The Fourth International Conference on Computational Systems Biology (ISB2010)
Suzhou, China, September 9–11, 2010

Drug Target Identification Based on Flux Balance Analysis of Metabolic Networks

Zhenping Li[1] Rui-Sheng Wang[2] Xiang-Sun Zhang[3,*]

[1]School of Information, Beijing Wuzi University, Beijing 101149, China
[2]School of Information, Renmin University of China, Beijing 100872, China
[3]Academy of Mathematics and Systems Science, CAS, Beijing 100190, China

Abstract Efficiently identifying drug targets with minimal side effects is a major challenge in new drug development. In this paper, we develop a method based on flux balance analysis (FBA) to identify drug targets in metabolic networks. The method, which is formulated into two linear programming models, first finds the steady flux of reactions and mass flow of metabolites in the pathologic state and determines the optimal flux and mass flow in the medication state such that the side effect caused by medication is minimized. Then drug targets can be identified by comparing the flux of reactions in both states and checking the reactions whose fluxes are changed. We give an illustrative example to show that the drug target identification problem can be solved effectively by our method.

Keywords Drug Target; Metabolic Networks; Flux Balance Analysis; Linear Programming

1 Introduction

Identification and validation of disease-causing genes as drug targets is an essential first step in new drug discovery and development. In pharmaceutics, drugs generally fail in the clinic for two reasons: they either do not work or are proved to be unsafe [14]. For example, if components other than intended targets are affected by a drug, toxicity or lack of efficacy will arise. Both of these problems lead to sloppy early target discovery and are among the main challenges in developing new drugs. Traditional drug development approaches focused more on the efficacy of drugs than their toxicity, which does not meet the increasing demand of public health on new drug development. In contrast, recent drug research in post-genomic era stresses on the identification of specific biological targets or gene products, such as enzymes or proteins for drugs, which can be manipulated to produce the desired effect of curing a disease with minimum disruptive side effects [14, 1]. With the complete sequencing of human and bacterial genomes and the subsequent accumulation of genomic, proteomic, and metabolomic data, systems biology approaches or network-based analyses hold great promise for identifying drug targets by analyzing the topological structure of biological networks, such as gene regulatory networks, metabolic networks and protein interaction networks [8, 5, 15, 6, 9, 7]. However, most of these

*Corresponding author: zxs@amt.ac.cn

network-based methods do not consider the factor of side effects, which may be the main reason why only modest results have been obtained so far.

Recently, a drug target identification model based on metabolic networks has been proposed by Sridhar et al. [12, 13], in which a set of enzymes (drug targets) is to be found to inhibit target compounds through drugs' action on these enzymes and meanwhile reduce the side effects caused to non-target compounds. In other words, inhibition of the identified drug targets will stop the production of a given set of target compounds, with eliminating a minimal number of non-target compounds. In their models, the side effect of a drug is defined as the number of non-target compounds eliminated while drugs inhibit the target compounds. They presented a scalable heuristic iterative algorithm as well as a branch-and-bound exact algorithm for solving the formulated drug target detection problem [12, 13]. Song et al. developed a double iterative optimization algorithm for the same problem [11]. Li et al. formulate this metabolic network-based drug target identification model as an integer linear programming (ILP) which ensures that optimal solutions can be exactly and efficiently obtained without any heuristic manipulation [10].

The drug target discovery model mentioned above is based on the logic biochemical relationships between reactions, enzymes and compounds: a reaction is inhibited if and only if at least one of its reactant metabolites is inhibited, and a product metabolite is inhibited if and only if all reactions producing this metabolite are inhibited. Although the definition of damage in this model reflects side effects to some extent, it is still too coarse and cannot capture the quantitative relationships among reactions, metabolites and enzymes. In the process of metabolism, the mass flow of metabolites and the flux of reactions satisfy balance relationships. If the target compounds are completely inhibited by manipulating drug targets, some non-target compounds may also be eliminated, which may make some other non-target compounds' concentrations changed. If the concentrations of these non-target compounds are out of healthy range, some symptom of side effects will appear. In fact, although the accumulation of target compounds in the sophisticated metabolic system may result in diseases, it is not reasonable to inhibit them completely. We only need to adjust their concentration or mass flow to a healthy range by medication strategies. For example, the healthy range of normal empty blood sugar concentration of a person is [0, 6.11] mmol/L. If his/her empty blood sugar concentration is larger than 7.0 mmol/L, then he/she may be diagnosed to be a diabetic patient. To cure diabetes, we need to reduce their empty blood sugar concentration to a healthy range. Sridhar et al.'s drug target identification model cannot handle this case. In [15], Vera et al. proposed a method called optimization program for drug discovery (OPDD) to identify enzyme targets in an enzymopathy. But this method needs to solve a large number of optimization programs and select the most feasible solution by additional criteria. Furthermore, it does not consider more about side effects.

In this paper, we give a new definition of damage to reflect side effects of drug action, which is quantitative and more reasonable. Then we propose a method to identify drug target based on flux balance analysis. The method is formulated into two linear programs: one is to find the steady flux of reactions and metabolites in the pathologic state, and the other is to determine the optimal flux in the medication state such that the side effect caused by medication is minimized. Then drug targets are identified by comparing the flux of reactions in both states and checking the reactions whose fluxes are changed. An illustrative example is given to show that the drug target identification problem can be

solved effectively by our method.

2 Drug Target Identification Based on FBA Models

In a metabolic network, enzymes catalyze reactions which take substrates and produce metabolites. Such processes constitute the whole metabolism system of a living organism. However, the malfunctions of some enzymes may lead to production of excessive concentration or mass flow of certain compounds in the sophisticated metabolic system, and thereby may result in diseases [2, 4]. Such compounds are generally considered as target compounds because they are directly related to the diseases. The remaining compounds in the metabolic system are all considered as non-target compounds. On the other hand, those enzymes are considered as drug targets if when manipulated by drugs the concentrations or mass flow of target compounds can be adjusted to healthy ranges. Hence, the drug target identification problem is to identify an enzyme set that can be manipulated by drugs to adjust the concentrations or mass flows of all target compounds to healthy ranges, while minimizing the gap between the concentration or mass flow of non-target compounds after medication and their healthy range. The sum of the gap between the concentration or mass flow of all non-target compounds in the medicine state and their healthy state range is defined as the side effects of the corresponding enzyme set (drug targets).

2.1 Metabolic Network Representation

A metabolic network is generally a biochemical network, in which chemical compounds are nodes and reactions catalyzed by one or several certain enzymes are denoted by directed edges. In order to make drug target identification easily understood, we develop another graphical representation of the metabolic network. This type of metabolic network is directed and has two types of nodes (bipartite digraph). One type represents chemical reactions and the other metabolites. A directed edge from a reaction to a metabolite means that the metabolite is a product of the reaction. A directed edge from a metabolite to a reaction represents that the metabolite is a reactant of the reaction. A reversible reaction is considered as two separate reactions corresponding to forward and backward reactions.

Suppose that there are m metabolites $\{A_1, A_2, \cdots, A_m\}$ and n reactions $\{R_1, R_2, \cdots, R_n\}$ in a metabolic network. $S = [s_{ij}]_{m\times n}$ and $T = [t_{ji}]_{n\times m}$ are the stoichiometric coefficient matrices of reactions. The kth column of matrix S denotes the coefficients of reactants in reaction R_k, while the kth row of matrix T denotes the coefficients of metabolites produced by reaction R_k. We can obtain the kth column of matrix S and the kth row of matrix T from the chemical equation of reaction R_k. In the contrary, the chemical equation of reaction R_k can be deduced from the kth column of matrix S and the kth row of matrix T. For example, the chemical equation of reaction R_k is

$$2A_1 + 3A_2 \rightarrow A_5 + 2A_6$$

Then $s_{1k} = 2$, $s_{2k} = 3$, $t_{k5} = 1$, $t_{k6} = 2$.

2.2 Determining Pathologic Fluxes of Metabolites and Reactions

Given a metabolic network in the pathologic state, in which we delete the reactions which cannot take place because its catalyzing enzyme is inhibited. Although the metabolic network is in the pathologic state, it still can produce as much biomass or energy (i.e. ATP) as possible so as to maintain tissue growth. So we can determine the flux of each reaction and mass flow of each metabolite in the pathologic state by a optimization model.

Let $v_j, j = 1,2,\cdots,n$ denote the flux of reaction R_j, $x_i, i = 1,2,\cdots,m$ denote the mass flow of metabolite A_i, that is, the mass of metabolite A_i produced or consumed by all the reactions it involves in the metabolic network. We use the following linear programming model of flux balance analysis (FBA) to determine the mass flow of metabolites and the flux of reactions in the pathologic state:

$$\max z = \sum_{i=1}^{m} t_{biomass,i} x_i \tag{1}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} s_{ij} v_j = \sum_{j=1}^{n} t_{ji} v_i \tag{2}$$

$$x_i = \sum_{j=1}^{n} s_{ij} v_j \tag{3}$$

$$x_i = \sum_{j=1}^{n} t_{ji} v_j \tag{4}$$

$$0 \le v_j \le U_j, \ j = 1,2,\cdots,n \tag{5}$$

$$0 \le x_i \le q_i, \ i = 1,2,\cdots,m \tag{6}$$

The objective function denotes the maximization of biomass produced by the metabolism process. Eq. (2) is a constraint that guarantees the mass balance of each intermediate metabolite. Eq. (3) defines that the mass flow of each metabolite is equal to sum of mass of this metabolite consumed by all reactions. Similarly, Eq.(4) guarantees the mass flow of each metabolite is equal to sum of mass of this metabolite produced by all reactions. Constraints (5) and (6) represent the capacity limits of fluxes and mass flows in the pathologic state, where U_j and q_i are the upper bounds of variable v_j and x_i respectively.

2.3 Determining Medication Fluxes of Metabolites and Reactions

In the pathologic state, the concentrations or mass flow of some metabolites are out of healthy ranges which result in the disease symptoms. For example, if the healthy range of the jth metabolite's mass flow is $[a_j, b_j]$, it means that x_j should satisfy $a_j \le x_j \le b_j$. If $x_j > b_j$ or $x_j < a_j$, we want to adjust some fluxes of chemical reactions by using drugs such that $x_j \in [a_j, b_j]$. In this adjustment process, the mass flows of some other non-target compounds may change to be out of their health ranges, which we define as the side effects of the drugs. A good drug should have minimal side effects. Aiming to minimize the side effects, we can find the concentration or mass flow of the metabolites and the flux of reactions in the medication state by using the following linear programming model:

$$\min \sum_{i\in N}(d_i^- + d_i^+) \tag{7}$$

$$\text{s.t.} \sum_{j=1}^{n} s_{ij}v_j = \sum_{j=1}^{n} t_{ji}v_i \tag{8}$$

$$x_i = \sum_{j=1}^{n} s_{ij}v_j \tag{9}$$

$$x_i = \sum_{j=1}^{n} t_{ji}v_j \tag{10}$$

$$0 \le v_j \le U_j,\ j = 1,2,\cdots,n \tag{11}$$

$$a_i \le x_i \le b_i,\quad i \in T \tag{12}$$

$$a_i \le x_i + d_i^- - d_i^+ \le b_i,\ \ i \in N \tag{13}$$

where N is the set of non-target compounds and T is the set of target compounds, and a_i, b_i are respectively the healthy lower and upper bounds of the mass flow of the ith compounds, d_i^-, d_i^+ are respectively the negative difference variable and the positive difference variable. Constraints (8),(9) and (10) are only for intermediate metabolites.

2.4 Identifying Drug Targets and Drug Dose for Diseases

By comparing the flux vector Y^0 in the pathologic state and the flux vector Y^1 in the medication state, we can easily find the reactions whose flux has been changed by medication. We can construct a sub-metabolic network by using all these reactions along with their reactants and products. All the compounds with zero in-degree are then deleted, that is, delete all the compounds which is not a product of any reaction in this subnetwork. These compounds come into the metabolism process from the outside of the system. We can find all the reactions without reactants in the resulting graph. These reactions are determined as drug targets. In other words, we can manipulate the concentration of enzymes that catalyze these reactions by drugs so as to adjust the fluxes of these reactions such that the mass flow of target compounds is in healthy range.

In [3], it is indicated that the flux of a reaction is correlated with the concentration level of the enzymes catalyzing this reaction. The concentration of enzymes can be controlled by drugs, so the drug dose can be determined according to the flux of reactions in both pathologic and medication state. We can also integrate the result with other successful methods such as primarily experimental methods for determining the suitable dose to cure the disease.

3 A Numerical Example

Figure 1 is an illustrative example of 12 metabolites and 8 reactions, where we assume that metabolites A_8, A_9, A_{11}, A_{12} are involved in the biomass reaction. In the pathologic state, the upper bounds of all reaction fluxes are taken as 10, and the upper bounds of mass flow of all metabolites are taken as infinity. The metabolic network in Figure 1 can be expressed by the following chemical reaction equations.

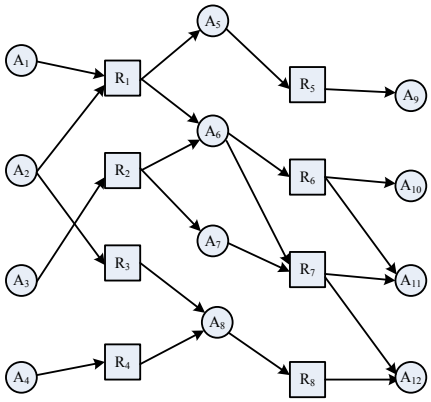

Figure 1: A metabolic network with 12 metabolites and 8 reactions.

$$
\begin{aligned}
R_1: &\quad 2A_1 + A_2 \to A_5 + A_6 \\
R_2: &\quad 4A_3 \to 3A_6 + A_7 \\
R_3: &\quad 3A_2 \to A_8 \\
R_4: &\quad A_4 \to 2A_8 \\
R_5: &\quad 2A_5 \to 3A_9 \\
R_6: &\quad 2A_6 \to A_{10} + 2A_{11} \\
R_7: &\quad A_6 + 3A_6 \to 2A_{11} + 3A_{12} \\
R_8: &\quad 2A_8 \to 3A_{12}
\end{aligned}
$$

Using Lingo software to solve the pathologic linear programming model, we can obtain the results expressed in Figure 2(a), where the optimal mass flow of metabolites is $X^0 = (20, 40, 15, 5, 10, 21.25, 3.75, 20, 15, 10, 22.5, 33.75)$, depicted beside the corresponding nodes, and the flux of reactions is $Y^0 = (10, 3.75, 10, 5, 5, 10, 1.25, 10)$, also depicted beside the corresponding nodes. Let

$$W = ([0,M],[0,M],[0,M],[0,M],[0,M],[0,M],[0,M],[10,15],[10,15],[10,15],[0,1])$$

be the healthy range of all the metabolites, and $U_j = 10$ for all $1 \le j \le 8$ be the upper bound of each reaction flux. we can find the optimal flux of reactions and mass flow of metabolites in medication state. The optimal mass flow of each metabolite is $X^1 = (20, 10, 4, 0, 10, 13, 1, 0, 15, 6.33, 13.33, 1)$ and the optimal flux of each reaction is $Y^1 = (10, 1, 0, 0, 5, 6.33, 0.33, 0)$, both shown in Figure 2(b). The side effect is 3.667 since the mass flow of metabolite A_{10} is 6.33 which is out of the healthy range $[10, 15]$.

By comparing Y^0 and Y^1, we construct a sub-metabolic network, shown in Figure 3(a). According to the method described in Section 2, drug targets are the enzymes which catalyze R_2, R_3, R_4 respectively. If we adjust the flux of R_2, R_3, R_4 respectively to be $1, 0, 0$ by drugs, then the concentration or mass flows of target compounds will be in healthy range with side effect 3.667. This result can also be obtained by the pathologic model with the constraints of reaction R_2, R_3, R_4's fluxes being $1, 0, 0$ respectively.

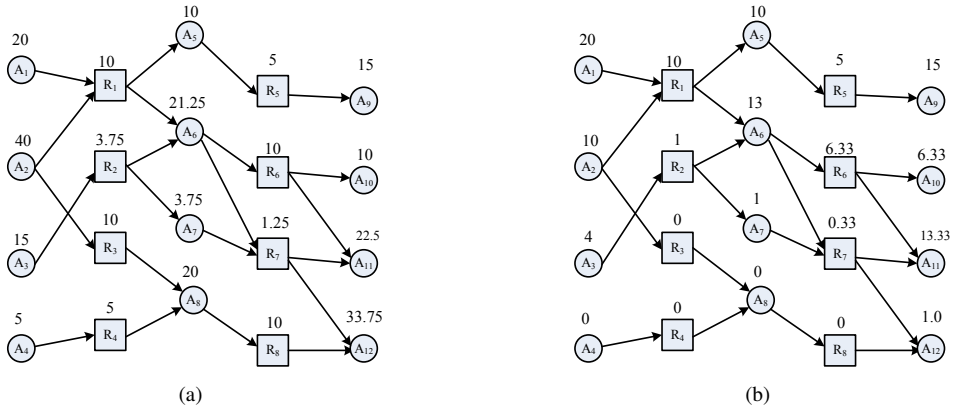

Figure 2: (a) The flux of reactions and mass flow of metabolites in the pathologic state. (b)The flux of reactions and mass flow of metabolites in the medication state.

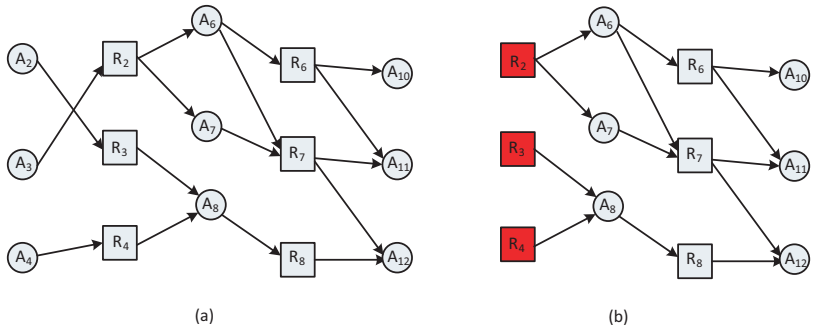

Figure 3: (a) The sub-metabolic network. (b) The reactions corresponding to drug targets are R_2, R_3, R_4.

4 Conclusion

Efficiently identifying drug targets with minimal side effects is a major challenge in new drug development. Previous models for identifying drug targets either are not quantitative or do not consider side effects. In this paper, we develop a quantitative method based on flux balance analysis to identify drug targets in metabolic networks. The method involves finding the steady flux of reactions and concentration or mass flow of metabolites in both pathologic state and medication state and also considers the side effects of drug action. We give an illustrative example to show that the drug target identification problem can be solved effectively by our method. In the future, we will explore the application of this method in the large human metabolic network.

Acknowledgments

This research is supported by the National Natural Science Foundation of China under Grant No.10631070, No.60873205, No.10701080 and the Beijing Natural Science Foundation under Grant No.1092011. It is also partially supported by the Foundation of Beijing Education Commission under Grant No. SM200910037005, the Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality (PHR201006217), and the Funding Project for Base Construction of Scientific Research of Beijing Municipal Commission of Education (WYJD200902).

References

[1] Campillos, M., Kuhn, M., Gavin, A.C., Jensen, L.J., Bork, P. Drug target identification using side-effect similarity. *Science*,321(5886): 263-266, 2008.

[2] Cascante, M., Boros, L.G., Comin-Anduix, B., de Atauri, P., Centelles, J.J., Lee, P.W. Metabolic control analysis in drug discovery and disease. *Nat Biotechnol.*, 20(3):243-249, 2002.

[3] De, R.K., Das, M., Mukhopadhyay, S. Incorporation of enzyme concentrations into FBA and identification of optimal metabolic pathway. *BMC Systems Biology*, 2:65, 2008.

[4] Faulon, J.L., Misra, M., Martin, S., Sale, K., Sapra, R. Genome scale enzyme-metabolite and drug-target interaction predictions using the signature molecular descriptor. *Bioinformatics*, 24(2): 225-233, 2008.

[5] Guimerà, R., Sales-Pardo, M., Amaral, L.A.N. A network-based method for target selection in metabolic networks. *Bioinformatics*, 23: 1616-1622, 2007.

[6] Hormozdiari, F., Salari, R., Bafna, V., Sahinalp, S.C. Protein-protein interaction network evaluation for identifying potential drug targets. *Journal of Computational Biollogy*, 17(5): 669-684, 2010.

[7] Karlebach, G., Shamir, R. Minimally perturbing a gene regulatory network to avoid a disease phenotype: the glioma network as a test case. *BMC Systems Biology*, 4: 15, 2010.

[8] Kell, D.B. Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discovery Today*, 11: 1085-1092, 2006.

[9] Kushwaha, S.K., Shakya, M. Protein interaction network analysis–approach for potential drug target identification in *Mycobacterium tuberculosis*. *Journal Theoretical Biology*, 262(2): 284-294, 2010.

[10] Li, Z., Wang, R.S., Zhang, X.S., Chen, L. Detecting drug targets with minimum side effects in metabolic networks. *IET Systems Biology*, 3(6): 523-533, 2009.

[11] Song, B., Sridhar, P., Kahveci, T., Ranka, S. Double iterative optimisation for metabolic network-based drug target identification. *International Journal of Data Mining and Bioinformatics*, 3(2): 124-144, 2009.

[12] Sridhar, P., Kahveciy, T., Ranka, S. An iterative algorithm for metabolic network-based drug target identification. *Pacific Symposium on Biocomputing*, 12: 88-99, 2007.

[13] Sridhar, P., Song, B., Kahveciy, T., Ranka, S. Mining metabolic network for optimal drug targets. *Pacific Symposium on Biocomputing*, 13: 291-302, 2008.

[14] Smith, C. Drug target validation: Hitting the target. *Nature*, 42: 341-347, 2003.

[15] Vera, J., Curto, R., Cascante, M.,Torres, N.V., Detection of potential enzyme targets by metabolic modelling and optimization: Application to a simple enzymopathy, *Bioinformatics*, 23(17): 2281-2289, 2007.