig. 3 is given, where reaction nodes are omitted. $\{A_{ex}, B_{ex}, C_{ex}, D_{ex}\}$ is a set of external compounds. In this network, there are 5 EMs, which are shown in Table 1. Although all values in Table 1