The Second International Symposium on Optimization and Systems Biology (OSB'08)
Lijiang, China, October 31– November 3, 2008

Mass flow model and essentiality of enzymes in 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 Understanding the metabolism mechanisms in living organisms is a major task for post-genomic biology. Cellular metabolism is usually represented by a complex network of reactants connected by chemical reactions catalyzed by enzymes. In this paper, a mass flow conservation model is proposed to describe the process of cellular metabolism, which can be formulated as a linear programming model. Based on such model, the importance of an enzyme is quantitatively defined in metabolic network according to the variation of mass flow and final products resulting from the deletion of the enzyme. This quantitative criterion can be used not only to predict the essentiality of enzyme but also to identify drug targets.

Keywords enzyme essentiality; mass flow; metabolic network; linear programming

1 Introduction

Metabolism is the biochemical modification of chemical compounds in living organisms. Cellular metabolism includes all chemical processes in a cell that produce energy and basic materials needed for important life processes. This includes the biosynthesis of complex organic molecules (anabolism) and their breakdown with release of energy (catabolism). A substance which participates a biochemical reaction is called a metabolite. A process in which two or more molecules (reactants) interact and produce a product, usually with the help of an enzyme, is called a biochemical reaction.

Cellular metabolism is usually represented by a complex network called metabolic networks, in which reactants are vertices and chemical reactions are edges. Biochemical reactions are usually catalyzed by an enzyme, which is usually a protein and is translated from a corresponding gene. Experimental determination the kinetic parameters in hundreds of reactions in a cell is a challenging problem. Therefore, analyzing the static structure of metabolic network to infer causal and physiological relationships is a promising approach. From a practical and biological point of view, it is very important to investigate the influence of enzymes on the metabolic network. Enzymes are subject to evolution and can be genetically engineered to change metabolic output. They can also be targets for drugs [5], so identifying important enzymes is a critical issue in drug design or pharmaceutical industry [12, 1].

An enzyme is considered as essential if the deletion of the gene coding for that enzyme has lethal effects on the organism under a given experimental condition. The experimental

identification of essential genes has been carried out in some bacteria and yeast [9, 11]. Several computational methods have been proposed to define the importance of enzymes in a metabolic network to predict enzyme essentiality, such as flux analysis method [7, 13], damage analysis method [2, 14], load point method [3], degree analyze method [4] and others [6, 10, 18, 17, 19, 20].

In this paper, we present a novel mass flow model for describing cellular metabolism, based on which we give a quantitative definition to determine the importance of enzymes for the survival of an organism. The method predicts quantitatively essentiality of an enzyme based on the variations of mass flow resulting from the deletion of the enzyme. Our quantitative criterion for evaluating enzyme importance is the deleterious effect of its removal from the network. Since the deletion of an enzyme may cause several chemical reactions being prohibited, so the mass flow of the metabolites as well as the final products in the network may change depending on the biological role of the enzyme. If the mass flow of an important final product vanishes, the organism will die; if the final mass flow decreases but not vanishes, the organism will exhibit some disease symptoms. Although living organisms have evolved to maximize their chances for survival [15], there still exist some enzymes whose deletion will be lethal. An important evaluation criterion is the change of the mass flow for some important metabolites. In this paper, we propose a quantitative criterion based on the mass flow change rate.

This paper is organized as follows: in section 2, a linear programming model based on the mass flow conservation for metabolic network is proposed; in section 3, the essentiality criterion of an enzyme is defined and a sample example is illustrated. Section 4 is conclusion.

2 Mass flow model for metabolic networks

Barabási and co-workers [8] have introduced a graph representation of the metabolic network, where the nodes and the links connecting the nodes denote metabolites and chemical reactions, respectively. In this paper, we develop another graphical representation of the metabolism[2] based on mass conservation condition. The metabolic network is directed and has two types of nodes (bipartite digraph [16]). One type represents chemical reactions and the other metabolites. A link between a reaction and a metabolite is directed towards the metabolite, if the metabolite is a product, and in the opposite direction, if the metabolite is a reactant. We treat reversible reactions as two separate 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. We can add a source node R_0 and a sink node R_{n+1} to the network. And add links with direction from the source node R_0 to the metabolite with indegree zero, where the metabolite with indegree zero means there is no chemical reaction producing it. Meanwhile, we can add links with direction from the metabolite with outdegree zero to the sink node R_{n+1}, where the metabolite with outdegree zero means that it is not a reactant of any chemical reaction and will leave the systems.

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & a_{1,n+1} \\ a_{21} & a_{22} & \cdots & a_{2n} & a_{2,n+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & a_{m,n+1} \end{pmatrix}$$

where A is the adjacent matrix from the metabolites to the chemical reactions in the metabolic network. For $1 \leq k \leq n$, $a_{ik} = 1$ if the metabolite A_i is the reactant of reaction R_k, and $a_{ik} = 0$ otherwise. For $k = n+1$, $a_{i,n+1} = 1$ if the metabolite A_i is not a reactant of any reaction; $a_{i,n+1} = 0$, otherwise.

$$B = \begin{pmatrix} b_{01} & b_{02} & \cdots & b_{0m} \\ b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix}$$

where B is the adjacent matrix from the chemical reactions to the the metabolites in the network. For $1 \leq k \leq n$, $b_{kj} = 1$ if the metabolite A_j is the product of reaction R_k, and $b_{kj} = 0$ otherwise. For $k = 0$, $b_{0j} = 1$ if metabolite A_j is not a product of any reaction, and $b_{0j} = 0$ otherwise.

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} & s_{1,n+1} \\ s_{21} & s_{22} & \cdots & s_{2n} & s_{2,n+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} & s_{m,n+1} \end{pmatrix}$$

$$T = \begin{pmatrix} t_{01} & t_{02} & \cdots & t_{0m} \\ t_{11} & t_{12} & \cdots & t_{1m} \\ t_{21} & t_{22} & \cdots & t_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ t_{n1} & t_{n2} & \cdots & t_{nm} \end{pmatrix}$$

S and T are the mass coefficient matrices of reactions. The kth column of matrix S denotes the mass coefficients of reactants in reaction R_k, while the kth row of matrix T denotes the mass coefficients of metabolites produced by reaction R_k. For example, we can obtain the kth column of matrix S and the kth row of matrix T from the conservation of mass equation in the reaction R_k. In the contrary, the conservation of mass equation in the reaction R_k can be deduced from the kth column of matrix S and the kth row of matrix T. That is, the conservation of mass equation in reaction R_k is

$$\sum_{i=1}^{m} s_{ik} A_i = \sum_{j=1}^{m} t_{kj} A_j$$

where s_{ik} (or t_{kj}) equals to the product of reactant A_i's (A_j) coefficient in the R_k's chemical equation and A_i's (A_j) molar mass.

For example, the chemical equation of reaction R_k is

$$2A_1 + 3A_2 \rightleftharpoons A_3 + 2A_4$$

The molar mass of A_1, A_2, A_3, A_4 are respectively $12, 20, 40, 22$. Then $s_{1k} = 2 \times 12 = 24$, $s_{2k} = 3 \times 20 = 60$, $t_{k3} = 40$, $t_{k4} = 2 \times 22 = 44$, and the conservation of mass equation corresponding to reaction R_k is as follows

$$24A_1 + 60A_2 = 40A_3 + 44A_4.$$

It can be simplified to

$$6A_1 + 15A_2 = 10A_3 + 11A_4$$

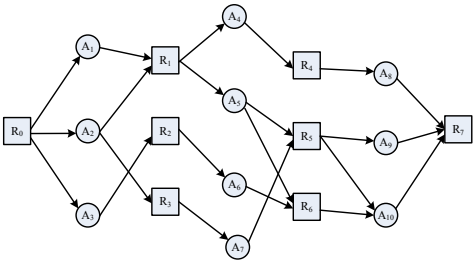

Figure 1: A metabolic network with 10 metabolites and 6 reactions. R_0 is the added source node and R_7 is the added sink node

Figure 1 is an illustrative example of 10 metabolites and 6 reactions, R_0 is the added source node and R_7 is the added sink node. R_i $(i = 1,2,\cdots,6)$ denote the reactions, for example, those which can be expressed by the following conservation of chemical reaction mass equations.

$$\begin{aligned}
R_1: &\quad 8A_1 + 4A_2 = 3A_4 + 9A_5 \\
R_2: &\quad A_3 = A_6 \\
R_3: &\quad A_2 = A_7 \\
R_4: &\quad A_4 = A_8 \\
R_5: &\quad 2A_5 + 3A_7 = A_9 + 4A_{10} \\
R_6: &\quad A_5 + A_6 = 2A_{10}
\end{aligned}$$

Let x_{ik} $(i = 1,2,\cdots,m; k = 1,2,\cdots,n)$ denote the mass of metabolite A_i participant in reaction R_k as reactant, and y_{kj} $(k = 1,2,\cdots,n; j = 1,2,\cdots,m)$ denote the mass of metabolite A_j produced by reaction R_k. The maximal flow model can be expressed by the following linear programming model.

$$\max z = \sum_{i=1}^{m} c_i x_{i,n+1} \tag{1}$$

$$
s.t.\begin{cases}
\sum_{k=0}^{n} b_{ki}y_{ki} = \sum_{j=1}^{n+1} a_{ij}x_{ij} \;\; for \;\; i = 1,2,\cdots,m & (2)\\
\sum_{i=1}^{m} a_{ik}x_{ik} = \sum_{j=1}^{m} b_{kj}y_{kj} \;\; for \;\; k = 1,2,\cdots,n & (3)\\
\frac{x_{ik}}{s_{ik}} = \frac{x_{jk}}{s_{jk}} \;\; for \;\; 1 \le k \le n;\; 1 \le i,j \le m,\; s_{ik} \ne 0,\; s_{jk} \ne 0 & (4)\\
\frac{y_{ki}}{t_{ki}} = \frac{y_{kj}}{t_{kj}} \;\; for \;\; 1 \le j \le n;\; 1 \le i,j \le m,\; t_{ki} \ne 0,\; t_{kj} \ne 0 & (5)\\
x_{ik} \le a_{ik}U_{ik} \; for \; 1 \le i \le m,\; 1 \le k \le n+1 & (6)\\
y_{kj} \le b_{kj}V_{kl} \; for \; 1 \le j \le m,\; 0 \le k \le n & (7)\\
x_{ik} \ge 0, i = 1,2,\cdots,m, k = 1,2,\cdots,n+1 & (8)\\
y_{kj} \ge 0, j = 1,2,\cdots,m, k = 0,1,2,\cdots,n & (9)
\end{cases}
$$

The objective function denotes the maximum combination mass of terminal compounds produced by the metabolism process, that is, the combination mass flow of the final metabolite should be maximum, where c_i is the coefficient of combination, the larger the c_i, the more important the compound is. Constraint (2) guarantees that mass of each metabolite is balanced. Constraint (3) guarantees the conservation of mass in each reaction, that is, the total mass of all the reactant is equal to the total mass of all the products. Constraints (4) and (5) guarantee that the conservation of reaction mass equation is satisfied, that is, the mass of reactants and products in each reaction satisfy the proportion relationship. Constraints (6) and (7) guarantee that all the mass flows along the arcs satisfy the capacity limit, where U_{ik} and V_{kl} are the upper bounds of variable x_{ik} and y_{kl} respectively. Constraints (8) and (9) ensure that all variables are nonnegative.

Since the mass of metabolites is proportional to its concentration, we can easily obtain the concentration of each metabolite using the value of its mass divided by its molar mass. On the contrary, if we use the concentration of each metabolite as the variables, we only need to revise the linear programming model. Hence, we can use the revised linear programming model to find the optimal concentrations or mass flow of each metabolite. For the metabolic network in Figure 1.

$$
A = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0\\
1 & 0 & 1 & 0 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 1 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0\\
0 & 0 & 0 & 0 & 1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

$$
B = \begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

The linear programming model of this metabolic network is as follows:

$$
\begin{aligned}
\max \quad & z = x_{8,7} + x_{9,7} + x_{10,7} \\
s.t. \quad & \sum_{k=0}^{6} b_{ki} y_{ki} = \sum_{j=1}^{7} a_{ij} x_{ij} \quad for \;\; i = 1, \cdots, 10 \\
& \sum_{i=1}^{10} a_{ik} x_{ik} = \sum_{j=1}^{m} b_{kj} y_{kj} \quad for \;\; k = 1, \cdots, 6 \\
& x_{1,2}/8 = x_{2,1}/4 \\
& y_{1,4}/3 = y_{1,5}/9 \\
& x_{5,5}/2 = x_{7,5}/3 \\
& y_{5,9}/1 = y_{5,10}/4 \\
& x_{5,6}/1 = x_{6,6}/1 \\
& x_{ik} \leq 10 a_{ik} \; for \; 1 \leq i \leq 10, \; 1 \leq k \leq 7 \\
& y_{kj} \leq 10 b_{kj} \; for \; 1 \leq j \leq 10, \; 0 \leq k \leq 6 \\
& x_{ik} \geq 0, i = 1, \cdots, 10, k = 1, \cdots, 7 \\
& y_{kj} \geq 0, j = 1, \cdots, 10, k = 0, 1, \cdots, 6
\end{aligned}
$$

Using Lingo software to solve this linear programming model, we can obtain the results expressed in Figure 2, where the mass flow of each metabolite is depicted beside its corresponding arc.

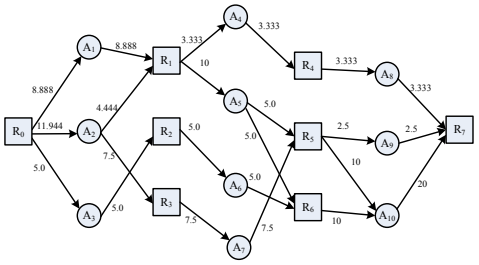

Figure 2: The mass flow of each arc in metabolic network of Figure 1. The maximal mass flow is 3.333+2.5+20=25.833. which means that the network can produce 25.833 total compounds.

3 Quantitative definition for enzyme importance

The metabolites produced by the network is essential for the organize's growth. If the mass of a metabolite is reduced, the concentration of this metabolite in organism will decrease, which will result in the slowly growth or death. The importance of an enzyme

can be defined by comparing the two values of mass flow before and after the enzyme is prohibited. Suppose that the mass flow of a metabolic network is z_0 before the enzyme is prohibited (or in wild state). When we prohibit the activity of an enzyme, the reactions catalyzed by this enzyme will be prohibited so that the mass flow will be changed. We can calculate the combinatorial mass flow in the new metabolic condition. This can be realized by simply adding a constraint to the linear programming model. The added constraint represents that the chemical reaction catalyzed by the enzyme is inhibitory. Supposing that the new mass flow value is z_1, then the importance of the enzyme can be defined by the following measure:

$$E = \frac{z_0 - z_1}{z_0} \tag{1}$$

For example, the combinatorial mass flow of the metabolic network in Figure 2 is 25.8333. After inhibiting the enzyme which catalyzes reaction R_2, the mass flow of the network is 14.1667 (see Figure 3). Hence, the essentiality of the enzyme which catalyze reaction R_2 is $E(R_2) = (25.833 - 14.1667)/25.8333 = 0.452$, which means that the combinatorial mass flow of the final metabolic in the new condition is equal to 54.8% of that in the wild condition.

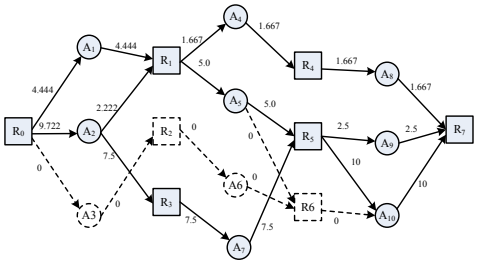

Figure 3: The maximal mass flow of the metabolic network in Figure 1. R_0 is the added source node and R_7 is the added sink node.

By the same way, we can calculate the importance of the enzymes which catalyze the other reactions, the result is as follows.

enzyme (reaction)	importance
R_1	1
R_2	0.452
R_3	0.548
R_4	1
R_5	0.548
R_6	0.452

$E(R_1) = (25.833 - 0)/25.833 = 1$ means that inhibiting the enzyme that catalyzes reaction R_1 is lethal.

4 Conclusion

In this paper, we first present a novel mass flow model to describe the process of cellular metabolism, and then gave an enzyme importance model to quantitatively predict essentiality of an enzyme from the variation of mass flow after prohibiting the chemical reactions catalyzed by that enzyme. Both models are clear and simple. They can be used not only to predict the essentiality of an enzyme in real metabolic networks but also to identify the drug targets. Comparing with flux balance analysis model, the mass flow model includes all the information of flux balance analysis model. It is much easier to understand the relationship between the metabolites and reactions in metabolic network by mass flow model.

Acknowledgments

This work is partly supported by Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality (PHR(IHLB)), National Natural Science Foundation of China under Grant No.60873205, No.10631070, No.60503004, No. 10701080, Scientific Research Base of Beijing Wuzi University.

References

[1] 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.

[2] Lemke, N., Her*é*dia, F., Barcellos, C.K., Reis, A.N., Mombach, J.C.M. Essentiality and damage in metabolic networks. *Bioinformatics*, 2004, 20, 115-119.

[3] Rahman, S.A., Schomburg, D. Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks. *Bioinformatics*, 2006, 22, 1767-1774.

[4] Samal, A., Singh,S., Giri, V., Krishna,S.,Raghuram, N.,Jain, S. Low degree metabolites explain essential reactions and enhance modularity in biological networks. *BMC Bioinformatics*, 2006, 7: 118.

[5] Karp, P.D., Krummenacker,M., Paley,S. and Wagg, J. Integrated pathway-genome databases and their role in drug discovery. *Trends Biotechnol.*, 1999, 17, 275-281.

[6] Motter, A.E., Gulbahce N., Almaas E., and Barabais A.L. Predicting synthetic rescues in metabolic networks. *Molecular Systems Biology* 2008, 4:168.

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

[8] Jeong, H., Tombor,B., Albert,R., Oltvai,Z.N. and Barab*á*si,A.-L. The large-scale organization of metabolic networks. *Nature*, 2000, 407, 651-654.

[9] Jeong, H. et al. Lethality and centrality in protein networks. *Nature*, 2001, **411**, 41-42.

[10] Jeong *et al*. Prediction of protein essentiality based on genomic data. *ComPlexUs*, 2003, **1**:19-28.

[11] Kato, J., Hashimoto, M. Construction of consecutive deletions of the Escherichia coli chromosome. *Mol Syst Biol*, 2007, **3**, 132.

[12] Kell, D.B. Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discovery Today*, 2006, **11**.

[13] Lmieli*ń*ski,M., Belta,C., Hal*á*sz,*Á*., Rubin,H., Inveatigating metabolite essentiality through genome-scale analysis of Escherichia coli production capabilities, *Bioinformatics*, 2005,**21**, 2008-2016.

[14] Mombach, J.C. *et al*. Bioinformatics analysis of mycoplasma metabolism: important enzymes, metabolic similarities, and redundancy. *Comput Biol Med*, 2005, **36**(5), 542-52.

[15] Darwin, C. The origin of species by means of natural selection. New York: Crowell, 1899.

[16] Chartrand, G. Introductory Graph Theory. Dover Publications, New York, 1977.

[17] Palumbo, M.C., Colosimo,A., Giuliani,A., Farina,L., Functional essentiality from topology features in metabolic networks: A case study in yeast, *Federation of European Biochemical Societies Letters*, 2005, **579**, 4642-4646.

[18] Yoon,J., Lee,K., Metabolic flux profiling of reaction modules in liver drug transformation, *Pacific Symposium on Biocomputing*, 2007, **12**: 193-204.

[19] Lee,J.M., Gianchandani,E.P., Papin,J., Flux balance analysis in the era of metabolomocs, *Briefings in Bioinformatics*, 2006 **7(2)**:140-150.

[20] Shlomi,T., Berkman,O., Ruppin,E., Regulatory on/off minimization of metabolic flux changes after genetic perturbations, *Proc. Natl. Acad. Sci. USA* , 2005, **102 (21)**: 7695-7700.